

THE VALIDITY ISSUE WITH RESPECT TO EVENT DATA*

Charles F. Hermann

Introduction

Anyone who has kept an eye on the emergence of event data in international and comparative politics recognizes that many individuals associated with the collection and use of these data have been sensitive to methodological and technical issues associated with the enterprise. Even without conducting a systematic review, the guess is safely ventured that two of the most extensively addressed subjects have been the measurement of conflict and the validity of sources for culling events. Given the prominence of the validity issue in the examination of event data, it may seem an exercise in overkill to address validity again. Yet it can be argued that despite the attention to sources of events, other troubling aspects of validity have been largely ignored.

Writing about cumulation and progress in international relations as a "pre-paradigmatic" science, Ashley (1976:155) referred to event data as a positive illustration:

The 'events data movement' as a generic phenomenon probably deserves to be called a developing research nucleus. It has been characterized by a very large measure of expansive cumulation over the years (e.g., McClelland, CREON, Azar, Moses-Brody, Rosecrance). And this might have suggested degeneration except that shared experiences (both successes and failures) are slowly giving rise to a widely shared adherence to a variety of selectively identified precepts that will increasingly guide future research.

What might we point to as the shared precepts with respect to event validity and what are their implications for future research? My personal answer is that there is less basis for consensus about validity than might appear at first glance. It does not seem too much to say that

* This essay was made possible by a grant to the author from the National Science Foundation (SOC76-83872) and from the Merston Center of Ohio State University. An outline of the present paper was circulated at a panel on international event data validity held in St. Louis, Missouri, at the annual meeting of the International Studies Association, March 20-23, 1974. The author gratefully acknowledges the persistent encouragement of Don Munton to transform the outline into this essay, and the comments of two reviewers.

whether the event data movement continues as a "developing research nucleus" depends to a significant degree on our collective treatment of such issues. To develop the reasons for this expressed concern, this essay will examine three aspects of event data validity--construct, internal, and source--as well as some associated issues.

Validity and Event Data

Validity is a rather awkward term to define. There is something slightly vacuous about suggesting that validity asks the question: Does something do or represent what it claims to do or represent? Nevertheless, the question appears at the heart of the matter. We might speak of the validity of a forecast based on its congruence with the occurrences it predicts. A pencil and paper measure of intelligence might be regarded as valid to the extent it differentiated subjects according to their ability to perform various operations which were independently determined to require abilities associated with our idea of intelligence. In a different realm, a foreign policy might be regarded as valid to the extent it produced the desired effects and the extent to which all the effects were anticipated. Elsewhere the present author and some associates have suggested "a theory is valid if it does what it purports to do" (Hermann, Phillips and Thorson, 1978).

Several observations about the concept of validity can be drawn from these examples. First, validity often seems to imply a standard or reference against which the accuracy of something can be assessed (e.g., comparing an I.Q. measure against some task requiring intelligence, or the forecast against the actual outcome). Second, it often seems more appropriate to consider the degree of validity or representativeness or goodness of fit rather than a dichotomy of valid or invalid. (A policy, measure, or theory may be more or less valid. Even the prediction of a distinct event that either occurred or failed to occur may be based on a forecasting procedure that may be only partially valid.) Third, validity is affected by the intended purpose of that which is being evaluated. (A measure of intelligence designed to assess performance in college may not do very well in measuring ability to solve problems in an alien culture. A policy designed to produce a certain kind of reaction in a particular regime may not be applicable when the regime changes.) The significance of these observations about validity will be more apparent as different kinds of event data validity are examined.

Construct Validity

The notion of construct validity is drawn from the psychological test and measurement literature (e.g., Cronbach and Meehl, 1955). The validity issue is whether the concept is adequately developed and specified and whether the operational measures of that concept capture the

intended meaning represented in the construct.¹ A concept from the small group literature offers an illustration of the significance of construct validity. The concept of group cohesiveness has been prominent in much of that literature and has been measured in a variety of ways in both laboratory and field studies. Unfortunately, Eisman (1959), among others, found little if any correlation among several measures of cohesion. Such a finding raises basic questions. Are the measures tapping different dimensions of the same concept? Are all or some of them actually measuring something different from cohesion? Is there really any meaningful phenomenon captured by the concept of confusion, or might analysts do better to decompose the term into several separate concepts?

As the illustration suggests, a test of validity can involve comparing various operational measures, all presumably derived from the same concept, with one another. Alternatively, one can compare them with measures of other constructs with which they logically should be highly correlated positively or negatively. Campbell and Fiske (1959) have formalized this latter approach in what they refer to as a multi-trait, multi-method strategy. On a less formal level, Cartwright and Zander (1960:70) in a more encouraging review of the group cohesion concept, note that a particular operationalization of cohesion "made sense" because it led to observations of behavior that one would expect to be associated with varying degrees of group cohesiveness.

What does construct validity mean for event data? The basic implication involves the concept of an event--the basic coding unit. At a very general level there would appear to be a consensus on the idea of what constitutes an event. Drawing either on the basic idea of communication (i.e., communicator, message, receiver) or Harold Lasswell's idea of "who gets what from whom," users generally agree upon the basic components of an event. An event involves an actor, action, and recipient. Beyond this point, however, there appears to be not only little agreement, but rather limited discussion and analysis.² It is likely that problems involving the level of disaggregation and boundaries between events exist as a result.³

¹ For a discussion and application of the construct validation approach, see the contribution of Don Munton in the last section of this volume.

² Several separate projects have devoted energy to the fuller conceptualization of an event, but they are not necessarily congruent with one another. See the statement relevant to WEIS by McClelland (1972); the one by Azar (1975) concerning COPDAD; and that by Hermann (1978) for CREON. Also see Hermann (1971) for a fuller discussion of this issue.

³ [Editor's note:] See the discussion by Munton in "Policy Makers, Public Records, and Reality" in this volume.

Consider the following hypothetical report constructed in classic journalism style of increasing specification as the story is developed.

Arab governments insisted today that Israel accept United Nations Resolution 242. Meeting in Amman, Jordan, the heads of state for Egypt, Lebanon, Saudi Arabia and Jordan discussed the Middle East situation. Leaving the afternoon session, President Sadat told reporters that he had urged the others to insist that Israel yield political control over all settlements in captured territories when these lands are returned. He states the others had agreed. A press representative of the Jordanian Government announced that a formal communique would be issued later in the day reaffirming the government's support for U.N. Resolution 242 and the need for Israel to accept it as a basis for further negotiation. Under questioning he acknowledged that the Lebanese delegation had already left Amman in disagreement over the Palestinian issue and would not be a party to the communique.

How many events does this statement contain? At the most general level, one could argue that there is only one action--a meeting between Egypt, Lebanon, Saudi Arabia and Jordan to exchange views on the Middle East. Alternatively, the report could be decomposed into a number of discrete events each with a separate action--Sadat's statement to others; disagreement over Palestinian question; a communique on U.N. Resolution 242; the departure of the Lebanese delegation. Most event data collections would have coding rules for resolving these questions. The resolutions, however, do not flow from the general consensus on the nature of an event and would not necessarily lead to the same conclusion. Moreover, some of these rules may be highly dependent upon syntax and stylistic features of the report that could be affected by rather minor revisions in presentation.

If this characterization is correct, coders using the coding rules of different event data collection systems would produce different results when given the identical coding material.⁴ The generic concept of event as agreed upon between different collections of data is probably insufficient to establish an unambiguous standard or referent for operational coding rules. Such a condition does not automatically raise questions of construct validity within a data set, but rather between different collections of events.⁵ Other tests would be needed to ex-

⁴ [Editor's note:] On this point, see the discussion and comparison of event data coding schemes by Russell Leng in this volume.

⁵ An example of a validity problem involving the concept of event occurred when Chan (1976) attempted to compare WEIS and CREON data set coverage of interactions between China, the United States, and the Soviet Union. The meaning of the comparison and its implications must be seriously questioned because the two data sets define and operationalize the construct "event" differently.

amine construct validity within a single data set--such as the correlation with similar and dissimilar properties. (For example, in a given data set the definition of an event might suggest there should be a high positive correlation between the number of separate votes a country's delegate cast in the United Nations and the number of events identified for that government as having occurred at the U.N. This illustration neglects the problem created by imperfect sources or conventions that a source might use in reporting U.N. activity.⁶

The construct validity of event data is not simply a technical issue to be examined among those who acquire such data. It has implications for the interpretations placed on analytical results. For example, both Rummel (1976:151-152) and Kegley, Salmore and Rosen (1974) report that the foreign activities of governments have a remarkably stable and relatively simple structure in terms of underlying dimensions. This observation is directly dependent upon both the concept of event and the nature of the coding schemes employed as well as the specific variables included in the analysis. If the concept of event is not very clear or different definitional schemes are used, then the underlying structure may shift as one performs the analysis on different data sets. Kegley, Salmore and Rosen (1974) did use different data sets, but most of the included collections are based upon the World Event Interaction Survey (WEIS) coding system. The one data set using a different event coding scheme yielded a different structure.

Internal Validity

Another kind of validity relevant to event data has been labeled by Campbell (1957) as internal validity. The more commonly used term for internal validity is reliability. The advantage of Campbell's terminology is that it calls attention to the close relationship between various forms of external validity (i.e., comparison to an external standard) and internal consistency. It is the internal consistency or stable qualities of a measure, type of data, or the relationships postulated by a model or theory that constitute internal validity. If repeated occurrences of items assumed to be the same appear with unexpected variability, then any comparison to other criteria becomes questionable. Which instance or set of instances of the varying phenomenon are the ones that should be compared? Because this question is normally unanswerable, some degree of internal validity or reliability must be established before exercises in external validity become worthwhile.

With respect to event data the issues of internal validity center on coding. Is the same material coded in the same way by different coders if they use the same encoding rules or by the same coder on

⁶ [Editor's note:] On this point, see the comparison of UN notes and events by Brian Tomlin in this volume.

several separate occasions? These are problems of inter-coder and intracoder reliability or--put another way--the degree of ambiguity in the coding instructions. The stability of a categorization or scale also can be called into question because the defining terms do not have the same meaning in different cultures. Scale consistency can also be affected by statistical procedures as, for example, when the state of a variable is assigned a certain weighting based on a factor score for a given sample of data. If the factor score changes when a new sample is introduced, then the consistency is called into question.

At least with respect to coder reliabilities, event data efforts have attended to internal validity in a minimal fashion. Intercoder reliabilities are often reported, but usually only as the percentage of agreement between several coders or perhaps a product-moment correlation. The appropriateness of a given procedure depends upon certain information which is infrequently given. In most cases, however, the percentage of agreement and correlation measures are not very satisfactory. For nominal sets of categories, simple percentages would not reveal the likelihood of agreement expected by chance. For example, consider coding an event for whether or not it involved global war. The likelihood of agreement in the affirmative by chance alone would be extremely low for most types of event data. Similarly, a high product-moment correlation between coders could occur even when actual agreement was not high if the two coders had differing average levels, or differing variabilities of judgment. More adequate statistics such as Scott's p_i (1955) or the Krippendorff agreement coefficient (1971) are available for particular situations, but are seldom used with event data.

Even more disturbing is the tendency to report a single average figure of agreement for an entire data set. Such procedure hides more than it reveals. It presumably combines scores for different coders on multiple categories or scales. If the scales are of different levels of measurement or are to be aggregated in alternative ways, the knowledge about internal validity provided by a single figure may be grossly misleading.

Moreover, it is not enough just to say how reliable are the individual variables. For any variable that is more complex than a dichotomy, it may be quite clear when some categories are to be coded or not coded and quite unclear when others are to be coded. Unreliability may be concentrated in one or a few categories of a measure, leaving the others very reliable. A single reliability score for the whole variable may be quite misleading. A more disaggregated discussion of reliability would be helpful if the user were interested only in certain categories (for example, for building an additional scale such as "commitment") or if the user were going to code some additional data and needed to know which parts of the coding rules required further specification.⁷

⁷ The author is indebted to Patrick Callahan (personal communication) on this point.

Source Validity

If those involved with event data have generally ignored construct validity and have tended to treat internal validity somewhat superficially, these charges do not apply to what we can call source validity. The requirement with this aspect of validity is to determine how adequate, comprehensive, and free of bias various alternative compilations of foreign policy and international events might be. At the heart of this validity issue is the possibility that an analyst's encoding of events may not adequately represent what actors actually do in world affairs because of distortions that exist in the source material. The usual sources for event data are newspapers, chronologies, or released government documents. Source validity is acute because these materials have been assembled for audiences and purposes different from that of the disinterested political analyst. Boulding (1966:74) has pointed to a comparable problem that plagued economics at an earlier point in time:

As long as the information system of the economy consisted of information collected as a by-product of other activities, for instance the taxation or the collection of customs duties, we got no clear view of the system as a whole. It was not until we began to collect information directly for its own sake . . . that we began to get a picture of the total economy which was reasonably well sampled and undistorted.

Regrettably, no analogous sources of event information have been compiled with the scientific researcher's need as the exclusive objective. Most current event data sets that have had a global focus have been assembled from one of a small number of newspapers (most notably, the *New York Times* or its index) or one of several public, continuous chronologies of world affairs (*Facts on File*, *Kessing's Contemporary Archives*, *Deadline Data*). Several major event data sets that have focused on one geographical region, such as the Middle East or Africa, have used multiple sources that specialize in that region, including newspapers, journals and specialized newspapers.⁸ Recognizing the potential validity problems involved in such sources, a number of authors have sought to address a variety of questions regarding source validity (e.g., Smith, 1969; Azar, *et al.*, 1972; Harle, 1972, Sigler, 1972, Doran, *et al.*, 1973; Laphier, 1975; Burrowes, 1974; Hoggard, 1974; Chan, 1976). Among the questions they have asked are:

- 1) Do regional data sources provide a more comprehensive and different array of events for a given region than

⁸ [Editor's note:] On the question of event data sources, see also the contributions in this volume of Russell Leng; Carl Jenks; Sophia Peterson; Barbara Salmore; Stephen Salmore and Fred Butler; Olav Knudsen, Tor Chr. Hildan, and Arve Thorvik; Timothy Shaw and Douglas Anglin; and Don Munton.

do sources that attempt global coverage?

- 2) What is the relative coverage in terms of number and types of events in various global sources?
- 3) How does the coverage of events for a country vary depending upon the political and cultural orientation of the data source?
- 4) In what ways do official government accounts of events differ from those in public sources?
- 5) Is there a tendency in American newspapers (perhaps in European as well) to report conflict events more readily than nonconflict events and to report only the most far-reaching occurrences for small, less economically developed, and/or politically closed systems while re-reporting much less significant events for large, more economically developed, and/or politically open systems?

Although some provisional answers to such questions emerge from these studies of source validity, they must be treated cautiously. One reason for caution is that some of the studies have been conducted without the attention to their designs that would provide high confidence in their findings.⁹ Even more important is the problem of purpose. At the outset of this essay it was noted that one must take into account the intended purpose of that which is to be validated. Thus, for instance, a simulation model designed for one purpose may be found to be relatively invalid when used for a different purpose (see Hermann, 1967). That observation hardly seems startling when stated directly, but sometimes is slighted when interpretations of findings are advanced.

⁹ A detailed critique of the existing source validity studies is beyond the scope of this essay, but certainly should be undertaken. One illustration of design problems must suffice (but also see note 3 above). Several source validity studies have included *Deadline Data on World Affairs* among the global sources to be examined. At least one of them (Burrowes, 1974) has used the normal or standard version of this chronology which is not satisfactory for constructing historical event data sets. The producers of *Deadline Data* periodically update their files to subscribers and instruct them to discard many older entries (which are replaced with shorter summaries) in order to make room for current material. The result is that if one of these files is used (as Burrowes has done), a highly telescoped chronology of events is provided with fewer and fewer events reported as one moves back in time. Obviously, such coverage does not compare favorably with sources that do not follow such a procedure. In order to avoid the difficulty, a set of *Deadline Data* must be acquired for which no older entries have been discarded.

One example may illustrate the ease with which even thoughtful researchers may slip into this trap. In an article designed to demonstrate the heterogeneity of different national decision processes, particularly with respect to the perception of crises, Bobrow, Chan, and Kringen (1977) intend to show how different cultural and political orientations affect crisis perception. They reproduce alternative lists of crises compiled by a Hungarian social scientist, an unidentifiable "team" in a publication from the People's Republic of China, and two American social scientists. The authors then display the number of events found in the CREON data set for the crises appearing in each list. Because CREON uses an American-produced chronology for its events, the test would appear to be a nice demonstration of bias in source coverage --or, more exactly, how western perceptions of what are important occurrences (crises) affect the adequacy of reporting. Unfortunately, the CREON data were not constructed to provide continuous coverage of all events for all countries. Instead, the data set includes events initiated by 36 national governments for randomly selected quarters (3-month intervals) in each of ten years. Thus, eleven out of the 28 crises on the Chinese list occurred or "peaked" during months when no CREON data were collected and others involved as principal actors some governments not included in the CREON list of actors. To compound the problem, the CREON project defines any event as the result of a political level decision to act. That conceptualization may not be an optimal definition for examining military crises because many separate military actions may flow from a single political level decision. The point is not to defend the CREON data set or condemn the essay on varying perceptions of crises, but rather to show the incongruity between the concerns of the authors with respect to a source coverage issue and the characteristics of a given data set.

Nor is the illustration above an isolated case in studies of event source validity. Projects requiring detailed accounts of a particular geographical cluster of nations would be handicapped by the more superficial coverage likely from global sources. Conversely, research concerned with external behaviors of actors from a wide variety of regions would be faced with considerable distortion if their sources were committed to coverage primarily of one region. Such varying purposes have not always been recognized in studies of event source validity. In sum, the source validity research to date may very well have raised more questions than it has answered.

Other Validity Issues With Respect to Event Data

The specific validity issues reviewed above also surface others which can be only briefly mentioned, but deserve fuller consideration elsewhere. One basic question is philosophical in nature, but the position one takes in answering it has critical importance for the interpretation of event validity. The query might be stated: Is there a knowable universe or population of international events? Frequently, one encounters research that seems to assume at least a theoretically finite set of events that have transpired at a given place and time and

which can be more or less accurately sampled. Consider the sketches of various positions:

- 1) Events have no reality other than in the eye of the beholder. Given N observers, there are N sets of events that do not exist independently of the experiences and mind sets of the observers.
- 2) Events are analytic constructs imposed on a "seamless web" of human behaviors, but they can be stipulated in such a fashion that any observer can distill from the seamless web the same set of events.
- 3) Some components of events--such as actors and recipients--are capable of common intersubjective identification, but others, such as the nature of action, are subjective and agreement on them is unlikely.
- 4) Only some types of events are knowable as perhaps when either the actor or the recipient (or both) acknowledge their existence or when multiple, independent, direct observers can confirm their occurrence or when there is some discernible residue as in acts of war or treaty formation.

A related issue is whether the idea of unbiased or undistorted events is a useful or necessary aspiration for much of the research on foreign policy and international relations. One could argue that for many purposes, the world as viewed by the *Washington Post*, the *London Times*, *Pravda*, *Le Monde*, *DieWelt*, *Asahi*, *Excelsior*, and the *Times of India* are useful surrogates for the views of their respective national policy makers. According to this argument, the fact that such sources differ can provide a lever for analysis, not merely a troublesome indication of bias. A variation on that position might be to argue that a half-dozen world news services (virtually all of which are directed by personnel of developed countries) structure most foreign and international events for most political elites around the world. Regardless of whether such an arrangement is desirable, their accounts of events are the ones that matter in understanding how political elites view events outside their country.

Still another question concerns what is left out of the concept of international events and the implications that may result. At least three classes of behavior seem beyond the reach of currently defined event data sets. First, there are decisions by actors to do nothing. Thus no action results. (No action also results when an actor fails to respond or remains indecisive in the face of a recognized occasion for decision. But in such cases, it is more problematical whether any behavior has occurred which event data sets can be said to miss.) A second category involves covert or otherwise secret acts that may never

be known to more than a very few people and which leave no artifacts--that is, no changes outside of the minds of those involved (e.g., an undisclosed verbal exchange between two heads of state who talk alone). Finally, there is the category that McClelland (1972) calls transactions as opposed to interactions. These are behaviors that are so routine and frequent as to go largely unnoticed at least at higher political levels and certainly in terms of public notice as reflected in media coverage. The boundary between these categories, however, may not be very stable either through time or across actors--particularly in an era when the traditional distinctions between high and low politics has eroded. There is also the problem of a transaction that retroactively becomes an interaction at the will of a policy maker or journalist. In assessing event validity it seems wise to consider what is being excluded--and with what degree of confidence--as well as what is supposed to be represented. After all, what is excluded may largely invalidate event data for some purposes.

Finally, brief note should be given to the standard, reference, or criterion against which an event data set is being evaluated. It is important to ask how confident we are of the standard as well as the object that is the subject of the validation exercise. In source validity, for example, one or more sources are being compared to one another, but which is the standard? In internal validity, the standard is consistency or replicability. In some cases it may be useful to ask whether under certain conditions, consistency should be expected in the reference system the data are to represent. In construct validity, we are comparing one concept to others to which we *assume* it should be associated or one operational measure with others that are *assumed* to be indicators of the same construct. These assumptions, as with any criteria in a validity program, should not be taken for granted.

This essay began with a quotation from Ashley (1976) about the event data movement. It should now be evident that the present author believes that more and better work on validity is needed if the event data movement is to contribute to international relations and foreign policy as a coherent research program. A concluding word of caution, however, seems appropriate. There is a danger of becoming too exclusively concerned with methodological issues surrounding event data. Of course, we must never ignore such matters, but neither must we ignore the need for clear demonstrations that event data can--along with other types of data--enhance our ability to deal with major substantive issues in comparative and world politics. One could argue that the number of major substantive studies in which event data have been critical still remains relatively small two decades after such data were beginning to be seriously collected. Part of the motivation for more and better validity studies will likely result when important substantive problems or puzzles are provisionally interpreted by research with event data.

- Ashley, Richard K. (1976). "Noticing Pre-paradigmatic Progress," pp. 150-157 in James N. Rosenau (ed) *In Search of Global Patterns*. New York: Free Press.
- Azar, Edward E., Stanley H. Cohen, Thomas O. Jukam, and James M. McCormick (1972). "The Problem of Source Coverage in the Use of International Events Data," *International Studies Quarterly*, 16, 373-388.
- Azar, Edward E. (1975). *Dimensions of Interaction*. Pittsburgh: International Studies Association Occasional Paper No. 8.
- Bobrow, Davis B., Steve Chan, and John A. Kringen (1977). "Understanding How Others Treat Crises," *International Studies Quarterly*, 21, 199-223.
- Boulding, Kenneth (1966). *The Impact of the Social Sciences*. New Brunswick, N.J.: Rutgers University Press.
- Burrowes, Robert (1974). "Mirror, Mirror, on the Wall . . . A Comparison of Event Data Sources," pp. 363-406 in James N. Rosenau (ed) *Comparing Foreign Policies*. New York: Wiley.
- Chan, Steve (1976). "An Assessment of Interstate Conflict Coverage Event Data Projects." College Park: University of Maryland, Crisis Warning and Management Project (mimeo).
- Campbell, Donald T. and Donald W. Fiske (1959). "Convergent and Divergent Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 56, 81-105.
- Campbell, Donald T. (1957). "Factors Relevant to the Validity of Experiments in Social Settings," *Psychological Bulletin*, 54, 297-312.
- Cartwright, Darwin, and Alvin Zander [eds] (1960). *Group Dynamics*. Evanston, Ill.: Row, Peterson. Second Edition.
- Cronback, Lee J., and Paul E. Meehl (1955). "Construct Validity in Psychological Tests," *Psychological Bulletin*, 52, 281-302.
- Doran, Charles F., Robert E. Pendley, George E. Antunes (1973). "A Test of Cross-National Event Reliability: Global Versus Regional Data Sources," *International Studies Quarterly*, 17, 175-204.
- Eisman, Bernice (1959). "Some Operational Measures of Cohesiveness and Their Interrelations," *Human Relations*, 12, 183-189.
- Harle, Vilho (1972). "Reliability of International Events Data Sources." Tampere, Finland: Research Institute of the University of Tampere (mimeo.).
- Hermann, Charles F., Warren R. Phillips, and Stuart J. Thorson (1978). "Validating International Relations Forecasts to Develop Theory," in Mazli Choucri and Thomas Robinson (eds) *Forecasting in International Relations*. San Francisco: Freeman, forthcoming.
- Hermann, Charles F. (1971). "What is a Foreign Policy Event," pp. 295-321 in Wolfram F. Hanrieder (ed) *Comparative Foreign Policy*. New York: McKay.
- _____ (1967). "Validation Problems in Games and Simulations with Special Reference to Models of International Politics," *Behavioral Science*, 12, 216-231.
- Hoggard, Gary (1974). "Differential Source Coverage in Foreign Policy Analysis," pp. 353-377 in James N. Rosenau (ed) *Comparing Foreign Policies*. New York: Wiley.
- Kegley, Charles W., Stephen A. Salmore, and David J. Rosen (1974). "Convergences in the Measurement of Interstate Behavior," pp. 309-342 in Patrick J. McGowan (ed) *Sage International Yearbook of Foreign Policy Studies*, Volume II. Beverly Hills: Sage.
- Krippendorff, F. K. (1971). "Reliability of Recording Instruments: Multivariate Agreement for Nominal Data," *Behavioral Science*, 16, 228-235.
- Lanphier, Vernard A. (1975). "Foreign Relations Indicator Project (FRIP)," pp. 161-174 in Edward E. Azar and Joseph D. Ben-Dak (eds) *Theory and Practice of Events Research*. London: Gordon and Breach.
- McClelland, Charles A. (1972). "Some Effects on Theory from the International Event Analysis Movement," in Edward E. Azar, Richard A. Brody, and Charles A. McClelland *International Events Interaction Analysis*. Beverly Hills: Sage Professional Papers in International Studies (02-001).
- Rummel, Rudolph J. (1976). "The Dimensionality of Nations Project," pp. 1949-1954 in Francis W. Hoole and Dina A. Zinnes (eds) *Quantitative International Politics*. New York: Praeger.
- Scott, William A. (1955). "Reliability of Content Analysis: The Case of Nominal Scale Coding," *Public Opinion Quarterly*, 19, 321-325.
- Sigler, John H. (1972). "Reliability Problems in the Measurement of International Events in the Elite Press," in John H. Sigler, John O. Field, and Murray L. Adelman *Applications of Events Data Analysis*. Beverly Hills: Sage Professional Papers in International Studies (02-002).
- Smith, Raymond F. (1969). "On the Structure of Foreign News: A Comparison of the *New York Times* and *Indian White Papers*," *Journal of Peace Research*, 11, 23-26.