

THE USE OF MICROCOMPUTER PROGRAMS TO
IMPROVE THE RELIABILITY AND VALIDITY
OF CONTENT ANALYSIS IN EVALUATION

Richard D. Frisbie, Ed. D.

Western Michigan University, 1986

The purpose of this study was to advance the body of knowledge about how practicing evaluators can use microcomputer programs to obtain reliable and valid content analyses of responses to open-ended survey questions. This is important because practitioners often analyze such responses as part of larger evaluation efforts. To address this general problem, content analysis experts have developed methods of survey and discovery to help delineate new category systems and obtain codes for responses based on existing categories. Fortunately, some of the methods can be implemented with general purpose microcomputer programs. For example, key words out of context--word lists--can be generated using some spelling checker programs. In addition, information retrieval--sorting responses by category--can be accomplished with data base management programs.

Reliability and validity of the categories developed or codes assigned were the dependent variables for two experiments. It was hypothesized: (a) participants who used the specialized output would create more reliable and valid category systems than those who did not use it (Experiment 1); and (b) participants who used the specialized output would produce more reliable and valid response codes than those who did not use it (Experiment 2).

The experiments were based on a simulated evaluation effort in which fictitious teachers responded to an open-ended survey question about their school's controversial accountability system. The participants, College of Education students, were asked to first create a category system for the responses and then code them into the final category system.

The Experiment 1 null hypotheses were retained and the Experiment 2 null hypotheses were rejected. In other words, the specialized output did not appear to help experimental participants create more reliable and valid category systems; but it did help them obtain more reliable and valid codes for the responses.

The results of Experiment 2 support using specialized output to help inexperienced analysts code relatively long responses using an established category system. Future studies can focus on factors possibly related to retention of the Experiment 1 null hypotheses. These factors include the types of specialized microcomputer output used, responses analyzed, and participants used.

THE GRADUATE COLLEGE
WESTERN MICHIGAN UNIVERSITY
KALAMAZOO, MICHIGAN

Date March 26, 1986

WE HEREBY APPROVE THE DISSERTATION SUBMITTED BY

Richard D. Frisbie

ENTITLED THE USE OF MICROCOMPUTER PROGRAMS TO IMPROVE THE
RELIABILITY AND VALIDITY OF CONTENT ANALYSIS IN EVALUATION

AS PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF Doctor of Education

Educational Leadership
(Department)

Mary Anne Bunda
Dissertation Review Committee Chair

James R. Sanders
Dissertation Review Committee Member

David L. Stupples
Dissertation Review Committee Member

Howard Poole
Dissertation Review Committee Member

APPROVED

Laure Brotzinger
Dean of The Graduate College

Date April 1986

THE USE OF MICROCOMPUTER PROGRAMS TO
IMPROVE THE RELIABILITY AND VALIDITY
OF CONTENT ANALYSIS IN EVALUATION

by

Richard D. Frisbie

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Education
Department of Educational Leadership

Western Michigan University
Kalamazoo, Michigan
April 1986

Dedicated to the memory of my father, Stanley K. Frisbie,
and to my mother, Norma E. Frisbie.

ACKNOWLEDGMENTS

This study could not have been completed without the hard work and support of a wide variety of individuals. These people provided resources, performed specialized tasks, supervised my work, and gave me the inspiration to continue.

Dan Stufflebeam gave me a place to "temporarily" sit down and think about evaluation, life, and how to play with those new toys--microcomputers. He also made the resources of the Evaluation Center available to me so I could actually perform this study. Because I was fortunate enough to win one of the Harold and Beulah McKee Graduate Research Fellowships, I also was able to more quickly complete the study and disseminate its results to the academic community.

The simulated evaluation project required the participation of several different groups of people. The Response Panel developed a set of responses to an open-ended survey question and created a conceptual framework for analyzing them. These panel members were Mary Anne Bunda, Ron Crowell, Richard Haring, and Tom Ryan. Dave Cowden and his students participated in the pilot study used to test the experimental procedures. Mary Anne Bunda, Joe Chapel, Ron Crowell, Richard Haring and their students participated in the full-scale simulation on which the experiments were based. The Hierarchy Panel analyzed the categories developed by the pilot study and experiment participants, and created a comprehensive framework for classifying all of them. Next to writing the dissertation, this was the

most difficult task of the whole study. The people on this panel were Cynthia Halderson, Dennis Shouse, and Eileen Stryker.

I was very fortunate to assemble a highly talented and knowledgeable doctoral committee. Simply asking myself how they would react to something I was about to do always made me work a little more and a little harder before they found out about it. When they were given the extra advantage of actually seeing something I had written, they always gave me solid advice on how to make it even better. Howard Poole, Jim Sanders, and Dan Stufflebeam all made this a better effort by their presence on my committee. Mary Anne Bunda, my doctoral committee chair, allowed me to follow my feelings or use my head, whenever either seemed appropriate. Not only that, she made sure I finished the dissertation within this lifetime.

Behind all this scholarly activity, my family stood in unwavering support. My puppies, Shanti and Akasha, were always ready to sit on my foot, steal my socks, or go for a good race through the snow. My babies--Amanda, Aleisha, and one to be named at a later date--gave me the best of reasons to be done and to get on with the rest of our lives. My brother and sisters--Stan, Shilah, Jo-Lynn, and Sandie--encouraged me to do well, and they convinced me I could. My parents, Stan and Norma, gave me life. They also gave me the spirit to strive for and become whatever I chose to be. Finally, my soul mate, Linda, gave me her love and showed me the path to a new awakening.

Richard D. Frisbie

TABLE OF CONTENTS

ACKNOWLEDGMENTS ii

LIST OF TABLES vii

LIST OF FIGURES ix

CHAPTER

I. INTRODUCTION 1

 Need for the Study 1

 Purpose and Objectives of the Study 4

II. REVIEW OF THE LITERATURE 8

 Introduction 8

 Key Concepts for Identifying the Relationships
 Between Evaluation and Content Analysis 9

 Key Components of Evaluation and
 Content Analysis Information 10

 Actions of Evaluation and Content Analysis 22

 Standards of Quality 27

 Working Definitions of Evaluation
 and Content Analysis 36

 An Overview of Evaluation 37

 Classification of Approaches 38

 Summary of Approaches 42

 A General Model for Conducting
 an Evaluation Effort 52

 An Overview of Content Analysis 57

 Classification of Uses 57

 Summary of Uses 59

Table of Contents--Continued

CHAPTER		
	Tasks of Content Analysis	74
	Performing Content Analysis	
	Tasks with Microcomputers	87
	Summary	92
III.	METHODOLOGY	96
	Introduction	96
	Overview of the Study	96
	Review of the Problem	96
	Simulation Activity	99
	Design of the Study	100
	Independent Variables	104
	Research Hypotheses	104
	Study Group and Sample	105
	Procedures	108
	Content Analysis Activities Performed by	
	the Researcher and Expert Panelists	109
	Experiments	111
	Pilot Study	121
	Data Analyses	125
	Measures of Dependent Variables	125
	Analyses of Dependent Variables	130
	Summary	131
IV.	RESULTS	133
	Introduction	133
	Effects of Microcomputer Output	
	on Category Development	134
	Category Reliability	135

Table of Contents--Continued

CHAPTER		
	Category Validity	138
	Effects of Microcomputer Output on Response Coding	140
	Coding Reliability	141
	Coding Validity	144
	Comparisons of Those Who Did or Did Not Complete Each Experiment	146
	Summary	152
V.	DISCUSSION	153
	Introduction	153
	Interpretation of the Results	153
	Limitations of the Study	154
	Design Limitations	154
	Plausible Explanations for the Retention of the Null Hypotheses of Experiment 1	155
	Plausible Explanations for Differential Participation	160
	Benefits of the Study	163
	Suggestions for Future Research	165
	Closing Comments	168
APPENDICES		
A.	Development of the Response Pool	170
B.	Microcomputer-Implemented Content Analysis Activities	184
C.	Development of the Category Hierarchy	195
D.	Simulation Materials for Session 1	205
E.	Researcher/Instructor Contract	228
	BIBLIOGRAPHY	233

LIST OF TABLES

1.	Surface and Underlying Contrasts for Evaluation and Content Analysis Information	15
2.	Four Types of Information Derived From Underlying Evaluation and Content Analysis Contrasts	15
3.	Six Basic Actions for Conducting an Evaluation or Content Analysis Effort	26
4.	Classifications of Reliability by Type of Agreement Represented	32
5.	Classifications of Validity	35
6.	Classification of Approaches for Conducting Evaluations Based on Epistemology, Major Perspective, and Orientation	41
7.	Summary of Approaches for Conducting Evaluations	44
8.	Uses of Content Analysis by Purpose, Communication Element, and Question	60
9.	Group Means and Standard Deviations for Hypothesis 1, Category Reliability	137
10.	Summary of t-Tests Between Observed Mean Scores and Corresponding Expected Scores Due to Chance Alone for Hypothesis 1, Category Reliability	137
11.	One-Way Analysis of Variance for Hypothesis 1, Category Reliability	138
12.	One-Way Analysis of Variance for Hypothesis 2, Category Validity	139
13.	Group Means and Standard Deviations for Hypothesis 2, Category Validity	139
14.	Group Means and Standard Deviations for Hypothesis 3, Coding Reliability	143
15.	Summary of t-Tests Between Observed Mean Scores and Corresponding Expected Scores Due to Chance Alone for Hypothesis 3, Coding Reliability	143

List of Tables--Continued

16.	One-Way Analysis of Variance for Hypothesis 3, Coding Reliability	144
17.	One-Way Analysis of Variance for Hypothesis 4, Coding Validity	145
18.	Group Means and Standard Deviations for Hypothesis 4, Coding Validity	145
19.	Characteristics of Students Enrolled in the Six Study Classes	147
20.	Summary of Chi-Square Tests of Differences on Selected Variables Between Those Who Did or Did Not Complete Experiment 1	150
21.	Summary of Analysis of Variance Tests of Differences on Selected Variables Between Those Who Did or Did Not Complete Experiment 1	150
22.	Summary of Chi-Square Tests of Differences on Selected Variables Between Those Who Did or Did Not Complete Experiment 2	151
23.	Summary of Analysis of Variance Tests of Differences on Selected Variables Between Those Who Did or Did Not Complete Experiment 2	151
24.	Summary of Analysis of Variance Tests for the Four Study Hypotheses	152

LIST OF FIGURES

1.	The Key Components of Evaluation and Content Analysis Information in Terms of Their Underlying Relationships . .	19
2.	The General Model for Conducting an Evaluation Effort . .	53
3.	Decision Network Using the General Model for Conducting an Evaluation Effort	55
4.	Uses of Content Analysis in Terms of the Working Definition and the Communication Paradigm	62
5.	Schematic Representation of the Two Experiments in Terms of the General Model for Conducting an Evaluation Effort	101
6.	Design of the Study for Two Procedurally Overlapping Experiments	103
7.	Relationships of the Experimental Designs to the Experimental Procedures	112