7-1974

Effects of evaluation on creative production

Dalia E. Katilius-Braun
University of Nebraska at Omaha

EFFECTS OF EVALUATION ON CREATIVE PRODUCTION

A Thesis

Presented to the

Department of Psychology

and the

Faculty of the Graduate College

University of Nebraska at Omaha

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Dalia E. Katilius-Braun

July, 1974

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Accepted for the faculty of The Graduate College of the University of Nebraska at Omaha, in partial fulfillment of the requirements for the degree Master of Arts.

Graduate Committee:	Name	Department
	James M. Thomas	Psychology
	Mary Y. Lanier	Psychology

Kenneth A. Deffenbacher
Chairman

July 25, 1974
Date

Gift
10-22-74

Table of Contents

List of Tables

Despite considerable interest in the problem of creativity, both in psychology and in other fields, there is still no standard definition of the term. Guilford (1967) defines creativity by making a distinction between convergent thinking and divergent thinking. In convergent thinking, according to Guilford, there is a single right answer or best answer to a problem, while in divergent thinking there is not. He uses the term divergent thinking interchangeably with creativity. His research indicates that of the factors identified as making up divergent thinking, ideational fluency, which Guilford defines as fairly rapid generation of units of verbal or semantic information, has least variance in common with intelligence. The correlation between ideational fluency and convergent thinking is .01 (Guilford, Frick, Christensen, & Merrifield, 1957). Since a number of studies have indicated fairly low correlations (around .3) between I.Q. scores and scores of creativity, at least in children (e.g., Getzels & Jackson, 1962), it would seem that ideational fluency may be an important component of creativity. Other writers approach the term somewhat differently. Parnes and Meadow (1959) use the criteria of uniqueness and usefulness as determiners of good ideas in their studies on creativity. Wallach and Kogan (1965) speak of total number of ideas generated as well as their uniqueness within a given subject sample. Though ideational output includes qualitative considerations as well as

quantitative ones, the present study is concerned only with the quantitative — number of ideas produced in a divergent thinking situation.

Most of the research in creativity to date falls into three areas: identification of creativity as a concept through a number of factor-analytic and psychometric studies; exploration of personality variables accounting for differences between creative and noncreative individuals; and techniques for increasing creativity in individuals, primarily through creativity training programs. The present study focuses on conditions affecting creative output. These conditions concern the degree of restriction placed on the subject during the process of ideational production.

While various writers describe the stages of creative effort in different terms, there is agreement that evaluation of the creative product should come last in the creative process. There is the implication that evaluation occurring earlier in the process might be detrimental to the process and possibly to the end product as well. Alamshah (1972) suggests that feelings of inferiority, stemming from criticism by self or others, are one of several common blocks to creativity. One can thus speculate that evaluation occurring early in the creative process might result in one or both of two events: acceptance of the ideas generated thus far as good or at least sufficient, in which case no further generation is needed; or self-criticism

of the ideas generated so far, in which case feelings of inadequacy may block further generation of ideas.

Turning to more experimental evidence, a number of studies have been conducted at the State University of New York at Buffalo, assessing the effects of creativity training and various instructions for creative production (Parnes & Noller, 1972). The work at Buffalo has consistently shown that significantly more good-quality ideas are produced by individuals under deferred-judgment instructions (where evaluation is to take place after production) than under concurrent-judgment instructions (where evaluation is to take place during production). They also report that there is a positive correlation between quantity and quality of ideas, and that deferred-judgment instructions, by encouraging a greater quantity of ideas, seem to increase also the number of good-quality ideas (Parnes & Meadow, 1959).

In developmental literature, where most of the work on creativity done so far appears, Wallach and Kogan (1965) claim that playful, nonevaluative conditions produce higher creativity scores. Kogan and Morgan (1969), on the other hand, tested creativity in children under "game-like" and "test-like" conditions and found no superiority for the "game" condition. This finding is supported by Warren and Luria (1972) who tested fifth and sixth graders under instructions designed to induce low, high and neutral evaluational sets, and found no differences among the different

evaluational sets. How can the findings of the latter two studies be reconciled with the results of the work at Buffalo, the research of Renner and Renner (1971) who also found non-evaluative, deferred-judgment instructions to be superior, and the myriad creativity training programs, all of which reiterate the need for deferring evaluation until after creative production?

An empirical answer must, of course, await further research, preferably research comparing adults and children with respect to the effects of evaluation on the creative process. Theoretically, however, the dilemma does not seem irreconcilable if one can entertain one or both of two possibilities. First, the notion of a playful, nonevaluative condition (to borrow Wallach and Kogan's term) may be very different and imply different consequences for adults and children. Piaget's position that the difference between adults and children is often qualitative may very well apply to attitudinal as well as cognitive phenomena. Thus, while in children "playful" tends to mean "different from the routine," especially in the typical educational environment, and the condition is essentially an environmental one, in adults "play" may be a much more internalized state, a freedom from the threat of serious consequences, especially negative ones. One wonders how serious the consequences of not doing well on a creativity test were to the fifth and sixth graders in Warren and Luria's study, and whether there

was any perceived essential difference between game and test conditions in Kogan and Morgan's children. Both game and test may well have been equally a break from school routine.

Related to this apparent discrepancy between adults and children in the effects of evaluation on the creative process is an idea suggested by the model of an innovative boundary formulated by Slevin (1971). He asserts that in an industrial setting there is a relationship between risk and value of a goal requiring innovative behavior which must be maintained at a proper balance for innovative behavior to occur. In children being tested in creativity, one might hypothesize both to be minimal. In college students, the usual subjects of the adult studies mentioned above, especially where the experiment is presented as part of course and student evaluation, as in the Buffalo studies, the consequences probably assume more importance, and hence the risk is greater. For other adults, outside a school environment, only a goal of real value might evercome the risk involved — the threat to self-esteem ensuing from evaluation. In fact, Alamshah (1967), in discussing conditions of creativity, claims that potential for creativity is enhanced by endeavor in an area that is high in the individual's value system, and to which that individual is willing to devote the greater part of his work effort. In an experimental situation, where the goals are rarely perceived as real or significant, one might assume that evaluative threat should be at a mini-

mum for creative production to emerge.

One further variable of concurrent evaluation that seems related, and that has been neglected in research on creativity, is that of bonuses or incentives for creative effort. While money is generally believed to be a powerful motivator, and motivation is conducive to achievement, how does this relate to creative effort? How do incentives affect concurrent judgement, and what are the effects on quantity and quality of ideational output?

The notion of goal definition and specificity, while not directly evaluative in the sense that it might threaten the self-image, may also relate to evaluation of the product. Mednick (1962) in somewhat cryptic fashion comments that when the problem solver knows the goal he is trying to reach, the elements of the goal become additional stimuli. In Mednick's view, additional stimuli appear to be desirable in the creative effort since they expand the range of raw materials from which novel associations can be made. In typical goal-oriented situations such as industry, however, goal definition often takes the form of limiting injunctions such as feasibility and usefulness. Whether the verbalization of these requirements in the instructions for creative production fit Mednick's notion of elements of the goal is unclear; the possibility of their effect on the creative effort should be explored.

This study, then, looks at the effects on ideational

output of instructions intended to convey nonevaluative, evaluative and incentive conditions, and goal only versus goal with feasibility and utility restrictions.

Method

Subjects and Design. Subjects were drawn from male Business Administration students at the University of Nebraska at Omaha (UNO) currently enrolled in four sections of Principles of Management. Female students were not used because the number of women in the Business Administration program at UNO was too small to permit matching across conditions. Subjects were randomly assigned to one of six conditions. All eighty students in the four classes were tested. Females and those who submitted incomplete responses were rejected (14 subjects). Then, since the smallest cell size obtained was ten, subjects were randomly rejected from each of the other cells to arrive at a cell size of ten on all six conditions.

The design was a 2 x 3 factorial, with two levels of Goal Specificity and three levels of Evaluation. The experimental task was the same for all subjects: to come up with specific new ideas for products and/or services that would be marketable to communities of senior citizens. In addition, one covariate was used: total number of ideas generated on the Patterns Test (PT), a measure of ideational output (Wallach & Kogan, 1965). This measure was used to control for

the effect of individual differences in ideational fluency.

Procedure. All subjects were tested in the classroom during a regular class period of Principles of Management. Their instructors had announced in the previous class that their next class would be an experiment, so that subjects were free to attend or not, though they were encouraged to do so.

All subjects were given identical oral instructions. Subjects in differential experimental conditions were given differential instructions in writing, typed prominently in their work booklets. Spot checking after the experiment indicated that subjects were not aware that their instructions differed from those of anyone else in the room. Subjects in different conditions were tested at the same time, in the same class session, to exclude differences other than written, instructional ones.

The testing sequence was as follows: All subjects were introduced to the PT, and were given about five minutes to practice on the first page. They were then given ten minutes for each of the three pages of the PT. (A copy of the PT is in Appendix A.) They were then told that the second part of the experiment is an applied task, similar to the kind of idea-generation tasks that occur in business. They were asked to read the instructions in their booklets carefully (differential instructions are listed in Appendix B), and not to discuss either the instructions or their ideas with

anyone else in the room. Thirty minutes were alloted to this part of the experiment. The experimenter observed that no one was still working at the end of the thirty minutes. Finally, the subjects were asked to fill out a questionnaire on the last page of their booklets concerning their attitudes about the experiment. (A copy of the questionnaire is in Appendix C.)

The ideational output of each subject was scored for total number of ideas by seven independent judges. (A copy of the instructions to the judges is in Appendix D.) The judges consisted of a sample of professors in the UNO School of Business Administration in the areas of Management, Marketing, Accounting, Finance and Decision Science — the academic equivalent of an industrial executive committee. Each judge scored the responses separately, avoiding the interactive bias of group evaluation. The responses from each subject were typed exactly as they had been written by the subject, so that the content remained the same, but judgment on such variables as neatness or handwriting was avoided. The response sheets were coded for subject and condition, so that the judges were not aware of which subject or condition they were evaluating at any time.

Results

Reliability and Analysis of Variance. Pearson r correlation coefficients were computed for each pair of judges. (See Table I.) Each of the Pearson r's were then converted

to z-scores, the z-scores averaged, and the average converted back to a Pearson *r*, yielding an average correlation coefficient of +0.93. High interjudge reliability was thus demonstrated.

Table I

Between-Judge Pearson *r* Matrix

	J1	J2	J3	J4	J5	J6
J2	.9260	--	--	--	--	--
J3	.9527	.9346	--	--	--	--
J4	.9286	.9659	.9351	--	--	--
J5	.9510	.9234	.9466	.9360	--	--
J6	.9273	.9428	.9429	.9311	.9324	--
J7	.8310	.9186	.8654	.9345	.8815	.9149

Judge variances were computed (see Table II), and an F_{max} test of heterogeneity of variance yielded an F_{max} value of 1.3984. Since this failed to exceed the tabulated F_{max} of 2.17 at $\alpha = .05$ (Kirk, 1968), it was decided that there was homogeneity of variance among the judges.

Table II

Judge Variances

J1	J2	J3	J4	J5	J6	J7
.125	.143	.172	.123	.128	.149	.131

Given high interjudge reliability and homogeneity of variance, the number of ideas for each subject were summed across judges. Data for the covariate were obtained by counting the number of ideas generated by each subject on the PT. A 2 x 3 factorial analysis of covariance was then performed. Adjusted cell means are displyed in Table III. (These means represent number of ideas times number of judges averaged across subjects.) While there were no

Table III

Adjusted Cell Means

	No Evaluation	Evaluation	Incentive
Goal Only	77.9245	22.3062	25.5721
Goal with Specifications	71.5974	27.3899	30.8098

significant effects found for either Goal Specificity or its interaction with Evaluation, Evaluation was found to be significant ($p < .0001$; see Table IV).

The Tukey A Test of Pairwise Multiple Comparisons was then performed on the three levels of Evaluation (see Table V). With df = 53, the difference between means must be 16.4059 to be significant at the .05 level, and 20.7712 to be significant at the .01 level (Kirk, 1968). The difference between Evaluation and Incentive was found to be nonsignificant. The difference between No Evaluation

Table IV

Analysis of Covariance Summary Table

Source	df	MS	F	p
Goal Specificity (A)	1	26.2970	0.0571	p < .8121
Evaluation (B)	2	15,528.8633	33.7414	p < .0001
A X B	2	200.6311	0.4359	p < .6490
Error	53	460.2313	---	

and Evaluation, and the difference between No Evaluation and Incentive were both significant.

Table V

Differences between Adjusted Means (Evaluation)

	No Evaluation (74.76095)	Evaluation (24.84805)	Incentive (28.19095)
No Evaluation	----	49.9129	46.5700
Evaluation	----	----	3.3429

Questionnaire Data. Questionnaire results revealed the subjects' mean age to be 26, with a median age of 25, and a range of 19 to 47. The mean number of years of military experience for all subjects was 4.2, with a median of 2 years, and a range of 0 to 22. Considering those subjects with more

than two years of military experience to be in the military, 26 subjects were in the military. There were no differences across conditions in either age or military experience.

No subject was aware of the real hypotheses of this study. The responses to Question Four ("Did you think of any ideas that you did not put down? Why?") resulted in a distribution of "yes" responses ranging from four to six in all six conditions. Thus all conditions appeared to respond similarly to this question.

Discussion

Some caution should be exercised in generalizing the results of this study to an industrial setting without replication of the study using subjects already in industry. At the same time, the student sample used here deviated from the typical college student sample. The average age of 26 is older than the usual average age for college students. More than half of the subjects are already in business or the military. A number of the remaining subjects have declared business administration as their major area of study. Thus the subject sample can be assumed to fall somewhere between college students and individuals in an organization. The judges represent a sample of areas found in the management of an organization (with the exception of manufacturing) and can be assumed to represent a point of view comparable to that of a decision-making body in an

organization. Given this subject/judge sample, then, the results should have more generalizability to business and organizational situations than studies using typical college student samples.

To summarize our results, no difference was found between Goal Only and Goal with Specifications. There was also no difference between Evaluation and Incentive. On the other hand, No Evaluation was significantly superior to both Evaluation and Incentive.

In idea-generating tasks in business, as in a marketing department, for example, it is common to have goal-limiting injunctions. Utility and feasibility are especially common requirements for new ideas. Yet our results indicate that these injunctions do not increase the number of ideas generated.

Bonuses as incentive to idea-generation are also believed to be effective. The wide-spread use of suggestion boxes is only one example of the use of monetary incentive for creative effort. Yet our results suggest that where an evaluative situation already exists, a bonus may not contribute anything toward overcoming inferior results. Within the limits of a $2.00 bonus, promising a bonus versus not promising anything made no difference. It would be useful to study the effects of incentive further by using a more substantial bonus. It would also be of interest to compare the effects of incentive in an evaluative versus a nonevaluative situation, as an

attempt to investigate whether the bonus itself implies evaluation to the subject, or whether it merely fails to overcome the inhibiting effects of an already expressed evaluative set.

A sizeable difference was found between nonevaluative instructions and evaluative instructions, either with or without incentive. The nonevaluative conditions were significantly superior in the number of ideas produced. This has implications for industry, if only to question the common practice of a superior's saying to his subordinate: "Why don't you come up with some <u>good</u> ideas for . . ." This finding also supports the results of the work done at Buffalo (Parnes & Noller, 1972) where deferred-judgment instructions were superior to concurrent-judgment instructions.

Further research needs to be done to evaluate the quality of ideas produced under these differential conditions, and the author is currently analyzing qualitative evaluations of the ideas generated by this sample of subjects. Results to date suggest that nonevaluative conditions might produce not only greater numbers of ideas, but greater numbers of high-quality ideas as well.

At this point it is not inappropriate to indulge in some speculation as to what subject variables may be operating to produce the superiority of the No Evaluation condition. Since it was the least restrictive condition that produced the most superior quantitative result, it seems

reasonable to hypothesize that the output under nonrestrictive conditions most closely approximates baseline creative ability, while the other conditions inhibit emergence of the baseline level. To put it simply, fostering creativity may be a matter of permitting a person to be creative, rather than helping or making him creative.

It is interesting to note that the questionnaire failed to show any real differences among conditions in ideas reported as generated but not written down. At this point, one can only speculate about the possible causes for the apparent discrepancy between experimental results and self-report data. Instructional manipulation produced a significant difference between No Evaluation and both Evaluation and Incentive, yet questionnaire data revealed no differences among conditions. It may be that the questionnaire was not sufficiently sensitive to show differences even though those differences existed. Another interpretation is that some factor such as social desirability was operating to produce essentially random results on the questionnaire. A third possible explanation, and one that would tie in with a theory of ego-threat in evaluative situations, is that subjects in the evaluative conditions were unwilling to admit that they had come up with "bad" ideas, or even suppressed the negative results of an evaluation that was part of the creative process since these negative results were ego-threatening, and hence the subjects were truly unaware of these rejected "bad" ideas.

Both Alamshah's philosophical speculations (1972) and Slevin's model of an innovative boundary (1971) suggest that risk, possibly of an ego-threatening sort, serves to inhibit ideational output or creativity. Further research should be done to investigate these hypotheses.

Finally, one might speculate whether the imposition of evaluation does not violate Guilford's definition of divergent thinking. When a subject is asked to come up with good ideas, there is an automatic implication of some criterion that differentiates good from bad. While this does not necessarily imply a single correct response, Guilford's distinguishing characteristic of convergent thinking, neither does it assure the absence of it. Might evaluation concomitant with creative production, then, be perceived as convergent thinking in divergent thinking's clothing?

In creative activity optimum motivational states may be quite different from those required for other types of cognitive behavior. Guilford's separation of thinking into convergent and divergent types is useful in starting to define the difference between mental ability and creative ability. But, whatever creativity is, how does it emerge? What motivational states, what personality factors, what situational conditions contribute to the emergence or suppression of creative productivity? This paper is an initial step toward studying this aspect of creativity.

References

Alamshah, W. H. The conditions of creativity. Journal of Creative Behavior, 1967, 1, 305-313.

Alamshah, W. H. Blockages to creativity. Journal of Creative Behavior, 1972, 6, 105-113.

Getzels, J. W., & Jackson, P. W. Creativity and intelligence. New York: Wiley & Sons, 1962.

Guilford, J. P. The nature of human intelligence. New York: McGraw-Hill, 1967.

Guilford, J. P., Frick, J. W., Christensen, P. R., & Merrifield, P. R. A factor-analytic study of flexibility in thinking. Psychological Laboratory Reports, 1957, No. 18. Los Angeles: University of Southern California.

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, Calif.: Brooks/Cole Publishing Co., 1968.

Kogan, N., & Morgan, F. T. Task and motivational influences in the assessment of creative and intellective ability in children. Genetic Psychology Monographs, 1969, 80, Part 1, 91-128.

Mednick, S. A. The associative basis of the creative process. Psychological Review, 1962, 69, 220-232.

Parnes, S. J., & Meadow, A. Effects of "brainstorming" instructions on creative problem solving by trained and untrained Ss. Journal of Educational Psychology, 1959, 50, 171-176.

Parnes, S. J., & Noller, R. B. Applied creativity: The creative studies project. Journal of Creative Behavior. 1972, 6, 11-22.

Renner, V., & Renner, J. C. Effects of a creativity training program on stimulus preferences. Perceptual and Motor Skills, 1971, 33, 872-874.

Slevin, D. P. The innovation boundary: A specific model and some empirical results. Administrative Science Quarterly, 1971, 16, 515-531.

Wallach, M. A., & Kogan, N. Modes of thinking in young children. New York: Holt, Rinehart & Winston, 1965.

Warren, G. H., & Luria, Z. Evaluational set and creativity. Perceptual and Motor Skills, 1972, 34, 435-438.

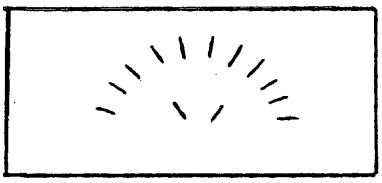
Appendix A

Patterns Test (PT)

You are going to look at some drawings. See how many different things you can think of that these drawings might be.

Here is an example.

What might this drawing be?



_____

_____

_____

_____

_____

_____

_____

Two things this might be are written in — it might be a porcupine, or the sun coming up. Think of some more things it could be and write them in the blank spaces.

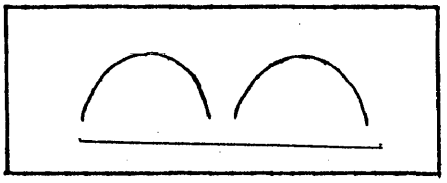
You don't have to write long answers — just enough so we know what your idea was.

When you are ready to start on the first drawing on the next page, put your pencil down and look up.

PLEASE DO NOT TURN THIS PAGE UNTIL YOU ARE ASKED TO.

1.

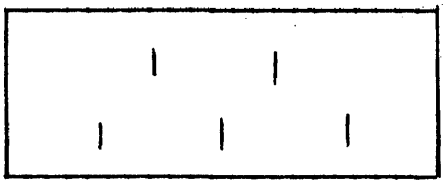
What do you think this drawing might be?



PLEASE DO NOT TURN THIS PAGE UNTIL YOU ARE ASKED TO.

2.

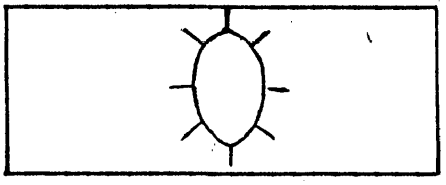
What do you think this drawing might be?



PLEASE DO NOT TURN THIS PAGE UNTIL YOU ARE ASKED TO.

3.

What do you think this drawing might be?



PLEASE DO NOT TURN THIS PAGE UNTIL YOU ARE ASKED TO.

Appendix B

Instructions for Differential Conditions

Condition 11 (Goal Only; No Evaluation):

> Please list all the ideas that come to your mind without judging them in any way. Forget about the quality of ideas entirely. The key point is only quantity on this task. Express any idea that comes to your mind.

Condition 12 (Goal Only; Evaluation):

> Please list all the good ideas you can think up. A group of judges from the Business Administration faculty will evaluate the quantity and quality of good ideas you produce. Don't put down any idea unless you feel it is a good one.

Condition 13 (Goal Only; Incentive):

> Please list all the good ideas you can think up. A group of judges from the Business Administration faculty will evaluate the quantity and quality of good ideas you produce. Don't put down any idea unless you feel it is a good one.
>
> In addition, a bonus of $2.00 will be awarded for each of the ten best ideas in the group. We will mail the bonuses for the best ideas within two weeks. Please write down your mailing address in the space below.

Condition 21 (Goal with Specifications; No Evaluation):

> Please list all the ideas that come to your mind that are feasible and useful. Forget about the quality of ideas entirely. The key point is only quantity on this task. Express any idea which comes to your mind that is feasible and useful.

Condition 22 (Goal with Specifications; Evaluation):

> Please list all the good ideas that are also feasible and useful that you can think up. A group of judges from the Business Administration faculty will evaluate the quantity and quality of good ideas you produce. Don't put down any idea, even a feasible and useful one, unless you feel it is a good idea.

Condition 23 (Goal with Specifications; Incentive):

Please list all the good ideas that are also feasible and useful that you can think up. A group of judges from the Business Administration faculty will evaluate the quantity and quality of good ideas you produce. Don't put down any idea, even a feasible and useful one, unless you feel it is a good idea.

In addition, a bonus of $2.00 will be awarded for each of the ten best ideas in the group. We will mail the bonuses for the best ideas within two weeks. Please write down your mailing address in the space below.

Appendix C

Biographical Information

NAME: _____ S. S. #: _____

AGE: _____ YEARS OF MILITARY EXPERIENCE: _____

. . . . . .

Questionnaire

1. Were you aware of what hypothesis was being tested in this study?

2. What do you think was being studied in this experiment?

3. What constraints or limitations did you feel were placed on your responses in this experiment?

4. Did you think of any ideas that you did not put down? Why?

5. How did you feel about participating in this experiment?

Appendix D

Instructions to Judges

Your task is to determine how many ideas there are on each sheet. THESE IDEAS ARE NOT TO BE EVALUATED IN ANY WAY BEYOND DETERMINING SEPARATE, DISTINCT IDEAS. There may be cases where a whole paragraph or more is only one idea with elaborations. On the other hand, only one sentence or even phrase may contain more than one separate idea.

Your judgments on quantity should follow these steps:

A. Read the entire set of responses for all subjects first. This will give you a feeling for the kind of response population you are dealing with, and will make subsequent discriminations easier.

B. Then decide how many separate ideas there are on a given response sheet and identify them from "1" to "n" on the lines at the left margin. Some of the responses have been given in the form of paragraphs rather than separate items. The lengthier ones have been double-spaced for easier reading and scoring. If there are no lines for identifying and marking the number of a separate idea on the left margin, please make them yourself.

C. Finally, indicate the total number of ideas for the given response sheet at the bottom of the sheet.