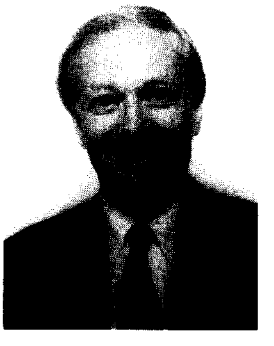
Reports should describe evaluation tools, products, and/or services found useful by practicing evaluators. Reports should also include the context of use and sufficient information for interested readers to contact the user directly for more detailed information; that is, name, address, and telephone number. Books, computer software and hardware, and new techniques—or old ones used in new ways—are just a few of the items that may be described.

A Technique for Objective Analysis of Qualitative Evaluation Data

HERBERT J. WALBERG, DAOXIN LU, RICHARD P. NIEMIEC and HERBERT J. WALBERG III



Walberg

ABSTRACT

Analysis of qualitative—usually verbal—information from interview transcripts, field notes, and other sources has required huge amounts of labor. Some critics, moreover, maintain that such analysis is less than reliable, objective and tabula rasa. A new computer program, however, allows precisely repeatable analysis without a priori frameworks and without tedious and subjective categorizing of passages. The program finds repetitive association of words or ideas that characterize the verbal research corpus which evaluators may display and interpret. This article explains the program and illustrates its use in an evaluation of a collaborative botanic education project sponsored by the National Science Foundation.

INTRODUCTION

Perhaps the most enduring debate in evaluation during the past few decades has been between the quantitative and qualitative camps. Many qualitative studies included data derived partic-

Herbert J. Walberg • University of Illinois at Chicago, College of Education, 1040 West Harrison Street, Chicago, IL 60607; Phone (312)996-8133; Fax: 996-6400, E-Mail: hwalberg@uic.edu.

Evaluation Practice, Vol. 18, No. 1, 1997, pp. 55–64.
ISSN: 0886-1633

ipant observations, interviews, focus groups, and responses to open-ended questionnaires. These were not entirely convincing to some quantitative advocates because such approaches seemed to lack objectivity. This article illustrates a new technique, which might be termed "objective" or "non-coded" content analysis that provides efficient, non-labor-intensive analysis of interview responses, focus group transcripts, archival documents, field notes, and other verbal material.

In the past, some content analysts appear to have compiled often extensive transcripts, coded passages according to preconceived categories, and counted the instances to determine the most common themes. They have typically stated their conclusions and often selected contextualized examples as illustrations. From such an approach, several problems ensue: subjectivity in formulating the set of categories, lack of reliability in coding or categorizing passages, and laborious effort in coding and analyzing the frequencies of the categories. The effort may be doubled or tripled if more than one coder is employed to check inter-coder agreement. If another researcher or evaluator were to repeat the work from the start, the categories chosen, codings made, and frequencies calculated might differ substantially—which would result in only partially reproducible results.

For these reasons, we employed an automated neural-network computer program named Galileo (Woelfel, Danielsen, & Woelfel, 1995) to summarize verbal material.[1] Unlike categorizing comments in traditional content analysis, the program has the capacity to find the strongest linkages among frequent words in the text. As explained further below, such analysis requires no a priori classification scheme. Such a "blank tablet" accords with current interest in qualitative methodology, made distinctive by open-ended inquiry, rather than employing prior hypotheses and investigator-imposed categories of data and analysis.

The common-sense premise of such analyses is that words that often appear close to one another are likely to be substantively related. If, for example, in a history of France, the words "Josephine" and "Napoleon" often appear together within a span of a few words (rather than separated by many paragraphs or pages), they are taken as related. Other nearby words or word combinations such as "courting," "longed for," "last letter," "missed her," "liaison," and "marriage" may suggest the nature of the relation even to people who never studied French history. Though a human might take months to count all such possible word co-occurrences in long texts, the computer can accomplish this in a few seconds—saving the evaluator's time to read and interpret the salient passages in context which the computer can readily find and display.

Such co-occurrence or word proximity analysis has several appealing features for qualitative research—one ideal of which is to investigate human discourse without investigator-imposed categories. As mentioned above, qualitative investigators typically identify features of verbal transcripts (or field notes); they then "triangulate" or look for repetitive confirmations of the features (and possibly lack of disconfirmations). These features can be subjectively identified (or "coded") and compiled to illustrate themes, variations on themes, and other systematic properties of texts. The quality of the results apparently depends on effort, experience, insight, and skill of analysts. Fielding and Lee's (1992) edited book offers varying ideals and methods. Skills and craft knowledge appear to be evolving, but much work is required to set forth agreed-upon standards and operational methods for sampling, analysis, and reporting.

Automated neural-network computer programs analyze word proximities and require no expert coding (though, as with older analytic methods, subsequent interpretation of results is required). The programs, which run on (IBM-compatible) personal computers, perform trillions of calculations. Both the input (text) and the output (word combinations or ideas), however, are qualitative. Finally, the programs yield perfect reliability; that is, identical results are obtained by different analysts using the same program with the same options on a given text (Danowski, 1980, 1993). For a more complete technical description of computer programs, see Danowski's publications.

If initial explorations of such programs prove fruitful, the method may help alleviate the massive work being undertaken by evaluators and other researchers. This article describes an automated method generally new to the mainstream of the social sciences which requires no category system and no coding. Two analysts employing the automated method would not produce close, but identical results in a few seconds. As in the case of previous content analyses, however, a transcript is required beforehand, and subjective interpretation of the results is required after the analysis (although some results are nearly self-interpreting). We employed the program in an evaluation of a National Science Foundation-sponsored, science project developed by a consortium of Midwest public gardens. After a brief overview of the program and our quantitative findings, we report the results of analyses of participants' open-ended comments.

THE MIDWEST GARDEN CONSORTIUM PROJECT

The consortium is a six-member group of public gardens that developed four- to six-week science units for kindergarten through sixth grade. The distinctive feature of the project was the employment of garden experiences as an integral parts of academic units; local flora displayed in the gardens, for example, are also depicted in the units and texts. After classroom study, participating students visited the gardens to carry out suggested activities with the guidance of garden staff. Follow-up activities then took place in their school classrooms.

To evaluate the project, we collected data via questionnaires given to the participating teachers and students and we interviewed their school principals. Data analyses of responses to the closed-ended items set the stage for our analyses of their open-ended comments. The teachers' responses to closed-ended Likert items (strongly disagree to strongly agree) were as follows:

- On aspects of the project surveyed, teacher satisfaction was high. Their overall quality-of-experience rating was 4.6 (out of 5).
- The teachers gave especially high ratings to hands-on and field activities.
- The project materials were highly rated. Teachers often cited the convenience of having all project materials in one place.
- Many teachers commented that there was insufficient time to cover the material and complete the activities. The time component also received the lowest rating of the ten aspects of curriculum.
- The strongly favorable results were highly consistent. The ratings of the units were not affected by unit, subject matter, or particular garden.

STUDENT AND TEACHER VERBAL RESPONSES

About 1,500 comments to open-ended questions were gathered from 570 students and 26 teachers. These questions (made explicit below) were formulated to gain an overall picture of the teachers' and students' experiences in the gardens and classroom work connected them. In answer to the six questions asked, the teachers wrote 3,996 words in all—the equivalent of 17 normal double-spaced pages. On average, each teacher wrote 154 words (or the equivalent of two-thirds of a normal typed page). Collectively, the students wrote 11,503 words (the equivalent of 49 typed pages). The average response length, however, was only 20 words to each of three questions.

The teachers' comments may provide the best insights for possible revisions in the unit redesign and use, but the reactions of the students may give the best insights on the states of mind of intended ultimate beneficiaries. Although the students on average wrote less than the teachers, there were many more of them. Both groups had much to say about the units. What could automated analysis tell us about the general patterns of meaning in both sets of responses?

Student Experiences

The computer program provides two major forms of output: a complete list of words used by all subjects in order of their frequency of use as an intermediate calculation, and "ideas" in the form of patterns of closely linked words that tended to co-occur in the respondents' answers. As in traditional content analysis, both output forms require some interpretation and explanation which may be best represented in an example. Consider the analysis of the student answers to the first student question: "What did you like best about the program?" (see Table 1).

TABLE 1
Frequently Used Words by Students in Response to the Question "What Did You Like Best about the Program?"

Word	*Percentage of All Words Used*
Liked	7.94
Plants	6.90
Going	4.55
Arboretum	4.11
Plant	2.74
Garden	2.68
Trip	2.14
Planting	2.08
Fun	1.97
Seeds	1.92
Field	1.31
Learning	1.26
Flowers	1.04
Best	.99
Experiments	.88
Everything	.88
Animals	.82
Making	.82
Growing	.82
Flower	.71
Total	46.56

As indicated by the total, these 20 most frequent unique words comprise nearly half of all the words in the students' answers. Although 1826 words were used (466 of them only once), these 20 words represent the most typical substantive responses.[2]

Though these are merely intermediate results, they are entirely straightforward, easy to understand, and offer insights into the most typical patterns of student responses—though, as in the case of all results, they require subjective interpretation. Several words in Table 1, for example, can be imagined as typical, single-word answers, such as "plants," "seeds," "flowers," "experiments," and [the] "trip." In addition, we may imagine sentences composed by the students and containing the most frequent words (the reader may want to try a few based on the above list before reading further). The first six words, for example, are contained in the following sentence: "I *liked going* to a *trip* to the *arboretum* [or *garden*] and seeing the *plants* [or a special *plant*] in the *garden*." Several of the next most frequent words are contained in: "*Planting seeds* in the *field* and *learning* about *flowers* were *best*."

The computer program does something like this, but objectively on the basis of frequency of co-occurrences of two or more words in the same passage. It directly yields sets of frequent, strongly linked words which may be considered predominant ideas expressed as quasi-sentences or phrases. For the first student question, the computer-generated patterns (in the order generated) can be seen in Table 2.

The numbering identifies successive computer runs. In the answers, adjacent words (joined with dashes) were most closely linked by the computer. Each new line indicates a set of associated words that are somewhat separate from words on the lines above and below. Lines in parentheses appeared to us as possibly idiosyncratic sets of words rather than meaningful ideas; they seemed less interpretable patterns than the other lines (readers may make their own inferences and judgements). These are included for the sake of objectivity and completeness.

The computer-generated patterns of responses suggest that students had favorable views that corroborate the closed-end responses. Examples of such words include: liked, fun, best, love, loved, different, and interesting. In addition, many of the verbs and adjectives suggest positively valued activities associated with the units and venues: learning, growing, planting, seeing, imaginary, different, search, working, visiting, experiment[ing], watching, measuring, drawing, smell[ing], hik[ing], talking, finding, know[ing], and tasting.

These two suggestive sets of words distinguish the unit and visit from conventional schooling, namely in providing favorable affect and direct, active experience. The content analysis corroborated the favorable responses to the closed-ended scales, but went beyond them in revealing the reasons, particularly with respect to specific qualities of experience. These were not confined to the statements composed by the evaluators on the questionnaires, but freely expressed in the students' own words.

Although generated by the computer, each line in Table 2 may be thought of as a sentence or phrase that omits common articles, prepositions, and passive verbs. The first line, for example, suggests sentences like the following: "*Going* to the *arboretum*, I *liked* the *plants* we saw and *learning* what *animals* eat them." The second line may suggest: "Seeing the *plant* on the *field trip* was *fun*." The third line may suggest the phrase: "*Planting* and *growing seeds*."

It is not our intention to compose sentences from the words of each line. It seems reason able, however, to call attention to the most frequent patterns concentrated in the first runs for each question, and to alert the reader to favorable or unfavorable judgmental comments. Some oddities are also worth noting. It seems odd at first glance that animals and bees were mentioned, but it could reveal intentional linkages that docents and teachers made between plants

TABLE 2
Computer-Generated Patterns of Student Responses

	Patterns
1.	Liked-plants-going-arboretum-learning-animals
	Plant-field-trip-fun
	Growing-planting-seeds
	Seeing-best-imaginary-flowers-making
	Learn-different-animal
	Flower-trees-search
	Groups-working-visiting
	(Walking-game-tree-experience)
	(Experiments-things)
2.	Green-house
	(Watching-peanuts)
	Measuring-animal-look-drawing
	Holden-trips
	Botanical-gardens
	Lots-kinds
	(Love-loved)
	(Smell-people)
	(Interesting-hike)
	(Dice-playing)
	(Forest-leaves-bugs-insects)
	(Peanut-bees)
3.	Taking-talk-roots
	Finding-crayons
	(Bus-places)
	(Know-interview-tasting-little)
	(Seen-man)

Note: The numbering identifies successive computer runs. The words are analyzed in order of frequency, that is, the first 50 words, the second 50 words, and so on until the program detects no further frequent patterns of co-occurrence.

and the animal kingdom. Why the word "peanut" came up in Midwestern venues may be attributable to the fact that it is a common snack, or to instructional emphasis on economic botany, or some other reason. Despite these anomalies, what is important are the students' overall positive attitudes toward the experience and their specific reasons for it, revealed by the automated analysis.

Teacher Experiences

Although there were fewer teachers, their written comments provide some additional perspectives. As shown in Tables 3 and 4, the teachers emphasized that the students enjoyed the program and that the students enjoyed walking with their groups. They also enjoyed the language activities including reading, the field trip, "the variety," and the multi-disciplinary approach.

TABLE 3
Frequently Used Words by Teachers in Response to the Question "What is the best thing about the program?"

Word	*Percentage of All Words Used*
Activities	4.93
Field	3.52
Hands	3.17
Trip	2.82
Enjoyed	1.76
Children	1.76
Garden	1.76
Variety	1.76
Program	1.76
Students	1.76
Learning	1.41
Materials	1.06
Really	1.06
Interesting	1.06
Experience	1.06
Opportunity	1.06
Great	1.06
Best	1.06
Kids	1.06
Multi	1.06

TABLE 4
Computer-Generated Patterns of Teacher Responses (Q1)

Patterns
Students-enjoyed-really
Loved-groups-working
Social-science-studies
Language-reading
Hands-activities
Field-trip
Program-variety-best-new
Learning-involvement
Better-become-scientist-student
Children-garden-opportunity
Multi-disciplinary-approach
Interesting-material-activity
Great-kids-good
Actually-observe-experience-results

TABLE 5
Frequently Used Words by Teachers in Response to the Question "What was the worst [thing about the program]?"

Word	*Percentage of All Words Listed*
Plants	3.85
Time	3.85
Students	2.35
Difficult	1.54
School	1.54
Worst	1.54
Class	1.54
Observe	1.15
Small	1.15
Measurement	1.15
Materials	.77
Interest	.77
Enough	.77
Look	.77
Really	.77
Garden	.77
Takes	.77
Student	.77
Writing	.77
Fragile	.77

TABLE 6
Computer-generated Patterns of Teacher Responses (Q2)

Patterns
Time-observe-enough
Look-takes-short
Writing-fragile-problem
Lot-measurement-measuring-activity
Understand-interaction
Everything-garden-seemed
(Unite-worst-part)
(Difficult-interesting-within)
(Plants-school-students)

The program also yielded information about what teachers didn't like about the program (see Tables 5 and 6). Many teachers, however, did not identify anything as "the worst." The only clear and repeated theme was the shortness of time to carry out the activities. The teachers wrote highly favorable comments and often praised the field trip experience. Many complimented the venue staff and spoke of the value of the extensive pre-visit preparation.

The written responses of the students and teachers confirm the ratings they made on the questionnaire and the interviews with principals and teachers. The unit materials, activities, and field trip were all very favorably rated—particularly with respect to learning, enjoyment, and awakening interest for further study of plants and science in general. The teachers in particular praised the complementary relation of school and garden activities and had many good words for the garden staff. A few teachers felt their students were insufficiently prepared academically for the unit, and many mentioned difficulties in fitting all the activities into the allotted time.

CONCLUSION

The present evaluation illustrates the use of automated content analysis in program evaluation. Although further examples and trials are in order, our experience suggests that the technique has much to offer. Automated analysis quickly yields objective, perfectly reproducible results. Running the program, especially using the well chosen and generally applicable defaults, could be accomplished by a minimally trained, non-expert staff worker. Although a computer-readable transcript is required, traditional content analysis typically also requires this. As in the case of traditional content analyses, the interpretation of the results is subjective, but many of the word frequencies and co-frequency patterns are nearly self-interpreting. Even when the patterns appear idiosyncratic, they can be presented separately (as in the present paper), so that evaluation consumers, meta-analysts, and others can assess the results and interpretations for themselves, and possibly interpret them differently. This separation of results from interpretation, of course, is desirable in much of scholarly and practical inquiry.

NOTES

1. If after reading this article, readers still want more information about how this program would serve their needs, write or call Herbert J. Walberg, University of Illinois at Chicago, College of Education, 1040 West Harrison Street, Chicago, IL 60607, phone (312) 996-8133 fax 996-6400 e-mail hwalberg@uic.edu. The program, called Galileo CATPAC 4 Windows, is available from Terra Research and Computing, 261 East Maple, Birmingham, MI, phone, (810) 258-9657.
2. A standard default file of words was employed to omit common articles, prepositions, passive verbs, and other words that add little substantive meaning to texts such as "was," "to," "in," "the," "an," and "first."

REFERENCES

Danowski, J. A. (1980). Message content co-occurrence in computer communication networks. Acapulco, Mexico: International Communication Association Workshop on Metric Multidimensional Scaling.

Danowski, J. A. (1993). Network analysis of message content. In W. D. Richards, Jr. (Ed), *Progress in communication sciences* (Vol. 12). Norwood, NJ: Ablex Publishing.

Fielding, N . G., & Lee, R. M. (1992). *Using computers in qualitative research.* London: Sage Publications.

Woelfel, J. K., Danielsen, S., & Woelfel, J. (1995). *Catpack user's manual.* Birmingham, MI: Terra Research.