Social Psychology Quarterly
1983, Vol. 46, No. 2, 163–168

An Experimental Comparison of Clustered and Scattered Scale Items

TOM W. SMITH
National Opinion Research Center

For at least the last 30 years social scientists have argued whether items in a scale should be clustered together or scattered apart. Psychologists designing personality inventories and social psychological scales have usually recommended the randomization of items. The Minnesota Multiphasic Personality Inventory for example contains 550 items made up of 9 psychiatric discriminant keys and four control or validity keys. Items are randomized throughout the test. The main purposes of randomization on these types of scales are to (1) help disguise the purpose of the scales and thus reduce intentional faking; (2) spread each scale across the test so transitory conditions (e.g., moods) and test reactions (e.g., fatigue) will fall equally on all scales; and (3) "eliminate" the possibility of order effects between scales by randomizing the constituent items.[1] Political scientists and sociologists on the other hand almost always cluster items together on their much shorter scales. This is usually done because of the face comprehensibility of presenting similarly formated and related items together and because complete randomization of items covering differing topics and using various formats does not seem practical.[2]

Despite these clear differences in ordering procedures between disciplines, general texts on survey methodology and test/questionnaire construction either do not address the question (Kidder, 1981; Nachmias and Nachmias, 1981) or present brief, general comments about the possibility of increasing response sets by clustering (Bailey, 1982) or the impracticality of randomization (Babbie, 1973). There is, however, a small literature of experimental work that has addressed the issue of clustering vs. scattering scale items.

Baehr (1953) gave a questionnaire covering 10 employee attitude scales to 454 retail store workers. A comparison of three different clustered vs. scattered experiments revealed no significant differences in the scale scores. Metzner and Mann (1953) administered paper-and-pencil tests to 844 white collar employees. The questionnaire covered attitudes toward their work. Split-halves were given either a sequenced form with scale items clustered together under topical labels or unsequenced forms with the scale items scattered. Significant differences in inter-item associations appeared between the forms but no clear pattern or direction to the differences was evident. Kirchner and Uphoff (1955) asked two union groups totaling approximately 100 members about their feelings towards various union groups. One form separately grouped six subscales while the other interspersed all items. Only one of 12 means was significantly different and no pattern was discerned. Inter-item associations were not inspected. Hayes (1964) conducted several experiments with college students from two universities. On Guttman scale items of mathematical ability and social anxiety he found no differences between dispersed and clustered forms. He probably examined both distributions and scalability, but details are not presented. Martin (1980) using the 1971 Detroit Area Study compared split-halves that either grouped four anomia items together or separated them with nonscale items. She found no reliable differences in either distributions or associations. Finally, Schuman and Presser (1981) reviewing the literature above, and one additional experiment involving the separation of two anomia items by an extraneous question found no evidence of clustering effects on either marginals or associations.

This research was done for the General Social Survey Project directed by James A. Davis and Tom W. Smith. The project is supported by the National Science Foundation, Grant No. SOC77-03279. GSS Technical Report No. 36.

[1] For a general discussion of psychological procedures see Cronbach, 1960; Nunnally, 1970; and Brown and Hernstein, 1975. Examples of many personality and social psychological scales can be found in Robinson and Shaver, 1973. In particular see Stagner's Facist Attitude scale (Robinson and Shaver, 1973, p. 363) where the purpose of the scale is hidden by (a) indirect questions, and (b) 15 unrelated filler questions among the 35 scale items.

[2] Examples of the prevailing practice in political science and sociology can be found in the various codebooks of the American National Election Studies of the University of Michigan and the cumulative codebook of the General Social Survey conducted by the National Opinion Research Center.

Table 1. Anomia by Form (Clustered *vs.* Scattered)

1A. Distributions

	Clustered	*Scattered*
A. Getting Worse (ANOMIA5)		
Agree	67.1%	63.0%
Don't Know	4.0	5.0
Disagree	28.8	32.0
	(742)	(757)
$\chi^2 = 2.96$, $p = .227$		
$\chi^2 = 1.98$, $p = .160$		
B. Not Fair to Have Child (ANOMIA6)		
Agree	41.0%	27.3%
Don't Know	3.7	3.3
Disagree	55.3	69.3
	(739)	(747)
$\chi^2 = 32.2$, $p = .0000$		
$\chi^2 = 31.5$, $p = .0000$		
C. Public Officials not Interested (ANOMIA7)		
Agree	68.0%	64.0%
Don't Know	2.3	4.5
Disagree	29.7	31.5
	(738)	(753)
$\chi^2 = 6.6$, $p = .036$		
$\chi^2 = 1.0$, $p = .318$		

1B. Inter-item Associations (gamma)

	Clustered	*Scattered*	*Clustered–Scattered*
ANOMIA5 × ANOMIA6	.578	.304	+.274*
ANOMIA5 × ANOMIA7	.478	.396	+.082
ANOMIA6 × ANOMIA7	.620	.474	+.146

Note: "Don't Knows" excluded from analysis.
* $p < .05$.

In contrast to the uniform negative findings of the specific clustering-scattering literature, other work on order effects in general (Schuman and Presser, 1981; Smith, unpubl.) and order effects within scale items (Hayes, 1964; Schuman and Presser, 1981) suggests, however, that order effects are not rare and usually occur between substantively related items. These findings suggest that order effects between scattered and clustered might be expected.

To examine the effects of clustering vs. scattering, two split ballots were employed on the 1982 General Social Survey. The 1982 GSS is a full probability survey of the adult, noninstitutionalized population of the conterminous United States. (For details see Davis and Smith, 1982.) On one form three items from the Srole anomia scale were clustered together.[3] On the other form the items were widely dispersed, separated by at least 26 intervening questions. Table 1A shows a substantial marginal shift for one of the three items. (For the first item we would not expect any shift related to clustering since in neither the clustered nor scattered condition is it constrained by prior scale items.) This effect is conditional on a respondent's answer to the first anomia question (Smith, unpubl.). Those who agreed that the average man is "less well-off" increased their agreement to the child item by 19.3 percentage points when this question immediately followed the "getting worse" item, while those who rejected the first statement showed no significant increase in acceptance of the child item (2.0 percentage point increase). Table 1B indicates that each inter-item correlation increases in the cluster condition with associations involving the marginal-sensitive child item showing greater increases. In brief, on the child question in particular and perhaps on the public officials question, it appears that clustering constrains response patterns and increases inter-item associations.

Since this finding is counter to the most relevant literature we tried to check it by looking at certain nonexperimental comparisons on the

[3] The agree-disagree items were: (1) "In spite of what some people say, the lot (situation/condition) of the average man is getting worse, not better," (2) "It's hardly fair to bring a child into the world with the way things look for the future," (3) "Most public officials (people in public office) are not really interested in the problems of the average man." For details on the Srole anomia scale, see Srole (1956).

Table 2. Anomia by Clustered and Semi-Clustered Years

2A. Distributions

	Clustered (1977, 1980, 1982x)	*Semi-Clustered (1973, 1974, 1976)*
A. Getting Worse (ANOMIA5)		
Agree	63.7%	58.7%
	(3522)	(4233)
$\chi^2 = 20.2$, $p = <.0001$		
B. Not Fair to Have Child (ANOMIA6)		
Agree	43.1%	39.0%
	(3522)	(4233)
$\chi^2 = 13.4$, $p = <.0005$		
C. Public Officials not Interested (ANOMIA5)		
Agree	68.9%	63.7%
	(3514)	(4256)
$\chi^2 = 23.0$, $p = <.0001$		

2B. Inter-item Associations (gamma)

	Clustered	*Semi-Clustered*	*Clustered–Semi-Clustered*
ANOMIA5 × ANOMIA6	.648	.777	−.129*
ANOMIA5 ×ANOMIA7	.521	.462	.059
ANOMIA6 × ANOMIA7	.620	.478	.144*

Note: ANOMIA5 always comes immediately before ANOMIA6. In 1973, 1974, and 1976 ANOMIA7 appeared in a separate grouping. This is the semi-clustered condition. In 1977, 1980, and 1982x ANOMIA7 appeared immediately after ANOMIA6. This is the clustered condition.

* $p < .001$.

GSS's that approximated the scattering and clustering experiment. Two examples were found, one using the same three anomia items and a second using a misanthropy scale. While conclusions are hampered by the nonexperimental nature of these comparisons, both lend support to the conclusions noted above. Table 2B compares GSS surveys on which the three anomia items were clustered (as in the experiment above) with years on which the public officials item was separated. There is unfortunately no control for time since the clustered surveys all come after the semi-clustered surveys. Since the significant marginal shifts are all in the same direction and affect both the consistently ordered items and the switched item, we doubt if any of these changes are related to clustering. The increase in associations for cross tabulations involving the public officials item may be a result of clustering however, and closely fits the pattern noted above in the experimental situation.

Table 3 compares GSS surveys using a semi-clustered and scattered versions of Rosenberg's misanthropy scale.[4] In 1973, 1975, and 1978 the "fair item" immediately followed the "helpful item" while in 1972, 1976, and 1980 the items were separated by 8 to 19 questions. The "trust item" was always separate from the other two. As in the other nonexperimental comparison, the marginal differences observed do not clearly relate to clustering. (The one large change occurs on the "helpful item" which does not follow other scale items on either group of surveys.) The associations on the other hand, seem to show a clustering effect with the association between fairness and helpfulness increasing in the semi-clustered condition. It thus appears that clustering frequently can increase the inter-item association between scale items and that marginal effects may occur although their regularity is less certain.

Given the positive results to our experimental and nonexperimental comparisons of scattered and clustered scale item and the negative results reported in the literature, we might wonder why the discrepancy occurs. One factor that may have repressed the effect in some earlier studies is the homogeneity of the topics and populations used. In the Baehr (1953), Metzner and Mann (1953), and Kirchner and Uphoff (1955) research the various scale items all dealt with a single topic (either work conditions or union affairs). While the

[4] The items are (1) "Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?" (2) "Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?" (3) "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?" For information on this scale see Robinson and Shaver (1973).

Table 3. Misanthropy by Semi-Clustered *vs.* Scattered Years

3A. Distributions

	Semi-Clustered (1973, 1975, 1978)	*Scattered (1972, 1976, 1980)*
A. Are People Fair? (FAIR)		
No	32.8%	35.0%
Depends	5.7	4.8
Yes	61.5	60.2
	(4469)	(4498)
$\chi^2 = 7.4$, $p = .025$ (Depends included)		
B. Are People Helpful? (HELPFUL)		
Yes	54.4%	46.5%
Depends	5.2	5.6
No	40.4	47.9
	(4469)	(4498)
$\chi^2 = 20.2$, $p = <.0001$ (Depends included)		
C. Are People Trusting? (TRUST)		
Yes	41.7%	45.5%
Depends	3.8	3.6
No	54.6	50.9
	(4482)	(4514)
$\chi^2 = 20.2$, $p = .002$ (Depends included)		

3B. Inter-item Associations (gamma)

	Clustered	*Scattered*	*Clustered–Scattered*
TRUST × HELPFUL	.603	.658	−.055*
TRUST × FAIR	−.690	−.687	.003
FAIR × HELPFUL	−.729	−.628	.106**

Note: In 1973, 1975, and 1978 HELPFUL immediately preceded FAIR with TRUST coming 20 questions before or after the clustered items. We class these years as semi-clustered. In 1972, 1976, and 1980 all items were separated by about 10 to 20 items.

* $p < .05$.

** $p < .001$.

subscales are more closely related, respondents focus their thoughts on the same topic in both forms. In addition, the experimental groups are expressing attitudes on situations of personal familiarity about which attitudes may be more reliable and integrated. In the case of the Hayes (1964), Martin (1980), and Schuman and Presser (1981) work, the lack of results may be because the scattered forms only separated the items by one to four questions. Recent research (Schuman, et al., forthcoming; Bishop, et al. unpubl.) indicates that context effects can occur even when there are numerous intervening questions. Thus in these experiments items may not be scattered enough to erase a clustering effect. While more experimentation will be needed to establish the generalizability of cluster effects, it appears that in surveys of a heterogeneous nature clustered scale items are more highly correlated than widely scattered scale items.

Given that clustering does increase the inter-item associations, is this good or bad? Proponents of one school of thought see the increase as reflecting more measurement error, artificially increasing associations through proximity effects (Webb, et al., 1966; Mason et al., 1981). Proponents of another perspective see this clustering as enhancing the reliability of the scales. By focusing respondents' attention on the domain being measured, more random noise is eliminated and more thoughtful and consistent results are obtained.[5] We subjected these conflicting hypotheses to a trial by construct validity, but came up with a hung jury. Using nine variables that have been found to be the strongest predictors of anomia (education, personal health, financial satisfaction, relative financial standing, church attendance, evaluation of life as exciting or dull, race, and voting in last two presidential elections—see Dodder and Astle, 1980; Hong, 1981; and Leonard, unpubl.), we compared the associations between these predictors and each anomia item individually and as a three-item additive anomia scale. On the 27 comparisons

[5] It has been demonstrated that more thoughtful and focused approaches result in more accurate results of factual and behavioral information (e.g., Belson, 1962; Cannell, et al., 1981; and Sparks, et al., 1977).

with individual items, 15 were higher when scattered and 12 when clustered. Similarly on the scale the split was five higher when scattered, four higher when clustered. Several individual differences were significant but no pattern was discernable. While more research will hopefully clarify this picture, it appears that scattering and clustering have no certain impact on the association between scale items and their predictors. Neither the pattern of association with predictor variables nor arguments about the impact of clustering vs. scattering demonstrate clear superiority.

Our findings suggest that clustering effects may be more common than the literature previously suggested. Distributions may be changed and associations are usually increased when clustering occurs. No clear impact occurs on associations between the scale items and other variables. Thus, when one is using a typical political or sociological scale in a model there is no clear preference for clustered vs. scattered items. Further research should clarify whether this applies generally. Replication does favor clustering, however. Since it will usually be impossible to closely duplicate the entire context of the scattered item, across-time and across-study comparability will be enhanced when the items are clustered and the context of that cluster is repeated. Clustering and scattering are not identical procedures, but since no clear superiority has been demonstrated, both approaches appear suitable in test/questionnaire construction.

REFERENCES

Babbie, E. R.
1973 Survey Research Methods. Belmont, CA: Wadsworth Publishing.

Baehr, M. E.
1953 "A simplified procedure for the measurement of employee attitudes." Journal of Applied Psychology 37:163–167.

Bailey, K. E.
1982 Methods of Social Research. New York: Free Press.

Belson, W. A.
1962 Studies in Readership. London: Business Publications.

Bishop, G. F., R. W. Oldendick, and A. J. Tuchfarber.
unpubl. "What Must My Interest in Politics Be If I Just Told You I Don't Know?" Paper presented to the American Association for Public Opinion Research, Hunt Valley, Maryland, May 1982.

Brown, R., and R. J. Hernstein
1975 Psychology. Boston: Little, Brown, and Co.

Cannell, C. F., P. V. Miller, and L. Oksenberg
1981 "Research on Interviewing Techniques." Sociological Methodology, 1981. Samuel Leinhardt (ed.), San Francisco: Jossey-Bass.

Cronbach, L. J.
1960 Essentials of Psychological Testing. New York: Harper and Row.

Davis, J. A., and T. W. Smith
1982 General Social Surveys, 1972–1982: Cumulative Codebook. Chicago: NORC.

Dodder, R. A., and D. J. Astle
1980 "A methodological analysis of Srole's nine-item anomia scale." Multivariate Behavioral Research 15:329–334.

Hayes, D. P.
1964 "Item order and Guttman scales." American Journal of Sociology 70:51–58.

Hong, L. K.
1981 "Anomia and religiosity: Some evidence for reconceptualization." Review of Religious Research 22:233–244.

Kidder, L. H.
1981 Selltiz, Wrightman, and Cook's Research Methods in Social Relations. New York: Holt, Rinehart, and Winston.

Kirchner, W. K., and W. H. Uphoff
1955 "The effect of grouping scale items in union attitude measurement." Journal of Applied Psychology 39:182–183.

Leonard, W. M., II
unpubl. "Sociological and Social Psychological Correlates of Anomia Among a Random Sample of Aged." Paper presented to the American Sociological Association, New York, August.

Martin, E.
1980 "The effects of item contiguity and probing on measures of anomia." Social Psychology Quarterly 43:116–120.

Mason, W. H., J. S. House, and S. S. Martin
1981 "Dimensions of political alienation in America: Theoretical and empirical." Population Studies Center, University of Michigan, February.

Metzner, H., and F. Mann
1953 "Effects of grouping related questions in questionnaires." Public Opinion Quarterly 17:136–141.

Nachmias, C., and D. Nachmias
1981 Research Methods in the Social Sciences. New York: St. Martin's Press.

Nunnally, J. C., Jr.
1970 Introduction to Psychological Measurement. New York: McGraw-Hill.

Robinson, J. P., and P. R. Shaver
1973 Measures of Social Psychological Attitudes. Ann Arbor: Institute for Social Research.

Schuman, H., and S. Presser
1981 Questions and Answers: Experiments on Question Form, Wording, and Context. New York: Academic Press.

Schuman, H., G. Kalton, and J. Ludwig
forthcoming "Context and Contiguity in Survey Questionnaires." Public Opinion Quarterly.

Smith, T. W.
unpubl. "Conditional Order Effects." Paper presented to the World Association for Public Opinion Research, Hunt Valley, Maryland, May 1982.

Sparks, R. F., H. G. Genn, and D. J. Dodd
1977 Surveying Victims: A Study of the Measurement of Criminal Victimization, Perceptions of Crime and Attitudes to Criminal Justice. Chichester, John Wiley and Sons.

Srole, L.
1956 "Social integration and certain corollaries: An exploratory study." American Sociological Review 21:709–716.

Webb, E. J., D. T. Campbell, R. D. Schwartz, and L. Sechrest
1966 Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago: Rand McNally.