

A Note on the Reliability of Accounting Lobbying Studies

Dr. Dale Buckmaster, Accounting, University of Delaware

Dr. David Durkee, Accounting, University of Delaware

Dr. Frederic M. Stiner, Accounting, University of Delaware

Abstract

Studies that are based on content analyses of portions of the Financial Accounting Standards Board Public Record have appeared regularly in accounting and business literature since 1978. Inter-rater reliability is a crucial determinant of the validity of content analyses, yet none of the studies based on content analysis of the Public Record report any measures of inter-rater reliability. This study provides some evidence of the degree of inter-rater reliability of these studies. Krippendorff's coefficient of agreement, a measure of inter-rater reliability is derived for each of eight issues from four raters performing a content analysis of respondent letters in the Public Record volume, Exposure Draft: Accounting for Certain Acquisitions of Banking or Thrift Institutions. In general, the coefficients indicated that extreme caution should be exercised in making inferences from studies based on content analyses of the Financial Accounting Standards Board Public Record.

Introduction

The purpose of this article is to provide some evidence of the reliability of studies based on content analyses in *Public Record* of the Financial Accounting Standards Board (FASB). Studies based on content analyses of the *Public Record* have appeared regularly in accounting and business literature since 1978. The first study to rely on a content analysis of the *Public Record* was Watts and Zimmerman's (1978) investigation of managements' motivations for accounting choices. Other studies of the same question using the *Public Record* have followed (For example, Kelly (1982), (1985); Dhaliwal (1982); Griffin (1982); King and O'Keefe (1986); Watts and Zimmerman (1986)). The *Public Record* provides the source of content analyses for another group of studies that examines the congruence of lobbying positions of classes of lobbyists and their influence on the rule-making body (Haring (1979); Brown (1981); Puro (1984), (1985); Feroz (1986); Allen and Buckmaster (1986); Buckmaster (1988); Buckmaster and Hall (1990)). Another group of studies investigates differences between FASB lobbyists and non-lobbyists (Kelly (1982), (1985); Griffin (1982), (1983); Dhaliwal (1982); Francis (1988); Feroz (1987), (1988)). King and O'Keefe (1986) use the *Public Record* to study lobbying as a signal of management behavior. Puro (1985) uses the *Public Record* to determine if Big Eight accounting firms use their influence to the disadvantage of smaller accounting firms. O'Keefe and Soloman (1985) use it to study managements' belief in the Efficient Markets Hypothesis.

Each volume of the FASB *Public Record* contains either a Discussion Memorandum, an Invitation to Comment, or an Exposure Draft and the written comments to the FASB in response to the document by interested parties (generally management and auditors). Discussion Memoranda and Invitations to comment are almost identical in substance and normally contain the set of recognizable, feasible accounting treatments for an accounting problem. Exposure Drafts are proposed accounting rules. The body of literature referenced above relies upon a content analysis of the written responses of the interested parties that classifies their responses as either supporting or opposing the FASB proposal or supporting or opposing specific provisions within the proposal. Classification of "support" or "oppose" positions requires a sophisticated understanding of the technical issues being analyzed as well as careful reading. Expert judgment must be utilized by the raters in most of the studies. Thus, the validity of FASB lobbying studies is obviously dependent upon the inter-rater classification reliability of the content analysis.

There is no evidence of the reliability of these studies. Feroz (1986, 1987) had a second reader classify the relevant responses in his two studies and reports agreement of .87 and .88 of the classifications respectively. But Krippendorff (1980) points out that agreement of two or more coders expressed as a percentage of total coding items "does not mean much at all. It says nothing about whether this is much or little or how it compares with

chance." (p. 133) In addition to the lack of existing evidence on studies published to date, second or multiple readers are expensive and most likely not feasible for many future studies. Assuming that FASB lobbying studies will continue, a cost-effective way of dealing with the reliability of future studies is to conduct a complimentary series of studies that specifically examine multi-reader coding agreement so that reasonable estimates of reliability are available. Also, evidence of the degree of reliability that one might expect will permit better evaluation of existing literature that uses the *Public Record* as a data source.

We provide some evidence of reliability in this paper. This paper reports a reliability statistic that is obtained from four readers independently analyzing the responses to the provisions of a randomly selected volume of the *Public Record*. Our results suggest that the inter-rater reliability of content analyses may be intolerably low for many purposes. Inferences should be made cautiously until more evidence is forthcoming on inter-rater reliability.

Methodology

The documents analyzed were the eighty written responses to the FASB *Exposure Draft: Accounting for Certain Acquisitions of Banking and Thrift Institutions* (1982). Three faculty coders met after becoming familiar with the exposure draft (ED) and divided the ED into eight sections (issues) that could be expected to elicit comment.¹ An "other" classification was added to include comments about issues not covered in the ED. The eight issues are identified in the Appendix.

Each of the three faculty coders plus one graduate student coder read the respondents' letters and, if the coder identified a respondent as taking a position on an issue, classified the comment as "support" or "oppose" for that particular issue. Then there are three possible classifica-

tions for each of eight issues: "support", "oppose", or "no position". The remaining issue (our Issue No. 1) was pervasive and central to the ED and all coders were to assign either a "support" or "oppose" classification to this issue.

The statistic we use to measure inter-rater reliability is the agreement coefficient statistic described by Krippendorff (1980).² This statistic measures the degree of agreement among raters greater than chance. Krippendorff argues that statistical tests of significance of this statistic are meaningless since an acceptable degree of reliability must be judged in relation to the seriousness of the consequences of making inferences from the particular study being conducted. Thus, our agreement coefficients are presented and discussed as descriptive rather than inferential statistics.

Results

Table 1 reports the agreement coefficients for all coders and the agreement coefficients for each combination of three coders. The combinations of three were computed to determine if ratings of any of the coders might be considered an outlier. Existence of an outlier will emphasize two sources of disagreement among coders: (1) coding disparity when coders have similar perceptions of issues, and (2) coding disparity when coders perceive issues differently. The likelihood of outlier coders makes the results of studies based on a single coder even more questionable.

There does not appear to be any particular threshold for acceptability of inter-coder reliability. Krippendorff suggests admitting variables only when the agreement coefficient exceeds .8 and using those with reliability between .67 and .8 only for "very tentative and cautious conclusions." He goes on to make the following statement (p. 147):

Table 1
Inter-Coder Agreement
AGREEMENT COEFFICIENT

	All Coders	Without Coder 1	Without Coder 2	Without Coder 3	Without Coder 4
Issue 1	0.6495	0.7410	0.6350	0.6303	<u>0.6266</u>
Issue 2	0.3246	<u>0.2045</u>	0.2904	0.5883	0.2712
Issue 3	0.4185	<u>0.3645</u>	0.5028	0.4336	0.3819
Issue 4	0.4255	<u>0.3822</u>	0.4192	0.4963	0.4218
Issue 5	0.5528	<u>0.5096</u>	0.5537	0.6004	0.5626
Issue 6	0.5734	0.5382	<u>0.5269</u>	0.6287	0.5874
Issue 7	0.5913	0.6126	0.4912	0.6833	<u>0.4843</u>
Issue 8	0.5470	0.5793	0.5552	0.6180	<u>0.4416</u>

Bold -- Highest Alpha
Underlined -- Lowest Alpha

Where possible, *standards for data reliability* should not be adopted *ad hoc*. They must be related to the validity requirements imposed upon research results, specifically to the costs of drawing wrong conclusions. If it were a matter of life and death, even a content analyst should not accept a standard for data reliability that would lead to an error in the result with a probability of, say, less than that of being killed in a car accident (which is what people seem to be willing to live with). If it is an exploratory study without serious consequences, that level may be relaxed considerably, but it should not be so low that the findings can no longer be taken seriously. (p. 147)

The highest and lowest agreement coefficients with all coders are .65 and .32 respectively. The highest and lowest agreement coefficients in the three-coder combinations are .74 and .20 respectively. In general, Issue 1 has the highest agreement coefficient of all the individual issues and Issue 2 has the lowest agreement coefficient except when the coefficient is computed omitting coder 3.

None of the agreement coefficients approach Krippendorff's criterion for acceptability and only two (question 1 without coder 2 and question 7 without coder 3) fall within the range that might be acceptable for inferences without dire consequences. Coder 3 appears to consistently provide outlier responses since the agreement among three coders is maximized for six of the eight issues when his responses are omitted.³ The coder 3 differences are particularly pronounced for question 2. When his responses are omitted for this issue, the agreement coefficient moves from the lowest in every case to near the mean for all agreement coefficients.

Discussion

The results of our reliability tests suggest caution in drawing conclusions from content analyses of the FASB *Public Record*. The evidence provided by this study suggests that such analyses do not appear to have an acceptable level of reliability. However, we do not suggest that such studies are not useful or that they should be discontinued.

First, conclusions that might be drawn from these studies are not life or death matters. Second, while inferences from a single study may not be appropriate, consistent results from a series of studies may provide enough evidence to overcome the reliability problem as well as other design problems. The reliability evidence available is limited to an analysis of a single *Public Record* volume. Even though the issues examined in this paper seem relatively straight-forward compared to other issues examined by the FASB, the converse may actually be the case. The issues examined may have actually contained unobserved characteristics that created rating difficulties not ordinarily present in other

issues contained in other volumes of the *Public Record*. A series of studies of inter-rater reliability is desirable and will provide information necessary to infer the expected level of reliability of content analyses of the FASB *Public Record*.

Footnotes

- Holsti (1960, p.135) cites research that indicates that training prior to coding can significantly increase the level of inter-coder agreement. However, he indicates that "nondirected discussion of categories and coding rules failed to result in significantly higher agreement, in part because discussions tended to focus on the utility of categories rather than on the ways in which they should be applied to data." Our discussion was of the "nondirected" type. Specific concentration on standardizing our responses would have provided greater inter-coder agreement, but would have reduced the validity of our measures given our goal of providing some indication of the reliability of accounting lobbying studies. Maximum independence among our coders seems most desirable.
- $\alpha = 1 - (\text{observed disagreement}/\text{expected disagreement})$, where α is the agreement coefficient. Krippendorff derives the computational form of the coefficient on pages 136 through 145.3. Coder 3 was one of the faculty coders, not the graduate student.

References

- Allen, Gary and Dale Buckmaster, "Making Accounting Rules: A Case Study," *Business and Public Policy*, Vol. 12, No. 2, pp. 23-26, Spring, 1986.
- Brown, Paul R., "A Descriptive Analysis of Select Input Bases of the Financial Accounting Standards Board," *Journal of Accounting Research*, Vol. 19, No. 1, pp. 232-246, Spring, 1981.
- Buckmaster, Dale, "The FASB and the Public Accounting Profession," *Business and Economic Perspectives*, pp. 1-9, Fall, 1988.
- Buckmaster, Dale and David Hall, "Lobbying with the FASB: The Case of Pension Terminations and Curtailments," *Journal of Applied Business Research*, Vol. 6, No. 2, Spring, 1990.
- Deakin, Edward B., "Rational Economic Behavior and Lobbying on Accounting Issues: Evidence from the Oil and Gas Industry," *Accounting Review*, Vol. 64, No. 1, pp. 137-151, January, 1989.
- Dhaliwal, Dan, "Some Economic Determinants of Management Lobbying for Alternative Methods of Accounting: Evidence from the Accounting for Interests Costs Issue," *Journal of Business Finance & Accounting*, Vol. 9, No. 2, pp. 255-265, 1982.
- Feroz, Ehsan H., "Corporate Demands and Changes in GPLA," *Journal of Business Finance and Accounting*, Vol. 14, No. 3, pp. 409-423, Autumn, 1987.
- Feroz, Ehsan, "Group Participation in Governmental Accounting Standards Setting: A Cluster Analysis," *Public Administration Quarterly*, Vol. 9, No. 4, pp. 470-485, Winter, 1986.
- Financial Accounting Standards Board, *Exposure Draft: Accounting for Certain Acquisitions of Banking or Thrift Institutions*, Financial Accounting Standards Board, Stamford, Conn., December 7, 1982.
- Francis, Jere R. "Lobbying Against Proposed Accounting Standards: The Case of Employees' Pension Accounting," *Journal of Accounting and Public Policy*, pp. 35-57, Spring, 1987.
- Griffin, Paul A., "Foreign Exchange Gains and Losses: Impact on Reported Earnings," *Abacus*, Vol. 18, No. 1, pp. 50-69, 1982.

12. Griffin, Paul A., "Management's Preferences for FASB Statement No.52: Predictive Ability Results," *Abacus*, Vol.19, No. 2, pp. 130-138, 1983.
13. Haring, J. R., "Accounting Rules and 'The Accounting Establishment,'" *Journal of Business*, Vol. 52, pp. 507-519. October, 1979.
14. Holsti, Ole R., *Content Analysis for the Social Sciences and Humanities*, Addison-Wesley, Reading, Mass., 1969.
15. Kelly, Lauren, "Corporate Lobbying and Changes in Financing or Operating Activities in Reaction to FAS No. 8," *Journal of Accounting and Public Policy*, Vol. 1, pp. 153-173, Winter, 1982.
16. Kelly, Lauren, "Corporate Management Lobbying on FAS No. 8: Some Further Evidence," *Journal of Accounting Research*, Vol. 23, No. 2, pp. 619-632, Autumn, 1985.
17. King, Raymond D. and Terrance B. O'Keefe, "Lobbying Activities and Insider Trading," *The Accounting Review*, Vol. 61, No. 1, pp. 76-90, January, 1986.
18. Krippendorff, Klaus, *Content Analysis: An Introduction to its Methodology*, Sage Publications, Beverly Hills, Calif., 1980.
19. O'Keefe, Terrance B. and Solomon Y. Solomon, "Do Managers Believe the Efficient Market Hypothesis? Additional Evidence," *Accounting and Business Research*, Vol. 15, pp. 67-79, Spring, 1985.
20. Puro, Marsha, "Audit Firm Lobbying Before the Financial Accounting Standards Board: An Empirical Study" *Journal of Accounting Research*, Vol. 22, No.2, pp. 624-646, Autumn, 1984.
21. Puro, Marsha, "Do Large Accounting Firms Collude in the Standard Setting Process?" *Journal of Accounting, Auditing, and Finance*, Vol. 8, No. 3, pp. 165-177, Spring, 1985.
22. Watts, Ross L. and Jerold L. Zimmerman, *Positive Accounting Theory* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1986, pp. 327-334.
23. Watts, Ross L. and Gerald L. Zimmerman, "Towards a Positive Theory of the Determination of Accounting Standards," *The Accounting Review* Vol. 53, No. 1, pp. 112-134, January, 1978.

APPENDIX CLASSIFICATION OF ISSUES

- | | |
|---------|--|
| Issue 1 | Specifies the amortization requirement for excess of liabilities assumed over the fair value of assets acquired in an acquisition of banking or thrift institutions. |
| Issue 2 | Specifies the criteria for adjusting (accelerating) the rate of goodwill amortization. |
| Issue 3 | Specifies the accounting treatment of determinable regulatory assistance arising after the merger. |
| Issue 4 | Specifies the accounting treatment of unanticipated regulatory assistance arising after merger. |
| Issue 5 | Specifies the accounting valuation and treatment of periodic assistance from regulatory agencies. |
| Issue 6 | Specifies the accounting treatment of contingent repayments of regulatory assistance. |
| Issue 7 | Specifies the disclosure requirements when a banking or thrift institution is acquired. |