Content Analysis Methodology and Applications to Curriculum Evaluation*

Amare Asgedom**

Introduction

Classification of educational research methods has always varied with authors and their inclinations. Some use analytical methods (Cates, 1985) as criteria for classification and others use type of research-designs for the same purpose (Best and Khan, 1989). The former emphasizes methods of analysis while the latter methods of data collection.

Current educational research scientists have often been heard applauding a dichotomous thinking, qualitative versus quantitative research, basic versus applied research, etc. Classification of educational research by type of design, however, seems to reflect the highest frequency in the textbooks of the field. Many would use the categories: Descriptive Methods, Historical Method and Experimental Methods.

The criterion of time has been a major difference in the former two. The latter is different from the former in methodological rigor and design. It can, however, be easily observed that a common criterion has not been used in all classifications, a weakness that needs to be redressed soon.

A preliminary study of the state of the art in the Ethiopian context including educational MA theses demonstrates overuse of the Descriptive Method of research. Although of the highest relevance in education, Experimental Research appears to be the least applied by Ethiopian educational researchers. The few that have appeared in the literature could be classified as pre-experimental studies and not as true-experimental designs.

The Descriptive Method is not a method as such as it involves different designs. These include Surveys, Content Analysis, Case Studies, Assessment Studies, Evaluation Studies, Comparative Studies and Developmental Studies.

---

* Presented to a Workshop, on Improving the Quality of Curriculum Materials, Mekelle, December, 1998.

** Director, Institute of Educational Research, AAU.

Content Analysis Methodology and Applications to Curriculum Evaluation*

Amare Asgedom**

Introduction

Classification of educational research methods has always varied with authors and their inclinations. Some use analytical methods (Cates, 1985) as criteria for classification and others use type of research-designs for the same purpose (Best and Khan, 1989). The former emphasizes methods of analysis while the latter methods of data collection.

Current educational research scientists have often been heard applauding a dichotomous thinking, qualitative versus quantitative research, basic versus applied research, etc. Classification of educational research by type of design, however, seems to reflect the highest frequency in the textbooks of the field. Many would use the categories: Descriptive Methods, Historical Method and Experimental Methods.

The criterion of time has been a major difference in the former two. The latter is different from the former in methodological rigor and design. It can, however, be easily observed that a common criterion has not been used in all classifications, a weakness that needs to be redressed soon.

A preliminary study of the state of the art in the Ethiopian context including educational MA theses demonstrates overuse of the Descriptive Method of research. Although of the highest relevance in education, Experimental Research appears to be the least applied by Ethiopian educational researchers. The few that have appeared in the literature could be classified as pre-experimental studies and not as true-experimental designs.

The Descriptive Method is not a method as such as it involves different designs. These include Surveys, Content Analysis, Case Studies, Assessment Studies, Evaluation Studies, Comparative Studies and Developmental Studies.

---

* Presented to a Workshop, on Improving the Quality of Curriculum Materials, Mekelle, December, 1998.

** Director, Institute of Educational Research, AAU.

Of all these designs, however, the method of Content Analysis has been neglected and inappropriately used in the Ethiopian context. It has always taken the form of making generalization without systematic and objective approaches. It is not an exaggeration to argue that it was used in the same way as a *Review of literature* which is just a part but not the whole of research.

This article uses available literature to present the methods, nature and application of Content Analysis in Education. Of course, this method is highly developed in the fields of communication but still underdeveloped and not well adapted to educational research, albeit its importance to the field.

What is Content Analysis?

The technical definition of Content Analysis varies from writer to writer. Stone and others (1966:5) have defined it as a research technique for making inferences by systematically and objectively analyzing specific characteristics within a *text*. In the same vein, Krippendorff (1980:20) said that it is a scientific method that yields inferences from essentially verbal, symbolic or communicative data. Those communicative data could be in written forms, broadcasts, visual or audio recordings or, they could also be computer texts. The media could be Print, CD-ROMS, diskettes, tapes, CDs, cassettes, etc.

Content Analysis is a formal system for doing something that we all do informally rather frequently, drawing conclusions from observations of content (Stempel III, 1981:119). We express, opinions about adequacy of various kinds of coverage by newspapers, magazines, radio stations and TV stations. Those opinions are based on what we observe as readers or listeners (Stempel III, 1981: 119).

As people, objects, vehicles, trees, could be the source of information in Survey Designs, words, phrases, sentences, paragraphs, or entire texts could be the source of data in Content Analysis research.

In survey studies, a questionnaire could be utilized to ask respondents to answer certain structured or unstructured questions. If respondents are illiterate their responses need to be recorded (without any alteration) by what we call personal interviewers. These people read the questions to them and just record their reactions however unsatisfactory these reactions might be. These interviewers are there only to mediate the problem of illiteracy. They are the hands (coders) of illiterate respondents.

In parallel to the questionnaire, a *Coding Sheet* is prepared in Content Analysis and coders are used to record their observations objectively. The Coding Sheet must, however, be highly structured in such a way that it minimally allows the opinion of the coder. In Content Analysis, the presence of coders is mandatory as written or broadcast materials cannot make self-reports in parallel with the use of personal interviewers (while using questionnaires to illiterate respondents).

This paper uses Berelson's definition (1952) of Content Analysis. *Content Analysis is a research technique for the objective, systematic and quantitative description of the manifest content of communication* (Berelson, 1952:18).

Objective

Objectivity means prevention from the influence of the opinion of the researcher or coder on the data collected. Impressions and subjectivity are controlled from affecting the data collection process. The researcher needs to be a detached but careful observer. Objectivity is achieved by having the *categories of analysis* defined so precisely that different persons can apply them to the same content and get the same results. If Content Analysis were subjective instead of objective, each person would have his/her own Content Analysis. The results need to depend upon the procedure and not the analyst.

Quantitative

Quantitative means simply the recording of numerical values or the frequencies with which the various defined types of content occur. Content Analysis does not exclude qualitative approaches; in fact, conventional Content Analysis studies have overemphasized the qualitative aspect. The intention here is to bridge the gap in Content Analysis studies. Large masses of content stuff can be systematically and scientifically studied with the help of statistics. The use of measurements, frequencies, and ranks help data management systems which otherwise could be too enormous to handle objectively. For instance, the assignment of nominal values to certain content characteristics such as assigning 3 to every lesson plan that has videotape or films; a 2 to every lesson plan that has still pictures; and a 1 to plans with no visuals. This approach highly simplifies the data management effort in testing the extent to which instructional materials appear in lesson plans of the teacher.

Manifest

Manifest content means the apparent content; that is, the content must be coded as it appears rather than as the content-analyst feels. If intent is analyzed, no matter how correct the interpretation is, the research will suffer from subjectivity. Another analyst might come up with different results. The outcome of such studies will depend more on the type of researcher rather than on the procedures of research. We might have as many outcomes as there are researchers. It is, however, often observed that studies that inquired into the meaning of contents are more frequent in the Ethiopian context. Yet, the common definitions of these meanings need to be agreed upon to minimize personal effects. Using Inferential and Evaluative coding systems (which will be discussed later in this paper), one can study with caution the intent and meaning of a given content.

Systematic

According to Stempel III (1981:120) systematic means: categories are set up; all relevant content is analyzed; and a set of procedure is applied in the same way to all content being analyzed. Four methodological issues stand out in Content Analysis. These are:

-Selection of Unit of Analysis
-Category Construction
-Sampling of Content
-Reliability of Coding

Selection of Unit of Analysis

Words, sentences, phrases, statements, paragraphs, or entire articles or books can serve as Unit of Analysis. Time and space are often used by communication researchers as Units of Analysis both for print and broadcast materials. In more sophisticated Content Analysis studies, bits and nits are also used as units and measurements of information (Amare, 1989).

In readability studies, words or sentences might be counted. Entire pages might be used for studying coverage. For instance, one would count pages or paragraphs to study the amount of material in a text--devoted to student activities versus teacher activities. Which type of Unit of Analysis to select is, therefore, dependent upon what kind of information is required from the Content Analysis study.

Category Construction

In curriculum evaluations, the concrete-abstract dichotomy can be used as two categories for study. In order to assess the extent to which a communication material is balanced in coverage, the researcher can use the categories, such as, Knowledge, Skill and Attitudes (testing balance among the components of the KAS model).

A graduate student at the Addis Ababa University used nine categories of teacher characteristics (competence, discipline, human, respect, responsible, right, strong, wise, unemotional, etc.) to study the image of the teaching profession in books (Demis, 1990). Another Graduate Student used the categories, love-content, spatial-content, political-content, cultural-content, personal-content while making a Content Analysis study of recorded cassettes in Ethiopia (Woubie, 1996). Of course, such studies transcend the descriptive type of coding and involve more interpretive and evaluative approaches but still leaving many questions unanswered.

In any case, formation of categories depends on the purpose of the study. Categories created by other professionals are also helpful. Attention must, however, be paid to three important points:

-Categories must be pertinent to the objective of the study
-Categories should be functional
-Categories must be manageable

These three overlap and blend to one another. Pertinence is achieved if the information generated by the category answers the research questions or permits the testing of the hypotheses of the study. If we are dealing with categories where we have no access to or do nothing to change them, then these categories are not functional.

There is always a need for making the number of categories manageable. Once coders are familiar with the set of categories, they should be able to operate without frequent reference back and forth to the list and the definitions. With less than 10 categories, it is more or less not difficult to operate but somewhere between 10 and 20, you exceed the capacity of the coder. If you use more than 30 categories, you will realize that the task is approximating impossibility. Categories must also be all-inclusive and mutually exclusive. Blending and exclusions might pollute the results.

Sampling of Content

Sampling of content is not any different from sampling in Surveys. The essential consideration is to ascertain that each unit of the population has an equal probability of being represented in the sample. The study of the content of books might, however, demand a hierarchical process of sampling. Using the principles of randomization, entire books could be sampled, chapters could be sampled, even pages could be sampled. The process might flow to paragraphs and sentences. In sampling of newspapers, the procedure might involve sampling of years, months, weeks, days and pages of newspapers. More-often-than-not, Content Analysis sampling uses Cluster or Area sampling techniques.

In curriculum evaluation, subject areas may not be sampled, as each subject is a subject of study. But syllabi, textbooks, teacher-guides can be sampled for making inferences from the sample. Readability, illustrativeness, attractiveness of page layouts, etc. can be inferred from samples. It is, however, mandatory that probability-sampling techniques must be strictly adhered to in order to make the sample representative of the population.

Although we know that a sampling error approaches zero with an increase in size of the sample, the quality of sampling doesn't depend more on size than representativeness. For instance, **The Gallup Poll** (An American polling company) has always taken a sample size of less than 3000 people to predict election of all American presidents and is always right contrary to **The Digest Straw Poll** (American Polling Firm) which once took a sample size of 10 million postcard ballots with a net output of 2 million responses and to predict the right president (Simon, 1980:198).

Reliability of Coding

Content Analysis needs to be systematic and objective. The researcher must also be concerned with reliability--consistency of classification. If we have two coders, we would expect them to agree on a number of times in the coding process. As a practical matter, we know that two coders will not agree completely, but unless we achieve some level of agreement, we obviously cannot claim that our study is systematic or objective.

Disagreement between coders is usually the result of either of inadequate definition of categories, or failure of coders to achieve a common frame of reference. It is not

uncommon in the early stages of a study for there to be more disagreement than agreement among coders.

To increase reliability, the researcher needs to provide clear definitions. It is possible to conduct trial runs and coder responses compared item by item continuously. The process should continue until a common frame of reference is successfully achieved.

Once coding has begun, spot checks should be made to make sure that the reliability level is not deteriorating. Briefings should be conducted to deal with problems that coders feel the definitions and instructions do not cover completely. If the researcher is doing all the coding for the study, the reliability problem remains the same. It is possible to use a second person work with him\her initially and also have that person do some spot-checking with the researcher.
For any content study, a reliability estimate ought to be calculated and reported. A minimum standard would be the selection of five to six passages to be coded by all coders in a pilot format.

Statistical Approaches to Reliability

There are two different statistical strategies in reliability studies. One approach is to test intercoder agreement by using a simple correlation technique. For instance, two coders might code the readability of a book as indicated in Table 1.

Table 1: Coding Process of Different Chapters of a Book

Chapter No	1	2	3	4	5	6	7	8
Coder A	1	1	1	1	0	0	0	0
Coder B	1	0	1	0	1	0	1	0

1= agreement; 0= disagreement

In the eight chapters of the book (Table 1), the two coders have agreed four times and disagreed again four times implying a 50 percent agreement (r =0.5). This agreement is not more than a chance agreement. It tells nothing to the reader about consistency of the coding system.

The second approach is to estimate the agreement level by taking the chance factor into account. The following formula estimates the coefficient of agreement:

$$II = \frac{P_o - P_e}{1 - P_e}$$

Where p_o is the observed percentage agreement, p_e is the percentage ag
expected by chance. Hence, the cited example in Table 1 demonstrates that th
no agreement between the two coders that exceeded the chance factor. Blinc
might have yielded the same results. Here, the coders have not demonstrated
in coding.

Of course, mere exceeding the chance figure is not sufficient. Many res
recommend a figure of 90 percent when the coders are only two. The 5C
could have been sufficient if the number of coders was more than two.

Training Coders

For studies using several coders, selection of trained coders could have been of paramount importance. But this possibility is not often practical. The people researchers recruit to code may, never have done any coding before. It is probably desirable to have people with somewhat similar academic backgrounds to be backed up by training.

Developing a common frame of reference is the challenge in training of coders. But this is not an easy task. However much carefully the researcher defines his/her categories, there will still be situations in which coders lack complete agreement. These situations need to be identified, discussed, and agreed upon. Bear in mind that the coder does not need to accept the rationale for handling a situation; he or she merely needs to know what the rule is. It is possible that one will have a coder who simply cannot develop the same frame of reference as the rest of the group. When that happens, of course, the coder should be dropped from the coding group. This happens usually because of major differences in background between one coder and the rest of the group.

The work of coders needs to be checked as the study progresses. Opportunities should be provided for coders to discuss problems. A major value of such discussions is simply to promote the common frame of reference.

Types of Coding

There are three types of coding: descriptive, inferential and evaluative.

Descriptive Coding: The content recorded is either described in words or reported in terms of tallies—marks which record each time that content occurred. The coder does not record anything except what he or she actually observed. The coder draws no

inferences and makes no judgements. This type of observation is usually easily quantifiable, and coders are usually quite reliable in their recording. In this paper, advocacy was made on the use of Descriptive type of coding.

Inferential Coding: The coder is required to consider what each observed content is indicative of and then record that meaning under a specified category. This type of coding generally entails the use of a categorized recording system which requires the assignment of each content to one of a number of categories. Such categorical assignment is markedly less reliable than simple descriptive coding but often supplies more detailed and informative data. With thorough training, coders can learn to make consistent and reliable inferential coding.

Evaluative Coding: The coder must judge the quality of the content and then record this judgment as an ordinal rating. Of the three types of coding, evaluative coding is usually the most difficult, since coders frequently differ in evaluative judgments. One person's *good* may easily be another's *average*. Fortunately, careful and thorough training of coders coupled with frequent exercises in which observers compare evaluative ratings can result in sufficiently reliable coding. Use of a few, clearly defined rank ordered-categories in which to record behavior can also assist in obtaining reliable evaluative codings. In general, the more ordinal categories among which the coder is asked to distribute content and the finer the distinction among categories; the lower is likely to be the reliability of evaluative codings.

Applications of Content Analysis

There could be many applications of the method of Content Analysis. Communication scholars apply it to test the communication model, WHO says WHAT to WHOM with WHAT EFFECT.

Coverage, balance, bias, etc., have always been the major content-analysis issues by newspaper journalists. Historians and sociologists have also inquired stereotypes in history books. Educators and linguists have made readability studies of textbooks. Content Analysis is also used for analyzing the written contents of depth interviews by many researchers. In the ensuing discussion, examples of Content Analysis studies that use different coding systems have been presented to demonstrate application of this method in educational research.

Readability of Textbooks—Descriptive Coding

Textbooks are prepared with the assumption that they will be read and understood by students. This assumption, however, is seldom true. Students have often found their texts hard and less interesting to read. It becomes, therefore, essential to engage in regular evaluation and improvement of textbooks. It is also possible to adopt Flesch's Readability and Human Interest models (Flesch, 1974) to assess ease (difficulty) in reading and human interest. These models are depicted in Table 2 and Table 3.

Readability Model

R.E. = 206.835 - 0.846WL - 1.015SL
RE = reading ease score
WL = number of syllables per 100 words
SL = average number of words per sentence.

The resulting score should range between 0 and 100 and can be looked up in Table 2.

Table 2
Chart for Interpreting Flesch Reading Ease Scores

Reading Ease Score	Description Of style	Estimated Reading Grade
90-100	Very Easy	5th grade
80-90	Easy	6th grade
70-80	Fairly Easy	7th grade
60-70	Standard	8th and 9th grade
50-60	Fairly Difficult	10th to 12th grade
30-50	Difficult	13th to 16th grade (college)
0-30	Very Difficult	Graduate Level

Source: *Communication Theories by Severin and Tankard, 1982. Pp.74-75.*

Table 3
Chart for Interpreting Flesch Human Interest Score

Human Interest Score	Description Of Style
0-10	Dull
10-20	Mildly Interesting
20-40	Interesting
40-60	Highly Interesting
60-100	Dramatic

Human Interest Model

To examine whether or not a text is interesting, Flesch has developed a Human Interest Formula (H.I.):

H.I. = 3.635PW + 0.314PS
H.I. = human interest score

PW = number of personal words per 100 words
PS = number of personal sentences per 100 sentences.

The scores also fall between 0 and 100. Their interpretations can be read from Table 3.

Flesch defines a word as a letter, number, symbol, or group of letters, numbers or symbols surrounded by white space (Severin & Tankard, 1982:76). For instance, 1998, U.S.A. and pre-literate would be counted as three words. A sentence is defined, as a unit of thought grammatically independent and usually marked by a period, question mark, exclamation point, semicolon, colon or dash. Syllables are counted by the way you would pronounce the word.

It is understandable that Flesch's models (formulae) were developed in and for the western society. Those models and their corresponding interpretations may not have direct applications in the Ethiopian situation. They may, however, help in conceptualizing the problem and through repeated research, we can develop locally applicable readability formulae. In the two formulae the variables might have direct applications; the constants, however, need to be locally adapted through repeated research. For instance, Taylor's (1953) Cloze Procedure can help to establish a standard of the reading abilities of Ethiopian students. The name is taken from the word closure which stands for the human tendency to complete a familiar but incomplete pattern.

To use cloze procedure, you need to take a sample of writing and mutilate it by replacing some of the words with blanks. A common way is to replace every fifth word with a blank. Then the mutilated passage is given to a test group of subjects who are asked to fill in the missing words. The cloze score becomes the number of blanks that are filled in correctly.

Different passages can, therefore, be developed and tested again and again to derive the major factors of readability. These can be compared with results of Flesch's readability formulae and develop a model that could be functional in the Ethiopian context.

Textbook Evaluations—Inferential and Evaluative Coding

Content Analysis studies in Curriculum Evaluation usually take the forms of inferential or evaluative coding. Fewer studies are made with descriptive coding involving simple tallying and less judgmental methods. Textbook page-layout, presence or absence of visuals, student-centered problems, projects, etc. can be studied with the latter. The type of materials/methods suggested in the textbooks can

be studied without using judgmental coding. For instance, the categories Demonstrations, visuals, Field Trips, Direct Experiences, Radio and Recordings, etc. can be used as categories and their frequency counted in the textbooks. By doing so the researcher establishes the extent to which textbooks reflect the necessary experiences and materials. The occurrence of teaching methods (lectures, discussions, student activities, etc) in text can also be tallied to establish the dominant teaching method implied in the system. Many Curriculum evaluators have often used categories that demand inferential or evaluative coding systems. They have also applied Semantic Differential Scales, such as, satisfactory-dissatisfactory, acceptable-unacceptable, relevant-irrelevant, etc. Examples of frequently used categories are indicated below:

1. **Organization of the text**
   - Sequence
   - Content
   - Layout

2. **Content**
   - Relevance of ideas
   - Methods
   - Materials

3. **Coverage**
   - Knowledge
   - Attitude
   - Skills

4. **Presentation**
   - Clarity
   - Sequence

5. **Illustrations**
   - Relevance
   - Clarity
   - Simplicity
   - Layout

6. **Values**
   - Cultural sensitivity
   - Gender sensitivity
   - Stereotyping

To this list can be added content organization such as **development of ideas** in the text which might take the forms: concrete-abstract; near-far; simple-complex; redundancy-uncertainty, etc.

Provided that clear definitions are provided to these categories and provided that coders develop a common frame of reference, these categories can serve as useful mechanism of data generation for evaluating all kinds of curriculum materials.

Computerized Content Analysis

Computerized procedures for Content Analysis have been developed and used in some studies. Under optimal conditions, such procedures can save time and provide highly reliable coding. At the same time, you should recognize that there are many situations in which the computerized procedures will not be helpful. Our main purpose in discussing computerized content analysis is to help the reader recognize the potential of computerized procedures.

The strength of the computer is its ability to do fairly simple tasks extremely fast. Practically, what this means to the content analyst is that the computer is valuable for a study that involves recognition of words or even syllables (Descriptive Coding). A good example of an efficient use of the computer for Content Analysis could be the visibility study of Ethiopian leaders in American media and American leaders in Ethiopian media. Such a study leads to counting and identification of the names of the leaders of the two countries from a large mass of materials.

References

Amare Asgedom (1989). *The Spatial Organization of the Population of Ethiopia;* **Ethiopian Journal of Development Research. 9:1-20.**

Berelson, B. (1952). **Content Analysis in Communication Research.** New York: The Free Press.

Best, J. W. and J. V. Kahn (1989). **Research in Education,** 6th. Ed. London: Prentice-Hall International.

Cates, W. M. (1985). **A Practical Guide to Educational Research**. London: Prentice-Hall International.

Demis Zergaw (1990). **Images of the Teaching Profession in Four Ethiopian Novels.** MA Term Paper. Addis Ababa University: Unpublished.

Flesch,R. (1974). **The Art of Readable Writing. New York: Collier Books.**

Krippendorff, K. (1980). **Content Analysis: An Introduction to Its Methodology**. London: Sage Publications.

Severin, W. J. and J. W. Tankard (1982). **Communication Theories: Origins Methods, And Uses**. New York: Hastings House.

Simon, R. (1980). **Public Relations: Concepts and Practices,** 2nd. Ed. Columbus: Grid Publishing, Inc.

Stempel, G. H. (1981). *Content Analysis* in G. H. Stempel III and B. H. Westley, eds. **Research Methods in Mass Communication**. London: Prentice-Hall. Pp. 119-131.

Stone, P. J., D. C. Dunphy, M. S. Smith, and D. M. Ogilvie (1966). **A Computer Approach to Content Analysis**. Cambridge: MIT Press.

Taylor, W. L. (1953). *Cloze Procedure: A New Tool for Measuring Readability.* **Journalism Quarterly.** 30:415

Woubie Kassaye (1995). **Content Analysis of Secular Amharic Songs (1974-1994)**. MA Thesis. Addis Ababa University: Unpublished.