MICHAEL F. ELEEY, GEORGE GERBNER, and
NANCY TEDESCO

Apples, Oranges, and the Kitchen Sink:

An Analysis and Guide to the Comparison of "Violence Ratings"

The "opposition" position in the dialogue begun on page three of this issue, is represented by Dr. George Gerbner, dean of the Annenberg School of Communications at the University of Pennsylvania, and Mr. Eleey and Ms. Tedesco, two research associates at the school. Readers are reminded that all four parts of this extended discussion are best read together, as constant reference is made to remarks in previous articles in the series.

NBC RESEARCHERS Coffin and Tuchman[1] are attempting not only to compare apples to oranges but also measures of mineral content to declarations of how the fruits taste. Studies that differ in purpose, method, rigor, and even phenomena observed cannot be directly compared without violating basic distinctions of communication research. Errors and ommissions further mar what might have been a valuable service if confined to reasonable parameters. And yet, despite the implausibility, when the correct comparative measures are used (or reported), the results turn out to be surprisingly different from what Coffin and Tuchman contend.

The Logic of Comparisons

The concept of a "violence rating" system discussed by Coffin and Tuchman can mean two things. Senators Pastore and Magnuson[2]

© JOURNAL OF BROADCASTING, 17:1 (Winter 1972-73)

have requested a scientific system of monitoring television violence so that "a report can be submitted to this Committee on the level of violence entering American homes." Congressman Joseph P. Addabbo (D-N.Y.)[8] has proposed legislation to establish a system of ratings for network shows in order to "warn viewers of physical violence or obscenity contained in such programs." The Congressman apparently has in mind a sort of consumer's guide to specific television series, while the Senate Committee presumably requires trend data to guide it in planning policy and generating relevant legislation. The distinction between these two is important: while they share the same public concern, their ultimate strategies and effects differ considerably.

The consumer-guide orientation seeks to rate individual program series so that viewers, especially parents, could isolate particularly offensive shows and regulate the family's viewing accordingly. This would entail some sort of series-by-series evaluation, and representative sampling of each program series would be necessary.

The public-policy orientation requires evidence of a different kind. It involves the evaluation of overall patterns in the entire TV schedule. Its particular data base would consist of samples representative not of individual programs or TV series, but of entire TV seasons.

The five approaches discussed by Coffin and Tuchman vary widely as to (1) analytical method, (2) documentation of procedures, (3) type of phenomena under direct observation, and (4) type of findings and appropriate applications. A convenient taxonomy is given in Table I.

Comparison of the approaches logically ought to be conducted within the parameters of such a scheme. For example, if tested, the findings of Gerbner ought to be somewhat consistent with those of Clark and Blankenburg. To compare across classifications may be interesting, but to demand convergent findings from approaches that differ in method and purpose is unwarranted and misleading. It appears helpful now to review these approaches in the context of their original designs.

TABLE I Approaches to "Rating" Violence: A Comparative Framework

	Method	Procedural Documentation	Type of Phenomena Directly Observed	Types of Possible Findings and Applications
Gerbner[a]	Formal content analysis	Complete	Samples of television programs	Overall patterns and complex images relevant to formulation of public policy
Clark and Blankenburg[b]	Formal content analysis	Complete	Samples of *TV Guide* synopses	Overall patterns and simple trends relevant to formulation of public policy
Greenberg and Gordon[c]	Formal survey	Complete	Audience (public and critics) perception and recall	Relation between viewing preference and sensitivity to violent content, etc.
Christian Science Monitor[d]	Informal tally	Incomplete	Samples of television programs	
NABB evaluations[e]	Critical review	Incomplete	Content of television series	Guidelines for family television viewing
NBC study of NABB evaluations[f]	Classification and tally	Incomplete	Critical reviews or "experts'" opinions	

[a] Gerbner, George, "Violence in Television Drama: Trends and Symbolic Functions." In George A. Comstock and Eli A. Rubinstein (eds.), *Television and Social Behavior,* vol. 1, *Media Content and Control.* (Washington: Government Printing Office, 1972), pp. 28–187. [b] Clark, David G., and William B. Blankenburg, "Trends in Violent Content in Selected Mass Media," In Comstock and Rubinstein (eds.), *op. cit.,* pp. 188–243. [c] Greenburg, Bradley S., and Thomas F. Gordon, "Perceptions of Violence in Television Programs: Critics and the Public." In Comstock and Rubinstein (eds.), *op. cit.,* pp. 244–258. [d] *Christian Science Monitor* (July 25, 1969), pp. 1, 3 and (October 14, 1968), pp. 1, 3. [e] National Association for Better Broadcasting, "Television for the Family." *In Better Radio and Television* [newsletter] (Los Angeles: National Association for Better Broadcasting), Annually since 1953. [f] Coffin, Thomas E., and Sam Tuchman, "Rating Television Programs for Violence: A Comparison of Five Surveys." JOURNAL OF BROADCASTING 17:3-21 (Winter 1972–73).

Content Analysis Approaches: Gerbner versus Clark and Blankenburg

The perspective of Gerbner's studies views the entire range of evening dramatic television programming as a dynamic symbol system which has important social consequences. To study the composition of such a mass-produced message system, the formal techniques of content analysis are employed to generate data concerning overall structural patterns generally unobservable to the casual viewer or critic.

In Gerbner's research, trained coders, working in pairs, view and code videotaped samples of dramatic programming. Coders are trained and tested in the application of the explicit recording instructions and definitions. All programs are coded by two such pairs working independently; in this way two codings of the sample are available for reliability analysis. The assessment of reliability is provided by an item-by-item statistical analysis of intercoder agreement.[4] High intercoder agreement demonstrates the replicability of the recording procedures and indicates that the data accurately reflect the phenomenon under investigation, and are not contaminated by coder error or bias.

Such an analytical approach provides reliable information about the overall patterns of unambiguous message elements found in whole systems, namely *all* network dramatic programs. This design does not generate data about specific programs. Coffin and Tuchman notwithstanding, the Gerbner studies never classified individual programs as "violent" or "nonviolent." They simply established the number of violent acts (if any) in every program, along with other dimensions of portrayal.

The Gerbner scheme includes a number of contextual, structural, and dynamic variables that measure various aspects of TV violence. Not only is violence measured in terms of the total number of violent actions found in each program, for example, but also in terms of its seriousness and overall significance to the plot and to characters in the drama.

Clark and Blankenburg attacked the problem of retrospectively constructing simple trends in violent content without having the origi-

nal programs available. They selected *TV Guide* synopses of program storylines as alternative source material, and instructed coders to record the degree of violence indicated by each synopsis. Although not identifical, their definition of violence and that of Gerbner are largely similar.

Clark and Blankenburg considered the direct coding method of Gerbner's 1967 and 1968 studies to be base-line indicators of the validity of the synopses codings. Comparing the results of the two methods, they found two possible sources of disagreement, the brevity of the synopses themselves, and the fact that comedies are difficult to code for violence. They concluded that "Violence indexed via synopses may be inherently low," and " . . . violence in all comedies may be underestimated in synopsis coding."[5] In spite of these problems, Clark and Blankenburg nevertheless found a high degree of correspondence between the two techniques: Scott's coefficient of agreement *pi* was .74.[6] Inexplicably, Coffin and Tuchman fail to report that finding.

Coffin and Tuchman present some dramatic differences between the two studies (see their Table V). The differences they note can only arise from an erroneous interpretation of the data. A correct interpretation rests on two important facts: (1) Clark and Blankenburg coded their synopses according to Gerbner's "violence significant to the plot" variable; Coffin and Tuchman however compare the former's results with figures from Gerbner that refer to *all* programs containing violence, not just those with violence significant to the plot. (2) As Clark and Blankenburg note, the evening time periods covered by the two samples are overlapping, not identical (Gerbner 7 to 10 P.M.; Clark and Blankenburg 8 to 11 P.M.). A comparison should have been restricted to those programs in the 8 to 10 P.M. period analyzed by both studies.

It is indeed difficult to imagine how these points could be overlooked: Clark and Blankenburg discuss them in depth. They expressly compare their results with Gerbner's data for 1967 and 1968 and report that the differences in percentage of programs containing violence are 7.3% in 1967 and 8.0% in 1968.[7] Neither is statistically significant. This picture is markedly different from Coffin and Tuchman's comparison of the two studies.

Viewers' Perceptions: Greenberg and Gordon

Greenberg and Gordon surveyed viewers' and critics' perceptions of the amount of violence aired during the 1969 season in Detroit. There is no reason to assume that such subjective impressions reveal more than recall or perhaps sensitivity; they may be used as evidence of the effect of exposure to violence, but not as accurate indicators of violent content. Such perceptions certainly cannot be compared to the observations of trained coders subjected to reliability tests and made according to strictly defined criteria.

Indeed, Greenberg and Gordon found (but Coffin and Tuchman do not note) that "There was partial support for the hypothesis that the amount of violence perceived in television programs was negatively correlated with regularity of watching those programs; the more frequent viewers judged the programs less violent, and vice versa."[8]

Greenberg and Gordon further found that "Giving a definition of television violence led to consistently larger estimates of program violence by the viewers."[9] These observations emphasize already serious questions about the wisdom of using subjective perceptions as measures of violent television content.

Nevertheless, if one were to compare these perceptions with Gerbner's objective measures, one would have to go about if differently than did Coffin and Tuchman. Greenberg and Gordon's respondents rated individual TV series on a five-point scale ranging from "no violence" to "a lot of violence." Mean ratings from both the public and the critics were compiled for each of the series. Coffin and Tuchman selected Gerbner's "number of violent acts in the program" as the appropriate continuous variable to correlate with Greenberg and Gordon's scaled data, after normalizing this variable by program length, apparently to obtain some index of violence saturation. A product-moment correlation was calculated between the rate of violent acts per half hour and the mean ratings of the public and critics. The results of .48 and .56, respectively, are offered by Coffin and Tuchman as evidence that the methods are in conflict. Several observations arise from this treatment:

First, the number of programs rated in common by Gerbner and Greenberg and Gordon was 48. With 46 degrees of freedom, the

product-moment correlation coefficients reported by Coffin and Tuchman are quite significant, far beyond the .01 level.

Second, Gerbner's data are properly reflective not of the individual series, but of the entire set of series taken as a whole. Thus, the appropriate test would consist of a correlation of the rank orders of the common programs.

Finally, since Greenberg and Gordon's respondents were not asked how much violence there was per half hour in the programs, there is no reason to normalize Gerbner's number-of-violent-acts data.

Therefore, a more appropriate technique would be to rank order the common programs by mean scores of public and critics (as in fact Greenberg and Gordon did), and by total number of violent acts for the series episodes analyzed by Gerbner. If this is done, the public's ranking shows a Spearman correlation with the Gerbner ranking of .69, and the critics' with Gerbner of .75. Both correlations are significant beyond .001.

If one feels that Gerbner's methodology requires validation by general public impressions (as Coffin and Tuchman do, but we do not), the significantly positive product-moment correlations, and the even stronger rank correlations ought to provide it.

The Christian Science Monitor Studies

The *Christian Science Monitor* staff conducted two "surveys" of television in 1968. The procedures used appear in their second report:

> *Thirty-one Monitor* staffers helped to compile the latest data, concentrating on what the networks call "prime-time"—7:30 P.M. to 11:00 P.M.—*as well as Saturday morning.*
>
> At least two tabulators, and often three, kept careful notes on every evening show *in which violence was likely to occur.* They recorded all killings, other incidents of violence, and threats of violence.
>
> *"Humorous" violence* such as that found on the "Jerry Lewis Show" or "Get Smart" *was not included* among the 254 incidents reported in this survey.[10]

Since the *Monitor's* findings have been offered by Coffin and Tuchman as scientific evidence, they should be evaluated in that context.

Without prejudicing the journalistic value of the studies, it is obvious that they fail to satisfy basic methodological criteria. (1) Although 31 "tabulators" recorded the data, no reliability estimates are presented. (2) It appears that some shows were excluded from the analysis because they were not programs "in which violence was likely to occur." It is impossible to determine exactly how many, or which, programs were thus eliminated. (3) Incidents of violence in humorous contexts were explicitly ommitted. (4) Although it appears that the violence information recorded by the tabulators was the number of violent "incidents" in each program, the charts that accompany the *Monitor* articles, and indeed the information used by Coffin and Tuchman, consisted of a three-category variable—Non-violent, Moderately Violent, and Violent. Precisely how the *Monitor's* continuous "number of violent incidents" variable was grouped into the three-category form is not explained.

In attempting to compare the *Monitor's* results with Gerbner's data, Coffin and Tuchman dichotomize two continuous dimensions—the former's three-category form and the latter's number of violent acts variable. If one could compare these two studies, certainly correlational techniques such as the point-biserial or tetrachoric correlation coefficients would have been more appropriate than simple percentage disagreement, as they take into account the continua underlying one or both of the variables.

Particularly interesting is the fact that nowhere in their paper do Coffin and Tuchman make any mention of Saturday morning children's programs. Both Gerbner's data and the *Monitor's* reports show that this time period probably constitutes the most violent block of programming on American television.

The NABB Evaluations

The National Association for Better Broadcasting publishes annual evaluations of television series, compiled by an anonymous panel of experts. While the organization discusses the criteria used in the evaluations, no evidence is presented that they have been uniformly applied. In the absence of such evidence, it cannot be maintained that these evaluations are comparable to the methodologically rigorous analysis.

Examination of these evaluations over the past few years also indicates that they are published in the beginning of each year and do not completely represent the Fall network seasons analyzed by Gerbner and Clark and Blankenburg. Furthermore, since it appears that some of the evaluations for the longer-lived programs do not change substantially from year to year, the evaluations may not be reliably sensitive to yearly changes that may occur within series. As such, they are inadequate sources of trend data. Finally, the evaluations exclude late-night series which the NABB feels hold little interest for children. The NABB, it is important to note, candidly warns that the evaluations are merely the opinions of their panel, and as such should be used as guides but not as definitive findings.

In view of this, the use of the NABB evaluations as measures of violent TV content cannot be justified. Nevertheless, Coffin and Tuchman, in having the NBC research staff code programs for violent content on the basis of the NABB material, tried to make objective data out of it. Needless to say, they did not report any reliability estimates for this coding task. It is obvious that any comparisons with proper content studies are inappropriate.

Conclusions

The preceding review illustrates the pitfalls one may encounter in trying to apply the five studies as alternate or complementary approaches to the "violence rating" system. Two possible functions of such a system are currently under discussion. From one perspective it should provide a consumer's guide for the television viewer; from the other it must supply overall trend indicators useful in planning public policy. An essential component of either type of rating system is the systematic, objective, and reliable inventory of the relevant aspects of television content. Such information is required to evaluate individual patterns across all programming for the public policy needs.

None of the studies discussed provides the sufficiently objective and systematic information on specific series needed for the consumer guide function. The two content analyses conducted by Clark and Blankenburg and Gerbner yield information on overall pattern and trends, and are potentially useful for the public policy purpose. Furthermore, on their common variable they demonstrate encourag-

ing cross-validation. The richer dimensionality of Gerbner's continuing analysis renders it more useful in terms of a range of measures to be included in a violence "profile" rather than a mere "rating."

Everyone would agree with Coffin and Tuchman that a vital step in the development of any rating system would require "research designed to discriminate between 'harmful' and 'harmless' [television] violence."[11] But there is a basic confusion running not only through their paper, but also through much of the current discussion about a rating system. It is the confusion between research on the effects of televised violence and the reliable determination of violent action in television programs. The latter can provide the basis for research about the role of symbolic functions of dramatic violence in real-life conceptions and behavior, but not the other way around.

To say that the proper course of action is first to understand the effects of televised violence, and *then* to measure it in terms of the particular psychological or sociological *effects* that matter is a prime example of this confusion. Scientific study should begin with the reliable determination of common message elements and not with presumed social, moral or policy effects of unknown or vaguely specified messages. For example, the speculation that violence in a humorous context has no serious effect is not borne out in any research. To weight slapstick violence less heavily than serious dramatic violence may seem to be conventional wisdom, but in fact it is scientifically unacceptable. Humor may be the best vehicle for the effective cultivation of certain definitions, images and attitudes.

In the meantime, the reports of the Surgeon General's Committee (of which Dr. Coffin was a member) have indicated a sufficiently strong relationship between violent tendencies and the sheer frequency of exposure to *any* TV violence to suggest a public health problem. Therefore, one can no longer argue that there is no information on the effects of even gross measures of violence.

What is needed for a more precise definition of the nature of these effects is the systematic and cumulative indexing of unambiguous message configurations which can then be related to (but not equated with) viewer perceptions and cognitions. This is what our studies attempt to provide.

Footnotes

[1] Coffin, Thomas E., and Sam Tuchman. "Rating Television Programs for Violence: A Comparison of Five Surveys." JOURNAL OF BROADCASTING, 17:3–20 (Winter 1972–73).

[2] U.S. Senate, Subcommittee on Communications of the Committee on Commerce, *Hearings on the Surgeon General's Report by the Scientific Advisory Committee on Television and Social Behavior,* 92nd Congress, March 21, 22, 23, and 24, 1972 (Washington: Government Printing Office, 1972, pp. 304–305).

[3] House of Representatives, 92nd Congress, H.R. 15264, June 1, 1972.

[4] Krippendorff, Klaus, "Bivariate Agreement Coefficients for Reliability of Data." In E. F. Borgatta and G. W. Bohrnstedt (eds.), *Sociological Methodology:* 1970 (San Francisco: Jossey-Bass, Inc., 1970).

[5] Clark, David G., and William B. Blankenburg, "Trends in Violent Content in Selected Mass Media," in George A. Comstock and Eli A. Rubenstein (eds.) Television and Social Behavior, vol. 1, *Media and Content Control* (Washington: Goverment Printing Office, 1972), p. 196.

[6] *Ibid.,* p. 196. For a discussion of the agreement coefficient see Scott, William A., "Reliability of Content Analysis: The Case of Nominal Scale Coding." *Public Opinion Quarterly,* 17:3:321–325 (1955).

[7] *Ibid.,* p. 196.

[8] Greenburg, Bradley S., and Thomas F. Gordon, "Perceptions of Violence in Television Programs: Critics and the Public," in Comstock and Rubinstein (eds.), *op. cit.,* pp. 244–258.

[9] *Ibid.,* p. 253.

[10] *Christian Science Monitor* (October 14, 1968), p. 3. Emphasis added.

[11] Coffin and Tuchman, *op. cit.,* p. 19.

THOMAS E. COFFIN and SAM TUCHMAN

A Question of Validity: Some Comments on "Apples, Oranges, and the Kitchen Sink"

WHILE Eleey, Gerbner, and Tedesco raise a number of technical issues about our paper (e.g., should we have employed rank-order correlations rather than product-moment correlations?),

these tend to be rather minor and peripheral points which do not get to the really essential differences between us. Accordingly, rather than try to respond to these various points individually in the limited space available, we feel that it is more useful to bring into sharper focus some of the more basic issues.

The central thrust of their comments seems to be that it was not appropriate for us to compare the Gerbner studies with other published research on violence content in television programs, since these latter studies lack the scientific objectivity, the measures of reliability, and the procedural documentation found in the Gerbner studies. We agree that the other studies lack the aura of scientism surrounding the Gerbner studies. However, we would suggest that the collective common-sense judgments of television programs obtained from parents, teachers, journalists, television critics, etc., should not be lightly dismissed merely because they are based on subjective evaluations rather than rigorous measurement procedures, or because their reports do not contain reliability coefficients or full procedural documentation.

This raises what is perhaps the most fundamental difference between us. Eleey, Gerbner, and Tedesco place particular emphasis on the high degree of *reliability* of the Gerbner study measures; the fact that there is good inter-rater correlation in rating programs. Assuredly, reliability is important. But even more important, in our estimation, is *validity*. Are the specific actions that Gerbner's analysts categorize as "violent" actually violent in a socially meaningful sense? Do these actions represent the kinds of violence on television which have prompted genuine social concern—shootings, knifings fistfights, etc.? It is in this critical area of validity that real questions must be raised about the Gerbner studies.

We first became aware of these problems when we delved behind the summary statistics in the earlier Gerbner studies to determine exactly which programs were classified as violent. Curiously, in their article Eleey, Gerbner, and Tedesco claim that their design "does not generate data about specific programs [p. 24]." In point of fact, however, data on "number of violent acts" are contained for specific programs in their 1967–1969 studies. Interestingly, though, after 1969 their published reports have omitted program-by-program re-

sults, so that it is not now possible for the researcher to go behind the averages.

Where we did have this opportunity, we found to our surprise that a substantial number of programs classified as violent (i.e., containing at least one "violent act") were actually *situation comedies*. For example, 13 of the 45 prime-time programs (29% that Gerbner classified as violent in 1969 turned out to be comedy shows. As we noted, one of the most violent shows that season, according to the Gerbner study, was the fantasy comedy, "I Dream of Jeannie," which is about the playful antics of a genie. Even "My World and Welcome to It," focusing on a gentle James Thurber character, was scored as containing two "violent acts" in a single episode.

Not only did such classifications seem anomalous from a common-sense point of view, they did not even square with Gerbner's own rather stringent definition of *violence* as "the overt expression of physical force against others or self, or the compelling of action against one's will on pain of being hurt or killed." All the coefficients of reliability in the world could not convince somebody really familiar, say, with the Thurberesque "My World and Welcome to It" that this is properly classified as a violent television show.

These initial doubts about the validity of the Gerbner violence index led us therefore to compare Gerbner with other published studies in this area. As our article indicates, the other studies are substantially out of joint with the Gerbner studies, while agreeing quite closely with one another. While these studies of course have various methodological limitations, we feel that the substantial discrepancies between them and the Gerbner studies confirm the initial common-sense doubts regarding the basic validity of Gerbner's data.

It should be emphasized that we would not be raising these questions about the validity of the Gerbner violence index if the index were being used as originally intended, as a sociological indicator of the role of television drama in reflecting power relationships in our society. However, the fact is that the Gerbner violence index has been utilized not merely as a sociological indicator, but also as a tool for social policy on television programming. For such purposes, we feel, it is essential that the data to be employed have *validity,* not merely reliability.

MICHAEL F. ELEEY, GEORGE GERBNER, and NANCY TEDESCO

Validity Indeed!

AUTHORS Coffin and Tuchman bring their discussion to rest on the consideration of validity, as they see it. Their argument, however, is not with the validity of our data, but really with the validity of the definition of violence we have employed. Their choice of validation criteria reveals once again a basic confusion between studies of content and investigations of effects, a confusion in which the logical order of this research is reversed. The task of content studies is to provide reliable assessments of images, which can then be related to studies of effects. To predicate analyses of content on the basis of presumed effects is an uncritical approach.

This confusion leads Coffin and Tuchman to maintain, for example, that violence in humorous contexts is of no concern, for its inclusion runs contrary to the popular wisdom of parents, teachers, journalists, critics, and the like. To be valid, they say, a definition of violence must correspond with those effects that "have prompted genuine social concern—shootings, knifings, fistfights, etc. [p. 32]." Why not relate symbolic violence to fear? To learning how to be a victim? To feelings of panic or power? Or to the peaceful *acceptance* of violence? These are equally plausible behavioral outcomes or effects.

Why not assume that by demonstrating power and differential risks in life, violent symbolic portrayals accomplish the tasks of real-life violence in a cheaper and more entertaining way? Why not assume that situation comedies and "gentle" Thurberesque characters (the repressed frustrations and violent fantasies of Walter Mitty come to mind) are especially effective devices for accomplishing the social functions of symbolic violence? These are equally plausible assumptions, but if one adopts Coffin and Tuchman's "common-sense" restrictions a priori, one could never determine which were

the most "likely" or important effects. We are all children of our culture, and only to the extent that we can critically examine its axioms will our research elucidate its dynamics and structure.

It is obvious that validity in communications research cannot be assumed to rest in a naive semantic correspondence between a symbol and a certain arbitrarily selected type of behavior. To do so is to beg the very question that communications researchers should try to answer. That question is: What types and ranges of conceptions and behavior (other than naive semantic equivalents) do symbolic representations in fact cultivate?

Coffin and Tuchman's "validity" is like that of the alchemist, so hypnotized by the elusive prospect of creating gold that he overlooks the more profound and varied results of his experiments. Symbolic violence may have a variety of functions, of which some may have no more to do with violent action than dreaming of falling has to do with gravity. If the researcher makes up his mind about this vital connection before he begins, he cannot possibly come to a sound—let alone valid—conclusion. We have attempted to avoid such unvalidated preconceptions in our studies, and thus have deliberately included symbolic violence in *all* its contexts. This is not to say that we ignore contextual aspects; in all our reports detailed breakdowns are provided separately for comedies and serious dramas, for cartoons and noncartoons, and for several other contextual categories.

To the extent that we avoid simplistic preconceptions and employ a stringent definition of violence, our studies should differ from others selected by Coffin and Tuchman to fit their presumptions. But here we leave it to the reader to decide whether the technical issues we raised earlier are really so minor as Coffin and Tuchman would have us believe. We only reiterate that closer attention to the methodological aspects involved leads to a comparative picture quite different from the one they have advanced.