This Week's Citation Classic™ CC/NUMBER 38 SEPTEMBER 17, 1984

Scott W A. Reliability of content analysis: the case of nominal scale coding. *Public Opin. Quart.* 19:321-5, 1955.
[Institute for Social Research, University of Michigan, Ann Arbor, MI]

This paper suggests an improved method of reporting the extent of interobserver agreement in assigning overt or verbal behavioral items to a set of categories. It was developed specifically for standard survey research coding operations, but it can be used in a wide variety of research situations to measure the reliability of classifying a large number of responses into nominal scale categories. The requirements are that the categories be mutually exclusive and that observations be duplicated on a random sample of the total set of responses being studied. [The *Science Citation Index*® (*SCI*®) and the *Social Sciences Citation Index*® (*SSCI*®) indicate that this paper has been cited in over 185 publications since 1955—one of the most cited for this journal.]

William A. Scott
Department of Psychology
Australian National University
Canberra 2600
Australia

July 24, 1984

"This was only my fourth experience, as a student, in publishing in a professional journal. I was quite surprised to learn that the article had attracted attention nearly 30 years later (see, for example, reference 1). Perhaps this is because it was an early attempt to confront a practical problem faced by investigators who found it appropriate to use open questions in their research. Also, it is shorter, more sharply focused, and more interestingly written than many of my subsequent papers. Perhaps it was easier for the relevant audience to understand than others' more complex treatments of the problem.

"We no longer have much occasion to use nominal-scale measures, but instead prefer to construct multiple-item summative scales to represent variables anticipated in advance. If responses are coded after the interview, we typically use a small number of categories that can readily be encompassed within some simple theoretical framework. Then we convert a nominal scale of k categories into k dichotomous variables. It is thus possible to represent the level of coding reliability for each category separately, using the intraclass correlation or *phi* coefficient.

"The preparation of this paper was encouraged in the stimulating environment of the University of Michigan's Survey Research Center, where I worked with colleagues and supervisors who had a marvelous capacity for giving ideas to each other, without concern for copyright. I still can't distinguish many of my own ideas from those of Steve Withey, George Belknap, Gerry Gurin, Libby Douvan, and Warren Miller, and this may account for my attitude toward plagiarism: it would be very annoying if someone tried to steal my last idea.

"The selection of *Public Opinion Quarterly* as an outlet stemmed from my previous experience publishing in that journal. Under encouragement from Ron Lippit, I had submitted a manuscript that grew out of a first-year graduate paper, reporting practical experience in the military government of Japan.[2] For 18 months I heard nothing, then back came a beautifully edited, shortened article in galley proofs. One hardly expects such assistance nowadays. It helped me adopt Don Campbell's advice in the face of many subsequent delays in production: 'Science is timeless.'

"I have recently learned of a PhD thesis in mathematics dealing with the *kappa* statistic[3-5] (*kappa* is a close relative of the *pi* statistic which I proposed). The eminent statistician who told me about it couldn't see how *kappa* was worth a thesis. But it seems it may have been worth four pages. As my paper was presented in the journal's section on 'Living Research,' it is gratifying to see that the baby has survived so long.

"On the occasion (sometime in the 1950s) of its radio broadcast of the 'Song of the Wood-Dove' from Arnold Schönberg's *Gurre-Lieder*, the New York Philharmonic invited the composer to comment. His reply went something like this: 'As I wrote this piece 50 years ago, your performance of it now gives me hope that, 50 years hence, your eminent orchestra might get around to playing something I am writing today.' That's not such bad encouragement, after all!"

1. Craig R T. Generalization of Scott's index of intercoder agreement. *Public Opin. Quart.* 45:260-4, 1981.
2. Scott W A. The information meeting as an instrument of social change in occupied Japan. *Public Opin. Quart.* 16:160-78, 1952.
3. Cohen J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20:37-46, 1960. (Cited 645 times since 1960.)
4. Fleiss J L. Measuring nominal scale agreement among many raters. *Psychol. Bull.* 76:378-82, 1971. (Cited 120 times.)
5. Landis J R & Koch G G. The measurement of observer agreement for categorical data. *Biometrics* 33:159-74, 1977. (Cited 75 times.)

