DOCUMENT RESUME

ED 275 713 TM 860 627

AUTHOR Fang, Wei Li; Covert, Robert W.
TITLE Assessing a Federal Approach to Evaluating Evaluations: The Joint Dissemination Review Panel. Can Approved Submissions Be Distinguished from Those Not Approved?
PUB DATE Apr 86
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, San Franciso, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; *Federal Programs; Higher Education; Predictor Variables; Program Costs; *Program Effectiveness; Program Evaluation; *Program Validation; *Validated Programs
IDENTIFIERS *Joint Dissemination Review Panel

ABSTRACT

Three studies identifying the characteristics of program submittals approved by the Joint Dissemintion Review Panel were examined. Between 1978-1985, 358 demonstration programs were submitted. For each of these years, between 51 percent and 83 percent of the submittals were approved. In this study, the submittals were reviewed on the basis of the Joint Panel's criteria for program effectiveness; evidence of impact; evidence of statistical reliability of effects; evidence that the effects are educationally meaningful; interpretability of tests; credibility of evidence; evidence that the effects are attributable to the intervention; and evidence of generalizability. The reviewers also considered the clarity and quality of the submittal itself. The program content areas most frequently included were basic skills (reading, mathematics, and writing), career education, and special education. Approved submittals, in addition to the official seven criteria, also provided adequate descriptions of the program components. Most approvals used quasi-experimental or experimental designs, and nonequivalent or randomized control groups. Although the number of submittals decreased, evaluation designs became more rigid in more recent years. Clear writing and presentation of results appeared to be a critical factor in acceptance. (GDC)

Assessing a Federal Approach to Evaluating Evaluations: The Joint Dissemination Review Panel

Can Approved Submissions be Distinguished from Those Not Approved?

Wei Li Fang, Ph.D.
and
Robert W. Covert, Ph.D.

Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, April 16-20, 1986

Wei Li Fang
School of Medicine
Box 382
University of Virginia
Charlottesville, VA 22908
804/924-2563

Robert W. Covert
Evaluation Research Center
School of Education
University of Virginia
405 Emmet
Charlottesville, VA 22903
804/924-0511

The Joint Dissemination Review Panel is a federal board, consisting of members of the Department of Education and the National Institute of Education. Since its establishment in 1973, the Panel has been responsible for reviewing educational products and practices on the basis of effectiveness criteria.

This paper presents information on three studies which were conducted on the 358 submittals which were reviewed by the Panel during the period 1978-85. The purpose of these studies was to identify those characteristics which distinguished approved submittals from disapproved ones. Each researcher examined the submittals on the basis of the Panel's effectiveness criteria. These effectiveness criteria were operationalized into variables so that the researcher could determine the presence or absence of the variable for each submittal.

Results indicated that one of the key variables in determining whether a submittal is approved is the quality of the presentation of the program components and the evaluation findings. Other predictors of approval included the type of evaluation design which was implemented and the cost of installation.

The purpose of this paper is to present information related to variables that may be used in differentiating between submittals which are approved or rejected by the Joint Dissemination Review Panel (JDRP). Three primary sources of information were used: a dissertation completed by Wei Li Fang in 1981, a dissertation completed by Kathleen Lynch in 1986, and a study completed by George Lam in 1986. Fang's dissertation covered submittals reviewed by JDRP in 1978 and 1979; Lynch's dissertation covered 1980-1983; and Lam's study covered 1984 and 1985. A total of 358 submittals were reviewed by the Panel during this eight-year time period. The number of submittals for each of these eight years is presented in Table 1 below.

Table 1

Number of Approvals and Disapprovals by Year

1978-1985

Year	Approvals		Disapprovals		Resubmittal Approved		Resubmittal Rejected		Total
	n	%	n	%	n	%	n	%	n
1978	23	55	14a	33	4	10	1	2	42
1979	29	55	15a	28	9	17	0		53
1980	31	69	14	31	0		0		45
1981	42	62	24	35	1	1	1	1	68
1982	31	51	25	41	4	7	1	2	61
1983	30	52	20	34	5	9	3	5	58
1984	13	68	6	32	0		0		19
1985	10	83	2	17	0	___	0	___	12
Totals	209	58	120	34	23	6	6	2	358

[a]Eleven of the 29 submittals which were reviewed in 1978 and 1979 were subsequently resubmitted and approved. The remaining 18 disapprovals were not resubmitted.

Table 1 indicates that the number of submittals to the Panel peaked in 1981, with a total of 68 submittals reviewed. In the last two years there has been a notable decrease in the number of submittals. According to the JDRP Executive Secretary, this has been due to decreased funding of the National Diffusion Network. Corresponding to the decrease in the number of submittals is an increase in the percentage of submittals approved. The Executive Secretary noted that the preparation and presentation of submittals have improved in the past two years.

In these three studies one of the major research questions was to identify those characteristics which distinguished approved submittals from those which were disapproved. In order to collect relevant data, JDRP effectiveness criteria were reviewed and operationalized into specific variables. A data collection form was then developed which contained these variables.

Effectiveness Criteria

According to the Ideabook, there are seven effectiveness criteria by which submittals are judged. These state that the submitting project has provided evidence of positive impact, that effects are statistically significant and educationally important, that measures are reliable and valid, that the observed effects can be attributed to the interventic that the evidence is credible, and that the intervention and its effec. .an be replicated in other sites. These criteria were operationalized in the following ways:

JDRP Criterion	Operational Definition of Variable
1. Evidence of impact	* Evaluation design
	* Type of comparison group
2. Evidence of statistical reliability of effects	* Statistical significance
3. Evidence that the effects are educationally meaningful	* Educational importance
4. Interpretability of measures	* Type of instrument (standardized or locally developed)
	* Reliability of the instrument
	* Validity of the instrument
5. Credibility of evidence	* Sampling procedures
	* Test administration procedures
	* Types of reported scores
	* Types of data analysis procedures
6. Evidence that the effects are attributable to the intervention	* Internal validity
7. Evidence of generalizability	* External validity
	* Replication

In addition to these seven criteria, another variable was added. This addressed the clarity or quality of the submittal itself.

Methodology in the three studies was similar. Each submittal was reviewed, and information on the seven JDRP criteria and two clarity variables was recorded on the data collection forms. Frequency data were tabulated on

each of the instrument's items. In the pages that follow, selected variables will be presented for the three studies.

Characteristics of Programs

Table 2 presents the number of programs by content area. The majority of submittals was in the area of basic skills (e.g., reading, math, writing). Of the 428 submittals reviewed by Fang and Lynch, over half (i.e., 224) addressed basic skills.[1] Two other areas that received attention were career education (48 submittals) and special education (36 submittals).

[1]Even though there was only a total of 232 submittals in 80-83, Lynch coded some submittals as having more than one content area. , she reported 345 programs for the 232 submittals.

Table 2

Number of Programs by

Content Area

	Fang			Lynch	
	Approv- als	Rejec- tions	Resub- mittals	Approv- als	Rejec- tions
Basic skills	15 (29)[b]	1 (6)	3 (23)	139 (63)	66 (53)
Career education	9 (17)	4 (22)	2 (15)	14 (6)	19 (15)
Computer literacy[a]					
Natural science[a]				11 (5)	8 (6)
Social Science[a]				11 (5)	6 (5)
Special Education	7 (13)		1 (8)	20 (9)	8 (6)
Health/Phys. Ed.[a]				11 (5)	4 (3)
Gifted Education[a]				4 (2)	3 (2)
Arts/Humanities[a]				3 (1)	0
Early Childhood Education	8 (15)	5 (28)	2 (15)		
Teacher Education	4 (8)	4 (22)	2 (15)	0	4 (3)
Other	9 (17)	4 (22)	3 (23)	8 (4)	6 (5)
TOTAL	52	18	13	221[c]	124[c]

Note: This information was not recorded by Lam.

[a]This was not a category used by Fang.

[b]Numbers in parentheses reflect percentages of column totals.

[c]Some submittals were coded in more than one content area. There was a total of 345 programs for 232 submittals.

The majority of programs that were submitted to JDRP during the eight-year period were for students from kindergarten through twelfth grade. These programs represented 87% of the submittals reviewed. This was a trend noticeable in each of the three studies.

Table 3

Number of Programs by

Educational Level

	Preschool	K-12	Post-Secondary	Total
Fang	12 (14)[a]	64 (77)	7 (8)	83
Lynch	12 (4)	299 (90)	23 (7)	334[b]
Lam	___	18 (95)	1 (5)	19
TOTAL	24	381	31	436

[a]Numbers in parentheses reflect percentages of row totals.

[b]Some submittals reported more than one educational level.

Even though JDRP emphasizes the seven effectiveness criteria, another essential ingredient of the submittal is a description of the components that comprise a program. When Fang reviewed submittals, she recorded whether none, some, or all of the components of the program were adequately described. Lynch used a somewhat different rating scale—she looked at whether some, most, or all of the components were described. These results are presented in Table 4. The majority of approved submittals either had most or all program components described.

Table 4

Number of Programs Which Include

Descriptions of Components

	Components Described				
	None	Some	Most[b]	All	Total
Fang					
Approvals	1[a] (2)[c]	23 (44)		28 (54)	52
Rejections		15 (83)		3 (17)	18
Rejections Later Resubmitted		12 (92)		1 (8)	13
Resubmittal Approved		6 (46)		7 (54)	13
Lynch					
Approvals		34 (24)	55 (38)	55 (38)	144
Rejections	___	34 (39)	39 (44)	15 (17)	88
TOTALS	1	124	94	104	328

Note: Lam did not look at this variable.

[a]While this project provided general information on the evaluation design, no specific information on the intervention's features was presented.

[b]This category was not used in Fang's study.

[c]Numbers in parentheses reflect percentages of row totals.

Findings

As noted previously, one of the JDRP criteria was evidence of impact. Evidence of impact was defined as the type of evaluation design that was implemented and the type of comparison group that was used in the design. In Table 5, four evaluation designs--pre-experimental, quasi-experimental, experimental, and qualitative--were identified by Fang, Lynch, and Lam. While the majority of approved submittals in 1978 and 1979 utilized either a pre-experimental or quasi-experimental design, the majority of approvals for the years 1980-83 were quasi-experimental designs. This trend changed in 1984 and 1985 when the majority of approvals were experimental designs. As time passes, it appears that studies with more rigid controls are being conducted and submitted to JDRP.

The type of comparison group reflects the type of evaluation design that was implemented. As is expected, the majority of the approvals in the years 1978 through 1983 used a nonequivalent control group. The majority of approvals in 1984 and 1985 used a randomized control group as the comparison group. These data are presented in Table 5.

Table 5

Type of Evaluation Design

	Pre-Experimental	Quasi-Exper'l	Exper'l Design	Qualitative	Total
Fang					
Approvals	24 (46)[a]	23 (44)	5 (10)	0	52
Rejections	12 (67)	5 (28)	1 (6)	0	18
Rejections Later Resubmitted	7 (54)	5 (39)	1 (8)	0	13
Resubmittals Approved	6 (46)	5 (39)	2 (15)	0	13
Lynch[b]					
Approvals	35 (18)	127 (65)	33 (17)	1 (1)	196
Rejections	15 (13)	88 (73)	17 (14)	0	120
Lam					
Approvals	10 (43)	0	13 (57)		23
Rejections	3 (38)	1 (13)	4 (50)		8
TOTAL					
Approvals	75	155	53	1	
Rejections	37	99	23	0	

[a]Numbers in parentheses reflect percentages of row totals.

[b]Lynch looked at all evaluation designs that were presented in the submittals. A total of 316 evaluation designs were reported in the 232 submittals.

Table 6

Type of Comparison Group

	Self/ mastery	National Norm	Pre-experimental	Non-equivalent	Randomized	Other	Total
Fang							
Approvals	4 (8)[a]	11 (21)	9 (17)	23 (44)	5 (10)		52
Rejections	2 (11)	6 (33)	4 (22)	5 (28)	1 (6)		18
Rejections Later Resubmitted	0	5 (39)	2 (15)	5 (38)	1 (8)		13
Resubmittals Approved	1 (8)	3 (23)	2 (15)	5 (39)	2 (15)		13
Lynch[b]							
Approvals		59 (25)	35 (15)	111 (46)	33 (14)		238
Rejections		29 (24)	15 (13)	59 (49)	17 (14)		120
Lam							
Approvals		4 (17)	3 (13)		14 (61)	2 (9)	23
Rejections	___	1 (12)	3 (38)	___	4 (50)	0	8
TOTAL	7	118	73	208	77	2	485

[a]Numbers in perentheses reflect percentages of row totals.

[b]Lynch looked at all the evaluation designs and types of comparison groups that were reported. A total of 358 comparison groups were reported in the 232 submittals.

A second criterion of effectiveness was the evidence of statistical reliability of the effects. This was operationally defined as the presentation of significance testing in the submittal's narrative. Both Fang and Lynch looked at this criterion and found that there was no difference between approvals and disapprovals on this variable. It appears that both included significance testing as a part of their data presentation.

Table 7

Inclusion of Significance Testing

	Descriptive Statistics Only	Significance Testing	Total
Fang			
Approvals	9 (17)	43 (83)	52
Rejections	3 (17)	15 (83)	18
Rejections Later Resubmitted	3 (23)	10 (77)	13
Resubmittals Approved		13 (100)	13
Lynch			
Approvals	4 (3)	140 (97)	144
Rejections	1 (1)	87 (99)	88

Note: Lam did not look at this variable.

Interpretability of measures was another criterion of effectiveness. This was operationally defined as the type of instrument that was administered (either standardized or locally-developed) and whether or not reliability and validity information was presented for the instrument. Table 8 presents the data on the type of instruments that were used by the program developers. Some programs administered both standardized and locally-developed instruments, which explains why the number of instruments exceeds the number of submittals. Lynch did not present these data by approved vs. disapproved submittals since she found that this variable made little difference in how a program was reviewed by JDRP. Reliability and validity information on each test is also omitted in this presentation since Fang and Lynch defined these variables somewhat differently.

Table 8

Types of Instruments Used

	Standardized	Locally-Developed	Modified[a]	Other	Total
Fang					
Approvals	42 (35)[b]	78 (64)		1 (1)	121
Rejections	14 (67)	7 (33)			21
Rejections Later Resubmitted	17 (63)	10 (37)			27
Resubmittal Approved	14 (58)	10 (42)			24
Lynch[c]	250 (62)	129 (32)	12 (3)	15 (4)	406
Lam					
Approvals	8 (35)	15 (65)			23
Rejections	7 (88)	1 (12)	___	___	8
TOTALS	352	250	12	16	630

Note: Lynch did not break these down into approvals and disapprovals.

[a]Fang and Lam did not use this category.

[b]Numbers in parentheses reflect percentages of row totals.

[c]Lynch looked at all instruments that were administered and reported. A total of 406 instruments were presented in the 232 submittals.

Other effectiveness criteria assessed evidence of educational importance, credibility of evidence, internal validity, and external validity.

Because Fang, Lynch, and Lam defined and/or presented these variables somewhat differently, these are not discussed here. However, evidence of generalizability, as defined by the inclusion of replication data, is presented in Table 9 below. Table 9 indicates that replication did occur, as demonstrated by data provided on more than one location or more than one time period. Lynch did not break these data down by approvals vs. disapprovals since she found that the majority of submittals included these replication data.

Table 9

Inclusion of Replication Data

Replication Occurred

by Locations or Time Periods

Fang	
Approvals	32 (62)[a]
Rejections	11 (61)
Rejections Later Resubmitted	7 (54)
Resubmittal Approved	9 (69)
Lynch	
Locations	29
Time periods	132
Lam	
Approvals	6 (26)
Rejections	0
TOTAL	226

[a]Numbers in parentheses reflect percentages of approved and disapproved submittals.

In addition to the effectiveness criteria, Fang, Lynch, and Lam looked at how clearly the submittal was written. Presentation of results appeared to be a critical factor in whether a submittal was approved or disapproved by JDRP. The majority of approved submittals were either clearly written (as rated on a five-point scale by Fang and Lam) or were of good quality (as rated on a three-point scale by Lynch). These findings are presented in Table 10.

Table 10

Clarity of Narrative

	Not Clear at All			Extremely Clear		
	1	2	3	4	5	Total
Fang						
Approvals		10 (19)[a]	19 (37)	22 (42)	1 (3)	52
Rejections		5 (28)	9 (50)	4 (22)		18
Rejections Later						
Resubmitted	1 (8)	6 (46)	5 (38)	1 (8)		13
Resubmittal						
Approved		1 (8)	5 (38)	6 (46)	1 (8)	13
Lam						
Approvals	2 (9)	3 (13)	3 (13)	10 (43)	5 (22)	23
Rejections		2 (25)	3 (38)	2 (25)	1 (13)	8

Quality of Submittal

	Poor	Fair	Good	Total
Lynch				
Approvals	5 (3)	50 (35)	89 (62)	144
Rejections	24 (27)	50 (57)	14 (16)	88

[a]Numbers in parentheses reflect percentages of row totals.

The importance of a well-written submittal was further illustrated by Fang and Lynch when they performed a stepwise discriminant analysis procedure on selected variables related to the JDRP review criteria. This procedure was initiated in order to determine whether approved submittals could be distinguished from disapproved submittals on the basis of JDRP review criteria. According to Fang's results, predictor variables included the clarity of the tables, a description of program components, the type of evaluation design, the number of instruments, and the elimination of selection and statistical regression effects. Lynch's results were similar; she found that predictor variables included the quality rating of the submittal, the amount of information that was included in the submittal, the type of evaluation design, and the cost of installation for the intervention. Both studies indicate that the presentation of program components and evaluation findings play a major role in the approval and disapproval of submittals.

Conclusions

1. The number of submittals has varied greatly over the eight-year period and is currently at an all-time low.

2. The quality of the evaluation designs has consistently improved, from roughly 10% in the 1970's to over 50% in 1984-85.

3. The content of the curricular area has varied over time and most likely reflects changes in priorities set by the National Diffusion Network.

4. One of the key variables in determining whether a submittal is approved is the quality of the presentation. This variable was identified as critical in each of the three studies.