DOCUMENT RESUME

ED 089 304 CS 201 101

AUTHOR Sager, Carol
TITLE Improving the Quality of Written Composition Through Pupil Use of Rating Scale.
PUB DATE Nov 73
NOTE 15p.; Paper presented at the Annual Meeting of the National Council of Teachers of English (63rd, Philadelphia, November 22-24, 1973)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Composition (Literary); Composition Skills (Literary); Evaluation Techniques; Grade 6; Intermediate Grades; *Language Research; *Program Effectiveness; Sex Differences; *Student Evaluation; Student Improvement; *Writing Skills; Written Language

ABSTRACT

The purpose of this study was to determine the effectiveness of a program designed to improve the quality of written composition by teaching students the components and use of a descriptive writing scale so that they could rate their own compositions and those of other students. The study consisted of two groups, an experimental group and a control group, comprised of sixth grade students from two schools in a suburb of Boston. The general hypothesis to be tested was that there would be no difference between the groups in the quality of writing as measured by the final-story scores. The conclusions were: the improvement made by the students in the experimental group was significantly greater than that made by the children in the control group; the scale proved to be a reliable instrument when used by either children or adults, and the program taught the children to become reliable raters; based on comparisons of total populations, girls improved significantly over boys in elaboration, structure, and total composition; according to the interaction results of program and sex, the program minimized sex differences in total composition, elaboration, organization, and structure; and correlations between IQ and sample story scores were too low to warrant further investigation. (RB)

Carol Sager
21 Wallis Road
Chestnut Hill, Ma. 02167
(617) 469-0258

IMPROVING THE QUALITY OF WRITTEN COMPOSITION THROUGH PUPIL USE OF RATING SCALE

In an attempt to improve the quality of writing, sixth-grade children were taught to use a descriptive writing scale consisting of four components. Sections on vocabulary, elaboration, organization, and structure were constructed along with a program for teaching. The children learned to rate their own stories and those of other children.

The data were analyzed to determine whether (1) the quality of composition would improve, and (2) the children could use the scale to rate compositions.

RATIONALE AND PURPOSE

Educators have long recognized the need to improve the quality of children's writing in the middle grades. Although there is a limited amount of research in this area, authors agree that children must become actively involved in diagnosing their own writing problems and evaluating their own compositions if improvement is to take place.

Contemporary authors are also in accord regarding the factors which they consider both important and appropriate for improving the quality of writing in the middle grades. These factors are: (1) vocabulary - the use of words to express a particular thought or idea, (2) elaboration - an abundance of related ideas which flow smoothly from one idea to the next, (3) organization - the arrangement of ideas in order, (4) structure - the way language forms are used to convey meaning.

In an attempt to subject these ideas to the rigors of scientific investigation, a program was designed to improve the quality of written composition by teaching children the components and use of a descriptive writing scale so that they could rate their own compositions and those of other children. The purpose of this

study was to determine the effectiveness of this program. More specifically, would children's writing improve in over-all quality when measured on the basis of vocabulary, elaboration, organization, and structure as a result of learning to rate compositions in these same areas?

CONSTRUCTION OF MATERIALS

To implement the purpose of this study it was necessary to construct: (1) a scale that could be used by middle-grade children, (2) a program to teach the components and use of the scale.

PROCEDURES

Selection of the Sample

This study was conducted in a suburb of Boston, Massachusetts. Two sixth grade classes from each of two schools participated in this experiment. The two schools chosen were previously paired by the system's guidance department on the basis of their similarity in regard to socio-economic level and general achievement of children. Although part of a wealthy community, the participating schools were classified as "inner-city" schools and qualified for Title I funds.

Description of Groups

The study consisted of two main groups - an experimental group and a control group - both of which were comprised of two sixth grade classes.

The experimental group followed the program designed to teach the components and use of the scale.

The control group studied the same four components of composition. They were given the definitions used by the experimental group but followed teaching procedures outlined in the school curriculum guide. The teachers in the control group were asked to keep weekly logs of the activities used.

Lessons in both groups were conducted for periods of 45 minutes, five days a week, for eight weeks. Both groups received the same motivations for creative writing and the

same amount of practice in writing.

Testing Procedures

This study extended for a period of ten weeks from October to December, 1970. Two weeks were reserved for testing.

During the first week, objective measures of IQ and writing ability and a sample composition were collected from each child.

During the final week of the study, a Final Story was collected.

In addition, in order to fulfill the purposes of this study, story ratings for ten stories were collected from the children in the experimental group to determine whether children could learn to be reliable raters.

Analyzing the Data

The two main problems presented by this study were: (1) Was the program designed for this study effective in improving the quality of children's writing as a whole as well as the quality of each of the components taught? (2) To what degree of proficiency could children be taught to use a descriptive writing scale to rate their own compositions and those of other children?

The general hypothesis to be tested was that there would be no difference between the groups in the quality of writing as measured by the Final Story scores. Since the groups could not be equated carefully at the beginning of the study, comparisons were made by an analysis of covariance which adjusted the Final Story scores in light of pre-experimental differences.

The proficiency of children as raters was determined by computing an estimate of reliability for the scale as a whole and for each of its four components from among children's ratings.

This study also investigated (1) reliability of the scale when used by adults, and (2) the effect of sex differences, IQ, interaction of program and sex on improvement in writing.

Collecting the Stories

Two story motivations were constructed and labeled Sample Story A and Sample Story B. To equalize any differences in motivations, half of both the experimental and control groups received Sample Story A while the other half of both groups received Sample Story B as the Preliminary Story. For the Final Story, this procedure was reversed.

No time limit was set. Children were allowed to write until they were finished.

Rating of Stories

The quality of written composition in both the Preliminary and Final Stories was measured by the scale constructed for this study. In this case the scale was used by adult raters who were trained to use the scale.

The stories were coded so that the raters would have no idea whether the stories were Preliminary or Final Stories or whether they belonged to the experimental or control group.

CONCLUSIONS

Within the limitations of this study, the following may be stated:

1. Teaching children the knowledge and use of a writing scale so that they could rate their own compositions and those of other children did improve the quality of children's writing. In every area tested, the improvement made by the children in the experimental group was significantly greater than that made by the children in the control group.

2. The estimated reliability of the scale when used by children was in the order of .99 with components ranging from .96 to .98.

3. The estimated reliability of the scale when used by adults was in the order of .97 with components ranging from .81 to .96.

These findings indicate that (a) the scale was a reliable instrument when used by either children or adults, and (b) the

program did teach the children to become reliable raters.

4. Based on comparisons of total populations, sex was a factor which favored the girls in the improvement of the quality of elaboration, structure, and total composition. Sex did not appear to influence improvement in vocabulary or organization.

5. According to the interaction results of program and sex, the program tended to minimize sex differences in the areas of total composition, elaboration, organization, and structure. While the boys in the control group tended to deteriorate in each of these areas and made approximately half the growth made by the control group girls, the boys in the experimental group improved in writing and even did slightly better than the experimental group girls.

6. Correlations between IQ and sample story scores were too low to warrant further investigation.

IMPLICATIONS FOR TEACHING

Statistical analyses have shown that the quality of written composition was indeed improved as a result of teaching children the knowledge and use of a descriptive writing scale. Yet, statistics do not tell the whole story. This study uncovered additional benefits which should also be noted.

In the first place, this program provided the child with the knowledge and means necessary to take responsibility for his own learning. By using the scale, the child could pinpoint his weaknesses, note what was necessary for improvement, and judge his own progress. This seemed to promote a sense of security and confidence which enabled the child to examine writing with a steely and critical eye. Throughout this study, children were able to attack their own stories and those of other children with a vigor and honesty that would be disastrous coming from a teacher. Yet no child felt the need to be defensive about his writing. No child seemed defeated by his mistakes.

Use of the scale also seemed to turn the idea of "writing as rewriting" into an automatic process. As children examined their own stories, comments such as "My sentences all begin the

same way," "I don't use any new and interesting words," "The conclusion doesn't go with the story," could be heard repeatedly. It did not seem to matter that some children were better judges of others than they were of themselves. They were still attempting to diagnose their own writing problems. And if some children were generous in assessing the degree to which they needed improvement, they were at least reviewing their writing from the point of view of making it better.

Another benefit of this program was that it provided a positive language arts experience for every child. This is not to say that every child improved the quality of his written composition, but every child did receive some measure of success, especially in the rating of stories. Even the most reluctant of students became eager, accurate, and vociferous judges. It was not at all unusual for heated debate over scoring to be continued on the playground or in the lunchroom long after the language arts period had ended. Quite often, those involved in these discussions were boys reputed to have little or no investment in school. Statistically, this enthusiasm for the program is somewhat captured by the interaction results.

Finally, this program offers certain benefits for teachers as well. Those who were trained as raters felt that learning to use the scale helped them to define specific student needs and made them more aware of the type of help they could offer. More important, perhaps, is that the program could free the teacher from the task of correcting children's stories. The implications from this alone are numerous and varied. Teachers could use this time to search for more fruitful ways of helping children, develop new practices, provide additional resources and activities. At the very least, they might be encouraged to provide additional time for the teaching of writing in the schools.

IMPLICATIONS FOR FURTHER RESEARCH

1. This study was carried out under a given set of circumstances. The question remains: To what extent can the findings be generalized beyond the particular population that participated in this study? Another question raised is: To what extent were

the results of the study dependent on the specific circumstances under which it was conducted? Therefore, the following are suggested for further research:

a. That this study be replicated using a broader socio-economic range and grade level to see if the results of this study are applicable to other socio-economic and/or grade levels.

b. That this study be replicated using trained teachers to direct the experimental group to see what effect this would have on the results of the studt.

c. That this study be replicated varying the order in which the different components of the scale are presented to see what effects this would have on the results of the study.

2. A problem basic to a rating scale such as the scale used in this study is that periodically it needs to be upgraded. Even during the eight weeks in which the children used the experimental materials, it was noted that in some cases the ceiling of the scale was too low to allow for much growth. Therefore, it is suggested that:

a. The scale be amended to include a higher ceiling and then used in a replication of this study to determine its effect on improving writing for high achievers.

b. Children be given the opportunity to upgrade the scale and a new study be designed and conducted.

FINDINGS AND CONCLUSIONS

Effect of Program on Quality of Written Composition

In order to determine the effectiveness of the program in improving the quality of writing, the adjusted Final Story means of the experimental group were compared with those of the control group.

Tables 1 and 2 present the status of the groups at the beginning of the study.

TABLE 1

GENERAL CHARACTERISTICS OF THE SAMPLE AT BEGINNING OF STUDY

	Experimental Group (N = 41)		Control Group (N = 42)	
	Mean	S.D.	Mean	S.D.
C.A. (mos.)	137.59	6.83	135.09	4.77
IQ	108.63	12.49	106.14	9.79
STEP	258.37	12.11	256.52	11.96

TABLE 2

MEANS AND STANDARD DEVIATIONS OF PRELIMINARY STORY SCORES

	Experimental Group (N = 41)		Control Group (N = 42)	
	Mean	S.D.	Mean	S.D.
Total Score[a]	5.68	2.66	5.67	2.71
Vocabulary[b]	1.02	0.78	1.21	0.71
Elaboration	1.39	0.76	1.36	0.72
Organization	1.71	0.83	1.57	0.85
Structure	1.56	0.73	1.52	0.82

[a] Maximum Total Score = 12.

[b] Maximum score for each subcategory = 3.

Although no attempt was made to equate groups statistically at the beginning of the study, the preliminary data shown above indicate that the groups scored very like in all factors tested.

By means of analysis of covariance, Final Story scores were adjusted for pre-experimental differences in IQ, writing ability, and Preliminary Story scores. The effect of the program on the quality of writing is shown in Table 3.

TABLE 3

EFFECT OF PROGRAM ON QUALITY OF WRITTEN COMPOSITION BASED ON FINAL STORY ADJUSTED MEANS

	Group		Difference	F	P
	Experimental (N=41)	Control (N=42)			
Total Score[a.]	7.89	4.86	3.03	40.91	.0001*
Vocabulary[b.]	1.89	0.98	0.91	34.16	.0001*
Elaboration	1.88	1.13	0.75	20.77	.0001*
Organization	2.20	1.38	0.82	24.58	.0001*
Structure	1.93	1.35	0.58	14.89	.0003*

* Significant at the .01 level or better

[a.] Maximum Total Score = 12.

[b.] Maximum score for each subcategory = 3.

Note: See Tables 9 - 13 for Source Tables.

As measured by the Final Story scores, the growth made by the children in the experimental group was greater than that made by the children in the control group to a statistically significant degree. These findings indicate that in over-all quality of composition, in vocabulary, elaboration, organization, and structure, the program designed for this study produced improvement that was statistically significant.

Thus, the general hypothesis on which this study was predicated, that there would be no difference between groups in the quality of writing, was disproved.

Reliability Among Raters

Using Ebel's formula [1] for intraclass correlation, an estimate of reliability for the scale as a whole and for each of the components was determined from children's ratings. Estimates of reliability were also determined from adult ratings. The results are shown in Table 4.

TABLE 4

ESTIMATE OF RELIABILITY AMONG RATERS

	Children (N=10)	Adults (N=3)
Total Scale	r = .99	r = .97
Vocabulary	r = .98	r = .96
Elaboration	r = .96	r = .87
Organization	r = .96	r = .81
Structure	r = .96	r = .87

These findings indicate (1) The scale was a reliable instrument when used by either children or adults (2) The program did teach the children to become reliable raters.

Effect of Sex Differences On Final Story Scores

To determine the effect of sex differences on the scores of the Final Story, a comparison was made between the Final Story adjusted mean scores of the total population of boys and the total population of girls.

Tables 5 and 6 show the status of the groups at the beginning of the study.

1./ Guilford, J.P. Psychometric Methods. 2nd ed. New York: McGraw-Hill Book Company, Inc., 1954. 395-397.

TABLE 5

GENERAL CHARACTERISTICS OF BOYS AND GIRLS AT BEGINNING OF STUDY

	Boys (N=36) Mean	S.D.	Girls (N=47) Mean	S.D.
C.A. (mos.)	137.84	7.37	135.92	5.45
IQ	106.60	11.75	108.05	10.67
STEP	254.20	11.93	259.65	12.29

The above means and standard deviations show the two groups were similar in these three factors before they were equated statistically.

TABLE 6

MEANS AND STANDARD DEVIATIONS OF PRELIMINARY STORY SCORES FOR BOYS AND GIRLS

	Boys (N=36) Mean	S.D.	Girls (N=47) Mean	S.D.
Total Score[a.]	5.00	2.85	6.19	2.56
Vocabulary[b.]	0.97	0.83	1.23	0.66
Elaboration	1.17	0.76	1.53	0.68
Organization	1.52	0.87	1.72	0.82
Structure	1.33	0.78	1.70	0.74

[a.] Maximum Total Score = 12.

[b.] Maximum score for each subcategory = 3.

An examination of the data in the above table indicates only small differences in score, but shows a very definite trend for the girls to score somewhat higher than the boys.

By means of analysis of covariance, Final Story scores were adjusted for pre-experimental differences and comparisons were made. The results are shown in Table 7.

TABLE 7

EFFECT OF SEX DIFFERENCES BASED ON FINAL STORY ADJUSTED MEANS

	Group		Diff-erence	F	p
	Boys (N=36)	Girls (N=47)			
Total Score[a]	5.88	6.87	0.99	4.72	.0333*
Vocabulary[b]	1.44	1.43	0.01	0.008	.9778
Elaboration	1.34	1.67	0.33	4.67	.03 *
Organization	1.64	1.94	0.30	3.67	.059
Structure	1.42	1.87	0.45	11.53	.0011*

* Significant at the .05 level or better.

[a] Maximum Total Score = 12.

[b] Maximum score for each subcategory = 3.

Note: See Tables 9 - 13 for Source Tables.

The difference between groups in the over-all quality of the Final Story as a whole, in elaboration and structure was statistically significant in favor of the girls.

There was no significant difference between groups in the over-all quality of vocabulary or organization.

These findings indicate that sex was not a factor which influenced the quality of vocabulary and organization. However, sex did appear to be a factor which favored the girls in the areas of elaboration, structure, and total composition.

Interaction Effects of Program and Sex On Quality of Written Composition

An analysis of covariance was also used to determine the joint effects of the program and sex on the quality of writing. Based on the adjusted mean scores for the Final Story, the difference in the Final Story scores between the boys and girls in the experimental group was compared to the difference in Final Story scores between the boys and girls of the control group. The results are shown in Table 8.

TABLE 8

EFFECT OF INTERACTION OF PROGRAM AND SEX ON THE QUALITY OF WRITTEN COMPOSITION BASED ON FINAL STORY ADJUSTED MEANS

	Experimental Group			Control Group		
	Boys (N=19)	Girls (N=22)	Difference	Boys (N=17)	Girls (N=25)	Difference
*Total Score[a]	8.29	7.48	0.81	3.46	6.25	2.79
Vocabulary[b]	1.95	1.83	0.12	0.92	1.03	0.89
*Elaboration	2.0	1.76	0.24	0.67	1.58	0.91
*Organization	2.35	2.05	0.30	0.94	1.82	0.88
*Structure	1.99	1.88	0.11	0.85	1.86	1.01

*Areas of significant interaction at the .01 level or better.

[a]Maximum Total Score = 12

[b]Maximum score for subcategories = 3.

Note: See Tables 9 - 13 for Source Tables.

Analysis of the data showed that there was no statistically significant interaction between Program and Sex in regard to the over-all quality of vocabulary. However, there was significant interaction between Program and Sex in regard to the over-all quality of elaboration, organization, structure, and total composition. In the experimental group, the boys did slightly better than the girls in each of these areas. Yet, in the control group, the boys did approximately half as well as the girls.

These findings indicate that while the program designed for this study was no better for boys than for girls in regard to the over-all quality of vocabulary, it tended to minimize the influence of sex differences on the over-all quality of written composition, as well as on the over-all quality of elaboration, organization, and structure. Stated another way, these findings indicate that the writing of the boys who did not use the experimental materials tended to deteriorate in over-all quality as well as in elaboration, organization, and structure.

Effect of IQ on Improvement in Composition

The correlation between IQ and the story scores were too low to warrant further meaningful investigation. For the purposes of this study, it may be said that IQ was not a factor which influenced improvement in composition.

TABLE 9

SOURCE TABLE FOR EFFECT OF PROGRAM, SEX, INTERACTION OF PROGRAM AND SEX ON TOTAL SCORE BASED ON FINAL STORY ADJUSTED MEANS

Source	df	Mean Square	F	P
Program	1	153.095	40.91	.0001
Sex	1	17.665	4.72	.0330
Interaction of Program x Sex	1	63.41	16.68	.002
Error	76	3.74	-	-

TABLE 10

SOURCE TABLE FOR EFFECT OF PROGRAM, SEX, INTERACTION OF PROGRAM AND SEX ON VOCABULARY BASED ON FINAL STORY ADJUSTED MEANS

Source	df	Mean Square	F	P
Program	1	15.96	34.16	.0001
Sex	1	.0004	.0008	.98
Interaction of Program x Sex	1	.26	.57	.45
Error	76	.47	-	-

TABLE 11

SOURCE TABLE FOR EFFECT OF PROGRAM, SEX, INTERACTION OF PROGRAM AND SEX ON ELABORATION BASED ON FINAL STORY ADJUSTED MEANS

Source	df	Mean Square	F	P
Program	1	8.92	20.77	.0001
Sex	1	1.98	4.62	.03
Interaction of Program x Sex	1	6.38	14.86	.0003
Error	76	.43	-	-

TABLE 12

SOURCE TABLE FOR EFFECT OF PROGRAM, SEX, INTERACTION OF PROGRAM AND SEX ON ORGANIZATION BASED ON FINAL STORY ADJUSTED MEANS

Source	df	Mean Square	F	P
Program	1	10.76	24.58	.0001
Sex	1	1.61	3.67	.059
Interaction of Program x Sex	1	6.78	15.49	.0002
Error	76	.44	-	-

TABLE 13

SOURCE TABLE FOR EFFECT OF PROGRAM, SEX, INTERACTION OF PROGRAM AND SEX ON STRUCTURE BASED ON FINAL STORY ADJUSTED MEANS

Source	df	Mean Square	F	P
Program	1	4.73	14.89	.0003
Sex	1	4.67	11.53	.0011
Interaction of Program x Sex	1	5.92	18.64	.0001
Error	76	.32	-	-