International Journal of Psychological Research, 2010. Vol. 3. No. 1.
ISSN impresa (printed) 2011-2084
ISSN electrónica (electronic) 2011-2079

Dutta Roy, D., (2010). Cluster analysis for test-retest Reliability. *International Journal of Psychological Research, 3* (1), 131-139.

Cluster analysis for test-retest Reliability.

Análisis de conglomerados para la fiabilidad de procedimientos de prueba-reprueba.

Debdulal Dutta Roy
Indian Statistical Institute

ABSTRACT

Conventional statistical model for assessing test-retest reliability of questionnaire is to compare composite scores of matched subtests or items between periods. Current study examined test-retest reliability in terms of both subtest and item wise comparison through paired t-test statistics. Data were collected from 72 students twice within 8 months intervals using reading motivation questionnaire (Dutta Roy, 2003). The questionnaire includes 3 intrinsic (rApp, rKnow,rAch) and 4 extrinsic (rAes, rRecog, rAff and rHarm) reading motivation subtests. Finding inconsistency between both subtest and item wise comparison, I propose a model of clustering items between periods. Hierarchical clustering with complete linkage shows that more than 55% of both test and retest items are assembled in the primary cluster suggesting high test retest reliability among 6 reading motivation variables.

.

Key words: Paired t-statistic**,** Hierarchical clustering, complete linkage, Dendograms, Reading motivation.

RESUMEN

Modelos estadísticos convencionales para evaluar la fiabilidad de procedimientos de prueba-reprueba (P-Rp) en cuestionarios buscan comparar puntuaciones combinadas de subpruebas pareadas o partes del cuestionario en la fase de prueba y la fase de reprueba. El presente artículo examina la fiabilidad de P-Rp en términos de subpruebas y comparaciones al nivel de partes del cuestionario a través de la pruebas t pareada. Datos de un cuestionario sobre motivación de la lectura tomados durante la prueba y la reprueba de 72 estudiantes, con un intervalo de 8 meses entre la prueba y la reprueba, se usan para ejemplificar el análisis de conglomerados (Dutta Roy, 2003). El cuestionario incluye 3 subpruebas intrínsecas y 3 subpruepas extrínsecas. Se propone un método de conglomerado de componentes del cuestionario para mostrar inconsistencias en las comparaciones entre las subpruebas y preguntas específicas que hacen parte del cuestionario.

Palabras clave: prueba t pareada, análisis jerárquico de conglomerados, agrupamiento por ligamiento completo, dendogramas, motivación para la lectura.

Article received/Artículo recibido: December 15, 2009/Diciembre 15, 2009, Article accepted/Artículo aceptado: March 15, 2009/Marzo 15/2009
Dirección correspondencia/Mail Address:
Debdulal Dutta Roy, Psychology Research Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata, West Bengal, INDIA-700108, Email: ddroy@isical.ac.in

INTERNATIONAL JOURNAL OF PSYCHOLOGICAL RESEARCH esta incluida en PSERINFO, CENTRO DE INFORMACION PSICOLOGICA DE COLOMBIA, OPEN JOURNAL SYSTEM, BIBLIOTECA VIRTUAL DE PSICOLOGIA (ULAPSY-BIREME), DIALNET y GOOGLE SCHOLARS. Algunos de sus articulos aparecen en SOCIAL SCIENCE RESEARCH NETWORK y está en proceso de inclusion en diversas fuentes y bases de datos internacionales.
INTERNATIONAL JOURNAL OF PSYCHOLOGICAL RESEARCH is included in PSERINFO, CENTRO DE INFORMACIÓN PSICOLÓGICA DE COLOMBIA, OPEN JOURNAL SYSTEM, BIBLIOTECA VIRTUAL DE PSICOLOGIA (ULAPSY-BIREME), DIALNET and GOOGLE SCHOLARS. Some of its articles are in SOCIAL SCIENCE RESEARCH NETWORK, and it is in the process of inclusion in a variety of sources and international databases.

International Journal of Psychological Research, 2010. Vol. 3. No. 1.
ISSN impresa (printed) 2011-2084
ISSN electrónica (electronic) 2011-2079

Dutta Roy, D., (2010). Cluster analysis for test-retest Reliability. *International Journal of Psychological Research, 3* (1), 131-139.

INTRODUCTION

The problem of assessing time consistency in test items has been the focus of extensive research and documentation in a variety of disciplines (Walker & Cosden, 2007; Duquette et al., 2005). It is a basic methodological step in the process of test development especially when there is a high probability in changing test item data. For example, Spielbergerôs state anxiety inventory was administered to Antarctica expeditioners across three periods of journey ï journey to, living in and return from the Antarctica. Out of 20 items, only 9 items differed significantly across periods (Dutta Roy, 1995). Chen and Small (2007) found poor response stability among stroke patients in assessing test-retest reliability of language imaging experiments. Usually, composite scores (sum of item weightage) across periods are compared in assessing test-retest reliability resulting difficulty to understand which sets of items are consistent over periods. This study proposes hierarchical cluster analysis model in order to overcome this limitation.

Cluster analysis

As a multivariate technique, cluster analysis helps to identify similar entities on the basis of characteristics they possess. It helps to classify objects or variables having functional homogeneity. The resulting object clusters exhibit high internal homogeneity (within cluster) and high external heterogeneity between any two clusters (Hair et. al., 2006). It is an inductive treatment and a purely empirical method of classification. There are two broad methods of cluster analysis ï hierarchical and non-hierarchical. In case of non-hierarchical cluster analysis, user has knowledge about the number of classes to be discovered or about the distance measure, on which, to base the classification. But in case of hierarchical clustering, subject has no such knowledge. Hierarchical cluster analysis is more explanatory than non-hierarchical.

Hierarchical cluster analysis provides a tree-like taxonomic system, in which at one end, every entity is a cluster and at the other end all entities are included in a common cluster. The hierarchical cluster configuration is usually represented by dendogram. Dendogram is a graphical representation (a tree graph) of the results of a clustering procedure in which the vertical axis consists of the objects or variables and the horizontal axis represents the number of clusters formed at each step of the clustering procedure.

Hierarchical cluster analysis is useful for exploring item cluster or item taxonomy. Item cluster refers to the group of homogenous items ï measuring some attributes of a particular component measured by corresponding psychological test. Dutta Roy (2003) using hierarchical clustering noted three item clusters of 12-items General health questionnaire or GHQ-12. The clusters are social dysfunction, psychological distress and self-esteem. No studies yet been conducted to examine consistency in item-clustering across periods. As a result, the model of test-retest reliability following item clustering is not known. Current study examines test-retest reliability of reading motivation questionnaire (Dutta Roy, 2003) using non-clustering and clustering approaches.

Reading motivation

Reading motivation refers to desire to put more effort on reading activities. Students motivate to read for extrinsic and intrinsic reasons. Extrinsic reasons mean reading for external pressures. On the contrary, intrinsic reasons mean reading for inner desires. Extrinsic motivation includes 3 things ï reading in order to be loved by others (rAff), to be recognized by others (rRecog), and to avoid otherôs punishment. Intrinsic reading motivation includes 4 things ï reading to develop mastery over reading (rAch), to acquire knowledge (rKnow), to apply knowledge (rApp) and to enjoy pictures and fonts (rAes). Reading motivation questionnaire (Dutta Roy, 2003) is a useful instrument to assess above 4 intrinsic and 3 extrinsic reading motivation variables. The study examined test-retest reliability of these seven variables through both paired t-test (Hamashima & Yoshida, 2002) and hierarchical cluster analysis. Insignificant difference in paired t-test indicates high test-retest reliability.

Hypothesis

There would be no significant mean differences between test and retest periods in seven reading motivation variables.

METHOD

Sample

Initially questionnaire was administered to 136 students (70 students of class III and 66 students of class IV) of one Government school. After eight months, the reading motivation questionnaire was re-administered to 90 samples as 46 students were not available during retest period. Among them 18 students gave incomplete data. So test-retest reliability for reading motivation questionnaire was assessed using 72 data (26 students of class III and 46 students of class IV). The mean age for grade III students was 8.5 years with SD 0.5 and same for the grade IV students was 9.5 years with SD 0.6.

Instrument

Reading motivation questionnaire (Dutta Roy, 2003) was used to examine test retest reliability of 3 intrinsic and 4 extrinsic reading motivation variables. Intrinsic variables are rAch, rKnow, rApp and rAes. rRecog, rAff and rHarm are extrinsic variables. There are 6 items in each subtest covering 42 items in total. For each item, one event is given with answers of two options. Selection of option reflects oneôs specific reading motivation. For example, the event is ñOn promotion to the new class, of the two books, which would you read first?ò And the options are: (a) Application of arithmetic in daily life and (b) Description of animals of different countries. First option reflects motivation for application and second option reflects motivation for knowledge. Highest score for each variable is 6 as there are 6 items to measure each subtest. Since scores are binary, item-total correlation for each subtest was computed using point biserial correlation and it was noted that all the coefficients were significant at 0.01 level suggesting good internal consistency among the items for 7 subtests (Dutta Roy, 2003).

Data analysis

Paired t-statistics and Cluster analysis were computed using the statistical software STATISTICAô99 (Kernel Release 5.5 A) of Stat Soft. Inc. Basic statistics menu was used for paired t-test and cluster analysis menu was used for construction of dendograms.

NON-CLUSTERING APPROACH

Paired t-tests and product moment correlations between test-retest measures are conventional statistical tools to assess test retest reliability in non-clustering approach. Paired t-test is used to compare means on the same or related subject over time or in differing circumstances. The observed data are from the same object or from a matched subject. Table 1 presents mean differences between test and retest periods in subtests and item wise paired t-tests

Subtests

7 subtests were used to measure 7 reading motivation variables. Averaging item weights of each subtest, subtest measures are determined. When test-retest reliability is high, subtest scores will not vary between the periods. Table 1 shows that 5 subtests scores did not vary significantly over periods. They were rApp ($t(71) = -0.22$, NS), rKnow ($t(71) = -0.27$, NS), rAes ($t(71) = -1.83$, NS), rAff ($t(71) = 1.56$, NS), and rHarm ($t(71) = 0.31$, NS) suggesting their stabilities in response consistency over periods. Significant mean differences were found in rAch ($t(71) = -3.45$, $p<0.001$) and rRecog ($t(71) = 3.40$, $p<0.001$) suggesting poor test-retest reliability of two subtests.

Item name	Application					
	Mean1	SD1	Mean2	SD2	t-statistic (df=71)	p-level
APH	0.75	0.44	0.65	0.48	1.41	0.16
APM	0.78	0.42	0.72	0.45	0.81	0.42
APN	0.96	0.20	0.92	0.28	1.00	0.32
APP	0.85	0.36	0.94	0.23	-1.84	0.07
APX	0.71	0.46	0.69	0.46	0.19	0.85
APAB	0.78	0.42	0.93	0.26	-2.99	0.00
Total	4.82	1.14	4.86	1.01	-0.22	0.82
		Knowledge				
KNH	0.25	0.44	0.35	0.48	-1.41	0.16
KNI	0.94	0.23	0.94	0.23	0.00	1.00
KNK	0.85	0.36	0.89	0.32	-0.83	41
KNU	0.89	0.32	0.86	0.35	0.53	0.60
KNW	0.79	0.41	0.63	0.49	2.43	0.02
KNZ	0.74	0.44	0.83	0.38	-1.41	0.16
Total	4.46	1.05	4.50	1.01	-0.27	0.78
		Achievement				
ACHJ	0.72	0.45	0.85	0.36	-1.83	0.07
ACHQ	0.74	0.44	0.78	0.42	-0.73	0.47
ACHR	0.82	0.39	0.83	0.38	-0.23	0.82
ACHW	0.21	0.41	0.38	0.49	-2.43	0.02
ACHX	0.29	0.46	0.31	0.46	-0.19	0.85
ACHAA	0.76	0.43	0.88	0.33	-1.92	0.06
Total	3.54	0.89	4.01	1.23	-3.45	0.00

Items

Table 1 shows item wise paired t-test results. All the items of rHarm did not vary significantly over periods suggesting very high stabilities. This is also supported by the subtest wise comparisons. Relatively less high reliability is noted in 5 sub tests as rApp, rKnow, rAch, rRecog and rAes as 5 out of 6 items (83%) did not differ over periods. Poor reliability was noted in rAff as 4 out of 6 items (66%) did not significantly differ. But following the subtest wise comparison, high test-retest reliability is noted in rAff.

To conclude, results revealed disparities in determining test-retest reliability through subtest and item wise paired comparison tests. Second limitation is non-consideration of association among all items. Therefore cluster analysis is proposed.

International Journal of Psychological Research, 2010. Vol. 3. No. 1.
ISSN impresa (printed) 2011-2084
ISSN electrónica (electronic) 2011-2079

Dutta Roy, D., (2010). Cluster analysis for test-retest Reliability. *International Journal of Psychological Research, 3* (1), 131-139.

Table 2. *Euclidean Distance matrix of seven reading motivation variables*

Application motivation (rApp)												
	APH	APM	APN	APP	APX	APAB	APH2	APM2	APN2	APP2	APX2	APAB2
APH	0.00	5.48	4.58	4.80	5.00	4.90	5.00	5.29	4.47	4.69	5.10	4.80
APM	5.48	0.00	4.36	4.58	5.00	4.69	5.39	4.90	4.69	4.24	5.66	3.87
APN	4.58	4.36	0.00	3.16	4.69	4.36	5.10	4.80	3.00	2.65	4.80	2.83
APP	4.80	4.58	3.16	0.00	4.24	4.12	5.83	5.00	3.87	3.87	5.57	3.46
APX	5.00	5.00	4.69	4.24	0.00	4.58	6.00	5.74	5.00	4.80	5.20	4.24
APAB	4.90	4.69	4.36	4.12	4.58	0.00	6.08	5.29	4.69	4.00	5.10	3.87
APH2	5.00	5.39	5.10	5.83	6.00	6.08	0.00	5.74	5.20	5.00	5.39	5.29
APM2	5.29	4.90	4.80	5.00	5.74	5.29	5.74	0.00	4.24	4.69	4.90	4.80
APN2	4.47	4.69	3.00	3.87	5.00	4.69	5.20	4.24	0.00	2.83	4.69	3.32
APP2	4.69	4.24	2.65	3.87	4.80	4.00	5.00	4.69	2.83	0.00	4.90	3.00
APX2	5.10	5.66	4.80	5.57	5.20	5.10	5.39	4.90	4.69	4.90	0.00	4.80
APAB2	4.80	3.87	2.83	3.46	4.24	3.87	5.29	4.80	3.32	3.00	4.80	0.00
Lowest	4.47	3.87	2.65	3.46	4.24	3.87	5.00	4.24	2.83	3.00	4.80	0.00
Knowledge motivation (rKnow)												
	KNH	KNI	KNK	KNU	KNW	KNZ	KNH2	KNI2	KNK2	KNU2	KNW2	KNZ2
KNH	0.00	7.21	7.00	6.93	7.00	6.56	5.00	7.21	7.35	6.78	6.24	6.78
KNI	7.21	0.00	3.32	2.83	3.87	4.12	6.71	2.45	3.16	3.74	5.20	4.00
KNK	7.00	3.32	0.00	3.87	4.69	4.69	6.32	3.61	3.61	4.12	5.83	4.58
KNU	6.93	2.83	3.87	0.00	4.12	4.36	6.71	2.83	3.16	3.74	5.20	447
KNW	7.00	3.87	4.69	4.12	0.00	5.48	6.63	3.87	4.12	4.36	5.10	5.00
KNZ	6.56	4.12	4.69	4.36	5.48	0.00	5.83	4.36	4.58	5.20	6.00	500
KNH2	5.00	6.71	6.32	6.71	6.63	5.83	0.00	6.56	6.86	6.71	5.66	6.86
KNI2	7.21	2.45	3.61	2.83	3.87	4.36	6.56	0.00	3.16	3.46	5.00	3.74
KNK2	7.35	3.16	3.61	3.16	4.12	4.58	6.86	3.16	0.00	3.46	5.57	4.24
KNU2	6.78	3.74	4.12	3.74	4.36	5.20	6.71	3.46	3.46	0.00	5.20	4.24
KNW2	6.24	5.20	5.83	5.20	5.10	6.00	5.66	5.00	5.57	5.20	0.00	5.74
KNZ2	6.78	4.00	4.58	4.47	5.00	5.00	6.86	3.74	4.24	4.24	5.74	0.00
Lowest	5.00	2.45	3.61	2.83	3.87	4.36	5.66	3.16	3.46	4.24	5.74	0.00
Affiliation motivation (rAff)												
	AFI	AFT	AFV	AFY	AFAA	AFAB	AFI2	AFT2	AFV2	AFY2	AFAA2	AFAB2
AFI	0.00	5.92	7.55	4.00	4.36	4.24	2.45	5.83	7.42	4.90	3.46	3.00
AFT	5.92	0.00	5.83	6.08	6.16	5.74	5.92	5.20	6.00	6.08	5.92	6.16
AFV	7.55	5.83	0.00	7.55	6.93	7.00	7.55	5.74	4.47	6.86	7.42	7.35
AFY	4.00	6.08	7.55	0.00	5.20	5.10	3.74	5.83	7.28	4.24	4.00	4.36
AFAA	4.36	6.16	6.93	5.20	0.00	4.36	4.12	5.92	6.93	5.39	4.36	4.47
AFAB	4.24	5.74	7.00	5.10	4.36	0.00	4.47	5.83	6.86	5.29	4.47	3.87
AFI2	2.45	5.92	7.55	3.74	4.12	4.47	0.00	5.48	7.55	4.47	2.45	3.00

AFT2	5.83	5.20	5.74	5.83	5.92	5.83	5.48	0.00	6.24	5.10	5.29	5.92
AFV2	7.42	6.00	4.47	7.28	6.93	6.86	7.55	6.24	0.00	7.14	7.42	7.35
AFY2	4.90	6.08	6.86	4.24	5.39	5.29	4.47	5.10	7.14	0.00	4.69	4.80
AFAA2	3.46	5.92	7.42	4.00	4.36	4.47	2.45	5.29	7.42	4.69	0.00	3.32
AFAB2	3.00	6.16	7.35	4.36	4.47	3.87	3.00	5.92	7.35	4.80	3.32	0.00
Lowest	2.45	5.20	4.47	3.74	4.12	3.87	2.45	5.10	7.14	4.69	3.32	0.00

Recognition motivation (rReco

	RCJ	RCL	RCP	RCS	RCY	RCZ	RCJ2	RCL2	RCP2	RCS2	RCY2	RCZ2
RCJ	0.00	7.42	4.58	6.08	7.07	5.57	5.00	6.93	4.69	5.29	6.48	5.29
RCL	7.42	0.00	7.48	5.48	4.12	6.93	7.21	4.58	7.42	6.86	5.20	7.00
RCP	4.58	7.48	0.00	6.32	7.28	4.47	3.74	7.14	3.87	4.80	6.71	4.58
RCS	6.08	5.48	6.32	0.00	6.08	6.00	6.48	6.08	6.08	5.92	5.74	6.08
RCY	7.07	4.12	7.28	6.08	0.00	7.00	7.14	4.69	7.48	6.78	4.24	6.93
RCZ	5.57	6.93	4.47	6.00	7.00	0.00	4.00	7.00	4.36	5.00	6.56	5.00
RCJ2	5.00	7.21	3.74	6.48	7.14	4.00	0.00	7.14	3.00	4.36	6.56	4.80
RCL2	6.93	4.58	7.14	6.08	4.69	7.00	7.14	0.00	7.48	6.32	4.90	7.07
RCP2	4.69	7.42	3.87	6.08	7.48	4.36	3.00	7.48	0.00	4.47	7.07	4.00
RCS2	5.29	6.86	4.80	5.92	6.78	5.00	4.36	6.32	4.47	0.00	6.32	5.29
RCY2	6.48	5.20	6.71	5.74	4.24	6.56	6.56	4.90	7.07	6.32	0.00	7.21
RCZ2	5.29	7.00	4.58	6.08	6.93	5.00	4.80	7.07	4.00	5.29	7.21	0.00
Lowest	4.58	4.12	3.74	5.74	4.24	4.00	3.00	4.90	4.00	5.29	7.21	0.00

Aesthetic motivation (rAes)

	AEK	AEM	AEO	AEQ	AES	AET	AEK2	AEM2	AEO2	AEQ2	AES2	AET2
AEK	0.00	5.00	7.07	4.69	5.48	6.48	3.61	4.58	7.42	4.58	7.00	6.08
AEM	5.00	0.00	7.14	5.00	6.24	5.74	4.00	4.90	7.21	5.29	6.63	6.00
AEO	7.07	7.14	0.00	7.21	5.83	6.00	7.28	6.71	3.32	6.86	5.39	5.74
AEQ	4.69	5.00	7.21	0.00	5.83	6.00	4.58	5.20	7.28	4.12	6.40	5.92
AES	5.48	6.24	5.83	5.83	0.00	6.63	5.39	5.39	5.74	5.92	5.92	5.92
AET	6.48	5.74	6.00	6.00	6.63	0.00	6.24	6.86	5.92	6.24	6.24	5.20
AEK2	3.61	4.00	7.28	4.58	5.39	6.24	0.00	4.00	7.62	4.24	6.78	5.83
AEM2	4.58	4.90	6.71	5.20	5.39	6.86	4.00	0.00	7.21	5.10	6.48	6.48
AEO2	7.42	7.21	3.32	7.28	5.74	5.92	7.62	7.21	0.00	7.35	4.69	6.00
AEQ2	4.58	5.29	6.86	4.12	5.92	6.24	4.24	5.10	7.35	0.00	6.48	6.16
AES2	7.00	6.63	5.39	6.40	5.92	6.24	6.78	6.48	4.69	6.48	0.00	6.32
AET2	6.08	6.00	5.74	5.92	5.92	5.20	5.83	6.48	6.00	6.16	6.32	0.00
Lowest	3.61	4.00	3.32	4.12	5.39	5.20	4.00	5.10	4.69	6.16	6.32	0.00

Harm avoidance motivation (rHarm)

	HAL	HAN	HAO	HAR	HAU	HAV	HAL2	HAN2	HAO2	HAR2	HAU2	HAV2
HAL	0.00	4.00	4.90	3.74	3.87	4.24	4.58	4.12	4.36	4.36	4.12	4.90
HAN	4.00	0.00	4.00	3.16	3.00	4.00	3.32	3.00	3.32	3.87	3.61	4.00
HAO	4.90	4.00	0.00	4.24	3.87	4.47	4.58	4.12	3.32	4.36	4.36	4.69

HAR	3.74	3.16	4.24	0.00	3.00	4.47	4.58	3.61	4.12	4.36	4.36	4.69
HAU	3.87	3.00	3.87	3.00	0.00	4.12	4.47	3.16	3.16	3.74	3.74	4.36
HAV	4.24	4.00	4.47	4.47	4.12	0.00	4.80	4.12	4.12	4.58	4.36	4.47
HAL2	4.58	3.32	4.58	4.58	4.47	4.80	0.00	4.00	4.00	4.69	4.24	4.36
HAN2	4.12	3.00	4.12	3.61	3.16	4.12	4.00	0.00	3.16	2.83	3.16	3.32
HAO2	4.36	3.32	3.32	4.12	3.16	4.12	4.00	3.16	0.00	3.74	3.46	4.12
HAR2	4.36	3.87	4.36	4.36	3.74	4.58	4.69	2.83	3.74	0.00	3.16	3.61
HAU2	4.12	3.61	4.36	4.36	3.74	4.36	4.24	3.16	3.46	3.16	0.00	3.87
HAV2	4.90	4.00	4.69	4.69	4.36	4.47	4.36	3.32	4.12	3.61	3.87	0.00
Lowest	3.74	3.00	3.32	3.00	3.16	4.12	4.00	2.83	3.46	3.16	3.87	0.00

Achievement motivation (rAch)

	ACHJ	ACHQ	ACHR	ACHW	ACHX	ACHAA	ACHJ2	ACHQ2	ACHR2	ACHW2	ACHX2	ACHAA2
ACHJ	0.00	5.20	4.80	6.56	6.86	5.39	5.00	5.10	4.69	6.08	6.48	4.58
ACHQ	5.20	0.00	4.47	7.07	6.93	5.29	4.90	4.12	4.80	6.00	6.40	4.69
ACHR	4.80	4.47	0.00	7.35	7.07	4.69	4.00	4.58	4.36	6.48	6.86	4.24
ACHW	6.56	7.07	7.35	0.00	5.48	6.93	7.21	6.86	7.00	5.10	5.39	7.07
ACHX	6.86	6.93	7.07	5.48	0.00	7.21	7.07	6.56	7.00	5.48	5.20	6.78
ACHAA	5.39	5.29	4.69	6.93	7.21	0.00	3.46	4.80	4.36	6.00	6.71	4.24
ACHJ2	5.00	4.90	4.00	7.21	7.07	3.46	0.00	4.12	3.61	6.48	7.00	3.46
ACHQ2	5.10	4.12	4.58	6.86	6.56	4.80	4.12	0.00	4.00	5.92	6.48	4.12
ACHR2	4.69	4.80	4.36	7.00	7.00	4.36	3.61	4.00	0.00	6.40	6.93	3.87
ACHW2	6.08	6.00	6.48	5.10	5.48	6.00	6.48	5.92	6.40	0.00	5.00	6.63
ACHX2	6.48	6.40	6.86	5.39	5.20	6.71	7.00	6.48	6.93	5.00	0.00	7.14
ACHAA2	4.58	4.69	4.24	7.07	6.78	4.24	3.46	4.12	3.87	6.63	7.14	0.00
Lowest	4.58	4.12	4.00	5.10	5.20	3.46	3.46	4.00	3.87	5.00	7.14	0.00

CLUSTERING APPROACH

Clustering approach follows principles of set formation. Set is formed with items having high within group homogeneity. The more within group homogeneity among the items, the less number of sub clusters or subsets are in a set. Sub clusters can be understood by analysis of dendogram (graphical plot of clusters). In this study, seven dendograms for seven subtests of reading motivation were prepared following complete- linkage method. Complete linkage method is a hierarchical clustering algorithm in which inter-object similarity is based on the maximum distance between objects in two clusters (the distance between the most dissimilar members of each cluster). At each stage of agglomeration, the two clusters with the smallest maximum distance (most similar) are combined. This technique eliminates the chaining problem identified with single-linkage and has been found to generate the most compact clustering solutions (Baeza-Yates, 1992). Even though it represents only one aspect of data (the farthest distance between numbers), many researchers find it the most appropriate for a wide range of clustering applications. (Jain & Dubes, 1988). As items of the questionnaire for each measure are correlated with each other, it is assumed that both test and retest items will be in one cluster.

Euclidean distance matrix

Euclidean distance measures distance between two points. Clustering starts with pair wise distances. Current study used Euclidean distance as it is assumed that all the item wise data are geometrically located in the psychological map of respondents. Low distance between items of test and retest periods indicates homogeneity between items or close neighbors in the psychological map. Table 2 shows Euclidean distance matrix for 7 subtest measures. This represents extent of distance within and between test and retest items. Suffix 2 in the dendograms is used to indicate retest items.

International Journal of Psychological Research, 2010. Vol. 3. No. 1.
ISSN impresa (printed) 2011-2084
ISSN electrónica (electronic) 2011-2079

Between items

Table 3 shows minimum Euclidian distance between test and retest items in 7 subtests. Namely rApp, rKnow, rAch, rAes, rRecog, rAff and rHarm. Among them, least distance is noted in rKnow (Dist=2.45) and in rAff (Dist=2.45). Again, only those two subtests possess same test and retest items. This suggests that both items of rKnow and rAff carried similar responses though items of respective subtests were presented across different periods with 8 months intervals.

Table 3. *Between items homogeneity*

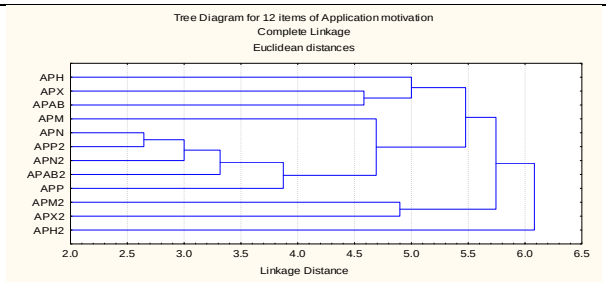
Variable	Items	Euclidian distance
rApp	APN, APP2	2.65
rKnow	KNI, KNI2	2.45
rAch	ACHAA, ACHJ2	3.46
rAes	AEO, AEO2	3.32
rRecog	RCJ2, RCP2	3.00
rAff	AFI, AFI2	2.45
	AFI2, AFAA2	2.45
rHarm	HAN2, HAR2	2.83

Note: 2 after the item name is used to indicate items presented at the retest period.

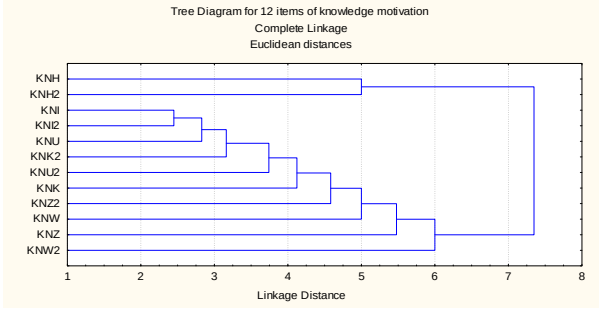
Sub divisions of Primary cluster

The cluster wherein dendogram centroid lies is called primary cluster. Primary cluster includes more number of items when test-retest reliability among item responses is high. Sub divisions in primary cluster are noted when primary cluster is divided into more sub clusters. Close watch to 7 dendograms

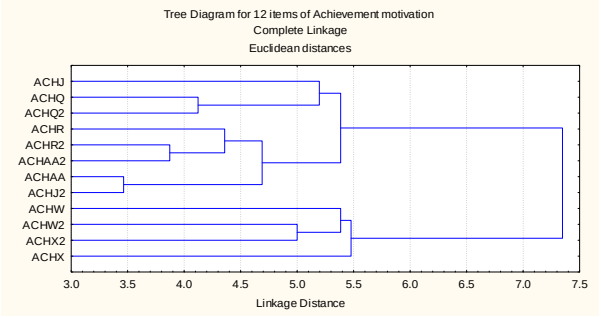
Figure 1. *Dendograms of 7 Reading motivation variables*



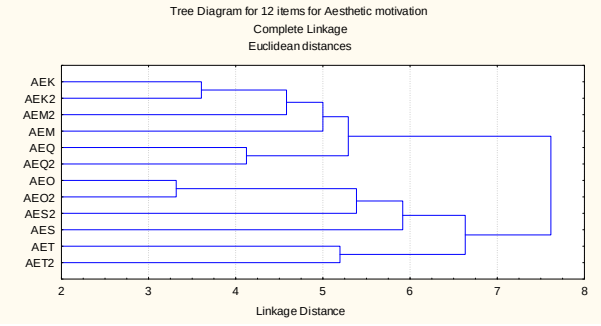
Application motivation



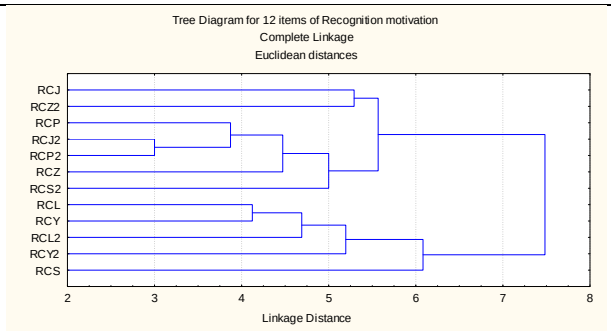
Knowledge motivation



Achievement motivation

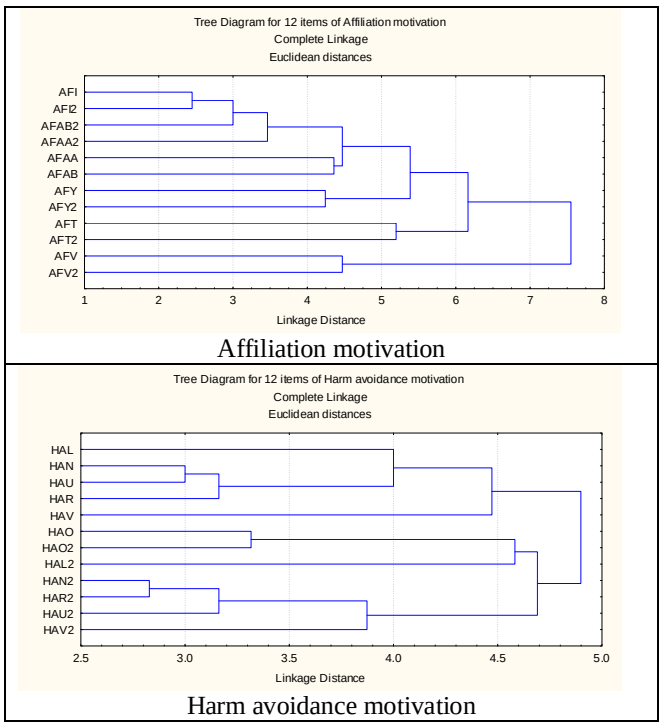


Aesthetic motivation



Recognition motivation

International Journal of Psychological Research, 2010. Vol. 3. No. 1.
ISSN impresa (printed) 2011-2084
ISSN electrónica (electronic) 2011-2079

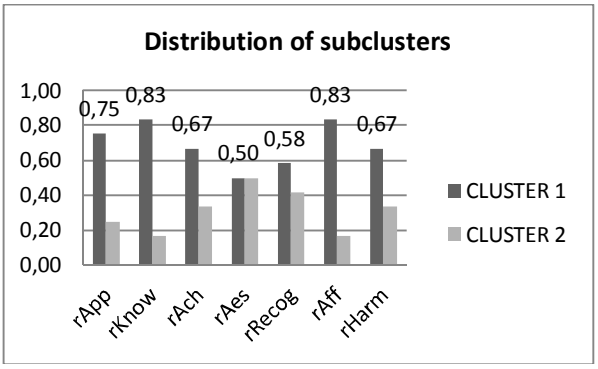


(Figure 1) reveals that there is only 1 division in primary cluster for rKnow and rAff variables suggesting high homogeneity in item responses vis-à-vis high test-retest reliability in both variables.

Primary and secondary clusters

When test-retest reliability is high, there are more items in primary in comparison with secondary clusters. Constellation of more items in one cluster indicates that both test and retest items are in the same cluster. Figure 2 presents constellation of items in both primary and secondary clusters.

Figure 2. *Distribution of members in both sub clusters. Cluster 1: Primary cluster, Cluster2: Secondary cluster*



More than 55% of items are in the primary clusters of rKnow (83%), rAff (83%), rAch (67%), rHarm (67%) and rRecog (58%). This suggests that except rAes (50%), all the subtests possessed high test retest reliability. Length of test-retest interval is an important factor affecting confidence in the stability assumption (Furr & Bacharach, 2008). Based on the results it can be concluded that reading motivation is relatively stable psychological construct in terms of time consistency.

DISCUSSION

Reliability of questionnaire in terms of time consistency was examined in this study using both non-clustering and clustering techniques. For non-clustering, paired t-test and for clustering, hierarchical cluster analysis were used.

In non-clustering technique, well recognized paired t-test was used in both subtests and item wise comparisons of test and retest measures. Following subtest wise comparison, low test-retest reliability was found in two subtests namely, rAch and rRecog as test ï retest measures of both significantly differed in both periods. But in item wise comparison, it is noted that only 1 item out of 6 for subtest rAch significantly differed suggesting relatively high test retest reliability. In item wise comparison, rAff variable is relatively poor in estimating test retest reliability as 2 out of 6 items differed significantly between periods. Subtests scores are actually total of item weightages. Based on the above findings, loss of item information is suspected during summing the item weightage.

On the contrary, clustering technique identifies 2 subtests (rKnow and rAff) having high test retest reliability. Clustering technique surpasses limitation of paired t-test technique as it considers only similar measures (matched subtest scores or item weightage) for purpose of comparison. But clustering technique considers set of items measuring same variable. Clustering technique clusters items in same group if the items are homogenous in measurement. Besides, it identifies the outlier with which investigator can think of later for any modification.

Results noted different parameters of dendogram to understand test-retest reliability of the questionnaire. They are between items homogeneity, constellations of items in primary and secondary clusters.

To sum up, the study shows importance of hierarchical clustering in determining test-retest reliability of questionnaire. This is also evident from the results that reading motivation questionnaire possesses good test-retest reliability.

International Journal of Psychological Research, 2010. Vol. 3. No. 1.
ISSN impresa (printed) 2011-2084
ISSN electrónica (electronic) 2011-2079

Dutta Roy, D., (2010). Cluster analysis for test-retest Reliability. *International Journal of Psychological Research, 3* (1), 131-139.

REFERENCES

Baeza-Yates,R.A. (1992). Introduction to data structures and algorithms related to information retrieval. In W.B. Frakes and R. Baeza-Yates (eds.), *Information retrieval: Data structures and algorithms* (pp. 13-27). Upper Saddle River, N.J: Prentice Hall.

Chen, E. E., & Small, S. L. (2007). Testï retest reliability in fMRI of language: Group and task effects. *Brain and Language, 102* (2), 176-185.

Duquette, J., McKinley, P.A., & Litwoski, J. (2005). Test-retest reliability and internal consistency of the Quebec-French version of the survey of pain attitudes. *Archives of Physical Medicine and Rehabilitation, 86* (4),782-788.

Dutta Roy, D. (2003). Cluster analysis of GHQ-12 items using Indian Antarctica expeditionersô responses. *Journal of Psychometry, 17* (1&2), 38-44.

Dutta Roy, D. (2003). *Development Of The Questionnaire For Assessment of Reading And Writing Motivation Of Boys And Girls Of Grades III And IV*. Unpublished project report submitted to the Indian Statistical Institute, Kolkata.

Dutta Roy, D., & Deb, N. C. (1999). Item-total score correlations of state anxiety inventory across different months in Antarctic expedition. *Psychological Studies, 44* (1&2), 43-45.

Furr, M. R., & Bacharach, V.R. (2008). *Psychometrics: An introduction. Singapore.* Sage publications.

Hair, J. F. Jr., Black,W. C., Bubin, B.J., Anderson, R. E., & Tatham, R.L. (2006). *Multivariate data analysis (6th Ed.).* New Delhi: Pearson Prentice Hall.

Hamashima, C., & Yoshida, K. (2002). Test-retest reliability of Japanese EuroQol instrument. *Paper presented at the Annual meeting of the International society of technology assessment in health care,* 18, Abstract no. 128.

Jain, A. K., & Dubes , R. C. (1988). *Algorithms for clustering data.* N. J: Prentice Hall.

Walker, S., & Cosden, M. (2007). Reliability of college student self-reported drinking behavior. *Journal of Substance Abuse Treatment, 33* (4),405-409.