DOCUMENT RESUME

ED 133 493 CE 009 331

AUTHOR Goody, Kenneth
TITLE Comprehensive Occupational Data Analysis Programs (CODAP): Use of REXALL to Identify Divergent Raters. Interim Report for Period May 1976-September 1976.
INSTITUTION Air Force Human Resources Lab., Lackland AFB, Tex. Occupational and Manpower Research Div.
REPORT NO AFHRL-TR-76-82
PUB DATE Oct 76
NOTE 19p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Computer Oriented Programs; *Data Analysis; *Individual Differences; *Job Skills; Research Methodology; *Supervisors; *Task Analysis; Task Performance; *Validity
IDENTIFIERS CODAP; Comprehensive Occupational Data Analysis Programs; *REXALL

ABSTRACT

REXALL is a flexible program within the CODAP (Comprehensive Occupational Data Analysis Programs) system designed primarily for analyzing judges' task-factor ratings, and may be used for identifying divergent raters. Divergent raters are those whose ratings are significantly different from the other raters' ratings. They may be the noncooperative raters who simply generate an arbitrary pattern of responses rather than try to follow the instructions, or they may invert the rating scale, or they may actually perceive the tasks differently. This report uses data from an actual study to show how REXALL is used to detect divergent raters, and to decide whether or not to delete them from the study since they may invalidate task means computed from ratings. It then uses the raw data from the study to verify the validity of the decisions made on the basis of the REXALL output. (Editor/WL)

AFHRL-TR-76-82

AIR FORCE HUMAN RESOURCES LABORATORY

COMPREHENSIVE OCCUPATIONAL DATA ANALYSIS PROGRAMS (CODAP): USE OF REXALL TO IDENTIFY DIVERGENT RATERS

By

Kenneth Goody, Sq Ldr, USAF/RAAF Exchange

OCCUPATION AND MANPOWER RESEARCH DIVISION
Lackland Air Force Base, Texas 78236

October 1976
Interim Report for Period May 1976 – September 1976

2

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This interim report was submitted by Occupation and Manpower Research Division, Air Force Human Resources Laboratory, Lackland Air Force Base, Texas 78236, under project 7734, with HQ Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base, Texas 78235.

This report has been reviewed and cleared for open publication and/or public release by the appropriate Office of Information (OI) in accordance with AFR 190-17 and DoDD 5230.9. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service (NTIS).

This technical report has been reviewed and is approved.

WILLIAM H. POPE, Lt Col, USAF
Chief, Occupation and Manpower Research Division

Approved for publication.

DAN D. FULGHAM, Colonel, USAF
Commander

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFHRL-TR-76-82	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) COMPREHENSIVE OCCUPATIONAL DATA ANALYSIS PROGRAMS (CODAP): USE OF REXALL TO IDENTIFY DIVERGENT RATERS		5. TYPE OF REPORT & PERIOD COVERED Interim May 1976 – September 1976 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Kenneth Goody		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Occupation and Manpower Research Division Air Force Human Resources Laboratory Lackland Air Force Base, Texas 78236		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62703F 77340116
11. CONTROLLING OFFICE NAME AND ADDRESS HQ Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base, Texas 78235		12. REPORT DATE October 1976 13. NUMBER OF PAGES 20
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

REXALL	Comprehensive Occupational Data Analysis Programs (CODAP)	
computer programs	rating scales	inverted ratings
job analysis	consequences of inadequate performance	
task analysis	divergent raters	
task factors	rater correlation table	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

REXALL is a very powerful and flexible program within the CODAP (Comprehensive Occupational Data Analysis Programs) system. It was designed primarily for analysing judges' task-factor ratings, and may be used for identifying divergent raters. Divergent raters are those whose ratings are significantly different from the other raters' ratings. They may be the non-cooperative raters who simply generate an arbitrary pattern of responses rather than try to follow the instructions, or they may invert the rating scale or they may actually perceive the tasks differently. Some divergent raters must be eliminated from the study to preserve the validity of the task means computed from the ratings. This report uses data from an actual study to show how REXALL is used to detect divergent raters, and to decide whether or not to delete them from the study. It then uses the raw data from the study to verify the validity of the decisions made on the basis of the REXALL output.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

PREFACE

This research was accomplished under project 7734, Development of Methods for Describing, Evaluating, and Structuring Air Force Occupations; 77340116, Development and Modification of Comprehensive Occupational Data Analysis Programs (CODAP).

The author, an officer of the Royal Australian Air Force, is serving with the United States Air Force under the Exchange Program. Particular appreciation is expressed to Dr. Raymond E. Christal for his guidance and advice.

The views expressed in this report are not necessarily the official views of the United States Air Force or the Department of Defense.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

COMPREHENSIVE OCCUPATIONAL DATA ANALYSIS PROGRAMS (CODAP): USE OF REXALL TO IDENTIFY DIVERGENT RATERS

I. INTRODUCTION

REXALL is a very powerful and flexible program within the Comprehensive Occupational Data Analysis Programs (CODAP) package, being designed primarily for analysing the inter-rater agreement among judges' task-factor ratings. One use of REXALL is for identifying divergent raters. After some general comments on REXALL, and a discussion of the term "divergent rater," this report uses data from an actual study to elaborate on this use of REXALL. The decisions made on the basis of the REXALL output will then be verified by reference to the raw data.

II. GENERAL DISCUSSION AND BACKGROUND

While REXALL was designed primarily for handling task factor ratings, it may be used for analysing and reporting data whenever a number of judges rate a set of items on some attribute. For example, it was applied when judges rated sets of officer job descriptions on various attributes (Christal, 1975), and when a sample of officers rated a set of education profiles on educational suitability for service in a particular utilization field (Watson & Goody, 1975). It may be applied to rankings as well as to ratings.

The items rated in the study used in this report were the 505 tasks in the task inventory for the Medical Service Specialist. A total of 93 first line supervisors rated the tasks on "Consequences of Inadequate Performance." The instructions page from the survey booklet is presented as Appendix A. It contains a definition of the task-factor involved, and describes the 9-point scale against which the judges were to make their ratings. The 9-point scale also appears as Figure 1. These ratings were gathered in order to obtain a measure, the mean rating, of Consequences of Inadequate Performance for each task. REXALL was used to determine whether any of the raters should be deleted from the study; and then to produce a card deck containing the mean task ratings, and to provide a measure of the inter-rater agreement on the ratings.

1. Minimal
2. Slight
3. Not very serious
4. Fairly serious
5. Serious
6. Very serious
7. Extremely serious
8. Almost disastrous
9. Disastrous

Figure 1. **9-point rating scale.**

A general description of REXALL and its relation to other CODAP programs has been documented by Christal and Weissmuller (1976). Figure 2 is a schematic presentation of the inputs to and outputs from the program. The outputs describing the items (tasks) that are rated are self-explanatory from Figure 2. Both outputs describing the raters were illustrated by Christal and Weissmuller (1976). The Inter-rater Reliability Table includes indexes of inter-rater reliability computed by the intraclass correlation formulas reported by Lindquist (1953). The Rater Correlation Table (Example, Table 1) is the tool for detecting divergent raters and will be treated in detail later in this report.

Before discussing Table 1 in detail, the term "divergent rater" needs clarification. This is simply a rater whose ratings are substantially different from those of the other raters. The most common of these are the non-cooperative raters who do not even try to follow the instructions, generating instead some arbitrary pattern of responses. Another type of divergent rater inverts the rating scale—instead of rating from low to

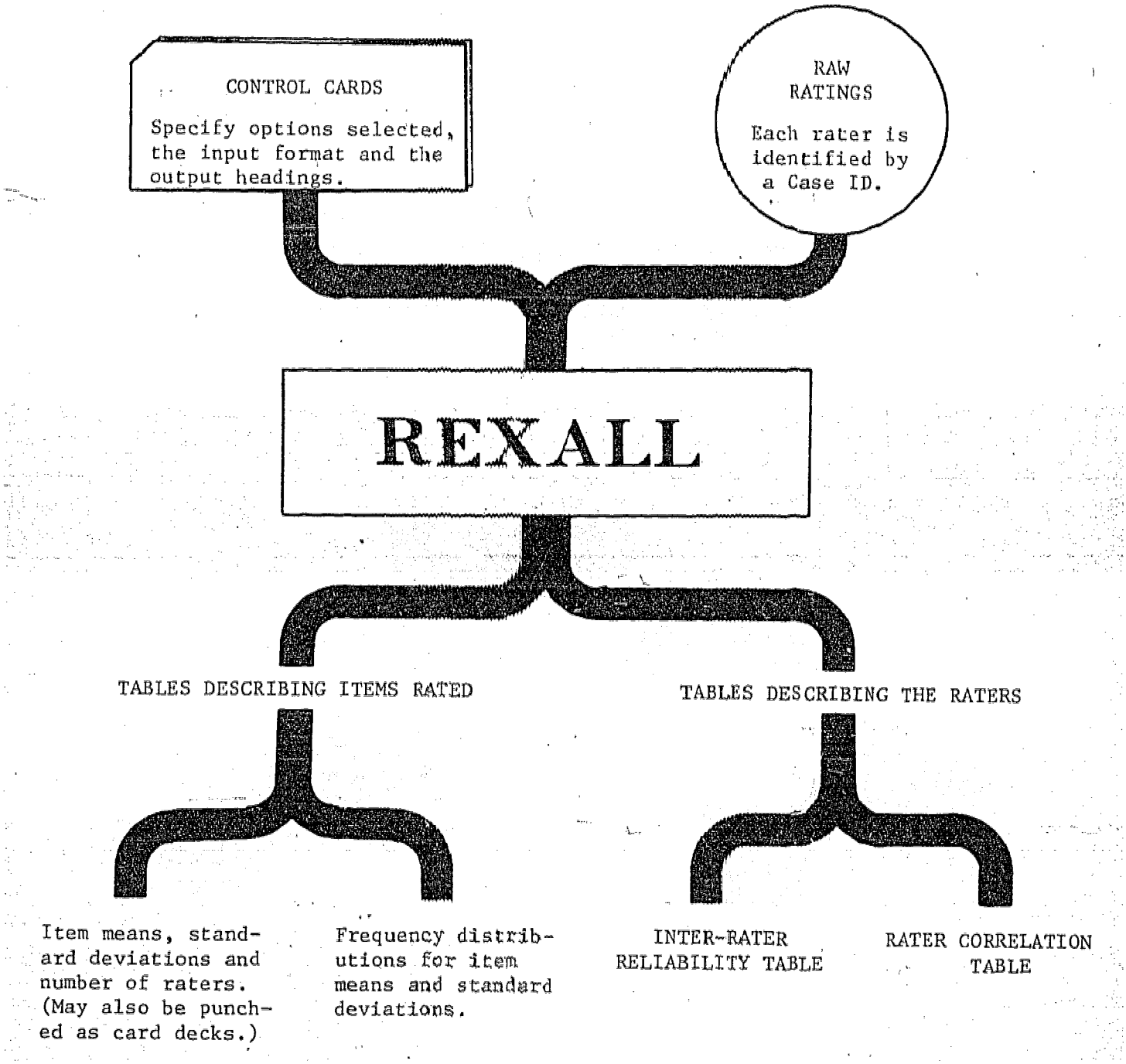


Figure 2. REXALL inputs and outputs.

Table 1. Rater Correlation Table (In Original Input Order)

Rater ID	Correlation	N Ratings	Mean	SD	Sample Mean	T-Value
01	.670	505	3.92	1.00	5.12	20.22
02	.617	446	5.52	1.93	5.12	16.53
03	.690	504	5.10	1.13	5.12	21.35
04	.654	484	7.05	1.19	5.11	18.98
05	.486	503	4.21	.98	5.12	12.45
06	.589	504	7.54	1.47	5.12	16.34
07	.667	505	5.62	1.61	5.12	20.05
08	.664	505	5.00	1.11	5.12	19.91
09	.703	505	5.13	1.62	5.12	22.18
10	.737	505	5.10	2.36	5.12	24.47
11	.680	502	4.83	1.33	5.12	20.74
12	.612	504	6.35	1.35	5.12	17.35
13	.648	505	5.18	1.26	5.12	19.10
14	.662	496	5.83	3.12	5.12	19.64
15	.733	503	4.29	1.58	5.12	24.11
16	.152	505	6.03	1.85	5.12	3.45
17	.561	504	5.20	2.17	5.12	15.19
18	.727	504	5.18	1.84	5.12	23.75
19	.544	505	5.10	1.68	5.12	14.53
20	.681	504	5.94	1.56	5.12	20.86
•	•	•	•	•	•	•
•	•	•	•	•	•	•
•	•	•	•	•	•	•
74	.496	503	5.88	1.34	5.12	12.80
75	.547	505	6.52	1.60	5.12	14.66
76	.511	480	4.64	1.25	5.16	12.99
77	.474	503	5.41	1.01	5.12	12.05
78	.704	505	4.92	1.00	5.12	22.24
79	.553	500	6.29	2.06	5.13	14.81
80	.707	501	6.14	1.11	5.11	22.32
81	.480	502	8.21	1.74	5.12	12.23
82	.727	504	5.92	1.70	5.12	23.75
83	.523	150	1.99	2.00	4.66	7.47
84	.474	95	4.71	.71	4.87	5.19
85	-.115	505	1.02	.15	5.12	- 2.61
86	.567	294	2.61	1.66	5.19	11.77
87	.000	505	9.00	.00	5.12	.00
88	.083	96	6.77	1.39	4.56	.81
89	.708	505	6.72	2.06	5.12	22.50
90	.705	503	4.73	.89	5.12	22.25
91	.518	503	4.50	.64	5.12	13.55
92	.562	504	3.87	.69	5.12	15.24
93	.554	496	3.10	1.08	5.12	14.80

high the rater rates from high to low. As will be demonstrated shortly, both types are easily identified from REXALL's Rater Correlation Table. The non-cooperative rater is usually dropped from the study, and inverted ratings can be arithmetically reversed or dropped from the study at the option of the researcher. Yet another type of divergent rater is not so clear cut. This is the rater whose perception of the items being rated is different from that of the other raters. This difference in perception may be very real or may in effect be a lack of discrimination power or of knowledge on the part of the rater. In any case, the decision on whether or not to retain such a rater in the study is rather subjective and each case must be judged on its merits.

III. USING THE RATER CORRELATION TABLE TO IDENTIFY DIVERGENT RATERS

The Rater Correlation Table produced by REXALL contains a row of data for each rater, printed in the same order in which the cases were input to the program. Table 1 is an extract from the table for the illustrative study being used in this report, raters 21 through 73 being omitted because of space limitations. The correlation column contains the correlations between each rater's ratings and the means for all raters on the same tasks, the extreme right-hand column being a "T Value" for determining the level of significance of deviation of this correlation coefficient from zero. The other four columns are the number of items (tasks) rated by each rater, the rater's mean and standard deviation and the "sample mean." The sample mean is the mean of all raters' ratings on the tasks rated by that rater.

It is this table that is used to decide which raters are divergent and whether or not they should be deleted from the study. The following paragraphs describe how these decisions were established in the study being used to illustrate this report. For ease of explanation, the table entries containing the 20 highest and the 20 lowest correlations have been extracted from the original table and are presented as Table 2. When the program is next revised, it is planned to provide an option to repeat Table 1, rearranged in descending order of correlation coefficient, and also on other variables contained in it.

From Table 2 it is evident that the majority of the correlation coefficients exceed .50 (in fact, 77% exceed .50). Even the "good" raters vary in number of tasks rated and in mean and standard deviation of ratings. However, their high correlations with the group indicate that they share similar perceptions of the relative consequences of inadequate performance of the tasks in the inventory.

The prime indicators of divergent raters are the correlation coefficients. A high negative coefficient indicates inverted ratings, while an insignificant coefficient is the result of a non-cooperative rater. The raters who have made a genuine attempt to obey the instructions, but whose perceptions of the tasks differ from those of the majority of the raters, should have a relatively low, but significant, correlation coefficient. Unfortunately, this condition may also typify the rater who cooperates for part of the inventory and then becomes non-cooperative.

Table 2 will now be used to identify the divergent raters in this study. There does not appear to be any cases of inverted ratings. The last six raters (in Table 2) are apparently non-cooperative raters, and at least the next two require special consideration as their perceptions of the tasks seem to be different from those of the majority of the raters. Each of the last eight raters in Table 2 will now be discussed individually.

Rater 85. The correlation coefficient (−.115) could, in isolation, suggest inverted ratings but the other data available suggests otherwise. All tasks were rated. The rater's mean of 1.02 and standard deviation of .15 suggest that almost every task has been rated 1, with only a few higher ratings. Probably by chance, these higher ratings were allocated to lower than average tasks, giving rise to the negative correlation coefficient. Rater 85 appears to have been non-cooperative and was deleted from the study.

Rater 41. A mean of 1.33 (standard deviation .81) indicates most tasks were rated 1 with a sprinkling of higher ratings, perhaps just to break up the pattern. Rater 41 seems to be another non-cooperative rater and was also dropped from the study.

Rater 87. The mean of 9.00 with a zero standard deviation indicates this rater rated every task 9. This is perhaps the most obvious type of lack of cooperation. Rater 87 was deleted.

Table 2. Rater Correlation Table (Re-ordered from High to Low Correlation)

Rater ID	Correlation	N Ratings	Mean	SD	Sample Mean	T-Value
67	.764	505	5.07	1.43	5.12	26.54
10	.737	505	5.10	2.36	5.12	24.47
15	.733	503	4.29	1.58	5.12	24.11
82	.727	504	5.92	1.70	5.12	23.75
18	.727	504	5.18	1.84	5.12	23.75
35	.726	503	4.49	1.44	5.12	23.66
44	.720	503	4.81	1.52	5.12	23.23
65	.720	494	4.19	1.07	5.12	23.00
89	.708	505	6.72	2.06	5.12	22.50
80	.707	501	6.14	1.11	5.11	22.32
90	.705	503	4.73	.89	5.12	22.25
78	.704	505	4.92	1.00	5.12	22.24
09	.703	505	5.13	1.62	5.12	22.18
21	.703	502	4.86	1.29	5.12	22.10
60	.699	479	4.41	1.40	5.14	21.35
72	.695	505	4.04	.84	5.12	21.69
34	.694	503	4.75	1.13	5.11	21.60
03	.690	504	5.10	1.13	5.12	21.35
64	.689	375	4.72	2.31	5.23	18.34
53	.687	505	6.13	1.70	5.12	21.18
•	•	•	•	•	•	•
•	•	•	•	•	•	•
•	•	•	•	•	•	•
74	.496	503	5.88	1.34	5.12	12.80
05	.486	503	4.21	.98	5.12	12.45
81	.480	502	8.21	1.74	5.12	12.23
77	.474	503	5.41	1.01	5.12	12.05
84	.474	95	4.71	.71	4.87	5.19
31	.454	479	8.44	1.14	5.14	11.12
27	.439	504	5.50	2.06	5.12	10.96
24	.396	502	3.62	1.25	5.12	9.64
52	.377	505	2.41	2.48	5.12	9.12
48	.358	505	4.98	.94	5.12	8.60
68	.327	505	3.93	.51	5.12	7.76
63	.245	384	3.02	2.64	5.09	4.94
54	.194	504	7.79	2.03	5.12	4.44
16	.152	505	6.03	1.85	5.12	3.45
61	.106	505	5.14	.96	5.12	2.38
88	.083	96	6.77	1.39	4.56	.81
22	.047	505	2.69	2.09	5.12	1.07
87	.000	505	9.00	.00	5.12	.00
41	-.012	505	1.33	.81	5.12	- .26
85	-.115	505	1.02	.15	5.12	- 2.61

Rater 22. This combination of very low correlation (.047), low mean (2.69) and high standard deviation (2.09), when the rater has responded to every task, suggests a propensity of 1 or 2 ratings, with a number of higher ratings (some up near the top of the scale) allocated indiscriminately among the other tasks. Rater 22 was treated as a non-cooperative rater and deleted.

Rater 88. This rater responded to only 19% of the tasks. Although this 19% of the tasks tended to be less demanding than average (sample mean = 4.56; overall mean = 5.12), this rater averaged 6.77 on them. The ratings that were provided are therefore very high, and a correlation of .083 suggests they are not very realistic. The insigificant correlation and low number of tasks rated justify deleting this rater on the grounds of lack of cooperation.

Rater 61. This case approaches the doubtful zone. The correlation is too low to believe the rater has been entirely cooperative; and it is too low to assume a simple difference in perception of the tasks. Rater 61 was dropped from the study as being non-cooperative.

Rater 16. Apart from the relatively low correlation coefficient, this rater's statistics seem fairly normal, except that the mean is perhaps a little high. There seem to be three possibilities: there is either a genuine difference in perception, or a lack of discrimination power, or a lack of complete cooperation. A clear decision cannot be made on the data available. As there are plenty of raters, and this one represents an isolated opinion if the ratings are genuine, the analyst chose to drop him from the study even though retention would not have seriously affected the task means and the inter-rater reliability statistics.

Rater 54. The analysis for this rater is very parallel to that for Rater 16. However, the very high mean (7.79) and standard deviation (2.03) indicates a very large number of 9 ratings, perhaps genuinely believed to be justified. This rating pattern has depleted the rater's discrimination power at the high end of the scale and thus caused the relatively low correlation coefficient. The decision on whether or not to delete this rater should have little effect on the objectives of the study. While an equally good case could be made for deletion, the analyst chose to retain this rater in the study.

The next six to ten raters above Rater 54 in Table 2 also have appreciably lower correlations than the majority of raters and could be classed as divergent raters. However, their ratings are considered genuine and the lower correlations can usually be explained. For example, the analysis for Rater 81 would parallel that for Rater 54; Raters 63 and 52 are the same except that by favoring the low end of the scale they have depleted their discrimination at that end of the scale. Raters 68, 48, and 84 demonstrate their low discrimination power by their lack of variation in ratings. As implied by the analysis for Rater 54, more than seven raters could have been deleted without significantly affecting the objectives of the study either way. Where to "draw the line" must remain, for the time being, a subjective decision on the part of the analyst. One more objective approach being examined is to have the program progressively eliminate the most divergent raters, one at a time, until the inter-rater reliability statistic for the stability of the item (task) means (R_{kk}) ceases to increase.

At this point, one further comment on divergent raters should be made. As is seen from Tables 1 and 2, there are considerable differences among the raters' means and among their standard deviations, caused by different perceptions of the words used to describe the levels on the 9-point rating scale. However, the magnitude of the correlations for the cases retained in the study are satisfactory evidence that these raters shared sufficiently similar perceptions of the relative consequences of inadequate performance of the tasks in the inventory. As it is a measure of the relative consequences that is required, no rater was declared divergent on the grounds of a high or low mean. In fact, these differences in means and standard deviations only add within task variance that is not justified when relative measures are being sought. Accordingly the standardization option discussed by Christal and Weissmuller (1976) was used in this study to remove the between rater variance.

The effects of deleting the seven divergent raters is presented in Table 3. Deleting the divergent raters, the ones who did not cooperate and those who perceived the relative values of the tasks differently, does appreciably increase the inter-rater reliability and therefore increases the stability of the task means. This applies whether or not the data are standardized.

Table 3. Effects of Deleting Seven Divergent Raters

	N	R_{1k}	R_{kk}
Raw ratings, 0 deletions	93	.151	.941
Raw ratings, 7 deletions	86	.195	.952
Standardized ratings, 0 deletions	93	.313	.976
Standardized ratings, 7 deletions	86	.355	.979

IV. VERIFICATION OF DECISIONS BY REFERENCE TO RAW DATA

Having made the above decisions based on the REXALL output, a copy of the raw data for each rater was obtained to examine their validity. Extracts of the raw ratings are presented in Appendixes B, C, and D as blocks of raw ratings, each block containing all the ratings made by one rater. Each block contains eight rows of digits, 69 digits in each of the first seven rows and 22 in the last row. This is a total of 505 digits, one for each task in the inventory. Each digit is a task rating by the rater to which the block refers, a 0 indicating the rater did not rate the task. The first digit on the first row is the rating for Task 1, the last on that row is Task 69, and so on, the last digit on the last row being the rating on Task 505.

Appendix B is the ratings for three "good" raters, provided for comparison purposes. It would appear that the raters have considered and rated each task individually. This does not mean, of course, that their ratings are identical. Because of differences in experience, work location, etc., there will be some differences in perception, and of course there will always be some error variance.

In contrast with Appendix B, Appendix C is the raw ratings for the six raters assessed as being non-cooperative. Each will now be discussed in turn, reflecting on the diagnoses previously made from Table 2 (the REXALL output).

Rater 85. As predicted, Rater 85 rated nearly every task 1. The seven 2 ratings and one 3 rating were all allocated to tasks which most raters considered less consequential than average. For example, the one task that was rated 3 was "schedule leaves or passes." This is hardly as consequential as, for example, tasks involving rendering emergency treatment to a patient, which were obviously rated lower by Rater 85.

Rater 41. The prediction was that most of this rater's ratings would be 1 with a sprinkling of higher ratings. Notice the three 9 ratings. These three tasks were: (a) plan records maintenance, (b) direct physical exercise or conditioning programs, and (c) direct preparation and maintenance of records or reports. Considering the number of tasks in the inventory that could result in the immediate death of a patient, it is evident this rater was non-cooperative. Perhaps he was trying to exercise his sense of humor.

Rater 87. Every task was rated 9 as anticipated.

Rater 22. A propensity of 1 and 2 ratings was predicted, with some high ratings allocated indiscriminately. This diagnosis is confirmed. The indiscriminate nature of the high responses can be illustrated by considering two tasks from the aeromedical evacuation duty: (a) 432. make up litters, and (b) 434. operate inflight emergency oxygen systems. The mean overall ratings on these two tasks were 3.56 and 6.30, respectively, but this rater rated the first task "almost disastrous" (an 8) and the second as "minimal" (a 1). It is probable that the highly rated tasks are those with which the rater is directly involved. Rater 22 did not comply with the instructions.

Rater 88. The raw ratings confirm the analysis from the data in Table 2. A faint-hearted attempt seems to have been made at the early tasks and the rater has left the rest of the booklet blank. This pattern is fairly common among non-cooperative raters who seem to believe no one will ever detect their lack of cooperation if the survey booklet appears to have been honestly completed.

Rater 61. The pattern of this rater's raw responses explains his low correlation. The ratings for the first 71 tasks are very consistent with those of the group. However, at that point the rater has become non-cooperative and rated all but one of the remaining tasks 5. To examine how good this rater's first 71 responses were, the table entries that would have resulted had tasks 72 through 505 not been rated at all, were computed. They appear as Table 4. These statistics are quite acceptable. It would have been far better if Rater 61 had left these remaining tasks blank.

Table 4. **Rater Correlation Table Entries for Rater Number 61 with Tasks 72-505 Treated as Blank**

Correlation	=	.637
N Ratings	=	71
Mean	=	6.03
Standard Deviation	=	2.34
Sample Mean	=	4.51
T-Value	=	6.87

The raw data for the two borderline cases (Raters 16 and 54) are presented in Appendix D. As suggested in the earlier analysis, their ratings should appear fairly normal unless they were partially cooperative raters (like Rater 61). Except for the abnormally large number of high ratings, particularly for Rater 54, both rating patterns seem reasonable. As already mentioned, the decision on whether to accept or reject these two raters is rather subjective, either course having little effect on the mean ratings of the tasks. Rater 54 has given more valid ratings than Rater 16, although they are far from perfect. There seems to be no reason to change the previous decision which was to accept one and reject the other.

V. CONCLUSION

All the decisions regarding divergent raters in this study, made on the basis of REXALL's Rater Correlation Table, have been verified by reference to the raw data. There were no cases of inverted ratings in this study, although arithmetical inversion of such ratings in another study conducted by the author did testify to the validity of treating high negative correlations as indicators of scale reversal. The interpretation of relatively low, but significant, correlation coefficients must remain somewhat subjective for the time being. More objective approaches to the handling of such cases are currently under consideration by the Air Force Human Resources Laboratory.

REFERENCES

Christal, R.E. *Systematic method for establishing officer grade requirements based upon job demands.* AFHRL-TR-75-36, AD-A015 756. Lackland AFB, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory, July 1975.

Christal, R.E., & Weissmuller, J.J. *New CODAP programs for analyzing task factor information.* AFHRL-TR-76-3, AD-A026 121. Lackland AFB, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory, May 1976.

Lindquist, E.G. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953, 359–361.

Watson, W.J., & Goody, K. *Matching job education requirements with candidates' educational attainments – A pilot methodological study.* AFHRL-TR-75-79, AD-A025 214. Lackland AFB, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory, December 1975.

APPENDIX A. INSTRUCTIONS USED BY JUDGES WHEN RATING TASKS IN MEDICAL SERVICES SPECIALIST INVENTORY ON CONSEQUENCES OF INADEQUATE PERFORMANCE

Explanation

This booklet contains a listing of tasks performed in your career ladder. You are asked to rate each task to indicate the Probable Consequences of Inadequate Performance of the task. In the Air Force, the consequences of inadequate performance of some tasks are much more serious than for other tasks. For example, if inadequate performance of a task will almost certainly cause an aircraft to crash, or a warehouse to burn down, or an airman to die, this would be more serious than inadequate performance of a task which merely causes inconvenience and irritation. As another example, the probable consequences of inadequate performance in responding to a fire alarm would be much more serious than the probable consequences of inadequate performance in folding hospital linen.

Definition

Consequences of Inadequate Performance is a measure of the seriousness of the probable consequences of inadequate performance of a task. It is measured in terms of possible injury or death, wasted supplies, damaged equipment, wasted man-hours of work, etc.

Your Task

Using the rating scale below, assign a numerical rating to each task in this booklet which you feel describes the probable consequences of inadequate performance of the task. Make your ratings by simply writing a number 1 through 9 in the column to the right of each task. Please attempt to rate all tasks.

Rating Scale

If the task is <u>not</u> done correctly, the probable consequences of inadequate performance would be:

1. Minimal (inadequate performance has minimal consequences)
2. Slight
3. Not very serious
4. Fairly serious
5. Serious
6. Very serious
7. Extremely serious
8. Almost disastrous
9. Disastrous (inadequate performance has disastrous consequences)

Your efforts in completing this booklet will be sincerely appreciated. When you have finished your ratings, please return this booklet to your CBPO/DPMPC.

APPENDIX B. RAW RATINGS FOR THREE TYPICAL "GOOD" RATERS

Rater Number 15

411322213034542434332432332234232243333323222444331433433533424333343
433343333434335334333233434333332233333324333433344343333333353333432 22
426343443346664554333335454466665333333236634555555555365665455434667
777773377764553444444556565544566775665555555566555545533433343334 43
547554436548555677944456654444339546545394655543433577544444444433 34
444434322344334332263344232234669777577525456569845945497496669064635
554454433345333473364755475433443766634555553333335476976645594455665
596557754555664499356 8

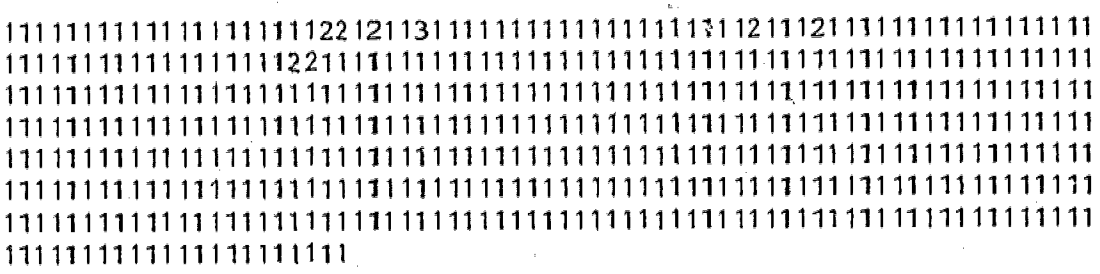
Rater Number 44

212455534543551445004332455224141544455211333356411445353512535455 15
435554555353555555333233333222135534524333533445663333333333663633 43
337555654457774844644367566456783665666344575554465557585445355333555
657583455565444333456466534544665654466665754567777466553355565665752
447633336646565676754356575765443854544568468854453376665655666552265
665355444466555533454455454446767555466455555567858845457577368344345
465554555556355653565565455554564766675555787777775867775567866555 88
595556785555775799335 7

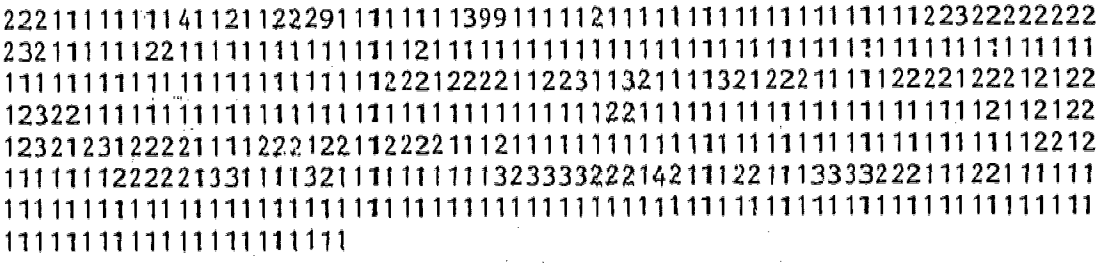
Rater Number 53

323132422543543437413472346345256355454433433244731644556573573636644
555654755652744635544556844544564777462866666655476555465555567555544
658666766668996855855566588589993675555468955776697695555555355655668
779995788985576766887888878888988887778887878888777788786667778777776
669777758858667888777856665888555975677879578787765688767787777745 87
777788555677776752775258563346888988566566878898788877897888897766 78
787775666776377883788877686857785866576558576678787859756855958778 87
797758954534477778557 7

Rater Number 85



Rater Number 41



Rater Number 87

Rater Number 22

415275221471971871112251172161131227326111624211111156666584475748324
544477557475844477447566242222225572642646421226212421112461111113142
111717555555221222111111311211524557252271511171122112113112116111111
112111111112115243331224342535221311134442142442555231111117541525112
431151112515252135141111225555522574111355412211572223222191244421152
111119111186117171814115151111111111131124551116111111111121666226617
114664711111111181111111111111131112111111121112111131111181122111
1111124411111111151111

Rater Number 88

656755566676556667877679155565679876755566678977799898988899877767679
867897777888977766665055670000907000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000

Rater Number 61

721475436984991474497772474655466648797971665899999794896683692755529
945555555555555555555555555555555555555555555555555555555555555555555
555555555555555555555555555555555555555555555555555555555555555555555
555555555555555555555555555555555555555555555555555555555555555555555
555555555555555555555555555555555555555555555555555555555555555155555
555555555555555555555555555555555555555555555555555555555555555555555
555555555555555555555555555555555555555555555555555555555555555555555
5555555555555555555555

APPENDIX D. RAW RATINGS FOR TWO BORDERLINE CASES

Rater Number 16

7578845569789849776688747866646788675544467787786477686887678587897 8
876887787868987778678865878999897988997977788898999989999989999999 9
999989999898999987988899898899889988888777897777798888887758577566577
789894566877786544565465633544475665476655868668877644454654557554763
537465326727336455454656464667653775564559578866454488754574454456456
667557445574555443454355353334646777556455674745755533464475557443334
544443464544344343444345464646454643445355454434455544555645654637545 54
473624563455433456345 5

Rater Number 54

947795569996995889777997989799588859999895977898988779988997987999999
955999889999889799976999789551415557995799999987999455599986699999498
991222996669999999999669999999999979999999999999999999999999799999
999999999999999999999922255929999999999779577777799977777799799997779
999995559999999999979999999975999999999997999999995999895999999995999
995777799979999755998965252998999997999999299999959559599175932599 45
999999997779195995995999999999949599997999999999999897729999902122 91
5677224922299927999298

☆U.S. GOVERNMENT PRINTING OFFICE: 1977-771-057/11