

Appendix A

Development of the Response Pool

Development of the Response Pool

The response pool was created for two basic purposes. First, it provided the central focus for designing the simulation activity used in both experiments. Second, the category system and codes for the responses created during the activity provided the basis for judging the validity of the participants' content analyses of the responses. The response pool development tasks were: (a) select a general source of information from which a response pool could be developed, (b) select one open-ended question and its related responses, (c) determine a basic classification framework, (d) prepare an oversized pool of potential responses, (e) recruit a panel of individuals familiar with the type of information from which the response pool would eventually be created, (f) have the panelists independently create a category system consistent with the basic classification framework selected and code a major portion of the oversized pool of potential responses, (g) process the panelists' work, (h) have the panelists meet to mutually define 10 categories and assign the set of about 200 responses to those categories, (i) select the final five categories and 100 responses, (j) divide the response pool into two groups with roughly equal numbers of responses from each category in each group and (k) produce a word count list of all 100 responses.

Select Source of Information

The first task was to select a general source of information from which a response pool could be developed. An ideal source of

responses would be a non-trivial, real-world evaluation report that contained at least one open-ended question and a large, diverse, unedited set of responses to such questions.

An evaluation report that meets these criteria is described by Patton (1980, pp. 23-30). The evaluation report included a summary of the findings from a mail survey given to the teachers of a public school district in the Midwest that used a controversial accountability system. The survey information was intended to help evaluate the accountability system from the perspective of the teachers. The questionnaire included a number of forced-choice questions and two open-ended questions. Three hundred seventy-three teachers (70 percent of those who responded to the questionnaire) responded to at least one of the open-ended questions. All of the comments written by the teachers were typed verbatim and included in the report, filling 101 single-spaced pages. The researcher obtained a copy of this evaluation report (Patton, French, & Perrone, 1976) and used it as the basis for developing the response pool.

Select Open-Ended Question

Next, one open-ended question was written for use in the simulation. The two questions used in the actual evaluation study and the responses elicited by them were sufficiently comparable to allow for one hybrid question to be used. This allowed all responses from the teachers to be considered when the simulation responses were developed. The two open-ended questions used in the study and the one open-ended question used in the simulation are presented in Table 1.

Table 1

Open-Ended Questions Used by Patton et al. (1976)
and Open-Ended Question Used in the Simulation

Item	Text
II. c. (p. 7)	Please use this space to make any further comments or recommendations concerning any components of the accountability system.
IX. (p. 15)	Comments--Finally, we'd like you to use this space to add any additional comments you'd like to make about any part of the [Hometown] accountability system.
Simula- tion	Please give us any comments or recommendations you would like to make about any part of the Hometown Public Schools accountability system.

Determine Basic Classification Framework

Because the emphasis of this study was on using output from commercially available microcomputer programs to improve the reliability and validity of content analysis, the general framework used for developing a category system in the simulation was intentionally kept simple. A highly complex framework would tend to draw attention away from the intended emphasis, especially if the participants found the framework too difficult to use. Osgood, Suci, and Tannenbaum (1957) found that people often make distinctions between objects along three dimensions: (a) evaluation, such as, good-bad or positive-negative; (b) potency, such as, hard-soft or strong-weak; and (c) activity, such as, active-passive or fast-slow (pp. 62-63). For simplicity, the dimension most relevant to an evaluation study, evaluation, was selected to be used along with an attitude object

dimension with components like teacher turnover or testing program. These two dimensions would then constitute a general framework for the panel and participants to create a category system.

Before the researcher read the set of responses in the actual evaluation report, the category system was to have the following structure for each category: [Evaluative] statement about [Object]; where [Evaluative] would be replaced with "Negative," "Neutral," or "Positive;" and [Object] would be replaced with a descriptive identifier for some aspect of the accountability system. It turns out that virtually every comment made by the 373 teachers about the accountability system was negative. This made the intended classification system inappropriate because one dimension would have no variability.

Instead, the general framework was modified so that each category was first given to represent a group of negative comments about the accountability system. Each category was then expected to have two parts: (1) an identifier--a brief, descriptive title for the category that indicates its negative nature and (2) a summary--an operational definition of the category complete enough to allow coders to decide whether a particular response should or should not be included in that category. It was also decided at that time to use 10 categories in the framework in order for the simulation problem to be challenging yet manageable for the participants. Because 10 categories and the related responses were eventually determined to be too many, the number of categories and responses was later cut in half. A graphic representation of the general framework for developing the category system is presented in Figure 1.

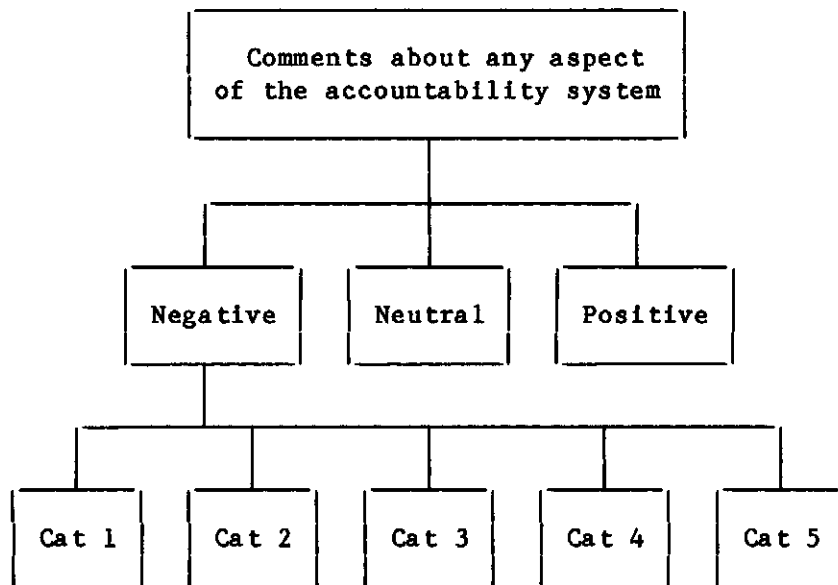


Figure 1. General Framework for Developing a Category System to Code Responses to the Open-Ended Simulation Question

Prepare Oversized Pool of Potential Responses

At this point, the potential responses for the panel to review were prepared. The goal of the panel was to select and code 200 responses that would fit into 10 categories they developed to comply with the general framework just described. One hundred responses from five categories were later selected as the final response pool. The researcher facilitated this process by preparing an oversized pool of specially designed responses for the panel to review. These responses were prepared through an iterative process as follows: (a) several category identifiers and summaries, based on the Patton et al. (1976) evaluation report, were drafted on numbered index cards; (b) responses or portions of responses that fit a category were

selected and modified, if necessary; (c) the respondent identification number was written on the card; (d) a data base of responses was constructed on a DECmate II list processing program with each record identifying the category number, respondent number, and text of each response; (e) the data base was occasionally sorted by category and printed for review and modification; (f) as a new category became apparent a card was prepared for it; (g) new categories and responses were added until 15 categories contained about 300 responses with about 20 responses per category; (h) 10 categories were then selected from the 15 and new responses were added until there were about 200 responses with roughly 20 responses per category.

The following criteria were used for writing and selecting the categories: (a) The categories should convey a negative evaluation about some aspect of the accountability system. Because neutral or positive responses were, by and large, not present in the set of responses, one content dimension and one evaluative dimension for each category was not needed. Both dimensions could be incorporated into a single, negatively weighted category. (b) The categories should be at roughly comparable levels of detail and fairly specific. This criterion should help raters decide into what category a response should be coded. (c) The categories should be exhaustive. All responses should fit somewhere. If this were not the case, a "garbage" category like "other" would be needed. This would not satisfy the other criteria. (d) The categories should be mutually exclusive. One best category for each response should exist. One category per response greatly simplifies analysis and interpretation.

(e) The categories should require more than a key word or phrase in most responses in order for a response to be properly coded. This criterion was intended to raise the difficulty of the tasks to a sufficiently high level to allow for differential performance to be demonstrated between the experimental and control groups. This was accomplished by having key words such as "teachers" be relevant to more than one category or by avoiding category definitions that could be operationally simplified to one key word.

A complementary set of criteria was used for writing and selecting the responses. The responses should: (a) convey a negative evaluation about some aspect of the accountability system, (b) have a clear best category into which it should be coded, (c) collectively not be codable by key words or phrases alone, and (d) be no more than a few sentences long. These criteria represent the iterative process necessary to produce a coherent yet realistic set of categories and responses at a usable level of difficulty for the experiments to follow.

Recruit a Panel

A group of four faculty members from Western Michigan University's College of Education was recruited to serve on the Response Panel. These faculty members included the Chair of the Departments of Educational Leadership, and Counseling and Personnel; one professor of Educational Leadership; one professor of Educational and Professional Development; and one associate professor of Educational and Professional Development. All panelists have had extensive

experience in public school settings and various types of educational research and evaluation activities.

Have Panelists Independently Create a Category System

The panelists received written instructions about how to proceed. They could also ask the researcher questions as needed. The panelists were given two weeks to create 10 categories and code 204 responses. Materials were returned to the researcher within 10 days.

The simulation materials used by the Response Panel were analogous to those used during the first task of the experiments. The main difference was Response Panel members were asked to create 10 categories and code about 200 responses while the participants were asked to create five categories and code 50 responses in Task 1.

The Response Panel received the following materials along with their written instruction: (a) Read Me First--an overview of the simulation and the simulation instructions, (b) Draft Introduction--a description of the background of the problem, (c) Practice Exercise--an in-class warm-up exercise, (d) Practice Exercise Answer Sheet--sample answers to the exercise, (e) Sample of Responses--204 responses to be coded, (f) Category Development Worksheets--worksheets for writing category identifiers and summaries, and (g) all 204 responses repeated on individual cards with 10 envelopes--a sorting aid not provided to the participants. Items (a) through (d) are presented in Appendix D. Items (e) and (f) are presented in Appendix B.

Process the Panelists' Work

The researcher processed the categories and response codes of the panelists and researcher in order to facilitate the group activity. This processing consisted of: (a) entering the category identifiers and summaries into a data base on a DECmate II; (b) mapping each of the panelists' categories onto the researcher's categories by labeling the researcher's categories "A" through "J" and then assigning each panelist's numeric category to the alphabetic label with the closest definition; (c) translating each panelist's numeric code for each response to the corresponding alphabetic code for the response; (d) entering the response codes onto another data base on the DECmate II; (e) sorting category identifiers and summaries by the alphabetic codes and printing them as the Response Panel Category Summary; (f) sorting the response codes by response identification number and by the code most used for each response; (g) printing both lists without the text of the responses as the Response Panel Classifications of Responses; (h) printing the alphabetic codes sorted by response identification number with the text of the responses; (i) constructing and printing a frequency coincidence matrix, and a related measure of agreement by categories, Krippendorff's alpha (Krippendorff, 1980, pp. 140-146), and, finally, (j) constructing and printing a percentage coincidence matrix by category, along with a corresponding Venn diagram. All printed materials were handed out to the panelists at the start of the group meeting.

Define Categories and Assign Responses

The four faculty members and the researcher met with the intention to finalize the category system and set of responses. However, because of the circumstances described in the next section, one more selection activity was needed. This group of five people constituted the Response Panel as far as decisions at the meeting were concerned. The first task was to finalize the category identifiers and summaries. Panelists discussed each category and the researcher wrote each identifier and summary pair on separate sheets of posted newsprint as each decision was made. The next task was to finalize the code and wording for 200 responses and discard four responses. Four out of five panelist had to agree on the disposition of a response before it was finalized. Responses with less than 80 percent agreement were highlighted on the appropriate handouts--26 out of 204 responses. After exactly 200 responses were approved the meeting was adjourned.

Select the Final Categories and Responses

The day after the panel meeting, the researcher cut the response pool down to 100 responses and five categories because of concern the tasks planned for the experiments would be too complex and time consuming for the participants. This decision was made because (a) Response Panel members expressed concern that the response pool probably was too large, (b) the first pilot study instructor dropped out of the study after seeing the size of the response pool, and (c) the day after the Response Panel meeting, the second pilot study

instructor, after seeing the size of the response pool, withheld consent to participate in the experiment until an indefinite later date. The criteria for selecting the final categories and responses were the same as stated above with the added criterion that the first task undertaken by participants in the experiments should take no more than one hour to complete. Time reports from the pilot study indicated five categories and 100 responses met this criterion.

Divide the Response Pool Into Two Groups

Responses were randomly assigned to two groups of 50 responses each. One group of responses was arbitrarily designated to be used with the first experiment. Both groups of responses were used in the second experiment.

Produce a Word Count List

The Word Count List was created from a microcomputer text file that contained the final 100 responses. This text file was processed using the "WORDFREQ" option of The Word Plus (Holder, 1982, p. 38) spelling checker program running on an Osborne 1 microcomputer. The resultant list was ordered in descending frequency of occurrence of each unique word. This list file was transferred to a DECmate II microcomputer where all words occurring only once were deleted, an appropriate heading was added, and the list was arranged in columns. The resultant document is presented as Exhibit 1.

Exhibit 1

Word Count List for Experimental Participants Only

ISU Accountability Study for Hometown Public Schools

WORD COUNT LIST OF WORDS OCCURRING MORE THAN ONCE IN FREQUENCY ORDER
FOR ALL RESPONSES TO THE OPEN-ENDED QUESTION

Here is a word count list I asked my secretary to put together from all of the responses to the open-ended question. He created it with one of the options on our spelling checker program on the word processor. The list contains all of the words that appeared more than once, sorted in order by frequency of occurrence. Maybe these lists will give you a few leads to follow when you start to develop the five categories of responses.

#	WORD	#	WORD	#	WORD
108	THE	9	WOULD	5	SO
72	TO	9	WHICH	5	EDUCATION
59	AND	9	THEY	5	VARIABLES
55	IN	9	PEER	5	MUST
53	A	9	THEIR	5	TESTS
47	OF	9	OUR	5	ETC
44	IS	9	OTHER	5	FAR
40	I	9	TOO	5	USED
36	ARE	8	BY	5	WORK
34	ACCOUNTABILITY	8	BEEN	5	AM
33	TEACHERS	8	WAS	5	S
32	NOT	8	WORKING	5	LEARNING
31	BE	8	ACCOUNTABLE	5	WANTS
28	AS	8	STUDENTS	5	ANY
28	IT	8	ADMINISTRATION	5	SEEMS
28	THAT	8	IF	4	RATE
26	SYSTEM	7	ONE	4	RATING
23	FOR	7	SOME	4	HOWEVER
19	WE	7	RATINGS	4	SHOULD
18	WITH	7	IDEA	4	BELIEVE
18	BUT	7	PEOPLE	4	SCORES
18	HAVE	7	WHEN	4	AT
16	GOOD	7	THIS	4	WILL
16	ON	6	MANNER	4	INDIVIDUAL
13	WHO	6	OR	4	USEFUL
13	HOMETOWN	6	MAKE	4	THOSE
12	HAS	6	DOES	4	CHILDREN
12	CAN	6	HE	4	TESTING
12	DO	6	BECAUSE	4	BAD
11	NO	6	SCHOOL	4	HIGH
11	THERE	6	OUT	4	JUST
10	ALL	6	WAY	4	EDUCATIONAL
10	AN	5	YOU	4	SEE
10	HUMAN	5	EXCELLENT	4	AMONG
10	TEACHER	5	THAN	4	EXPECTED
10	FEEL	5	LIKE	4	HAT
10	MANY	5	PROGRAM	4	US
10	EACH	5	WERE	4	FROM

#	WORD	#	WORD	#	WORD
4	ANOTHER	3	BECOME	2	UPON
4	BEST	3	LEVEL	2	DIDN'T
4	EVEN	3	GOAT	2	ITS
4	TEACHING	3	MOST	2	COMPONENTS
4	INTO	3	I'M	2	WHAT
4	STUDENT	3	RESULT	2	HELD
4	TAKE	3	LOWER	2	FIGHTING
3	TOGETHER	3	VERY	2	GONE
3	SHARING	3	IDEAS	2	DECENCY
3	HIS	3	THEM	2	STANDARDIZED
3	IT'S	2	CHECK	2	DEMEANING
3	OBJECTIVES	2	ABOUT	2	PROBLEM
3	RATED	2	SELF	2	DURING
3	NONE	2	HIM	2	AFRAID
3	UP	2	NEED	2	PROGRESS
3	THINK	2	PRODUCT	2	DOESN'T
3	SHOW	2	ACHIEVE	2	RESENT
3	ONLY	2	SAME	2	JOB
3	DONE	2	EVALUATIONS	2	JOKE
3	CONTROL	2	CONSIDERATION	2	INCREASING
3	PRINCIPAL	2	B	2	CLASSES
3	PRESSURE	2	IMPLEMENTED	2	NEVER
3	THINGS	2	BUSINESS	2	JUDGING
3	CONCERNED	2	ATTITUDES	2	TOOL
3	DEALING	2	BACKGROUNDS	2	EVALUATION
3	FORCE	2	ORDER	2	LOST
3	CAUSE	2	MIGHT	2	ROOM
3	AGAINST	2	SIMPLY	2	WELL
3	ADMINISTRATORS	2	SLOW	2	TIME
3	CHILD	2	POSITIVE	2	YEAR'S
3	BETWEEN	2	HUMANITY	2	CARE
3	GROWTH	2	GET	2	CLASS
3	BEINGS	2	POSSIBLE	2	WHOLE
3	MORE	2	RESPECT	2	UNDER
3	COULD	2	DON'T	2	HOME
3	ME	2	TRYING	2	SINCE
3	MY	2	EVERYONE	2	PLAY
3	EVER	2	PROBLEMS	2	LOW
3	SUPERINTENDENT	2	MANAGEMENT	2	MERIT
3	WRONG	2	TRY	2	ECONOMIC
3	COMPETITION	2	SHORT	2	PAY
3	RATHER	2	THING	2	BETTER
3	STAFF	2	VALUE	2	SHARE
3	CANNOT	2	IMPORTANT	2	THEMSELVES
3	GROUP	2	SITUATION	2	PART
3	HERE	2	MATERIALS	2	MEETING
3	LITTLE	2	EXPERIENCE	2	RESPONSIBLE
3	MODEL	2	MAY	2	OTHERS
3	USE	2	CLASSROOM	2	PRESENTLY
3	FEELING	2	FACTORS		
3	HOW	2	VIEWED		
3	BEYOND	2	POOR		
3	MUCH	2	BASED		

Appendix B

Microcomputer-Implemented Content Analysis Activities

Microcomputer-Implemented Content Analysis Activities

Four microcomputer-implemented content analysis activities were conducted by the researcher alone as part of the experimental procedures. These activities were used to: (a) independently process information generated by each participant during a previous task, and (b) prepare individualized materials for each of them to use during the next task, if one followed. The activities are described in more detail in the next four sections.

Content Analysis Activity 1

The purposes of this activity were to process the information generated by the participants during Task 1 and prepare the materials for them to use during Task 2. Processing involved entering each participant's identification information, category identifiers, summaries, and response codes onto the microcomputer data base.

Three documents were then prepared for each participant. Versions for fictitious participants are used in this appendix. The first document, Exhibit 1, was a personalized (addressed to the individual participant) form memo that was otherwise the same for all participants. It was used to provide the instructions for Task 2. The second document was different for each participant but it was prepared by the same process for all. It amounted to a typed version of the participant's Category Development Worksheet. An example is displayed as Exhibit 2. The third document, the Sample of Responses, was different for each participant and it was prepared one way for

Exhibit 1

Personalized Form Memo for Task 2

M E M O

October 25, 1984

TO: Audry Farber

FROM: Dick Frisbie

SUBJECT: HPS Accountability Study

Thank you for working on developing a category system for coding responses to the open-ended question of the HPS Accountability Study. I have processed your categories and entered your codes for the sample of responses.

Please check over the materials I have returned to you, Audry. Make any changes you think are needed and return the final categories and codes for the sample of responses to me on **November 1**. Dr. Powerful has not yet returned from her meeting, but I would like to have these materials ready for her when she returns.

experimental participants and another way for control participants. The documents for both groups had the codes given to the responses typed in a column next to each response with another column added to record a change of codes, if desired. In addition, the experimental group had their responses sorted into five groups arranged by the codes assigned to each response. The identifier and summary of the applicable category was also placed at the beginning of each group of responses. Exhibit 3 is an example. The control group had their responses in the same order as Task 1 with no category identifiers or summaries placed anywhere in the document.

Exhibit 2

Processed Category Development Worksheet

ISU Accountability Study for Hometown Public Schools

CATEGORY DEVELOPMENT WORKSHEET

ID: 0 **NAME:** Nick Danger

Cat # 1 Identifier: QUANTIFYING EFFECTIVE TEACHING

Summary: There are variables involved in classroom instruction which make it difficult to fairly evaluate all teachers by constant standards.

Cat # 2 Identifier: NEGATIVISM OF ADMINISTRATION

Summary: It is difficult to expect success for a controversial evaluation system if management cannot refrain from damaging teacher morale.

Cat # 3 Identifier: OBJECTIVITY OF PEER EVALUATION

Summary: Teachers will respond to a chance to evaluate one another by rating colleagues highly in return for high ratings.

Cat # 4 Identifier: COMPETITIVENESS AMONG STAFF

Summary: This system promotes unproductive rivalries and tensions among staff members who feel pressured.

Cat # 5 Identifier: SUPPORT CONCEPT, REJECT THIS DESIGN

Summary: Accountability systems are based on sound principles, but this one is not suitable.

 Responses Sorted by Category for Experimental Participants Only

Independent State University

 USING A MAIL SURVEY TO ASSESS THE ACCOUNTABILITY SYSTEM
 FROM THE PERSPECTIVE OF THE TEACHERS

SAMPLE OF RESPONSES TO THE OPEN-ENDED QUESTION

ID: 0 NAME: Professor Valery Powerful
 TIME NEEDED TO VERIFY CATEGORIES AND RESPONSES: _____ HRS. _____ MIN.

QUESTION: Please give us any comments or recommendations you would like to make about any part of the Hometown Public Schools accountability system.

Category
 # Old New Response

- Cat #1 Identifier: SOUND CONCEPT INAPPROPRIATELY IMPLEMENTED
 Summary: Accountability is important and valuable but not as it has been devised for use in Hometown.
- 2 A ___ Any of the components could have been utilized effectively had they been presented in a positive, professional manner.
- 4 A ___ GOATs are nothing more than good organization which no one can argue against but the manner in which it was devised and implemented in Hometown leaves much to be desired.
- 7 A ___ Accountability seems a good thing to me. Testing seems to be a good thing. But the way they are implemented and pushed on Hometown teachers is wrong.
- 14 A ___ As I see the system as a whole, it is very good in design. However, it is not being used to upgrade the level of achievement, but rather to do just the opposite.
- 22 A ___ I feel there should be some type of accountability system but none like we are presently using.
- 26 A ___ Accountability can be a useful measurement tool. However, the system here will ultimately fail because of how it has been run.
- 31 A ___ Accountability, when used in a positive manner, could be useful. When an accountability model like that in Hometown is used, this defeats the purpose of teaching in the classroom.
- 34 A ___ I'm sure the system has some merit. However, there are many kinks which need to be ironed out.
- 37 A ___ A good idea gone wrong because of dissention between the teaching staff and those in high administrative positions. As a result, the students and accountability system have become of little use to each other and unpleasantness has replaced harmony.
- 38 A ___ The Hometown accountability system must be viewed in its totality and not just in the individual component parts of it. In toto it is opperssive and stifling.

- 45 A ___ Accountability is important, but not as a fear developing tool. I was among the four who resigned.
- 47 A ___ The accountability system is a good idea gone bad.
- 50 A ___ The accountability system falls short when measuring some of the most important facets in life - honesty, getting along with others, and learning to be a winner and loser gracefully, self control, etc.

Cat #2 Identifier: DIVISIVENESS AMONG INSTRUCTIONAL STAFF
Summary: Implementation of the system has created tension and division among instructional staff.

- 5 B ___ Competition is increasing for high scores on HAT tests and good ratings by principals.
- 13 B ___ Principal evaluation of teachers became a "report card comparison" among teachers causing jealousies, pickiness, and accusations of "browning." I have seen a once-unified staff become polarized and unhappy.
- 15 B ___ Teachers stay in room and do not share.
- 19 B ___ Unfortunately the end result of the accountability system has been the tension and division between teachers, rather than the progress and development of our students.
- 23 B ___ It seems to imply that we must be in competition with our colleagues in order to be good teachers.
- 46 B ___ A merit pay system would be a mistake as it would force even the good teachers to become concerned only about themselves. Each would be trying to outdo the other and thus would cause limited sharing and exchanges of ideas and materials among teachers.
- 49 B ___ Accountability here is backbiting, and dividing (as he wants) teachers.

Cat #3 Identifier: LACKS PROVISION FOR CONTEXT VARIABLES
Summary: Student variability and other context variables are ignored in the system.

- 6 C ___ Someone who is in the classroom dealing with all types of kids, some who cannot read, some who hardly ever come to school, some who are in and out of jail, this teacher can see that, and the rigid accountability model that neglects the above mentioned problems is pure "B*****."
- 8 C ___ There are too many variables that enter in to make it work.
- 9 C ___ There are too many variables in the educational system for accountability to work.
- 18 C ___ Anytime you deal with young adults many variables are involved. I do not feel we can force teachers to accept unstable variables to play a part in evaluating.

-
- 29 C ___ The system--in no way--considers the various elements beyond testing that goes into the make-up of individual classes.
- 32 C ___ No one wants to take low students in their room anymore because the principal will look at their scores and think they are poor teachers because their students scored lower - terrible!
- 35 C ___ OUR system presently seems to be under the illusion that we have total control of the educational processes for each child. WE are responsible! B***** - we are partially responsible but not over home and peer group.
- 39 C ___ It doesn't take into consideration that some children have different socio-economic, emotional, and educational backgrounds and support from parents that keep them from learning.
- 41 C ___ The accountability system has little or no provisions for low I.Q.s, drugs, liquor, sex, home problems, lack of interest in school by student and/or family, etc.,--but teachers still have to produce!
- 43 C ___ Under the accountability system all teachers are rated on same standards and all classes are expected to make 1 year's growth, even if records show that group has never shown 1 year's growth.

Cat #4 Identifier: COLLUSION ON PEER REVIEWS

Summary: Teachers deliberately give peers high ratings in order to protect each other.

- 1 D ___ Peer ratings of a teacher in this system becomes an exercise of writing "5 for excellent."
- 11 D ___ Peer ratings are a joke!! All teachers rate each other straight 5's - Excellent.
- 21 D ___ Peer ratings: in my experience, have been done with the highest rating on each point.
- 24 D ___ As to teacher peer ratings we have an agreement in our building that no one is rated lower than "good."
- 25 D ___ We all got together in our school and rated each other No. 5 on the scale (excellent).

Cat #5 Identifier: SYSTEM INHUMANE

Summary: The system is viewed as lacking the human element and ignores human relations.

- 3 E ___ The administration was quick to criticize, demand, and put pressure on us, but slow (if ever) to recognize, praise, and encourage us as human beings.
- 10 E ___ The superintendent is too heavy handed and relies on threats when he wants to sell a program instead of working with us.

Content Analysis Activity 2

The purposes of this activity were to process the information generated by the participants during Task 2 and prepare the materials for them to use during Task 3. Relatively little work was required to process the information generated during Task 2. Only changes to existing identifiers, summaries, or codes for specific responses were made, if so indicated on a participant's Category Development Worksheet or Sample of Responses. Otherwise, the information existing on the data base was retained. This concluded the procedures for the category experiment. All the remaining procedures apply only to the coding experiment.

Three documents were prepared for each participant to be used during Task 3. The first document was a form memo, individually addressed, that was otherwise the same for all participants. It was used to provide instructions for Task 3. The second document, the Official Categories Summary, Exhibit 4, was identical for all participants. It contained the final set of categories to be used by all participants to code responses for the remainder of the simulation. The third document, the Complete Set of Responses, was different for each participant, and it was prepared one way for experimental participants and another way for control participants. The Complete Set of Responses contained 100 responses, the 50 responses used during the first two tasks plus 50 new responses. The document was prepared the same way the Sample of Responses was prepared for each group during Content Analysis Activity 1. An individual participant's

Exhibit 4

Official Categories Summary

ISU Accountability Study for Hometown Public Schools
Dr. Valery Powerful, CENTER Director

OFFICIAL CATEGORIES SUMMARY

Cat. A Identifier: SOUND CONCEPT INAPPROPRIATELY IMPLEMENTED
Summary: Accountability is important and valuable but not as it has been devised for use in Hometown.

Cat. B Identifier: DIVISIVENESS AMONG INSTRUCTIONAL STAFF
Summary: Implementation of the system has created tension and division among instructional staff.

Cat. C Identifier: LACKS PROVISION FOR CONTEXT VARIABLES
Summary: Student variability and other context variables are ignored in the system.

Cat. D Identifier: COLLUSION ON PEER REVIEWS
Summary: Teachers deliberately give peers high ratings in order to protect each other.

Cat. E Identifier: SYSTEM INHUMANE
Summary: The system is viewed as lacking the human element and ignores human relations.

Complete Set of Responses looked different than the Sample of Responses to the extent that changes were made to specific response codes plus 50 uncoded responses were added after the 50 coded responses. The result was that the experimental participants had a Complete Set of Responses with 50 coded responses sorted by their own five categories and headed by their own identifiers and summaries. The 50 uncoded responses followed. The control participants had a Complete

Set of Responses with 50 coded responses in the same order as in Task 1 plus 50 uncoded responses added after the first 50.

Content Analysis Activity 3

The purposes of this activity were to process the information generated by the participants during Task 3 and prepare the materials for them to use during Task 4. Processing only involved the codes for the responses, and all 100 responses were processed for each participant. For the first 50 responses, numeric codes were replaced by alphabetic codes. For the second 50 responses, previously uncoded responses were given alphabetic codes.

Three documents were prepared for each participant to be used during Task 4. The first document was a form memo, individually addressed, that was otherwise the same for all participants. It was used to provide instructions for Task 4. The second document was another Official Categories Summary. It was provided to ensure each participant had a copy available for the task. The third document was an updated Complete Set of Responses. It was different for each participant, and it was prepared one way for experimental participants and another way for control participants. For experimental participants, the 100 coded responses were sorted by the five Dr. Powerful (Response Panel) categories, and each group of responses was headed by the appropriate Dr. Powerful identifier and summary. For the control participants, the 100 coded responses printed in the same order as they were for Task 3, and no category identifiers or summaries were printed anywhere on the document.

Content Analysis Activity 4

The purpose of this activity was to process the information generated by the participants during Task 4. Relatively little work was required to process this information. Only changes to existing codes for specific responses were made, if so indicated on a participant's Complete Set of Responses. Otherwise, the information existing on the data base was retained. This concluded the procedures for the coding experiment.

Appendix C
Development of the Category Hierarchy

Development of the Category Hierarchy

The category hierarchy was created for two basic purposes. First, it provided the means for deriving the measures of two dependent variables, category reliability and category validity. Second, it provided a qualitative framework for characterizing the categories developed by the participants. The tasks performed to develop the category hierarchy were to: (a) recruit a panel, (b) have the panelists independently create a hierarchy and classify the categories generated by the pilot study participants, (c) process the panelists' work on the pilot study-generated categories, (d) have the panelists cooperatively determine the final framework of the pilot study category hierarchy and assign participant categories to their proper location in the framework, (e) have the panelists independently create a hierarchy and classify the categories generated by the experiment participants, (f) process the panelists' work on the experiment-generated categories, and (g) have the panelists cooperatively determine the final framework of the experiment category hierarchy and assign participant categories to their proper place in the framework.

Recruit a Panel

A group of three doctoral students from Western Michigan University's Department of Educational Leadership were recruited to serve on the Hierarchy Panel. All panelists had direct experience planning and conducting evaluation studies. They also had extensive work experience ranging from elementary schools to colleges.

Have Panelists Independently Classify
the Pilot Study-Generated Categories

The task of the panelists was to build upon the basic classification framework used in the Hometown Public Schools simulation activity described in Read Me First (see Appendix D) in such a way that all categories generated during the pilot study could be assigned to a single position in the framework. The five categories created by the Response Panel were used as the initial categories in the framework. If broader, narrower, related, or completely unrelated categories needed to be added to the original framework, it was up to each panelist to do so.

The panelists were instructed to write identifier and summary pairs for each new category and to locate each new category in the existing structure by drawing a new representation of the hierarchy. They were also instructed to assign each pilot study category to the synonymous hierarchy category.

Written instructions were given to the panelists. They could also ask the researcher questions as needed. In addition, they received the Pilot Study-Generated Categories, a set of 45 categories developed by those participants, and a set of Hierarchy Development Worksheets similar to the participants' Category Development Worksheets (see Appendix B). The panelists were asked to perform the task in one week. However, the last packet was returned about two weeks later.

Process the Panelists' Work on the Pilot Study Categories

The researcher processed the panelists' categories and assignments of participants' categories to the hierarchy in order to facilitate the group activity to follow. Two new documents were produced for distribution at the meeting: (1) the Hierarchy Panel Pilot Study Category Summary, a listing of all the original and new categories of the hierarchy; and (2) the Hierarchy Panel Classifications of Pilot Study-Generated Categories, a listing of all categories generated by the pilot study participants with the panelists' corresponding classification codes.

This processing consisted of: (a) entering the panelists' category identifiers and summaries into a data base on a DECmate II, (b) mapping the panelists' categories onto each other by matching roughly synonymous summaries and then changing each panelist's numeric category label to an alphabetic label, (c) translating each panelist's numeric code for a participant's category to the corresponding alphabetic code for the category, (d) entering each panelist's alphabetic code and all the corresponding participant categories into another data base on the DECmate II, (e) sorting the panelists' category identifiers and summaries by the alphabetic labels and printing them as the Hierarchy Panel Pilot Study Category Summary, and (f) sorting the participants' categories by the most used panelist code and printing the categories and codes as the Hierarchy Panel Classifications of Pilot Study-Generated Categories.

Cooperatively Determine the Pilot Study Category Hierarchy

The three panelists and the researcher met to determine the final definitions and framework of the pilot study category hierarchy and to assign each participant category to its proper location in order to verify that the hierarchy could be used to classify all the categories. This group of four people constituted the hierarchy panel as far as decisions at the meeting were concerned. At least three out of four people had to agree before a decision was final.

The first task was to select new categories, finalize their identifiers and summaries, and locate them in the framework. Each panelist was given a copy of the Hierarchy Panel Pilot Study Category Summary, the Hierarchy Panel Classifications of Pilot Study-Generated Categories, and photocopies of each hierarchy framework developed by the panelists. Panelists discussed each category and selected the unique additions to the framework. The researcher wrote each identifier and summary pair on sheets of posted newsprint as each decision was made. The researcher also redrew the framework on a chalkboard as each new category was added. When this task was completed, each category in the framework was assigned an identification letter, and the framework was copied to a sheet of newsprint.

The second task was to assign the 45 participant-generated categories to their appropriate locations in the newly created framework. The group considered and recoded each participant category that was not unanimously coded into the same alphabetic hierarchy category by the three panelists. After completion of this task, the panel meeting ended.

Have Panelists Independently Classify the Experiment-Generated Categories

The task of the panel here was to build upon the basic classification framework they had developed from the previous activity in such a way that all categories generated during the experiment could be assigned to a single position in the framework. The process they were instructed to use was directly analogous to that used with the pilot study-generated categories. The basic differences were that they had a more complete framework to start with and they had a new set of categories to use.

The panelists again received: (a) written instructions; (b) the Experiment-Generated Categories, a set of 370 categories generated by 75 participants; and (c) the Hierarchy Development Worksheets. They were asked to perform the task in one week. However, because of the size of the task and the intervening Winter break, the last packet was returned about eight weeks later.

Process the Panelists' Work on the Experiment Categories

This activity was directly analogous to processing the panelists' work on the pilot study categories. The two documents produced for distribution at the forthcoming meeting were: (1) the Hierarchy Panel Experiment Category Summary, a listing of all the previous and new categories of the hierarchy; and (2) the Hierarchy Panel Classifications of Experiment-Generated Categories, a listing of all categories generated by the experiment participants with the panelists' corresponding classification codes.

Cooperatively Determine the Experiment Category Hierarchy

The panelists and the researcher met to determine the final definitions and framework of the category hierarchy for the experiment and to assign each participant category to its proper location in the hierarchy. These assignments became the raw data for the dependent variables, category reliability and category validity. The procedures used were analogous to those used to cooperatively develop the pilot study hierarchy and classify those participant-generated categories. The basic differences were that the panel had the more complete hierarchy to start with and they had a new set of categories to classify. Two three-hour sessions, one week apart, were needed to create the final hierarchy and code the participant categories in relation to it.

The text of the final hierarchy is presented in Table 1. A graphic representation of the final hierarchy is presented in Figure 1. Categories G and M were initially created by the panel, but no participant categories were given these codes in the end. Seven hybrid categories were also created because many participant categories contained important features of two Hierarchy Panel categories.

Table 1

Identifiers and Summaries for the Final Category Hierarchy

Category A	Identifier: SOUND CONCEPT INAPPROPRIATELY IMPLEMENTED
Summary:	Accountability is important and valuable but not as it has been devised for use in Hometown.
Category B	Identifier: DIVISIVENESS AMONG INSTRUCTIONAL STAFF
Summary:	Implementation of the system has created tension and division among instructional staff.
Category C	Identifier: LACKS PROVISION FOR CONTEXT VARIABLES
Summary:	Student variability and other context variables are ignored in the system.
Category D	Identifier: COLLUSION ON PEER REVIEWS
Summary:	Teachers deliberately give peers high ratings in order to protect each other.
Category E	Identifier: SYSTEM INHUMANE
Summary:	The system is viewed as lacking the human element and ignores human relations.
Category F	Identifier: UNSOUND CONCEPT
Summary:	Accountability is an unworkable concept.
Category G	Identifier: BIASED PUBLIC COMMUNICATIONS--not used alone
Summary:	The Accountability System promotes a biased view of school to the public.
Category H	Identifier: ACCOUNTABILITY SYSTEM DETRACTS FROM INSTRUCTIONAL ACTIVITIES
Summary:	Attention to system detracts from instructional and related activities.
Category I	Identifier: ILL-DEFINED, NEGATIVE
Summary:	Negative in tone but lacking in specificity, and uninterpretable.
Category J	Identifier: NEUTRAL COMMENT
Summary:	Comment reports about some aspect of the system in a neutral tone and without suggestion or implication.
Category K	Identifier: ADMINISTRATION ABUSES TEACHERS
Summary:	Administration is abusive to teachers.
Category L	Identifier: PERFORMANCE APPRAISAL INVALID
Summary:	The system for appraising performance is invalid.

Table 1 (continued)

Category N	Identifier: SYSTEM DETRIMENTAL TO STUDENTS
Summary:	The Accountability System is detrimental to student learning and progress.
Category AH	Identifier: A & H
Summary:	Categories A & H Combined.
Category AB	Identifier: A & B
Summary:	Categories A & B Combined.
Category BD	Identifier: B & D
Summary:	Categories B & D Combined.
Category BE	Identifier: B & E
Summary:	Categories B & E Combined.
Category BL	Identifier: B & L
Summary:	Categories B & L Combined.
Category CE	Identifier: C & E
Summary:	Categories C & E Combined.
Category GK	Identifier: G & K
Summary:	Categories G & K Combined.

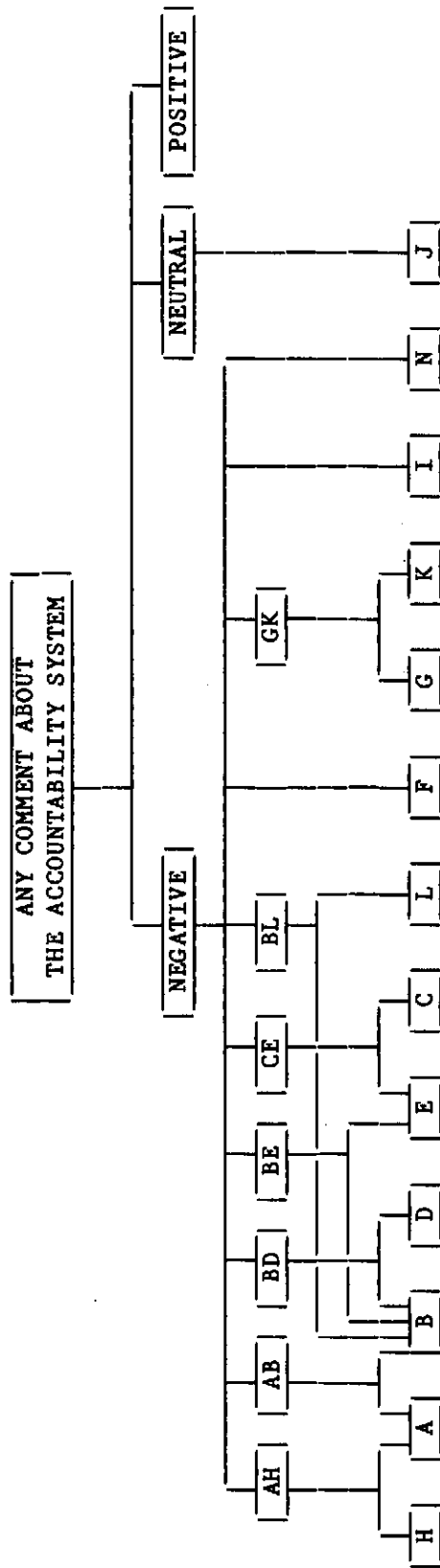


Figure 1. Final Category Hierarchy

Appendix D

Simulation Materials for Session 1

Simulation Materials for Session 1

The three purposes of the Session 1 were to: (1) introduce the study, (2) provide a classroom lecture on content analysis, and (3) start the simulation activity used as the organizer for the content analysis task to follow. The researcher conducted the session according to a detailed script (Exhibit 1) organized by the three purposes identified above. The script ensured a degree of consistency for Session 1 between classrooms. It was not distributed.

The participants were given a number of handouts during the session to help them with the first task. In order, they received: (a) Content Analysis: Answers to Four Practical Questions (Exhibit 2); (b) Read Me First (Exhibit 3); (c) a Draft Introduction of an evaluation report (Exhibit 4); (d) a Practice Exercise (Exhibit 5); (e) a Practice Exercise Answer Sheet (Exhibit 6); and (f) the final group of handouts in a sealed envelope (discussed in Appendices A and B). A detailed description of the first session in relation to the rest of the procedures is discussed in the Experiments section of Chapter 3.

Exhibit 1

Script for Session 1

I. INTRODUCTION

MY NAME IS DICK FRISBIE. I'M HERE TO ASK YOU TO PARTICIPATE IN A STUDY ABOUT USING THE OUTPUT FROM MICROCOMPUTER PROGRAMS TO HELP CONDUCT CONTENT ANALYSES OF RESPONSES TO OPEN ENDED SURVEY QUESTIONS. THIS STUDY IS PART OF MY DOCTORAL DISSERTATION AT WESTERN MICHIGAN UNIVERSITY.

YOU ARE ABOUT TO PARTICIPATE IN AN EXPERIMENT TO DETERMINE IF THE OUTPUT FROM COMMERCIALY AVAILABLE MICROCOMPUTER PROGRAMS CAN HELP PEOPLE DO A BETTER JOB AT CONDUCTING A CONTENT ANALYSIS OF RESPONSES TO OPEN-ENDED SURVEY QUESTIONS. YOU WILL NOT BE USING ANY MICROCOMPUTERS YOURSELF, BUT ONLY SOME OF THE OUTPUTS THEY PRODUCE.

WHAT I WILL DO NEXT IS PROVIDE AN OVERVIEW OF THE SIMULATION, PRESENT A BRIEF LECTURE ON CONTENT ANALYSIS, AND THEN START THE SIMULATION ACTIVITY WHICH IS THE CORE OF THE STUDY.

THE SIMULATION PROBLEM IS BASED ON A NON-TRIVIAL, REAL-WORLD EVALUATION STUDY THAT USED A MAIL SURVEY OF PUBLIC SCHOOL TEACHERS TO ASSESS THEIR REACTIONS TO A CONTROVERSIAL ACCOUNTABILITY SYSTEM. THE ACTUAL MAIL SURVEY IS NOT DIRECTLY RELATED TO MY STUDY OF CONTENT ANALYSIS EXCEPT THAT IT PROVIDES AN INTERESTING AND REALISTIC SETTING FOR THE SIMULATION.

FOUR OUT-OF-CLASS TASKS MUST BE PERFORMED BY THE PARTICIPANTS IN THIS STUDY. YOU MUST FIRST DEVELOP A CATEGORY SYSTEM INTO WHICH THE RESPONSES TO THE OPEN-ENDED QUESTION CAN BE CODED. THIS CATEGORY SYSTEM SHOULD BE BASED ON THE BACKGROUND MATERIAL PROVIDED AND A SAMPLE OF ACTUAL RESPONSES. I WILL ASK YOU TO TAKE NO MORE THAN ABOUT ONE HOUR TO COMPLETE THIS TASK. A GROUP OF EXPERTS IN THIS FIELD COMPLETED A SIMILAR TASK WITH FOUR TIMES AS MANY ITEMS AS YOU WILL HAVE IN AN AVERAGE OF ONE TO TWO HOURS.

THE SECOND TASK WILL BE TO VERIFY THE SYSTEM PREVIOUSLY DEVELOPED. THIS IS A SORT OF "LAST CHANCE TO CHECK YOUR WORK" TASK. IT SHOULD TAKE NO MORE THAN ABOUT A HALF HOUR TO COMPLETE.

THE THIRD TASK WILL BE TO CODE A SET OF RESPONSES BASED ON AN EXISTING SET OF CATEGORIES. THESE CATEGORIES MAY OR MAY NOT BE THE SAME AS THE ONES YOU DEVELOP. IT SHOULD TAKE NO MORE THAN ABOUT ONE HOUR TO COMPLETE.

THE FOURTH AND FINAL TASK WILL BE TO VERIFY THE CODES GIVEN TO RESPONSES IN THE THIRD TASK--ANOTHER "LAST CHANCE TO CHECK YOUR WORK" TASK. IT SHOULD TAKE NO MORE THAN ABOUT A HALF HOUR TO COMPLETE.

Exhibit 1 (continued)

FOR THE NEXT EIGHT WEEKS YOU WILL BE GIVEN A TASK EVERY OTHER WEEK. I WILL USE THE ALTERNATE WEEK TO PROCESS YOUR WORK AND PREPARE THE MATERIALS FOR YOUR NEXT TASK, IF ONE FOLLOWS. AFTER EVERYONE HAS TURNED IN THEIR FINAL TASK, I WILL ATTEMPT TO ANSWER ANY QUESTIONS YOU MAY HAVE ABOUT THE STUDY.

PARTICIPATION IN THE STUDY IS COMPLETELY VOLUNTARY, BUT THE TASK IS INTERESTING AND RELATIVELY PAINLESS. IT PROVIDES AN EXCELLENT OPPORTUNITY TO LEARN AND PRACTICE CONTENT ANALYSIS TECHNIQUES ON A VERY COMMON TYPE OF INFORMATION COLLECTED BY HUMAN SERVICE ORGANIZATIONS--RESPONSES TO OPEN-ENDED SURVEY QUESTIONS. HOPEFULLY, PARTICIPATION IN THE STUDY WILL HELP YOU TO BECOME BETTER EVALUATORS OR RESEARCHERS WHEN THE NEED ARISES AND BETTER PARTICIPANTS WHEN ASKED TO OFFER YOUR INPUT AGAIN. IN ADDITION [YOUR INSTRUCTOR] WILL BE GIVING YOU SOME INCENTIVES TO PARTICIPATE IN THIS STUDY. FIRST, YOU WILL BE GIVEN MUCH ENCOURAGEMENT AND MORAL SUPPORT TO COMPLETE THE TASKS. IN ADDITION, [LIST CONTRACTUAL INCENTIVES].

ARE THERE ANY QUESTIONS BEFORE I BEGIN THE LECTURE ON CONTENT ANALYSIS?

II. LECTURE

THIS LECTURE ON CONTENT ANALYSIS WILL FOCUS ON PROVIDING ANSWERS TO FOUR PRACTICAL QUESTIONS ABOUT CONTENT ANALYSIS. **THE QUESTIONS ARE: 1. WHAT IS IT? 2. WHAT ARE ITS USES? 3. WHEN CONDUCTING SURVEYS, WHEN SHOULD IT BE USED WITH OPEN-ENDED QUESTIONS, INSTEAD OF USING QUANTITATIVE ANALYSIS WITH FORCED-CHOICE QUESTIONS? AND 4. HOW IS IT DONE?**

EACH QUESTION COULD BE ANSWERED DIFFERENTLY THAN THE ONES I WILL PROVIDE AND MANY OTHER RELEVANT QUESTIONS COULD BE ADDRESSED AS WELL; BUT THIS LECTURE WILL GIVE YOU A GOOD BACKGROUND ON SOME GENERAL CHARACTERISTICS OF CONTENT ANALYSIS AND POINT OUT HOW THE SIMULATION RELATES TO A BROADER UNDERSTANDING OF CONTENT ANALYSIS.

HERE IS A HANDOUT THAT SUMMARIZES MY ANSWERS TO THE FOUR QUESTIONS. I WON'T DISCUSS ALL OF THE INFORMATION ON THE HANDOUT NOW, BUT I HAVE PROVIDED IT FOR YOUR FUTURE REFERENCE. RIGHT NOW I'LL JUST POINT OUT HOW THE QUESTIONS AND ANSWERS ARE RELATED TO THE SIMULATION PROBLEM YOU ARE ABOUT TO BEGIN. [HO]

1. WHAT IS IT?

[READ DEFINITION.] [POINT OUT KEY ELEMENTS OF THE DEFINITION IN THE HANDOUT.]

THIS DEFINITION OF CONTENT ANALYSIS FITS VERY WELL WITH THE

Exhibit 1 (continued)

SIMULATION BECAUSE IT IS DIRECTED YET FLEXIBLE ENOUGH TO SUPPORT THE TYPES OF VALUE-BASED ASSESSMENTS USUALLY FOUND IN AN EVALUATION STUDY SUCH AS THE ONE ON WHICH THE SIMULATION IS BASED. THIS SIMULATION IS TYPICAL OF MANY EVALUATION STUDIES CONDUCTED IN EVALUATION.

2. WHAT ARE ITS USES?

I HAVE PROVIDED SOME GENERAL USES OF CONTENT ANALYSIS ALONG WITH SOME MORE SPECIFIC EXAMPLES. THE SPECIFIC EXAMPLES INDICATE THE WIDE RANGE OF USES TO WHICH CONTENT ANALYSIS HAS BEEN PUT; BUT I EXPECT ANY ONE OF US COULD ADD AT LEAST A FEW MORE EXAMPLES TO THE LIST. IT IS A VERY VERSATILE TECHNIQUE.

GENERALLY SPEAKING, THE SIMULATION WILL USE CONTENT ANALYSIS TO DESCRIBE CHARACTERISTICS OF COMMUNICATION. SPECIFICALLY, IT WILL HAVE A COMBINED EMPHASIS AS AN AID IN TECHNICAL RESEARCH OPERATIONS (TO CODE RESPONSES TO AN OPEN-ENDED QUESTION IN A MAIL SURVEY) IN ORDER TO REFLECT THE ATTITUDES, INTERESTS, AND VALUES OF A GROUP OF PEOPLE (TEACHERS WORKING IN A PARTICULAR PUBLIC SCHOOL DISTRICT). THIS IS ONE OF THE MOST COMMON USES OF CONTENT ANALYSIS IN EDUCATIONAL EVALUATION.

3. WHEN CONDUCTING SURVEYS, WHEN SHOULD IT BE USED WITH OPEN-ENDED QUESTIONS, INSTEAD OF USING QUANTITATIVE ANALYSIS WITH FORCED-CHOICE QUESTIONS?

THE THIRD QUESTION AND ANSWER HAVE BEEN PROVIDED BECAUSE A COMMON USE OF CONTENT ANALYSIS IN SOCIAL SCIENCE AND HUMAN SERVICE ORGANIZATIONS IS TO ANALYZE RESPONSES TO OPEN-ENDED SURVEY OR INTERVIEW QUESTIONS. AS I JUST STATED, THE SIMULATION USES CONTENT ANALYSIS OF RESPONSES TO AN OPEN-ENDED MAIL SURVEY QUESTION.

THE CLASSIC COMPETITOR WITH CONTENT ANALYSIS OF RESPONSES TO OPEN-ENDED QUESTIONS IS SOME TYPE OF QUANTITATIVE ANALYSIS OF RESPONSES TO FORCED-CHOICE QUESTIONS. SUCH QUESTIONS HAVE A LIMITED AND KNOWN SET OF RESPONSES FROM WHICH A RESPONDENT MUST CHOOSE. THE LIBERAL, IF ANY, LIMITATIONS ON RESPONSES TO OPEN-ENDED QUESTIONS PERMIT AN ALMOST LIMITLESS NUMBER OF DIFFERENT RESPONSES.

MY ANSWER TO THIS QUESTION INCLUDES THE ADVANTAGES AND DISADVANTAGES OF BOTH FORCED-CHOICE AND OPEN-ENDED QUESTIONS. I ALSO SUGGEST UNDER WHAT CONDITIONS WHICH TYPE OF QUESTION AND ANALYSIS SHOULD BE USED. [READ RECOMMENDATIONS WITH EMPHASIS ON OR'S & AND.]

4. HOW IS IT DONE?

FINALLY, I ASK HOW CONTENT ANALYSIS IS DONE AND PROVIDE A GENERAL OUTLINE FOR HOW TO DO IT. MOST PEOPLE WILL NOT BE DIRECTLY INVOLVED

Exhibit 1 (continued)

IN ALL OF THE STEPS LISTED HERE. THEY WILL USUALLY BE INVOLVED IN ONLY A FEW OF THE ACTIVITIES. YOU WILL SPEND MOST OF YOUR TIME DEVELOPING CODING INSTRUCTIONS BY CREATING A CATEGORY SYSTEM, AND YOU WILL TRANSFORM, OR CODE, DATA--THAT IS, ASSIGN CATEGORY CODE NUMBERS OR LETTERS TO INDIVIDUAL RESPONSES.

THE FINAL PAGE SHOWS SOME COMMON WAYS OF REPORTING THE RESULTS OF A CONTENT ANALYSIS. THIS EXAMPLE IS BASED ON ONE QUESTION A SCHOOL DISTRICT MIGHT ASK A SAMPLE OF REGISTERED VOTERS AS PART OF A COMMUNITY SURVEY. THE DISTRICT MAY PREFER TO HAVE THIS BE AN OPEN-ENDED QUESTION RATHER THAN A FORCED-CHOICE QUESTION IF THEY DON'T WANT TO "LEAD" THE VOTERS TO GIVE ANY PARTICULAR RESPONSES, EVEN IF THEY HAVE A GOOD IDEA IN ADVANCE HOW THE RESPONSES WILL BE PUT INTO CATEGORIES. THE PURPOSES OF THIS ANALYSIS WERE TO DESCRIBE THE VOTERS' RESPONSES TO THE QUESTION THIS YEAR AND COMPARE THEM TO HOW THE VOTERS RESPONDED TO THE SAME QUESTION LAST YEAR.

THE FIRST SECTION REPRESENTS A NARRATIVE APPROACH TO SUMMARIZING THE RESPONSES. TYPICAL RESPONSES ARE PRESENTED FOR EACH CATEGORY WITH SOME SUMMARY STATISTICS. A LONGER VERSION OF THIS SECTION WOULD INCLUDE ELABORATED DEFINITIONS OF THE CATEGORIES, A FEW MORE ACTUAL RESPONSES, AND A SYNTHESIS OF AN OVERALL IMPRESSION OF THE GROUP OF RESPONSES.

THE SECOND SECTION REPRESENTS AN EXCLUSIVELY STATISTICAL APPROACH TO SUMMARIZING THE RESPONSES. THE FIRST TABLE SHOWS THE FREQUENCIES OF RESPONSES IN EACH CATEGORY FOR THE TWO YEARS. THE SECOND TABLE SHOWS THE CONTRIBUTIONS OF THE CELLS TO THE VALUE OF CHI SQUARE, AN INFERENCE STATISTIC. THE CONCLUSION IS THAT THE VOTERS DID, INDEED, RESPOND DIFFERENTLY TO THE QUESTION BETWEEN 1983 AND 1984. AN EXAMINATION OF THE CONTRIBUTIONS TO THE VALUE OF CHI SQUARE SUGGESTS THIS IS DUE TO A LOWER CONCERN FOR FINANCES AND A HIGHER CONCERN FOR THE 3 R'S FROM 1983 TO 1984.

THE THIRD SECTION REPRESENTS A GRAPHIC APPROACH TO DESCRIBING THE RESPONSES. THE FREQUENCIES FROM THE FIRST TABLE WERE CONVERTED TO PROPORTIONS AND PERCENTS THEN PRESENTED IN SIDE-BY-SIDE BAR GRAPHS. THIS PRESENTATION MAKES IT RATHER EASY TO NOTICE THE RELATIVELY LARGE DROP IN CONCERN FOR FINANCES AND RISE IN CONCERN FOR THE 3 R'S INDICATED IN THE CHI SQUARE ANALYSIS. IT ALSO MAKES IT EASIER TO NOTICE THE MOST IMPORTANT ISSUE IN 1983 WAS FINANCES, BUT THE MOST IMPORTANT ISSUE IN 1984 WAS DISCIPLINE.

ARE THERE ANY QUESTIONS BEFORE WE GO ON TO THE SIMULATION?

Exhibit 1 (continued)

III. SIMULATION

THIS SIMULATION IS BASED ON AN ACTUAL EVALUATION STUDY CONDUCTED BY A GROUP OF INDEPENDENT EVALUATORS FOR A TEACHERS' UNION OF A PUBLIC SCHOOL DISTRICT. THE UNION AND THE ADMINISTRATION WERE HOTLY DIVIDED OVER THE DISTRICT'S ACCOUNTABILITY SYSTEM. THE PURPOSE OF THE **EVALUATION STUDY** WAS TO ASSESS THE ACCOUNTABILITY SYSTEM FROM THE PERSPECTIVE OF THE TEACHERS. THE PURPOSE OF **THIS SIMULATION** IS TO PROVIDE AN INTERESTING BUT REALISTIC BACKDROP TO TEST SOME CONTENT ANALYSIS TECHNIQUES IMPLEMENTED WITH THE USE OF OUTPUT FROM MICROCOMPUTER PROGRAMS. THE SIMULATION COULD HAVE BEEN BASED ON OTHER STUDIES IN OTHER FIELDS AND STILL BE USED TO TEST THE CONTENT ANALYSIS TECHNIQUES OF INTEREST.

READ ME FIRST

THE SIMULATION BEGINS WITH **READ ME FIRST**. PLEASE LOOK OVER THIS HANDOUT AS I EXPLAIN IT TO YOU. **[HAND OUT] READ ME FIRST** SETS THE STAGE FOR THE SIMULATION AND GIVES YOU BASIC INSTRUCTIONS FOR THE FIRST TASK. THE FIRST PAGE SETS UP THE SIMULATION. YOU ARE NOW A STUDENT RESEARCHER WHO MUST FILL IN FOR THE DIRECTOR OF THE CENTER AT ISU, DR. POWERFUL. YOUR JOB IS TO CREATE A FIVE CATEGORY CLASSIFICATION SYSTEM CONSISTENT WITH THE GENERAL FRAMEWORK PICTURED IN FIGURE 1 ON THE THIRD PAGE. YOUR INSTRUCTIONS ARE ON PAGE 2. FIRST, YOU WILL READ SOME MORE INFORMATION HERE AND COMPLETE A PRACTICE EXERCISE. THEN YOU WILL BE GIVEN A PACKET OF MATERIALS LIKE THE PRACTICE EXERCISE TO WORK ON AT HOME. BASICALLY, YOUR JOB IS TO CREATE FIVE NEW CATEGORIES, CODE THE RESPONSES ACCORDING TO THOSE CATEGORIES, AND RETURN YOUR WORK TO ME NEXT WEEK. PLEASE READ ALL OF THE DETAILS IN THE INSTRUCTIONS RIGHT BEFORE YOU BEGIN YOUR WORK AT HOME.

DRAFT INTRODUCTION

HERE IS THE **DRAFT INTRODUCTION** OF THE EVALUATION REPORT PREPARED BY DR. POWERFUL. IT CONTAINS A "BACKGROUND OF THE PROBLEM" AS SEEN BY DR. POWERFUL. IT IS INTENDED TO GIVE YOU A FEEL FOR THE CONTEXT IN WHICH THE MAIL SURVEY WAS CONDUCTED. PLEASE TAKE A MINUTE TO LOOK AT IT NOW AND SPEAK OUT IF YOU HAVE ANY QUESTIONS. **[HAND OUT]**

PRACTICE EXERCISE

HERE IS A **PRACTICE EXERCISE** DEVELOPED BY DR. POWERFUL. IT SHOULD GIVE YOU AN IDEA ABOUT WHAT KINDS OF CATEGORY IDENTIFIERS AND SUMMARIES SHE WANTS YOU TO WRITE. AFTER EVERYONE HAS COMPLETED THE EXERCISE, I WILL HAND OUT AN ANSWER KEY AND WE WILL DISCUSS THE EXERCISE. FEEL FREE TO ASK QUESTIONS AS YOU GO ALONG. **[HAND OUT]**

Exhibit 1 (continued)

PRACTICE EXERCISE ANSWER KEY

HERE IS THE ANSWER KEY TO THE EXERCISE. THE ANSWERS ARE PRINTED IN BOLDFACE TYPE. READ THE SUGGESTED ANSWERS AND LET ME KNOW IF YOU HAVE ANY QUESTIONS. **[HAND OUT]**

TASK 1 PACKET

THE PACKETS I AM ABOUT TO HAND OUT INCLUDE A TEAR-OFF IDENTIFICATION SHEET AND THE SET OF MATERIALS FOR YOUR FIRST TASK. **PLEASE!!! DO NOT OPEN YOUR PACKETS AROUND ANY OF YOUR CLASSMATES AND PLEASE DO NOT DISCUSS THE DETAILS OF YOUR ACTIVITIES RELATED TO THE SIMULATION WITH ANY OF YOU CLASSMATES UNTIL EVERYONE HAS COMPLETED THE FOURTH TASK.**

THE INFORMATION ON THE ID SHEETS WILL BE HELD STRICTLY CONFIDENTIAL. IT WILL ONLY BE USED FOR DATA IDENTIFICATION, COMPARING DEMOGRAPHIC INFORMATION BETWEEN **GROUPS** OF PARTICIPANTS, OR FOR ME TO CONTACT YOU IF THAT BECOMES NECESSARY LATER ON. PLEASE FILL OUT THE ID SHEET NOW AND PASS IT UP TO ME. **PRINT** YOUR NAME HOW YOU PREFER TO BE CALLED, GIVE A PHONE NUMBER WHERE YOU CAN MOST EASILY BE REACHED, AND FILL IN THE REMAINDER OF THE FORM TO THE BEST OF YOUR KNOWLEDGE OR JUDGMENT. **[HAND OUT]**

FROM NOW ON, EACH OF YOU WILL HAVE A SLIGHTLY DIFFERENT SIMULATION EXPERIENCE, BASED ON THE MATERIALS YOU RECEIVE **AND** ON HOW YOU PERFORM EACH OF THE FOUR TASKS. BECAUSE THIS STUDY IS INTENDED TO TEST A SET OF SPECIFIC RESEARCH HYPOTHESES, SHARING YOUR MATERIALS OR EXPERIENCES WITH OTHER PARTICIPANTS FROM NOW UNTIL THE FOURTH TASK IS COMPLETED WOULD THREATEN THE VALIDITY OF THE FINDINGS. AFTER **EVERYONE** HAS TURNED IN THEIR LAST SET OF MATERIALS, YOU SHOULD FEEL FREE TO DISCUSS THE STUDY WITH EACH OTHER. I WILL ALSO TRY TO ANSWER ANY QUESTIONS I CAN AFTER THAT TIME.

THAT'S ALL WE HAVE TO DO FOR NOW. REMEMBER, YOUR FIRST TASK IS TO DEVELOP A **FIVE** CATEGORY SYSTEM TO CODE THE RESPONSES TO THE OPEN-ENDED QUESTION. YOUR BASIC INSTRUCTIONS ARE ON THE SECOND PAGE OF **READ ME FIRST**. ITEM 12 LISTS THE TWO THINGS YOU NEED TO BRING BACK TO ME AT THE **NEXT CLASS MEETING** ON _____ :

1) THE **CATEGORY DEVELOPMENT WORKSHEET** WITH ONE IDENTIFIER AND ONE SUMMARY FOR **EACH** OF THE **FIVE** CATEGORIES YOU DEVELOP; AND 2) THE **SAMPLE OF RESPONSES** WITH EACH RESPONSE CODED INTO **ONLY ONE** OF YOUR FIVE CATEGORIES.

THANKS FOR YOUR HELP WITH THIS STUDY. I'LL SEE YOU AT THE NEXT CLASS MEETING!

Exhibit 2

Content Analysis
Answers to Four Practical Questions

1. What is it?

"**Content analysis** is a research technique for making replicable and valid inferences from data to their context" (Krippendorff, 1980).

The main purpose of using content analysis is to make **inferences--draw conclusions--**from a set of **data**, such as, documents, conversations, survey responses, paintings, or photographs, to their **context--**the original source and situation in which the data were created. As a **research technique**, content analysis has its own special procedures for processing information. These procedures are designed to encourage inferences made to be both **replicable--reliable**, or reproducible by different researchers at different times and places--and **valid--justifiable** in terms of a well-defined, symbolic interpretation of the data.

2. What are its uses?**General Uses** (Holsti, 1969)

- * to describe characteristics of communication--asking **what, how,** and to **whom** something is said
- * to make inferences as to the antecedents of communication--asking **why** something is said
- * to make inferences as to the effects of communication--asking with **what effects** something is said

Specific Examples (Berelson, 1952)

- * to describe trends in communication content
- * to trace the development of scholarship
- * to disclose international differences in communication content
- * to compare media or "levels" of communication
- * to audit communication against objectives
- * to construct and apply communication standards
- * to aid in technical research operations (to code open-ended questions in survey interviews)
- * to expose propaganda techniques
- * to measure the "readability" of communication materials
- * to discover stylistic features
- * to identify the intentions and other characteristics of the communicators
- * to determine the psychological state of persons or groups
- * to detect the existence of propaganda (primarily for legal purposes)

Exhibit 2 (continued)

-
- * to secure political and military intelligence
 - * to reflect attitudes, interests and values ("cultural patterns") of population groups
 - * to reveal the focus of attention
 - * to describe attitudinal and behavioral responses to communications

3. When conducting surveys, when should it be used with open-ended questions, instead of using quantitative analysis with forced-choice questions?

Advantages, Disadvantages, & Guidelines for Choosing Between Forced-Choice and Open-Ended Questions (Demaline & Quinn, 1979)

Forced-Choice Questions

Advantages

1. It is easier for respondent to answer.
2. Focuses respondent's answer on issues and data of importance to you. Respondents categorize themselves instead of you categorizing them.
3. More questions can be asked because time is saved by the respondent simply checking.
4. Precoded answers are easily analyzed.

Disadvantages

1. It requires advance information about possible response categories that may be given.
2. It may bias responses by suggesting answers.
3. It does not allow for diversity and richness in individual expression.

Open-Ended Questions

Advantages

1. It can easily be formulated without knowing the full range of answers that may be given.
2. It can accommodate questions for which a wide range of different answers will be given.
3. It does not condition or bias the answer as much as the forced-choice question.

Disadvantages

1. It requires the respondent to write a lot. Communication skills may influence the answer, in addition [to] the other respondent characteristics.
2. The respondent may address different facets of the question in which you may not be interested or may not give complete information in answering the question.
3. Fewer questions can be asked in a questionnaire because answers may be lengthy and time consuming to give.

Exhibit 2 (continued)

-
4. Responses are difficult to analyze. The investigator must devise a coding scheme and then categorize responses based on this scheme. The diversity and richness of responses are usually reduced by this process, and it is time consuming.

Use Forced-Choice Questions & Quantitative Analysis If

1. it is important that the effort or verbal skills of the respondents be kept relatively low, OR
2. there is a clear understanding of what the likely or important responses will be, OR
3. a large number of questions need to be asked in relation to the time available, OR
4. it is important that the responses be easy to code for analysis.

Use Open-Ended Questions & Content Analysis If

1. the full range of likely and valid responses is not known or a wide range of responses is expected, OR
2. there is concern for biasing respondents if a set of possible answers is given, AND
3. the skills and time needed for coding potentially complex responses are available.

4. How is it done?**Steps in Content Analysis (Krippendorff, 1980)****Design (aspects)**

- * applying the framework for content analysis
- * searching for suitable data
- * searching for contextual knowledge
- * developing plans for unitizing and sampling
- * developing coding instructions
- * searching for contextually justifiable procedures
- * deciding on qualitative standards
- * budgeting and resource allocation

Execution (contains one or more of the following)

- * sampling by sampling units until the sample can be judged sufficiently representative of the population
- * identification and description of recording units which must be reproducible and satisfy criteria of semantical validity where applicable
- * data reduction and transformation of data into a form required for analysis, retaining all relevant information
- * application of context-sensitive analytical procedures (analytical constructs) to yield inferences
- * analysis, identification of pattern within inferences, testing hypotheses regarding relations between inferences, and results obtained by the methods and pragmatic validation of findings

Exhibit 2 (continued)

Report (should be specific about some or all of the following)

- * a statement of the general problem to which the research pertains
- * an account of the background of the problem
- * a statement of the specific objectives of the content analysis
- * a justification of the choice of data, methods and design
- * a description of the procedures actually followed
- * a presentation of the findings
- * a self-critical appraisal of the procedures followed and the results obtained

References

- Berelson, B. (1952). **Content analysis in communications research**. New York: Free Press.
- Demaline, R. E., & Quinn, D. W. (1979). **Hints for planning and conducting a survey and a bibliography of survey methods**. Kalamazoo, MI: Western Michigan University, The Evaluation Center.
- Holsti, O. R. (1969). **Content analysis for the social sciences and humanities**. Reading, MA: Addison-Wesley.
- Krippendorff, K. (1980). **Content analysis: An introduction to its methodology**. Beverly Hills, CA: Sage.
-

Exhibit 2 (continued)

**A Comparison of 1983 vs. 1984 Responses to the Question:
What is the most important issue facing the schools today?**

	<u>1983</u>	<u>1984</u>
A. Finances:	39%	32%
	e.g., The cost of a good education is getting too high.	
B. Discipline:	36%	39%
	e.g., Kids don't take responsibility for their actions any more.	
C. 3 R's:	10%	16%
	e.g., It's time to get back to the basics.	
D. Other:	15%	13%
	e.g., You have no right to close the school in our neighborhood! We need fewer administrators and more teachers.	

**Most Important School Issue?
(Number of Voters)**

	<u>Finances</u>	<u>Discipline</u>	<u>3 R's</u>	<u>Other</u>	<u>Total</u>
1983	279	256	70	104	709
1984	231	281	111	89	712

(Contributions to Value of Chi Square)

1983	2.36	0.56	5.89	0.57	
1984	<u>2.61</u>	<u>0.51</u>	<u>3.72</u>	<u>0.67</u>	
Total	4.97*	1.07	9.61*	1.24	16.89 = Chi Square

* Largest contributions to value of Chi Square sig. at alpha = 0.01

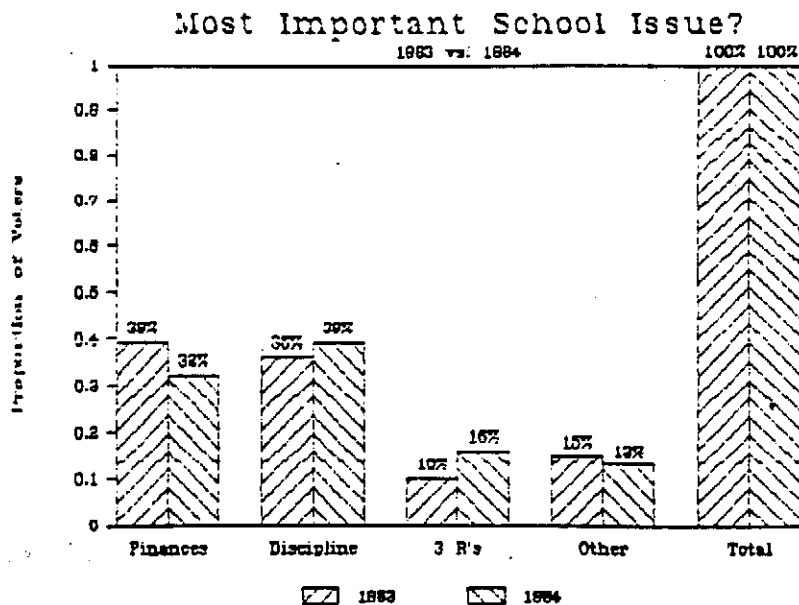


Exhibit 3

Read Me First

You are a student employee at the Center for Evaluating New Techniques in Educational Research (CENTER) at Independent State University (ISU). Your job is to assist the director, Dr. Valery Powerful, in whatever tasks she needs to accomplish. Because you have shown exceptional talent in past performances, Dr. Powerful has much confidence in your abilities--often giving you a chance to work on very difficult assignments.

The newest project at the CENTER involves assessing a controversial accountability system of a moderately large public school district in a nearby state. This project is jointly funded by the local, state, and national education associations. Because of several setbacks in devising a design agreeable to all parties concerned, particularly the superintendent, Dr. Powerful and the sponsors finally decided to use teacher responses to a mail questionnaire as the major data source. Consequently, this study has become an attempt to review the accountability system from the perspective of the teachers.

So far, you have not worked on this project because of your extensive involvement in another project, A Content Analysis of Homework Assignments for Introductory Research Courses at the Graduate Level. However, Dr. Powerful has just been called away for four weeks to attend an emergency meeting of the National Network of CENTER Directors in Honolulu. Before she went to the meeting, she put together some materials she was working on and left you this note:

Please take over the work on this accountability study in my absence. All of the questionnaires have been returned. Responses to the forced-choice questions have already been coded so don't worry about those. I would like you to spend your time working on the responses to the open-ended questions. My secretary has already typed a draft of my introduction to the report, a sample of the responses to the open-ended question, and some worksheets I had planned to use.

I should warn you that the responses I have read so far have been overwhelmingly negative. This means that categories like "Negative Comments About the Accountability System" or "Negative Comments About the Superintendent" would not be particularly helpful. Instead, I want you to create a set of five categories, each of which captures a particular set of negative comments about the accountability system. These categories should be a) exhaustive--all

Exhibit 3 (continued)

categories fit somewhere, b) mutually exclusive--there is one best category for each response, and c) consistent with the general framework in Figure 1--all categories are at the same level of detail and they all represent one group of similar negative comments about the accountability system. I made up a practice exercise for you to do before you start working on the actual responses. This should give you an idea of the level of detail and tone I want the categories to capture. **DO NOT** use any of these categories when you make up five of your own. My secretary will help you with the practice exercise when you are ready to begin.

Here is how I would like you to proceed:

1. **READ** the **DRAFT INTRODUCTION** of the report to get a flavor of the study.
2. **COMPLETE** the **PRACTICE EXERCISE**.
3. **GO HOME** and work on this **ALONE!** I want an **INDEPENDENTLY DEVELOPED CATEGORY SYSTEM**. If you have any questions about how to work on this assignment, call my secretary, Dick Frisbie, at 383-8166 (work) or 375-4271 (home). I had a chance to talk to him about how I wanted this done before I left for the emergency meeting.
4. **FIND A LARGE, CLEAR WORK AREA TO SPREAD OUT YOUR MATERIALS**.
5. **READ** the **SAMPLE OF RESPONSES** to the question.
6. **CREATE EXACTLY FIVE CATEGORIES**, no more, no less, **INTO WHICH ALL RESPONSES MAY BE CODED**. [Powerful people are sometimes arbitrary.]
7. **USE A "GARBAGE" CATEGORY** like "Other" or "Miscellaneous" **IF YOU DECIDE ONE IS NECESSARY**.
8. Use whatever techniques you think would work best for deciding what the five categories should be, except, **DO NOT CUT UP THE CATEGORY DEVELOPMENT WORKSHEETS** or **SAMPLE OF RESPONSES**. My secretary needs to use them after you complete your task.
9. **EACH CATEGORY MUST HAVE** a) a **NUMBER**--use 1 through 5, b) an **IDENTIFIER**--a brief, descriptive title for the category which indicates its negative nature, and c) a **SUMMARY**--an

Exhibit 3 (continued)

operational definition of the category complete enough to allow people besides yourself to decide whether a particular response should or should not be included in that category.

10. **WRITE ONE IDENTIFIER AND SUMMARY FOR EACH NUMBER ON THE CATEGORY DEVELOPMENT WORKSHEETS.**
11. **PUT** the appropriate **CATEGORY NUMBER NEXT TO EACH RESPONSE** on the **SAMPLE OF RESPONSES**.
12. **RETURN** the **CATEGORY DEVELOPMENT WORKSHEETS** and the **SUMMARY OF RESPONSES** to my secretary **ON JULY 19, 1984**. He will then process your work and return it to you for a final review before we start coding all of the responses.
13. Because my secretary will return your work to you after he has processed it, **DO NOT SPEND MORE THAN ABOUT ONE HOUR TO COMPLETE THIS TASK**. You will get another chance to review your work and make any changes you think are necessary.

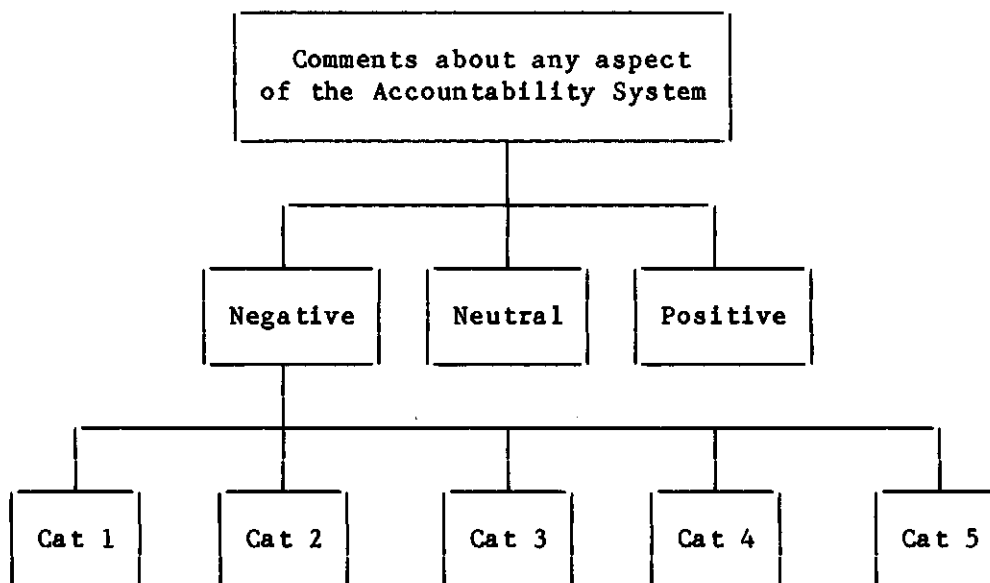


Figure 1. A general framework for developing a category system to code responses to the open-ended question, "Please give us any comments or recommendations you would like to make about any aspect of the Hometown Public Schools accountability system."

Exhibit 4

Draft Introduction

DRAFT

The CENTER, Independent State University

USING A MAIL SURVEY TO ASSESS THE ACCOUNTABILITY SYSTEM
FROM THE PERSPECTIVE OF THE TEACHERS

Introduction

The report which follows is an evaluation of the Hometown Public Schools accountability system from the point of view of the teachers. At the time of this study the Hometown accountability system had been operating for four years under the leadership of the current superintendent. In a report issued by the School Administration two years ago the following background was provided:

Hometown Public Schools, in an effort to plan logically for improving academic achievement of all students and, at the same time, increase efficiency of school operation, organized a Department of Research and Development one school year ago.

Hometown students in grades one through nine were tested in September and May one school year ago, using the nationally standardized Hometown Achievement Tests (HAT). Students in all grades are being tested during the current school year.

Teachers receive test scores of individual students, composite scores for their class groups, and an analysis showing which items on the test were most frequently missed by their students.

Teachers also administer their own Goals & Objectives for Achievement on Teacher-made (GOAT) tests to determine how well students are learning the classroom material and/or skills.

This continual flow of information helps teachers to monitor their own performance and guides them as they plan lessons to move each student toward greater achievement and the conquering of weaknesses which inhibit that achievement.

Exhibit 4 (continued)

The salaries of all Hometown school administrators, including the superintendent, are based on annual performance evaluations.

At the beginning of each school year, administrators develop performance objectives. At the end of the school year, a "percentage of accomplishment" is determined and ratings by several relevant groups are considered.

Performance profiles are developed for all teachers. The profile, which is shared with the teacher as a source of information for improving teacher performance, includes views of students, the principal, other teachers, parents, and the teacher's self-evaluation plus extensive student achievement data.

Information and comments are gathered, interpreted, and reported with the help of Research and Development staff members. The collected information is used to guide school personnel toward improving their performance.

The Hometown accountability system has received considerable national attention. For example, the **Journal of American School Boards** reported last year that by this summer "Hometown schools will probably have one of the most comprehensive computerized systems of personnel evaluation and accountability yet devised" (p. 34).

On the other hand, conflict over the school system's accountability program has been high for several years; charges and countercharges have been exchanged regularly between the Hometown Education Association (HEA) and the Office of Superintendent of Schools. The HEA, for example, has charged that teachers were being demoralized; the superintendent has argued that teachers don't want to be accountable. Politically charged statements have flowed with increasing frequency from both sources since the beginning of the accountability system--making constructive dialog increasingly more difficult.

This spring the HEA sought assistance from the state and national education associations to engage in a review of the accountability system. Because repeated attempts to enlist the participation of the Hometown Public Schools Administration have failed, this study has used as its major data source teacher responses to a mail questionnaire administered during the first week of June.

As the conflict between the administration and teachers has grown (including unsigned contracts, lawsuits, and countersuits), the HEA has asserted that the accountability system as implemented in

Exhibit 4 (continued)

Hometown is destructive to the educational system. This survey of teacher attitudes about the experiences with the accountability system was undertaken in large measure to test that assertion. Therefore, the report which follows is an evaluation of the Hometown Public Schools accountability system from the perspective of the teachers.

Note. Blind adaptation from Patton et al. (1976).

Exhibit 5

Practice Exercise

ISU Accountability Study for Hometown Public Schools

PRACTICE EXERCISE

1. Using the sample responses as input, write an Identifier and two Summaries to complete the category system.
2. Using the category system you completed, assign category code numbers to the responses which do not yet have a code.

Categories

Category #1. Identifier: Teaching to the Test

Summary: Teachers drill their students with the actual test questions and answers prior to administration of the test.

Category #2. Identifier: Reprisals Against Teachers

Summary: _____

Category #3. Identifier: _____

Summary: _____

Responses

No. Cat. Response

1. ___ If a teacher voices an opinion that is in disagreement with the administration, the teacher is subject to transfer and is given the most difficult students to try to teach. This has happened to me.
2. ___ We have lost focus on what kids in Hometown need to learn, and are teaching to a test.
3. ___ If you disagree, don't say so or you will be reprimanded.

Exhibit 5 (continued)

-
4. ___ It is my opinion that Hometown uses the "Accountability" program that they do in order to present a front to the public.
 5. ___ As an untenured teacher I feel that my job is in jeopardy if I voice my opinion contrary to policy or the superintendent's feelings.
 6. ___ Some teachers start to rehearse their students in January for the spring testing in May.
 7. ___ This entire accountability system is a "game" designed to impress the community.
 8. ___ The worst thing about accountability systems is that you find yourself teaching to the test.
 9. 3 As you may well be aware, the program has resulted in a situation in which a huge amount of time and effort is spent in a "whitewash" attempt to appear highly successful.
-

Exhibit 6

Practice Exercise Answer Sheet

ISU Accountability Study for Hometown Public Schools

PRACTICE EXERCISE: ANSWER SHEET

1. Using the sample responses as input, write an Identifier and two Summaries to complete the category system.
2. Using the category system you completed, assign category code numbers to the responses which do not yet have a code.

Categories**Category #1. Identifier: Teaching to the Test**

Summary: Teachers drill their students with the actual test questions and answers prior to administration of the test.

Category #2. Identifier: Reprisals Against Teachers

Summary: Teacher challenges to individual administrators or the accountability system in general are met with involuntary transfers, reprimands, dismissals, and other types of reprisals--real or anticipated.

Category #3. Identifier: Administration Misleading the Public

Summary: The administration deliberately misrepresents the implementation process and effects of the accountability system.

Responses**No. Cat. Response**

1. 2 If a teacher voices an opinion that is in disagreement with the administration, the teacher is subject to transfer and is given the most difficult students to try to teach. This has happened to me.
2. 1 We have lost focus on what kids in Hometown need to learn, and are teaching to a test.
3. 2 If you disagree, don't say so or you will be reprimanded.

Exhibit 6 (continued)

-
4. 3 It is my opinion that Hometown uses the "Accountability" program that they do in order to present a front to the public.
 5. 2 As an untenured teacher I feel that my job is in jeopardy if I voice my opinion contrary to policy or the superintendent's feelings.
 6. 1 Some teachers start to rehearse their students in January for the spring testing in May.
 7. 3 This entire accountability system is a "game" designed to impress the community.
 8. 1 The worst thing about accountability systems is that you find yourself teaching to the test.
 9. 3 As you may well be aware, the program has resulted in a situation in which a huge amount of time and effort is spent in a "whitewash" attempt to appear highly successful.
-

Appendix E
Researcher/Instructor Contract

AGREEMENT TO PARTICIPATE IN A CONTENT ANALYSIS STUDY

PARTIES TO THE AGREEMENT

This agreement is between Richard D. Frisbie, hereinafter referred to as the Researcher, and _____, hereinafter referred to as the Instructor. The Instructor agrees to participate in a content analysis study conducted during the Fall Semester, 1984, at Western Michigan University, by the Researcher. Instructor participation includes permitting the Researcher to recruit the students, hereinafter referred to as the Participants, enrolled in the Instructor's class, _____: _____, conducted during Fall Semester, 1984, at WMU.

OVERVIEW OF THE STUDY

The purpose of the study is to advance the body of knowledge concerning the extent to which computer outputs, produced from commercially available general purpose programs for microcomputers, and based on techniques for survey and discovery in content analysis, can be used to improve the reliability and validity of content analyses conducted on responses to open-ended questions found in evaluation studies. A set of four research hypotheses has been developed to assert that such techniques will indeed improve both the reliability and validity of category system development and coding responses in relation to an existing category system. The experimental study the Instructor hereto agrees to participate in is designed to test these hypotheses.

The experiment will use a simulation problem which is based on a non-trivial, real-world evaluation study. The study used a mail questionnaire to obtain the reactions of teachers to a controversial accountability system of a public school system in the Midwest. Some of the information obtained from the teachers was in the form of responses to a set of open-ended questions. The focus of the experiment will be to determine if certain types of computer output related to these responses will help experimental participants produce more reliable and valid category systems as well as, code responses with more reliability and validity, when compared to control participants.

SUMMARY OF THE PROCEDURES

The Participants of the experiment will receive a classroom lecture about the theories and methods of content analysis from the Researcher. The Participants then must independently perform a content analysis of 100 responses in four parts: (1) develop a

category system; (2) verify the system; (3) code a set of responses; and (4) verify the codes. The experiment will require eight classroom sessions to allow sufficient time for the Researcher and Participants to perform their respective tasks.

The schedule of group sessions (S_n) and Participants' out-of-class tasks (T_n) are summarized below:

	<u>Date</u>	<u>Summary</u>	<u>(Time Estimate)</u>
S1	_____	Lecture, Setup, Ss receive 1st task materials	(1 hr)
T1	_____	Prtcpts perform 1st task: develop a category system	(1 hr)
S2	_____	Participants return 1st task materials	(10 min)
S3	_____	Participants receive 2nd task materials	(10 min)
T2	_____	Prtcpts perform 2nd task: verify category system	(1/2 hr)
S4	_____	Participants return 2nd task materials	(10 min)
S5	_____	Participants receive 3rd task materials	(10 min)
T3	_____	Prtcpts perform 3rd task: code a set of responses	(1 hr)
S6	_____	Participants return 3rd task materials	(10 min)
S7	_____	Participants receive 4th task materials	(10 min)
T4	_____	Participants perform 4th task: verify the codes	(1/2 hr)
S8	_____	Participants return 4th task materials	(10 min)

The Researcher also agrees to provide an in-class debriefing on the study on _____. Because of the timing of the debriefing, only preliminary findings will be presented.

NON-DISCLOSURE

Because the study uses a "double-blind" design, the Instructor agrees to refrain from (a) seeking information about assignments to treatments or (b) disclosing to the Participants the details of the problem, research design, variables, assignments to treatments, procedures, analyses, or any other information which would jeopardize the integrity or validity of the experiment until after all data is collected. The Instructor will also discourage Participants from discussing any details of the experiment until all data has been collected.

In addition, the Instructor agrees to refrain from distributing any descriptions of the study or its findings: (a) at public meetings, such as, conferences, workshops, or conventions; or (b) in written forms, such as, personal communications, monographs, or sections in newsletters, magazines, journals, or books; until after the successful defense of the related dissertation by the Researcher.

NON-INTERFERENCE

The Instructor agrees to refrain from helping the Participants to perform any of their assigned tasks related to the experiment, will discourage Participants from working together, and will discourage Participants from seeking help from anyone other than the Researcher. All questions about the study should be directed to the Researcher at 383-8166 (work) or 375-4271 (home).

COSTS OF MATERIALS

The Researcher will arrange to pay for all handouts and materials.

INFORMED CONSENT

The Researcher agrees to provide the Participants with an explanation of (a) the nature and purpose of the study in general terms, (b) the procedures to which the Participants will be exposed, (c) the extent of Participant confidentiality and anonymity and (d) the clear statements that a Participant may, without prejudice to him/her, withdraw from the study, cease participation, and/or have his/her data destroyed at any time of his/her choosing.

INCENTIVES FOR THE PARTICIPANTS TO COMPLETE THE STUDY

The Instructor agrees to actively encourage the Participants to complete the study. Examples of the types of phrases the Instructor will use to encourage the Participants to complete the study include: (a) "Research is very important and you are making a very important contribution to this particular research study." (b) "You are helping to answer some very important questions about the practical uses of computers for conducting content analysis." (c) "Participating in research studies helps advance the body of knowledge for a field and leads to the development of new and innovative practices."

The Instructor also agrees to give specific incentives for Participants who complete all four tasks of the experiment (Check those that apply.)

1. Replace one regular assignment. Participants who do not complete all four tasks must complete the assignment.
2. Provide _____ Bonus points;
3. Award the higher of two grades in "borderline" cases.
4. _____

MODIFICATIONS TO THE AGREEMENT

This agreement constitutes a general framework for participation in a content analysis study. Procedures not explicitly described in this agreement shall be conducted according to the Methods section of the dissertation proposal related to this study. Modifications to the explicitly stated sections of this agreement may be made only with the mutual consent of the Instructor and the Researcher. Such modifications shall be documented as a signed and dated amendment to this agreement.

SUMMARY OF BENEFITS TO THE PARTIES

The Instructor/Participants will receive: (a) a classroom lecture on the theories and methods of content analysis; (b) experience with a real-world content analysis problem; (c) experience with a true experimental design research study; and (d) a debriefing about the study after data collection has been completed. The Researcher will receive a portion of the sample of participants needed to conduct a dissertation-related experiment.

SIGNATURES

Instructor	Date	Researcher	Date