

SCI coverage is identical to a cumulation of the weekly ASCA coverage.

The total error control which we employ is, in practice, inseparable from the total system of production. We consider it an important guiding principle that humans and machines be coordinated in error-checking systems—the two used in complementary capacities.

LITERATURE CITED

- (1) Sher, I. H., Foeman, G. H., Baus, E. H., "A Slide Rule for Calculation of the Number of Double Bonds and Hydrogen Atoms;" paper presented at the 145th National Meeting of the American Chemical Society, Division of Chemical Literature, New York, September 1963.
- (2) Elias, A. W., Garfield, E., Foeman, G. H., Revesz, G., "Mechanization of Chemical Information Publications and Services;" paper presented at the 149th National Meeting of the American Chemical Society, Division of Chemical Literature, 149th Meeting, ACS, Detroit, April 1965.
- (3) Standley, C. L., *J. Appl. Phys.* **35**, 1530 (1964).
- (4) Mead, C. A., *Phys. Rev.* **128**, 2088 (1962).
- (5) Garfield, E., Sher, I. H., *Science Citation Index 1964*, Annual Cumulation, Citation Index, Part 3, Column 8001, Institute for Scientific Information, Philadelphia, Pa., 1965.
- (6) *Ibid.*, **1965**, Annual Cumulation, Citation Index, Part 1, Column 314, Institute for Scientific Information, Philadelphia, Pa., 1966.
- (7) Hirata, M., *Hiroshima J. Med. Sci.* **13**(1), 29 (1964).
- (8) Altshuler, C. H., Angevine, D. M., *Am. J. Pathol.* **25**, 1061 (1949).

Effects of Errors in the Chemical Literature on the Compilation of Critically Evaluated Data*

WILLIAM H. EVANS
National Bureau of Standards, Washington, D. C.

Received May 11, 1966

The types of errors encountered in the chemical literature while compiling critically evaluated data are discussed.

In any program for the collection and compilation of data to obtain a set of critically evaluated tables, the investigator is plagued by errors in the literature he must use. These errors often seem trivial to an outsider, but to the person who is trying to obtain the "best" values possible, they are extremely bothersome. Such errors make it difficult to assess the true worth of the experimental measurements, which are often of high reliability; they cast shadows on the results, and may, in severe cases, cause the measurements to be discarded.

In our program at the National Bureau of Standards on the preparation of the tables of Selected Values of Chemical Thermodynamic Properties, we have encountered our share of these errors. The errors we are concerned with here are not those in experimental measurements due to systematic or random factors, or even to the misinterpretation of the measurements; they are the small typographic and calculational errors and errors due to careless writing. It is probable that these cause us as much trouble as the problems from the uncertainties in the measure-

ments themselves—and cast doubt on many excellent sets of experimental data.

Perhaps a few examples out of the many we have encountered will indicate the types of "errors" that occur.

Ideally, a paper should give all of the experimental details and results, and the auxiliary data used, so that an evaluation of the results may be made. Unfortunately, because of space limitations, this is not possible today. (Even 75 years ago, when articles would run to 40 or 50 pages, the details were still not always given.) Therefore, a selection of the more pertinent results must be made. During this selection, during the recalculation of results into final form, during the preparation of the manuscript and the transformation into the printed page, there are many chances for errors to occur: errors of omission, of transposition, of inconsistent values.

It may be as simple as using the wrong molecular weight to convert from the specific quantities actually measured to molar quantities; but if the value used or the basic specific quantities measured are not given, this error is hidden. As an example, a recent paper from a reputable laboratory on the low-temperature specific heat of a compound gives the mass of sample used, the corresponding number of moles, and the molecular weight. The experimental results have been converted to the molar

*Presented before the Division of Chemical Literature, Symposium on Error Control in the Chemical Literature, 151st ACS National Meeting of the American Chemical Society, Pittsburgh, Pa., March 23, 1966.

basis. However, if the sample mass given is divided by the molecular weight, a different number of moles is obtained. In addition, the molecular weight given does not correspond to that on the atomic mass scale in use at that date, or anywhere near that date. Use of the correct molecular weight does not eliminate the difference. This leaves the evaluator with the problem: To what mole do the reported molar values refer?

Similar questions occur where values of the thermodynamic functions $(G^\circ - H_0^\circ)/T$, $(H^\circ - H_0^\circ)/T$, and S° are tabulated. Occasionally sets will be found in which the fundamental relationship $(G^\circ - H_0^\circ)/T = (H^\circ - H_0^\circ)/T - S^\circ$ is not met. This leaves the evaluator the problem of deciding which function in error. He is forced to repeat the calculation, which may not be easy because of a lack of other details.

In tabular material the omitted or transposed line is frequent. One case involved a table of thermal functions over a range of temperatures. In the middle of the table one set of functions was omitted, and those for the next higher temperature were substituted. This dislocation continued to the last entry, where the last set of values was repeated. Without this warning signal, it would have been unlikely that the skip would have been noticed. Another table from the same laboratory was worse; one temperature and another set of values were omitted so that the table seemed correct.

The practice of many journals of requiring experimental data, such as vapor pressures, to be represented by only a small graph and an equation leads to problems. I have had cases where a vapor pressure equation was given along with one measured point (to show the purity, perhaps, by comparison with other measurements) which could not be obtained from the equation. This again dumps the problem in the lap of the evaluator: Is the equation correct?

Then there are solutions in which the concentration as given is not represented by the composition described.

This is all included in the general class of errors arising from giving too little or too much information. If too little information is given, there is no way to detect the single transposed digit, a common typo. The results must be taken on faith and hope. If too much is given, then the problem is to locate the wrong value among several interrelated quantities.

Related to these errors are those cases where there are ambiguities in sign, units, scales, etc. These usually arise because the author doesn't see any question. (At the time he writes, the exposition is crystal clear; 20 years later, it may not be.) High temperatures are reported with no indication of degrees C. or K. Concentrations are given as molar and the symbol for molal is used, or vice versa. This type of confusion is very evident in reporting heat data. In the past (and too often today) authors will talk about heat and enthalpy as though they were interchangeable. If the heat of combustion of a hydrocarbon is measured, there is not much question: q is heat evolved, and is positive; ΔH is the enthalpy change for the system, and is negative. But for such quantities as heats of solution, in which heat may be either absorbed or evolved, the problem is acute. For instance, does a value of -1.3 kcal. mole for the heat of solution of compound X mean q or ΔH ? There is one article in which

the fragmentary experimental details indicate a value of ΔH as negative; the authors label it as q with a negative sign but proceed to use it as ΔH !

There are also the experimenters who, apparently, fall into their own traps; they use the appropriate auxiliary corrections, but apply them with the wrong sign.

These, in brief, are some of the errors in the chemical literature that make the life of a critical data evaluator complicated. What can be done, or what is being done to correct such errors? The answer, unfortunately, is little.

A few journals, such as the *Journal of the American Chemical Society*, publish annual errata sections, containing corrections which the authors have submitted. Others—for example, the *Journal of Chemical Physics*—publish errata each issue. These are often so brief that the reader can't decide, in many cases, if the original article was of interest to him. They are covered by *Chemical Abstracts*, at least, but only by title, which gives little clue as to content.

Other journals publish lists of corrections in the form, "Pg. 2950, line 10, read 0.1 instead of 1.0." This is worse. The only way they are of any use is to have someone, usually the librarian, systematically enter the corrections in the articles. If this is done, future reference to the paper will uncover the correction. If the paper has already been abstracted before the correction is made, the correction is lost to the evaluator. (The same is true, of course, of the *Journal of the American Chemical Society* type of correction.) Worse still are the small slips of paper, often gummed, included with later issues of the journal. If these are not inserted in their proper place, they may actually be lost. It is somewhat disconcerting, though, to find them bound with the journal issue in which they were sent out.

Even more obscure are the corrections buried in articles. An author may introduce a correction in the text or as a footnote in a later article. This may or may not be caught by a thorough reader; it will never make the abstract. Or a correction may appear as a "private communication" footnote in another author's paper. Here there is no hope even of tracing the correction through the author in the indexes. And, finally, there is the large group of journals that never publish any errata.

This whole system of errata publication assumes that the authors will send in corrections. Some do. Others apparently never read their papers carefully.

Even such useful tools as the *Citation Index* won't help much. If an error is suspected, this can be of aid in running it down. Unfortunately there are too many articles to check each article each time it is used.

What can be done to improve the situation? For recent papers, the evaluator can write the author, and perhaps prod him into publishing a correction. With papers published 25 or 50 or 75 years ago, this is impossible. Even if a correction is published, it is difficult to link it to the original article. So the evaluator is left with the problem of errors and discrepancies that must be resolved before the experimental work can be evaluated. In all too many cases nothing can be done, and the whole work must be down-graded or discarded. What the evaluator really needs is a perfect memory—to link errata with the originals (and this may be possible with mechanized searching)—and an ability to read between the lines.