Reprinted from
JOURNAL OF PERSONALITY ASSESSMENT
Vol. 39 1975 No. 1

A Multiple-Choice Version of the Sentence Completion Method

RON SHOUVAL, EDMUND DUEK, and AVITAL GINTON
Tel Aviv University

Summary: A sentence completion (SC) measure of children's autonomy was adapted to a multiple choice (MC) form. Criteria were proposed which MC should meet in order to be reasonably equivalent to SC. MC should not be obvious to the child; thus, the means and variances of the two tests should be similar, and MC should not show a congregation around the "good" answers, nor should it correlate positively with a measure of social desirability. Both tests should correlate significantly with one another. Any factor structure present in SC, which suggests a differentiation within the concept of autonomy should be duplicated in MC. Most of these criteria were met, and it was concluded that a multiple choice form corresponding to a sentence completion measure, testing clearly defined personality areas, could be a reasonable alternative for many purposes.

The value of the sentence completion method in personality assessment has been stressed in recent surveys. Goldberg (1965), summarizing considerable research literature, says that the method compares favorably to other instruments. Murstein (1965), in assessing the data accumulated by Goldberg, states the method to be generally valid, and that it is probably the most valid of all the projective techniques reported in the literature. Sechrest (1968) is less encouraging, but adds that in the "rather dismal context" (p. 603) of demonstrated validity of other similar measures, such as the Rorschach and the TAT, and coupled with the economy of the sentence completion method, the relative validity of the method becomes impressive.

No agreement exists among researchers as to the typological classification of the technique (Goldberg, 1965). Rohde (1946) sees it as a projective device, and Campbell (1957) uses the term but qualifies it to relate to matters of which the respondent is aware. Forer (1950) sees it as a controlled projective test, and Hanfmann and Getzels (1953), as half way between a projective technique and a questionnaire. Most of those cited above would probably agree with Rotter and Rafferty (1950) that the respondent's answers reflect his own wishes, desires, fears and attitudes, although these researchers also stress the willingness of the subject to provide information. This problem of the degree of control which the subject exercises over his responses is recognized as a problem in all projective techniques, and does not disqualify the use of the term, although it might help to place the method along a continuum of "levels" of awareness (Goldberg, 1965).

Inherent in the projective method has been the technique of leaving the response to the subject. Stimuli have been employed of more or less ambiguity or directedness, but the subject has been free to respond with words of his own choosing. This very element of spontaneity has been a hallmark of the method. As the subject talks, with minimal clues from the situation, what emerges is a truly individual product guided from within himself, and some amalgam of his idiosyncratic wishes, fears, and attitudes is presented to the tester. Recently, researchers have stressed that the subject is not really without a clue as to what is expected from him in the testing situation (Schachtel, 1967; Schafer, 1954). What the subject expresses is a subtle interplay of his interpretation of the meaning of the total testing situation and of his own more prevalent needs or attitudes. The tester receives the private productions of the subject as his source material while seeking to become as aware as he can of situational pressures on the subject in order to use these to good interpretive account.

Closed, multiple-choice tests have

Thanks are due to Dan Zacai and Yemima Halfon for assistance in data assessment.

Copyright 1975 by the Society for Personality Assessment, Inc.

many advantages of economy of scoring. Tests can be scored clerically in a fraction of the time it would take trained psychologists, and the whole problem can be sidestepped which involves training assessors to use a scoring manual, of achieving high interjudge reliability, and even of constantly checking testers' standards for "fading" caused by fatigue or by subtle changes in categorization and emphasis over time. An open-ended method maintains its advantages particularly in providing for the idiosyncratic and the unexpected. There are certain situations, however, including large research projects, where it can be too costly and then practical considerations may make the use of a more formal method more suitable. This is so especially when the parameters of a test have been defined in advance. The closed method may even be preferable when one is seeking "fidelity" rather than "bandwidth" (Cronbach & Gleser, 1957) and is concerned only with a specific dimension or dimensions and with greater reliability of the delimited value being studied.

Traditionally, closed questionnaires, with ready-made answers from among which the subject chooses, have not been considered to belong to the projective category. It is recognized, however, that the subject's response is a personal decision in the sense that he expresses his wishes, needs, misunderstandings, and intentions. In this sense, however structured the format of the question may be and of the available multiple choice answers, each individual makes his own choice in the way he understands both the questions and answers. True, a list of ready answers from which to choose can be very confining. Individuality may be lost, since these are not the subjects' own spontaneous answers and the choices presented may be both restricted and restrictive. They may, in addition, be choices that are too obvious and a respondent may make a clever and defensive selection which conceals other pertinent personal data. These ready-made answers may thus allow an easy intrusion of set, such as social desirability. When one develops a structured sentence completion test, it is necessary to provide answers appropriate to the answers the subjects themselves would give. This would not be the same as allowing spontaneity and something will be lost as the method becomes formal. The ready answers can only be a pale and decimated approximation of what the person would say of his own accord, but one could seek at least to approximate the world of people and their responses, freely given, to an open-ended sentence completion test. Thus the multiple choice answers would best be chosen from typical answers. The idiosyncratic response could not be represented in the choices offered, but the answers would cover various degrees of whatever dimension the sentence usually evokes, and the respondent might choose an answer approximating what he might have said. It becomes an empirical problem whether the carefully selected ready-made answers and spontaneous answers of the same person approximate each other. One would not be looking for identity of words but for similarity in scoring, along a continuum of whatever is being measured (such as anxiety or dependency). Obviously, the open-ended answer would contain extra clinical information; however, if one's intention is to get at specific information in the most economical way, and to avoid at the same time problems of interjudge-reliability, a multiple choice format may have a certain value. It would also be necessary to show that the "obvious" multiple choice format is not so obvious in the sense of being highly correlated with social desirability. On the surface, such a structured method seems far more likely to be susceptible to the intrusive effects of social desirability than the open format. A child should easily be able to pick out the "good" answer, when response possibilities are spread out before him in writing. Thus, if the same child responds to both forms, we would expect him to receive a higher, more autonomous score on the multiple choice. He may spontaneously write "I'll ask my father to do it," but when he is told specifically of other ways to respond – such as the far more adult "I'll fix it myself," he will surely sense

that such a response is more desirable.

Specific limitations and expectations are now proposed for a multiple-choice personality measure which aims at a certain correspondence to a sentence completion measure. If MC is an obvious test, most children might be expected to avoid the low autonomy answers and congregate around the "good" answers at the upper end of the scale. Thus, first, the mean of MC would be significantly higher than that for SC and, second, the variance of answers with MC would be significantly smaller. Even if there were such differences, however, they would not necessarily invalidate MC, provided MC answers did not bunch up at the ceiling of the test, with a relatively higher mean and a truncated distribution. This comparison of means of the two tests should be done while taking into consideration possible order effects, depending on which form was administered earlier.

Third, we would expect the MC form to show much the same correlation with a measure of social desirability as does the SC form. During test construction, the SC form was found to correlate negatively with social desirability $r = .3$, i.e., the test, to the contrary, was not one which led the child with a high approval motive to choose answers indicating autonomy. The MC form might conceivably reverse the correlation so that children tending to choose socially desirable answers would choose high autonomy answers. The social desirability measure is also considered a measure of conformity, so that the negative correlation with SC gives the SC measure of autonomy a certain validity – the more autonomous, the less conformist. If the correlation of social desirability with MC is positive, the validity of MC is rendered dubious.

Fourth, the MC measure, as well as the SC form, should stand in much the same relation to teachers' and peers' estimates of the child's autonomy. What is gained in efficiency and economy should not be lost in validity. In earlier studies, correlations between teacher and SC varied between .11 and .54, averaging about .35. Similarly the correlations between peers and SC averaged .29. The correlation of MC with teachers and schoolmates should be roughly of this order, much the same as the correlation of SC with these external criteria on the same data.

Fifth, MC should correlate highly with SC, if the two tests are to be seen as interchangeable. What sizes of correlation should be demanded between the open and closed versions? The test-retest correlation for SC is 0.8. MC is not the same test since it contains numerous strictures in its response availabilities. It seems that something less than $r = 0.8$ between MC and SC would be obtained. While even a low correlation, if significant, could be an indication of utility of the structured test, especially in research, and where economy considerations prevail (Cronbach, 1970, p. 135), obviously the greater the correlation, the more the tests can be considered equivalent.

The magnitude of the correlation between SC and MC should be considered in the light of typical test-retest correlations obtained using other sentence completion tests. These have not usually been very satisfactory. Churchill and Crandall (1955) report correlations over a period of from one to three years varying from .38 to .54 using the Rotter Incomplete Sentence Blank. Burwen, Campbell, and Kidd (1956) report a test-retest correlation over one year on a measure of attitude to authority of $r = .12$. Fiske and Van Buskirk (1959) and Osterweil and Fiske (1956) report that the content of the great majority of responses on a sentence completion test was changed on retesting to a greater or lesser degree. The correlation between the open-ended and multiple-choice tests is to be studied against this background.

One final restriction must be put on the MC form. This more economic method may wash out or obliterate such finer subtleties of the open-ended form (and of the general concept of autonomy), as can be shown to exist on the basis of a factor analysis. The closed form may be more obvious or more likely to arouse a general response set (to choose consistently high or low autonomy answers) regardless of possible orthogonal factors that can be shown to emerge in factor analyses of the open-ended test. Thus, the SC, while constructed as uni-

dimensional, may emerge as multidimensional, and the MC as unidimensional. Studies of the SC which show a factorial structure will be presented separately in detail. For purposes of this presentation, it can be said that on SC three or four orthogonal factors emerged, which suggests that different kinds of autonomy exist with little if any relation between them. Sentence stems that suggest independence in the face of obstacles (toy breaks, forget bag on bus, button torn, come home hungry, etc.) clustered together and showed little weighting on other factors. Sentence stems that suggest independence of parental pressure (mother opposes my buying the sweater I want, etc.) formed a second orthogonal factor. Items that suggest independence when faced by trauma or anxiety (fall and scratch hand, wake in fright, friend's insult, etc.) clustered together and may belong more to a concept of emotional rather than instrumental dependency. A fourth factor, which has perhaps something to do with independent peer pressure, was only suggested by the data. Thus the concept of autonomy differentiates into at least these factors, and a situation-specific or agent-specific, rather than trait-consistent approach seems advisable to this behavior.

The MC test must be measured against a test which achieves this kind of differentiation by agents and situations. Its efficacy will be shown to be sufficient not only if it meets the other restrictions (similar means and variance, minimal relation to social desirability, similar predictive values, high intercorrelation between SC and MC) but also this requirement of factorial structure.

Method

A sentence-completion (SC) measure of autonomy in children has been constructed (Shouval & Duek, Note 1), and this paper presents work done on the test to produce an alternate multiple-choice format (MC). The open-ended (SC) test was constructed in such a way as to measure, as far as is possible, only one dimension of behavior. All items were discarded that did not lead to responses which could be scored on a 3-point scale of high-low autonomy. Sentence stems involved the presentation of a situation of conflict or frustration with the outcome determinable either by a solution, on the one hand, of self-help or self-determination, or, on the other hand, by a reliance on others or a surrender to the pressure of others. A manual for scoring items was prepared, which provided examples of answers received from children, and categorized these as to whether they reflected high, medium, or low autonomy. On the basis of this manual, a closed (MC) test was constructed[1] to illustrate typical answers received from children in the course of using the test. The sentence-endings chosen are the children's responses, not the expectations of the experimenters. Thus a sentence stem reads: "If I forget my school bag on the bus, I . . . " and the child writes in his conclusion. In the multiple-choice version, three alternative answers are presented:

a. I will go to the police so that they can find it.

b. I will go to the bus terminal and ask if anyone found the bag.

c. I will tell my mother and she will go and look for it for me.

The autonomous answer is b., while "to go to the police" is considered a little more dependent but less so than just to tell "my mother."

Both tests were administered to various groups of children ranging in age from eight to thirteen. The SC and MC tests were given in that order to 188 children with a one-month interval, and to 171 children in randomized order at a two-month interval. This latter group came from three schools and a total of seven classrooms, and the children of each classroom were randomly divided with half being given each order of forms. In some of the classes, children were given a measure of social desirability (Crowne-Marlowe, 1964, Hebrew adaptation for children), and teacher and peer

[1] The Autonomy Multiple Choice Measure (AUTMC) is available upon request from the author.

Table 1

Relations Between Sentence Completion and Multiple Choice Autonomy Forms – Means, Standard Deviations, Intercorrelations and Homogeneity of Variance

Age 13	Sentence Completion		Multiple Choice		*N*	*r*	*F* ratio
	X	S.D.	X	S.D.			
School A[a]	57.86	3.79	57.76	4.59	34	.69	1.46
School B[a]	56.85	4.38	58.08	4.65	33	.56	1.13
Schools C, D, E[a]	57.26	4.35	60.92	4.35	87	.56	1.00
Schools C, D, E[b]	59.84	4.20	59.33	3.35	84	.34	1.57*
School F[a]	57.54	4.36	61.96	3.40	31	.52	1.64*
Age 11							
School F[a]	54.63	4.59	55.54	4.42	22	.56	1.08
School G[a]	54.62	3.96	52.28	4.58	29	.53	1.48
Age 10							
School G[a]	48.17	5.07	49.50	5.30	18	.55	1.09
Age 8							
School F[a]	47.66	5.70	50.09	5.59	21	.70	1.04

[a] Order of forms Open–Closed, 1–2 month interval.
[b] Order of forms Closed–Open, 1–2 month interval.
* $p < .05$

ratings of autonomy were gathered. Using definitions of autonomy as set forth in the SC manual, teachers presented estimates of the children ranking them from high to low autonomous on a 5-point scale. The child's classmates gave sociometric ratings on a similar scale.

Means and standard deviations for each form were computed, as well as the intercorrelation between the two forms and a measure of homogeneity of variance between the forms. The significance of difference between the means was computed in an analysis of variance design with repeated measurements, with form, order, and classroom as main effects and form x order as the interaction. Correlations were computed together with teacher and peer ratings of the child's autonomy. The factor structure of these two tests was then compared.

Results

The means and standard deviations for children given both the Sentence Completion and Multiple Choice Autonomy forms appear in Table 1. The data are presented for nine groups of children arranged according to schools and ages. In

Table 2

Analysis of Variance[a] of Sentence Completion and Multiple-Choice Autonomy Forms

(Age 13, Schools C, D, E; 7 Classrooms)

Source of Variation	*S.S.*	*d.f.*	*M.S.*	*F*	*p*
Test Form	262.37	1	262.37	15.87	0.01
Order	49.67	1	49.67	3.00	–
Class	52.92	6	8.82	0.53	–
Form x Order	314.86	1	314.86	19.04	0.01
Error	5488.05	332	16.53	–	–

[a] Repeated measurements, seven school classes with order randomized, two-month interval.

schools A, B, F, and G, the tests were given in the order, SC – MC. An examination of the means shows increase of autonomy with age in both forms of the test. At each age, the means for both forms approximate each other, but those for MC tend to be slightly higher. The significance of difference of means was computed, however, only for schools C, D, and E. Here "order" was randomized with half of the children of the seven school classes being given order SC – MC, and half MC – SC. The results of the analysis of variance (See Table 2) for the data presented for schools C, D, and E show a significant main effect for the test forms. Children receive a higher mean score with MC. The order effect is not significant, nor are there significant differences between the seven classes. A significant interaction effect was recorded for form x order. As can be seen by examination of the means in Table 1 for schools C, D, and E, this interaction arises from a test form (either SC or MC) being given as the second test and showing a higher mean score than the same test when it is administered first. The first limitation we set for the MC was that the mean score approximates the MC. This condition was not fulfilled; however, this is not critical since the scores do not bunch up at the top end of the scoring range and thus limit differentiation among persons. Further, the increase of autonomy with age, as noted over ages 8, 11, and 13 in earlier reports (Shouval & Duek, Note 1), is repeated here with new data and occurs also for the MC. Thus, although the means may be slightly higher, the new measure maintains differentiative power.

The second limitation was that the standard deviation of MC be similar to that of SC. A very small standard deviation for MC would suggest decreased differentiation by the test. The homogeneity of variance between the two forms of the test is given in Table 1. Of nine *f* ratios computed, there is an equal division among them as to whether the standard deviation is higher for SC or for MC. However, two *f* ratios are significant ($p < .05$), and in both these cases, the standard deviation for MC is smaller. The MC form thus shows a slightly smaller standard deviation. This finding is not crucial since MC, nevertheless, shows a variation of scores sufficient to enable differentiation of persons.

The third limitation pertained to the possible relation of MC to a social desirability measure. In previous studies, the SC measured was not found to be affec-

Table 3

Correlation of Sentence Completion (AUTSC-1) and Multiple-Choice (AUTMC-1) Forms of Autonomy Test with Social Desirability

	AUTSC			AUTMC		
	Male	Female	Total	Male	Female	Total
School A	-.07	.00	.11	.03	-.25	-.08
N	15	19	34	15	19	34
School B	.10	.16	-.10	-.04	-.05	-.13
N	19	14	33	19	14	33

Note. Social Desirability (Crowne-Marlowe) – a negative correlation indicates more Autonomy, less Social Desirability.

ted by this response set, and it is critical for MC that it be similarly free of this effect. Correlations of MC and SC with social desirability for the same children (See Table 3) show that the relation of social desirability to either form, SC or MC, is insignificant and close to zero. The significant negative correlations found earlier were not repeated here with this sample; however, it is clear that MC is not saturated with this response set, a finding which would have rendered it unsuitable.

The fourth limitation was that the two forms should stand in much the same relation to teachers' and to peers' nomination ratings of autonomy. The correlations are given in Tables 4 and 5. Those with teachers' ratings are lower than in previous findings; however, the correlations with MC are certainly not lower than these for SC. Similarly for peer nominations, the MC form is no less an effective measure given this particular criterion.

The fifth restriction that could limit the possible effectiveness of MC was that of the intercorrelation with MC. The correlations for the various classes average about .55 (See Table 1). This correlation is to be seen against the test-retest of SC which is $r = .80$, and test-retest of other sentence-completion tests which range (over longer periods of time) from 0.12 to 0.54. It might be said that the tests approximate each other also with respect to this requirement of intercorrelations. The relationship between the two forms is reasonably, albeit not, especially satisfactory.

The final requirement related to the factor structure. MC and SC should be similarly differentiated into factors. Full data will be presented in a separate report on various factor analyses performed on various versions of the tests. At this point it will suffice to report that a factor analysis of MC on 273 children aged 13-14 shows the same factor structure as in SC. In both forms, there is a differentiation of the concept of autonomy into perhaps four independent factors covering instrumental autonomy or autonomy when faced by obstacles of nature, autonomy in the face of parents' pressures, autonomy when in traumatic situations, and possibly autonomy in the face of peer pressures.

Discussion

Multiple-choice and sentence completion measures of children's autonomy yielded the following items of comparison. Means of the MC were higher, and standard deviations smaller. Only small

Table 4

Correlations of Autonomy Sentence Completion and Autonomy Multiple-Choice Tests with Teachers' Ratings of Child's Autonomy

	AUTSC-1			AUTMC-1		
	Male	Female	Total	Male	Female	Total
School $\frac{A}{1}$	.32	−.03	.13	.25	.25	.24
N	15	19	34	15	19	34
School $\frac{B}{2}$	.32	.21	.16	.14	.58	.26
N	19	14	33	19	14	33

Table 5

Correlations of Autonomy Sentence Completion and Autonomy Multiple-Choice Tests with Peer Nominations of Autonomy

	AUTSC-1			AUTMC-1		
	Male	Female	Total	Male	Female	Total
School A	.58	.33	.45	.62	.50	.56
N	15	19	34	15	19	34
School B	.26	.01	.19	.05	.48	.25
N	19	14	33	19	14	33

numerical differences were reflected, and MC shows no particular ceiling effect or truncation for the ages tested. MC was not affected by any tendency to give socially desirable answers. Both tests stood in much the same relationship to external criteria, teachers' and peers' ratings. The tests intercorrelated with each other on an average $r = .55$, which is lower than the test-retest of this sentence-completion test of .80, and as good as the test-retests reported of other sentence completion measures. The factor structure which suggests that autonomy is a mutidimensional concept is clear in both open and closed versions of the test.

No claim is made for a general rèplacement of open-ended sentence completion measures by multiple-choice measures. What is suggested is that when multiple-choice answers are chosen from the world of children's spontaneous responses, when the test is designed rigorously as a "fidelity" measure of one variable (autonomy, in this instance) or a group of conceptually related variables (the various kinds of autonomy), and when scoring of the open-ended test is done using a comp-

rehensive manual and the definitions in the manual are used for constructing the multiple choices, then sentence completion and multiple-choice tests can be made reasonably comparable. The sentence completion format retains its importance for individual clinical assessment, especially of the idiosyncratic, but, for large-scale testing, the closed, multiple-choice format seems no less adequate and certainly far more economical for clearly defined diagnostic purposes — such as the assessment of autonomy.

REFERENCES

Burwen, L. S., Campbell, D. T., & Kidd, J. The use of a sentence completion test in measuring attitudes towards superiors and subordinates. *Journal of Applied Psychology*, 1956, *40*, 248-250.

Campbell, D. T. A typology of tests, projective or otherwise. *Journal of Consulting Psychology*, 1957, *21*, 207-210.

Churchill, R., & Crandall, V. J. The reliability and validity of the Rotter Incomplete Sentences Test. *Journal of Consulting Psychology*, 1955, *19*, 345-356.

Cronbach, L. J. *Essentials of psychological testing* (3rd ed.). New York: Harper and Row, 1970.

Cronbach, L. J., & Gleser, G. C. *Psychological tests and personnel decisions.* Urbana: University of Illinois Press, 1957.

Crowne, D. P., & Marlowe, D. *The approval motive: Studies in evaluative independence.* New York: Wiley, 1964.

Fiske, D. W., & Van Buskirk, C. The stability of interpretation of sentence completion tests. *Journal of Consulting Psychology*, 1959, *23*, 177-180.

Forer, B. R. A structured sentence completion test. *Journal of Projective Techniques*, 1950, *14*, 15-29.

Goldberg, P. A. A review of sentence completion methods in personality assessment. *Journal of Projective Techniques and Personality Assessment*, 1965, *29*, 12-45.

Hanfmann, E., & Getzels, J. W. Studies of the sentence completion test. *Journal of Projective Techniques*, 1953, *17*, 280-294.

Murstein, B. I. *Handbook of projective techniques.* New York: Basic Books, 1965.

Osterweil, J., & Fiske, D. W. Intra-individual variability in sentence completion responses. *Journal of Abnormal Social Psychology*, 1956, *52*, 195-199.

Rohde, A. R. Explorations in personality by the sentence completion method. *Journal of Applied Psychology*, 1946, *30*, 169-181.

Rotter, J. B., & Rafferty, J. E. *Manual: The Rotter Incomplete Sentences Blank.* New York: Psychological Corporation, 1950.

Schachtel, E. G. *Experiential foundations of Rorschach's test.* London: Tavistock Publications, 1967.

Schafer, R. *Psychoanalytic interpretation in Rorschach testing.* New York: Grune and Stratton, 1954.

Sechrest, L. Testing, measuring and assessing people. In E. F. Borgatta and W. W. Lambert (Eds.), *Handbook of personality theory and research.* Chicago: Rand McNally, 1968.

Reference Note

1. Shouval, R., & Duek, E. Children's autonomy: A sentence completion measure. Unpublished manuscript, 1973.

Ron Shouval, PhD
Department of Psychology
Tel-Aviv University
Tel-Aviv, Israel

Received: June 8, 1973