INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

PHYSICAL CHEMISTRY DIVISION

COMMISSION ON PHYSICOCHEMICAL MEASUREMENTS AND STANDARDS*

AN ANNOTATED BIBLIOGRAPHY ON ACCURACY IN MEASUREMENT

Prepared for publication by

J. P. CALI[1] and K. N. MARSH[2]

[1]National Bureau of Standards, Washington DC 20234, USA

[2]Department of Chemistry, University of New England, Armidale, NSW 2351, Australia

AN ANNOTATED BIBLIOGRAPHY ON ACCURACY IN MEASUREMENT

Prepared for publication by J. P. Cali[a] and K. N. Marsh[b]

[a]United States Bureau of Standards, Washington D.C. 20234, USA
[b]Department of Chemistry, University of New England, Armidale,
NSW, 2351, Australia

The Comité International des Poids et Mesures has made a study on the
problems associated with the evaluation and presentation of uncertainties
of the results of measurement. It was clear from that study that there was
no uniformity of opinion with regard to the major questions raised. As a
contribution to the continuing discussions this Annotated Bibliography on
Accuracy in Measurement was prepared. The Bibliography does not take any
particular position with regard to the proper meaning or interpretation or
application of accuracy in measurement. The Bibliography is primarily
directed at those scientists who wish to review the most significant
contributions made over many years on the subject of accuracy in
measurement. Large treatises dealing with statistics and statistical
analysis have not been included. It can be argued that accuracy is not
basically a statistical phenomena, even though it is agreed that the
investigation of measurement error and the determination of the magnitude
of errors cannot be accomplished unless the measurement system is under
statistical control. Rather accuracy, in some ways, touches upon, perhaps
profoundly, the basic philosophical questions that surround the
measurement process itself.

INTRODUCTION

During the last ten years members of IUPAC Commission 1.4, Physicochemical Measurements
and Standards, have often wrestled with the various meanings and interpretations of
accuracy as it is applied to measurement problems. That there are many meanings of
accuracy is clearly and quickly evident to any scientist who interacts with fellow
scientists making similar measurements. The Comité International des Poids et Mesures
(CIPM) decided at its 1977 Fall meeting to study the problems associated with the
evaluation and presentation of the uncertainties of the results of measurement. A
questionnaire was prepared and sent to all members of CIPM, and the replies were tabulated
and commented upon in a BIPM report. (See bibliography for report citation and contents).
A quotation from the conclusions section of this report is most instructive.

> "The diversity of the replies received shows clearly that a uniformity of opinion
> is not yet reached. On the other hand, the probing questions may have stirred up
> the minds of some participants and led them to question things which are usually
> considered as well established. Indeed, the process of thinking over some of the
> basic problems seems to be well under way. Perhaps the most remarkable outcome of
> the questionnaire lies in the simple fact that the majority of the participants
> seem to have no final opinion on most of the problems raised: they are realizing
> the difficulties involved and are, one has the impression, waiting for sound
> proposals. This should be a favourable situation for coming to some agreement
> which we hope, will then be acceptable to a large majority."

It is not the intention of this bibliography to take any particular position with regard
to the meaning or interpretation or application of accuracy in measurement. Indeed, as
evidenced by the studies within the CIPM and other international bodies, it would be
presumptuous to do so. For example, the CIPM study group is composed of scientists whose
everyday work is intimately concerned in one way or another with measurement at the most
basic level. Thus, this bibliography is directed toward those scientists who wish to
review the most significant contributions made over many years on the subject of accuracy
in measurement. We have deliberately omitted most large treatises which deal in a general
way with statistics and statistical analysis. It is the view of this Commission that
accuracy is not basically a statistical phenomenon, although it is readily agreed that the
investigation of measurement error and the determination of the magnitude of these errors
cannot be successfully accomplished unless the measurement system is under statistical
control. Instead accuracy touches on, perhaps profoundly, the basic philosophical
questions surrounding the measurement process itself. We leave the user of this
bibliography to make his own judgement in this regard.

ANNOTATED BIBLIOGRAPHY

The Bibliography is arranged in alphabetical order of the name of the first author or editor or the institution.

A. Allisy,
Les erreurs aléatoires,
Conservatoire National des Arts et Métiers, Paris (1975), (in French).

R. L. Ackoff (Ed.),
Scientific Method: Optimising Applied Research Decisions,
John Wiley, New York (1962).
An excellent text covering many fundamental topics concerned with measurement from a broad philosophical basis. Of special interest are: Chapter 1: The nature of science and methodology; Chapter 3: Formulating the problem; Chapter 4: Models; Chapter 5: Definitions; Chapter 6: Measurement. Other topics of lesser interest include sampling, estimation, testing hypotheses, and the implementation and organization of research. This work is highly recommended as a starting point for the scientist who wants to study the topic of measurement accuracy from basic principles.

G. B. Airy,
On the Algebraic and Numerical Theory of Errors
of Observations and the Combination of Observations,
Macmillan, London (1861).
One of the classical texts on probabilities and errors.

American Society for Testing and Materials (ASTM),
ASTM Manual on Quality Control of Materials, ASTM Comm. E-11,
ASTM Spec. Tech. Publ. 15-C (1951).
Consists of three parts: (1) presentation of data; (2) presenting plus and minus($\pm$) limits of uncertainty of an observed average; (3) control chart method of analysis and presentation of data.

American Society of Testing and Materials (ASTM),
Use of the Terms Precision and Accuracy as Applied to
Measurement of a Property of a Material,
ASTM Standards, Part II, 1758-66 (1961).

Anon.,
Round-table Discussion on Statement of Data and Errors,
Nucl. Inst. and Methods 112, 391-95 (1973).
An informal discussion held during the First International Summer School of Radionuclide Metrology in 1972. All the papers and much of the discussion following the various sessions are recorded. The conference, although limited to radionuclide metrology, addressed problems associated with many aspects of error detection and correction of systematic errors. Chapter 1 (pp 1-47) has several good review papers on these aspects. Chapter 8 is devoted entirely to the Statement of Data and Errors and is followed by the round table discussion. Papers in Chapter 8 include: Statement of Results of Experiments and Their Accuracy, by A. Williams, P. J. Campion and J. E. Burns; Statistical Methods Applicable to Counting Experiments and Evaluation of Experimental Data by H. H. Ku; and Treatment of Errors in Low-Activity Measurements by M. Hillaire.

F. J. Anscombe and J. W. Tukey,
The Examination and Analysis of Residuals,
Technometrics 5, 141 (1963).
Abstract: "A number of methods for examining the residuals remaining after a conventional analysis of variance or least-squares fitting have been explored during the past few years. These give information on various questions of interest, and in particular, aid in assessing the validity or appropriateness of the conventional analysis. The purpose of this paper is to make a variety of these techniques more easily available, so that they can be tried out more widely.

Techniques of analysis, some graphical, some wholly numerical, and others mixed, are discussed in terms of the residuals that result from fitting row and column means to entries in a two-way array (or in several two-way arrays). Extensions to more complex situations, and some of the uses of the results of examination, are indicated."

D. C. Baird,
Experimentation: An Introduction to Measurement Theory and
Experimental Design,
Prentice Hall, New Jersey (1962).
This book is suitable as an introductory text for first and second year physics courses. It includes a section on the nature of measurement and the propagation of uncertainties. The author treats uncertainty as a statistic which is used to signify outer limits of confidence within which we are almost certain (ie. perhaps 99 per cent certain) that the measurement lies. The detection of systematic error, the estimation of its magnitude, and its incorporation in the reported result is considered only when systematic errors are random and can be treated statistically.

N. C. Barford,
Experimental Measurements: Precision, Error and Truth,
Addison-Wesley, Reading, Mass. (1967).
A straight-forward empirical approach to the subject. This book would form the basis for a lecture course for first year university students.

M. S. Bartlett,
Probability and Chance in the Theory of Statistics,
Proc. Roy. Soc. A141, 518-34 (1933).
Frequency laws in statistics mean laws of chance. The distinction between chance and probability is emphasized. An attempt is made to show why exact arguments about chance are more fundamental as a mathematical basis for statistical theory and inference than the formula of inverse probability which, in statistics, is a hybrid of exact chance. Unknown formal theory can lead to misleading assumptions. Examples of arguments about chance include Fischer's methods of maximum likelihood and fiducial probability. Such arguments do not pretend to abolish the judgment and common sense necessary when we use general theory to help us make, from a particular sample, an inference about the population.

Y. Beers,
Introduction to the Theory of Error,
Addison-Wesley, Cambridge, Mass. (1953).

C. A. Bennett,
Application of Test for Randomness,
Ind. Eng. Chem. 43, 2063-67 (1951).
The study of runs and of mean square successive differences are recommended as tests for non-randomness. Their use may indicate factors systematically affecting results.

P. R. Bevington,
Data Reduction and Error Analysis for the Physical Sciences,
McGraw-Hill, New York (1969).

C. A. Bicking,
The Reliability of Measured Values - Part II An Illustrative Example,
Photogrammetric Eng. 18, 554-58 (1952).
The definition of reliability recognizes that the total variation in measurements may be separated into two parts, one termed inherent probability and the other termed overall reproducibility. The paper discusses the difference between results obtained by multiple observations by a single operator and the results obtained by many operators.

R. T. Birge
The Propagation of Errors,
Amer. Phys. Teacher 7, 351-57 (1939).
This article is an instructional exposition of Birge's work on propagation of errors. He stated the problem thus, "Given a definitely assigned uncertainty for each of a set of independently measured quantities, what is the resulting uncertainty in any specified function of these quantities?"

R. T. Birge
<u>Phys. Rev.</u> <u>40</u>,   207-227 (1932).
            <u>40</u>,   228-261 (1932).
            <u>42</u>,   736 (1932).
            <u>48</u>,   918 (1935).
            <u>54</u>,   972-973 (1938).
            <u>55</u>,   1119 (1939).
            <u>57</u>,   250 (1940).
            <u>58</u>,   658-659 (1940).
            <u>60</u>,   766-785 (1941).
These papers present some of Birge's work on the calculation of errors by the method of
least squares as applied to the determination of the best values for several of the basic
physical constants. The author presents examples from the (then) recent past showing
inconsistencies and even errors in the treatment of systematic error analysis. This
series of papers is quoted in such detail because they provide an extremely rich source of
comments (together with appropriate bibliographic references) by Birge on the treatment of
errors by several different schools of thought at that time (first third of 20th century).

P. W. Bridgman,
<u>The Logic of Modern Physics</u>,
Macmillan, New York (1928).
See comments under A. S. Eddington Reference

S. S. Brown, M. J. R. Healy, and M. Kearns,
Report on the Inter-laboratory Trial of the Reference Method
for the Determination of Total Calcium in Serum Part I and II
<u>J. Clin. Chem. Clin. Biochem.</u> <u>19</u>, 395-412, 413-426 (1981).
This report discusses in detail the results of various inter-laboratory trials on the
determination of calcium in serum. A detailed analysis is made of the precision, accuracy,
and sources of variability.

Bureau International des Poids et Mesures (BIPM),
<u>Report on the BIPM Enquiry on Error Statements</u>,
Rapport BIPM - 80/3 (1980).
Procès-Verbaux CIPM, <u>48</u>, C1-C36(1980), (in French).
This is an essential document summarizing the current situation with regard to error
statements. That there is no unanimity on questions concerning such things
as accuracy, uncertainty, and the differentiation of random and systematic error is
painfully evident from even a cursory reading. The fact that experts from metrological
laboratories of 21 countries cannot agree on these basic questions supports the contention
that there is no easy solution to the accuracy question. Some of the questions asked in
the world-wide survey include: is there an essential difference between random and
systematic errors? Should one recommend practical rules for (a) the expression of
systematic errors (b) combining random and systematic errors and (c) expressing the final
uncertainty?
The bibliography is also useful. One section gives regulations and guidelines regarding
accuracy statements. The section contains 22 references on special studies on error
statements, several of which are included in this bibliography.
This report is probably the best starting point for those planning an in-depth study of
accuracy since it clearly states the pitfalls and the present confusion with regard to the
use of various terms in describing accuracy in measurement.

J. P. Cali,
An Idea Whose Time has Come,
<u>Clin. Chem.</u> <u>19</u>, 291-93 (1973).
This was a guest editorial which stressed the importance of achieving accuracy in methods
used in clinical chemistry. The author based his argument on the concept that measurement
compatibility in clinical measurements (especially important as these measurements need to
be, for every individual, compatible over both distance and time) is best achieved through
accuracy. Five requirements for achieving accuracy were delineated: (1) agreement on a
consistent set of measurement units; (2) well-characterized reference materials; (3)
reference methods of demonstrated accuracy; (4) fixed (or routine) methods used in
everyday practice and tested for accuracy via (2) and (3); and, (5) quality assurance
programs to insure long-term reliability of the measurement system.

J. P. Cali,
Problems of Standardization in Clinical Chemistry,
Bull. World. Health. Org. 48, 721-26 (1973).
This article discusses the problems that require solution for there to be reliable
measurements in clinical chemistry on a world-wide basis.
In order for analytical results in clinical chemistry to be accurate, precise, and
specific, a systematic approach is necessary. Furthermore, because these systems are so
complex and the need for standardization is so widespread, it will require international
coordination. Agreement on the units of measurement, the production and certification of
standard reference materials, and the development of reference methods of demonstrated
accuracy will require the support of all segments of clinical chemistry.

J. P. Cali,
A Systematic Approach to Accuracy in Clinical Chemistry,
Med. Instrum. 8, 17-21 (1974).
When measurements made in clinical chemistry laboratories are meaningful, the values
obtained are accurate, precise, and specific. The latter two characteristics, which are
related to reproducibility and singularity respectively, represent no great problem in
clinical chemistry measurements. Accuracy, which is related to the true value, however,
remains a somewhat elusive goal. Unless a measurement system is based on accuracy,
comparison of results obtained over time and distance in different laboratories may lead
to doubtful or misleading conclusions. A meaningful measurement system consists of five
parts: (1) a rational, self-consistent, agreed-on system of units of measurement; (2) well
characterized materials used in conjunction with (3) referee methods of known accuracy to
realize in practice the base units and their derivative; (4) field or applied methods of
measurement, assessed for accuracy via parts 2 and 3; and (5) a process whereby the
long-term integrity of the measurement system is assured.

J. P. Cali and W. P. Reed,
The Role of NBS Standard Reference Materials
in Accurate Trace Analysis,
Nat. Bur. Stand. Spec. Publ. 422, 41-63 (1976).

J. P. Cali and C. L. Stanley,
Measurement Compatibility and Standard Reference Materials,
Ann. Rev. Matl. Sci. 5, 329-43 (1975).

J. M. Cameron,
Measurement Assurance,
Nat. Bur. Stand. Int. Rep., 77-1240 (1977).
The procedures by which one establishes that the uncertainty of individual measurements is
adequate to their needs has been titled measurement assurance. This note discusses the
factors involved in achieving measurement assurance, beginning with a base which serves as
a standard or reference. The paper then discusses the determination of the uncertainty
relative to this base and the need for control of the measurement process to assure the
continuing validity of the accepted process parameters.

J. M. Cameron,
Measurement Assurance,
J. Qual. Technol. 8, 53-55 (1976).
There is a need for measurements in the fields of health, safety, environmental control
and the nuclear safety area to be adequate for their intended purpose. Further their
uncertainty should be small enough to only negligibly affect the decisions and performance
of the processes of which they are a part. This is, of course, no less true for most other
measurements in science and industry. This note discusses the procedures by which one
obtains measurement assurance, the analogue for measurement processes of industrial
quality assurance.

P. J. Campion, J. E. Burns, and A. Williams,
A Code of Practice for the Detailed Statement of Accuracy,
National Physical Laboratory, UK, Her Majesty's Stationery Office, London (1973).
We quote directly from the Preface of this key monograph: "The main purpose of this Code
of Practice is to put forward recommendations as to how uncertainty can be expressed so as
to avoid ambiguity. Further, it also discusses some of the ways in which estimates of
uncertainty can be derived from individual measurements, including the separation into

random and systematic categories, the procedures that can be adopted for combining the individual uncertainties in each category and the various ways in which acceptable statements of accuracy may be made. It is not a manual of statistics, nor does it deal with the subject of limits of error as used in instrument specification. It is primarily concerned with the detailed reporting of scientific measurements of the highest quality, but the recommendations are such that they may be used for any situation where a statement of uncertainty is required, e.g. in certain calibration certificates, etc. However for those routine measurements in which the uncertainty of an instrument calibration is small, bearing in mind its subsequent use, it may be unnecessary to state the uncertainty in the detail required by this Code".

K. L. Churney and G. T. Armstrong,
Studies in Bomb Calorimetry,
J. Res. Natl. Bur. Stand. 72A, 453-65 (1968).
This paper describes an experimental study on the determination of the energy of combustion of benzoic acid. It is quoted (primarily) to illustrate how highly accurate experimental work is designed and carried out. It also illustrates how systematic errors are estimated and their uncertainties used to estimate the final overall uncertainty.

Codata Task Group on Publication of Data in the Primary Literature,
(D. Garvin, T. Golashvili, H. V. Kehiaian, N. Kurti, E. F. Westrum Jr. (chairman)),
Guide for the Presentation in the Primary Literature of Numerical
Data Derived from Experiments,
Codata Newsletter no. 8 (1972); Codata Bulletin 9 (1973); NSRDS News (Feb. 1974).
A statement of the minimum information needed to ensure that the reader can understand the quantitative data, assess their precision and accuracy, and recalculate the results when values for auxiliary data change.

Committee E-11 on Quality Control of Materials (ASTM),
ASTM Manual for Conducting an Interlaboratory Study of a Test Method,
ASTM Spec. Tech. Publ. 335, American Society for Testing and Materials,
Philadelphia, Pa. (1963).
The preface states: "The procedure (i.e., this manual) is intended for the evaluation of well-defined physical or chemical testing processes which yield measurement of properties. Thus, the procedure covers interlaboratory evaluation of test methods, but not interlaboratory evaluations of materials." Included is a brief discussion of the measurement process. Basically the manual is concerned with precision, although some statements concerning accuracy such as "Systematic laboratory difference can be essentially eliminated by arranging for the laboratories to test a reference material at the same time they test the unknown." are made.

N. H. Cook and E. Robinowicz,
Physical Measurement and Analysis,
Addison-Wesley, Reading, Mass. (1963).
A comparatively elementary text based on a course for graduate students. The majority of the book is devoted to an analysis of the possible sources of error associated with a wide range of physical measurements made with simple apparatus. Numerical problems of a very practical nature are given.

E. L. Crow,
An Analysis of the Accumulated Error in a Hierarchy of Calibrations,
IRE Trans. Instr. I-9, 105-14 (1960).
This study is aimed at allocating errors (i.e., degree of accuracy to be attained) at various levels in a measurement hierarchical network so as to minimize costs over the entire network. Two questions are posed: (1) how do errors accumulate from echelon to echelon in a hierarchy of calibrations, and (2) if a certain accuracy must be achieved at the final echelon level, then what is the optimum allocation of errors among echelons?

L. A. Currie,
Sources of Error and the Approach to Accuracy in Analytical Chemistry,
Chap. 4, Part I, Sec. B, 95-242,
Treatise on Analytical Chemistry, 2nd Ed. I. M. Kolthoff and P. J. Elving (eds.)
John Wiley, New York (1978).
One of the best, most recent discussions of accuracy. It draws heavily on examples from nuclear- and radio-chemistry. An extensive bibliography (185 references) covers much

pertinent literature in the period 1965 to present. Topics covered include: importance
of accuracy and its impact in the real world; underlying assumptions surrounding accuracy
and the study of errors; the physicochemical model and the chemical measurement process
including sources of error, characterization of procedures and results, validation and
standards; and exploratory techniques.

     O. L. Davies and P. L. Goldsmitth
     <u>Statistical Methods in Research and Production</u>,
     4th Edn., Oliver and Boyd, Edinburgh (1972).
This book, written and revised by members of the staff of Imperial Chemical Industries
Ltd., presents statistical methods in the context of the chemical industry.

     W. E. Deming and R. T. Birge,
     On the Statistical Theory of Errors,
     <u>Rev. Modern Phys</u>. <u>6</u>, 119-61 (1934).
This is a key paper dealing primarily with the statistical treatment of random errors in
measurement. The detection of systematic errors (or their likelihood) is discussed by
Birge in <u>Phys. Rev</u>. <u>40</u>, 207-61 (1932). Systematic errors (specifically) are not
addressed, but several interesting observations are made concerning them and their
treatment, for example "if there were no systematic errors present, the mean of the parent
population would be the true value of the quantity being measured. The effect of a
systematic error is to displace the mean of the parent population of observations above or
below the true value. This correction, if ever isolated and evaluated, can be added to or
subtracted from the mean of the parent population to give the true value", and, "the true
value of the quantity being measured is approached by correcting for systematic errors,
one after another. The effect of accidental errors can be reduced as far as desired by
taking enough observations. The measurement of each systematic correction presents a
problem in statistics, for a correction cannot be intelligently applied unless its
precision is stated."
The following topics are discussed: specification of the parent population; the
distribution of certain properties of samples drawn from a normal (Gaussian) parent
distribution; and the estimation of the probable error (by three different techniques;
maximum likelihood, empirical estimates, and the posterior method).

     J. R. DeVoe, (ed.),
     Validation of the Measurement Process,
     <u>ACS Symposium Series</u> No. 63, American Chemical Society, Washington (1977).
These are the papers presented at a symposium on the measurement process held in 1976.
The authors of the six chapters are all active in various aspects of analytical chemistry,
including statistics. Of special interest are the first four chapters, especially chapter
4. The six chapters are: (1) Statistical control of measurement processes, G. Wernimont;
(2) Testing basic assumption in the measurement process, J. J. Filliben; (3) Systematic
error in chemical analysis, L. A. Currie and J. R. DeVoe; (4) Role of reference materials
and reference methods in the measurement process, G. A. Uriano and J. P. Cali; (5)
Optimization of experimental parameters in chemical analysis, S. N. Deming; and (6)
Components of variation in chemical analysis, R. C. Rhodes.

     DIN (Deutsche Industrie-Norm),
     <u>Begriffe der Qualitätssicherung und Statistik</u>,
     DIN 55350 (especially parts. 13, 21, 22, and 24), (1979) (in German).
This is the basic standard for the Federal Republic of Germany with regard to terms,
definitions, applications of statistics to measurement and quality control. This document
follows closely the ISO 3534 standard.

     DIN (Deutsche Industrie-Norm),
     <u>Statistische Auswertungen</u>,
     DIN 53804 Part. 1, (1981) (in German).
     Statistical evaluation and measurable (continuous) characteristics.

     DIN (Deutsche Industrie-Norm),
     <u>Präzision von Prüfverfahren, Bestimmung von Wiederholbarkeit und Vergleichbarkeit</u>,
     DIN-ISO 5725, (1981) (in German).
This is the German Standard for the precision of test methods and the determination of
repeatability and reproducibility by interlaboratory tests. The Standard provides

numerical definitions for the repeatability and the reproducibility of the results of a
standard test method. It discusses the implications of these definitions and presents
some practical rules for the interpretation of the various terms. It also describes the
organization and analysis of interlaboratory experiments for this purpose. This Standard
is identical to ISO Standard 5725.

K. Doerffel,
Statistik in der Analytischen Chemie,
VEB-Verlag, Leipzig (1966) (in German).
This publication is specially for the analytical chemist without a high degree of
knowledge in mathematics. The author relies on practical experiences in the evaluation of
the results of an analysis, and the reader is introduced to all the important aspects of
statistical analysis.

T. J. Dols and B. H. Armbrecht,
Assessment of Analytical Method Performance Characteristics: Systematic Error,
J.Assoc. Off. Anal. Chem., 60, 940-945 (1977).
Abstract: "A random review of the analytical literature shows the need to define more
clearly the terms for techniques that are used to assess the merits of analytical methods
for a particular purpose. One such performance characteristic is the systematic error or
bias of a method. This term is defined and contrasted with other terms commonly used in
method assessment. Components of the systematic error are described and techniques are
given for their measurement."

N. E. Dorsey and C. Eisenhart,
On Absolute Measurement,
Sci. Monthly 77, 103-09 (1953).
A philosophical discussion of the bases of measurement. Discussed are theory of errors,
averaging, quaesitum, definitive value, dubiety, measurement procedure, measurement
report, and some miscellaneous thoughts and observations. A key paper, essential for an
understanding of measurement accuracy.

J. W. M. DuMond and E. R. Cohen,
Our Knowledge of the Atomic Constants F, N, m, and h in 1947, and
of Other Constants Derivable Therefrom,
Rev. Mod. Phys. 20, 82-108 (1948).
An excellent paper illustrating the interdependence of the basic physical constants with
regards to their values and their uncertainties. Especially interesting is a discussion
of the Faraday Constant, F, and the divergence of the values of F when arrived at by two
different methods. The authors recommend a.review of methods used for determining the
value of the Faraday. This paper illustrates how physicists, and more specifically
metrologists, view systematic error in measurement, and use the knowledge of small
uncertainties in some of the constants to infer that systematic errors must be present in
other related constants.

K. Eckschlager,
Errors, Measurements and Results in Chemical Analysis,
van Nostrand Reinhold, London (1969).
The title is self-explanatory; errors are treated from a chemical standpoint and specific
problems of weighing, titration, quantitative precipitation and numerous instrumental
analytical techniques are discussed in relation to the theory of errors and application of
statistical methods.

K. Eckschlager,
Criterion for Judging the Acceptability of Analytical Methods,
Anal. Chem. 44, 878-79 (1972).
A short note which illustrates the difficulties encountered when systematic error
estimates are combined in one overall equation with random errors, especially when no
regard is given to the end-use for which the measurement is made. Such a criterion also
depends upon there being known the true or correct value for the quantity under test.

A. S. Eddington,
The Nature of the Physical World,
Macmillan, New York (1928).
Although more modern texts are available, this book (plus those of P. W. Bridgman and B. Russell, q.v.) gives clear and reasoned arguments to explain the difficulties involved in the interpretation of physical phenomena. Discussions in each of these books show how reliance on probability theory is necessary for understanding physical phenomena. These texts are of interest to those scientists who wish to understand the philosophical basis of measurement.

H. Egan and T. S. West (eds.),
Collaborative Interlaboratory Studies in Chemical Analysis,
Pergamon, Oxford (1982).
This book contains the majority of papers presented at the International Symposium on Harmonization of Collaborative Analytical Studies in Helsinki in 1981. Recently there has been considerable interest throughout the chemical sciences in the standardization and validation of those analytical methods which are the basis of specification, and the legal enforcement of standards covering all aspects of human health and endeavour. It is obviously desirable that such standards and analytical methods should be compatible. There are now several national and international research institutions whose primary goal is to develop such methods throughout the world. This volume summarizes the state of progress in this area.

C. Eisenhart,
The Reliability of Measured Values - Fundamental Concepts,
Photogrammetric Eng. 18, 542-54 and 558-65 (1950).

C. Eisenhart,
Realistic Evaluation of the Precision and Accuracy of Instrument
Calibration Systems,
J. Res. Nat. Bur. Stds. 67C, 161-87 (1963).
One of the fundamental papers on measurement theory and the philosophy of measurement. This paper is of such value and importance that its abstract is given in full:
"Calibration of instruments and standards is a refined form of measurement. Measurement of some property of a thing is an operation that yields as an end result a number that indicates how much of the property the thing has. Measurement is ordinarily a repeatable operation, so that it is appropriate to regard measurement as a production process, the product being the numbers, i.e., the measurements, that it yields; and to apply to measurement processes in the laboratory the concepts and techniques of statistical process control that have proved so useful in the quality control of industrial production. Viewed thus it becomes evident that a particular measurement operation cannot be regarded as constituting a measurement process unless statistical stability of the type known as a state of statistical control has been attained. In order to determine whether a particular measurement operation is, or is not, in a state of statistical control, it is necessary to be definite on what variations of procedure, apparatus, environmental conditions, observers, operators, etc., are allowable in repeated applications of what will be considered to be the same measurement process applied to the measurement of the same quantity under the same conditions. To be realistic, the allowable variations must be sufficient in scope to bracket the circumstances likely to be met in practice. Furthermore, any experimental program that aims to determine the standard deviation of a measurement process as an indication of its precision, must be based on appropriate random sampling of this likely range of circumstances.
Ordinarily the accuracy of a measurement process may be characterized by giving (a) the standard deviation of the process and (b) credible bounds to its likely overall systematic error. Determination of credible bounds to the combined effect of recognized potential sources of systematic error always involves some arbitrariness, not only in the placing of reasonable bounds of the systematic error likely to be contributed by each particular assignable cause, but also in the manner in which these individual contributions are combined. Consequently, the inaccuracy of end results of measurement cannot be expressed by confidence limits corresponding to a definite numerical confidence level, except in those rare instances in which the possible overall systematic error of a final result is negligible in comparison with its imprecision".

C. Eisenhart,
Expression of the Uncertainties of Final Results,
Science 160, 1201-04 (1968).
This paper gives a succinct discussion and recommendations for expressing the results of measurements. It strongly presses for statements that keep separate and distinct uncertainties resulting from systematic and random error. The author argues for four distinct forms of expressing the final results: (1) systematic error and imprecision both negligible; (2) systematic error not negligible, imprecision negligible; (3) neither systematic error nor imprecision negligible; (4) systematic error negligible, imprecision not negligible.

R. P. Ekins,
Basic Concepts in Quality Control, in
Radioimmunoassay and Related Procedures in Medicine 1977 6-20,
International Atomic Energy Agency, Vienna (1978).
A critical discussion of the philosophical and conceptual problems associated with the measurement of molecularly heterogeneous analytes, with particular reference to radioimmunoassay.

L. B. Eppstein and G. B. Levy,
Misinterpretation of Statistical Intercept Values,
Clin. Chem. 24 1286-87 (1978).
The reporting of slope and intercept when comparing a new method can lead to errors if the intercept is taken as a measure of the "offset" between the methods. It is suggested that better results are obtained if the data are transformed by subtracting the "normal" values from each. More accurate values for the "offsets" are thus obtained, as demonstrated by results for Sodium ion in serum, determined with the use of three automatic electrolyte analysers and a manual flame-photometric method.

R. W. Fennell and T. S. West,
Recommendations for the Presentation of the Results of Chemical Analysis,
Pure Appl. Chem. 18, 439-42 (1969).
IUPAC recommendations (Analytical Chemistry Division) for presenting results of chemical analysis. Terms defined include accuracy, error and bias. True value is not discussed nor is there a recommendation for reporting the overall uncertainty of the final results.

D. J. Finney,
Statistical Method in Biological Assay,
Charles Griffin and Co Ltd, London (1978).
A classical text on the statistical principles of biological assay with specific discussion on assay validity and its evaluation. The subject of Radioimmunoassay is reviewed within the general framework of biological assay permitting the clinical analyst to draw on the exerience gained in the Bioassay field in confronting the problems of accuracy.

H. Frehse and G. Timme,
Quantitative Residue Analytical Reliability:
Beatitude through Application of Latitude,
Residue Review 73 27 (1980).
This article discusses the accuracy and reproducibility of various analytical techniques used to determine pesticide residue.

W. O. Fitzgibbons,
Accuracy – An Industrial Viewpoint,
Methods and Standards for Environmental Measurements, SP464,
3-8, US Government Printing Office, Washington (1977).

D. Garvin,
Guidelines for the Reporting of Numerical Data and Experimental Procedures,
J. Res. Nat. Bur. Stand. 76A, 67-70 (1972).
General recommendations for the reporting of results, particularly in the area of physical chemistry, are given. Stressed is the importance of estimating the inaccuracy of the measurements. Likely sources of systematic error are briefly mentioned. A short bibliography is included with sections on: symbols, units, and nomenclature; atomic weights; energy; physical constants; precision and accuracy; and several other physico-chemical quantities.

R. W. Gerard,
Quantification in Biology,
ISIS 52 334-352 (1961).
A broad ranging essay on the nature of the measurement process and its role in the aquisition of knowledge.

L. Gonnella,
Proposal for a Revision of Measurement Theory and Terminology,
Alta Frequenza 44 622-28 (1975) (in Italian).
The unsatisfactory situation with regard to the usage of terms and standards concerning measurement and measuring instruments is due to the inadequacies of some basic notions. The author calls for a theoretical reformulation and asks that the very concepts of error and true value be given up. In the measurement of a parameter a full value-span, i.e. a set (rather, a fuzzy subset) of numbers tied to a unit of measure, is meant to represent as a whole, a parameter. The central term of the set serves as a reference value for the parameter, while its half-width embodies the uncertainty of the measurement. The results may be given a statistical treatment and they are so assigned as to ensure that various measurements of the same parameter are congruent, i.e. their value-spans overlap. Any parameter has a certain intrinsic uncertainty below which a measurement cannot be assigned a value without undoing the very definition of the parameter itself. Different types of measurable quantities must be recognized with respect to the way their value-span may be expressed and treated. In a measurement the aim is not to judge how far from the true value is the value indicated by the instrument, but to convert the instrument's reading into a proper measurement after accounting for the influence quantities. This is carried out through a calibration operation, defined operatively, which translates the reading-value (expressed in the instrument's own output format) into the central term of the value-span, and assigns to it an instrumental uncertainty, as functions of values of the influence quantities. The inexactness vested in a measurement deed is thus integrated into the very definition of the measure, instead of appearing as a judgment on its validity. This approach is claimed to give a sound base for describing the comparison of results as actually carried out in practice, without hazy distinctions between systematic and random errors. It is necessary to consider what one can and wants to correct and what is left uncertain. It is claimed that a coherent treatment can be given for all quantities and instruments, even those as yet unamenable to systematic handling.

M. M. Gupta, G. N. Saridis and B. R. Gaines, (eds.),
Fuzzy Automata and Decision Processes,
Elsevier-North Holland, New York (1977).
The twenty four chapters by various authors provides a broad introduction to fuzzy set theory and its applications. A comprehensive bibliography lists 763 papers. The theory that is described for handling measurement uncertainties owes nothing to classical statistical concepts and the gaussian error function.

A. T. J. Hayward,
Repeatability and Accuracy,
Mechanical Eng. Pubs., London, UK (1977).
A standard procedure is proposed for measuring the repeatability and estimating the accuracy of industrial measuring instruments on the basis that "It is more important to be nearly right and understandable than to be academically accurate and imcomprehensible." - which Hayward gives as a quotation, author unknown.

L. Hogben,
Statistial Theory
Allen and Unwin, London (1957).
Insofar as this book discusses the philosophy behind the use of statistical theory in several branches of science, it belongs with the books by Bridgman, Eddington, and Russell, but it is also more relevant, in that it attempts to determine the validity of the assumptions made on a case by case basis.

W. G. Hunter and W. F. Lamboy,
A Bayesian Analysis of the Linear Calibration Problem,
Technometrics 23, 323-28 (1981).
A complete review of the statistical problems associated with calibration. Extensive references are given. Pages 329 to 350 of the above-referenced journal includes a discussion by various contributors and a reply by the authors.

W. Horwitz,
The Inevitability of Variability in Pesticide Residue Analysis, in
Advances in Pesticide Science, H. Geissbüchler (ed.),
Pergamon, Oxford (1979).
Discusses the accuracy, reproducibility and limitations of various techniques used to
determine pesticide residue.

K. Iizuka,
The Accuracy and Permissible Tolerance in Industrial Measurement,
Transactions of the Japan Society of Mechanical Engineers 77, 595 (1974),
(in Japanese).
Definitions of the terms relevant to the errors of measurement and the methods of
evaluating the repeatability and reproducibility of measurement are described. A standard
procedure to determine the tolerances after due consideration of the measuring accuracy is
also presented.

International Organization of Legal Metrology (OIML),
(Organization International de Métrologie Légale),
Vocabulary of Legal Metrology,
British Standards Institution, PD6461 (1971).
(Note: This version, in English, is an unofficial translation from the French. OIML
currently recognizes only the French version as official)
This vocabulary of measurement terms and definitions is recognized by OIML member
adherents as authorative and it is indispensable for those working in the legal
metrological areas of measurement. The vocabulary is comprehensive and internally
consistent. The ten chapters cover the following topics: Chapter 0, the definition of
legal metrology and its scope; Chapter 1, organizations and services; Chapter 2,
activities of the services of legal metrology; Chapter 3, documentation and marking;
Chapter 4, quantities and units of measurement; Chapter 5, measurements; Chapter 6,
measuring instruments and their classification; Chapter 7, measuring instrument
construction and component parts; Chapter 8, errors in the results of measurement and
errors of measuring instruments; Chapter 9, conditions of use and metrological properties
of measuring instruments.
With regard to the subject matter of this publication, most topics of interest are to be
found in chapters 8 and 9.

International Union of Pure and Applied Chemistry,
A report of Commission I.2 on Thermodynamics by G. Olofsson,
S. Angus, G. T. Armstrong and A. N. Kornilov,
Assignment and presentation of uncertainties of the numerical results
of thermodynamic measurements,
Pure Appl. Chem. 53, 1805-25 (1981); J. Chem. Thermodynamics 13, 603-22 (1981).
This paper elaborates on an earlier IUPAC recommendation that "estimates of the precision
indices and probable accuracy of the data should be given by the authors. The various
sources of uncertainty should be rigorously described with clear separation of measurement
imprecisions, numerical analysis deviations, and possible systematic biases. The methods
and assumptions for the statistical analyses should be indicated. Possible sources and
magnitudes of systematic errors should be identified and enumerated".

K. Ishikawa, T. Fujimori, and H. Kume,
Introduction to the Statistical Method of Error Analysis for
Chemists and Chemical Engineers,
Tokyo Kagaku Dojin, Tokyo (1964), (in Japanese).
A practical and detailed explanation of error analysis in quantitative chemical analysis.
This book includes the definition of relevant technical terms, treatment of measuring and
sampling errors, statistical estimates and tests, and regression analysis.

Japan Industrial Standards Z 8402,
General Rules for Permissible Tolerance of Chemical Analysis
and Physical Test,
Japan Standards Association (1974), (in Japanese).
Definition of terms used, the methods for determining permissible tolerances and the usage
of the tolerance values are described. Procedures for improving accuracy, methods of
treating the accuracies and biases, experimental methods for checking accuracy, and the
treatment of unusual values are also discussed in the appendices.

N. K. Jerne and E. C. Wood,
The Validity and Meaning of the Results of Biological Assays,
Biometrics 5, 273-99 (1949).
A detailed review of the assumptions underlying biological assay systems that is lucidly written and has a great deal to contribute to the discussion of the intellectual and practical problems of physico-chemical measurement.

J. M. Keynes,
A Treatise on Probability, Chapter 29,
Macmillan, London (1921).
The use of a priori probabilities for the prediction of statistical frequency and the theorems of Bernoulli, Poisson, and Tchebycheff are outlined. The limitations to the applicability of these theorems is discussed.

S. J. Kline and F. A. McClintock,
Describing Uncertainties in Single-sample Experiments,
Mech. Eng. 75, 3-8 (1953).
In this contribution an attempt is made to describe and place a value on uncertainty by fixing odds that an experimenter would be willing to give (or state) that the value reported lies within his stated uncertainty. Basic to the premise is the concept that uncertainties within an experiment have an estimatable distribution. Thus, under this view, systematic errors can (and are) treated statistically. The mathematics underlying these concepts are described and recommendations given for reporting results.

W. H. Kruskal,
Some Remarks on Wild Observations,
Technometrics 2, 1-3 (1960).
A simple, down to earth discussion of outliers, (some of which are extremely important in providing clues to systematic errors) with practical suggestions as to when results should be included or rejected from a series of measurements or observations. Eight references which will lead the reader to most of the relevant literature on the subject of outliers, rejection criteria, etc. are given.

H. H. Ku,
Notes on the Use of Propagation of Error Formulas,
J. Res. Natl. Bur. Stand. 70C, 263-73 (1966).
An expository review of propagation of error, theory, and practices in light of current practice and theory. Examples such as the accuracy of approximations and the reporting of the uncertainties of final results are discussed.

H. H. Ku,
Expressions of Imprecision, Systematic Error, and Uncertainty
Associated with a Reported Value,
Measurements and Data 72-77 (1968).
Three tables giving recommended imprecision, systematic error, and uncertainty statements based on a reported value and the index or measure of error. Highly recommended as a good summary, in practical form, of how reported values and their uncertainties should be stated.

H. H. Ku, (ed.),
Precision Measurement and Calibration - Statistical Concepts and
Procedures,
Nat. Bur. Stand. Spec. Pub. 300, Vol. 1,
US Govt. Printing Office, Washington (1969).
The one single most comprehensive volume on measurement in all its complexity. The 400 plus pages are in 7 sections: (1) The measurement process - precision, systematic error and accuracy; (2) design of experiments in calibration; (3) interlaboratory tests; (4) functional relationships; (5) statistical treatment of measurement data; (6) miscellaneous topics; and (7) abstracts of recent publications.
Although over 90 per cent of the papers collected in this volume are contributions of statisticians employed at the US-NBS, the treatment is broad in scope and not narrow like many statistical treatises. The authors do not exist in a pure statistical world, but are ever drawn back to a pragmatic reality through their interactions with the measurement scientists of the NBS.
Also included are extensive references to the work of other measurement scientists. The period that most of the work reported covers are the two and a half decades 1940-1965.

Zh. F. Kudrjashova, S. K. Rabinovich, and K. A. Reznik,
Recommendations of Methods of Evaluation of Measuring Results on
Direct Measurements,
Izdatjelstvo standartov, Trudy metrologicheskikh institutov,
edit. 134 (194) (1972), (in Russian).
This paper deals with the fundamental statistical methods for the critical evaluation of
measurements and for the characterization of accuracy and precision. Statistical methods
are given for different cases, for example, for the checking of normal and other
distributions of results, for homogeneity of different groups of measuring results, and
for the discovery and elimination of systematic errors.

Zh. F. Kudrjashova and S. G. Rabinovich,
Methods of Evaluation of Measuring Results on Indirect
Measurements,
Izdatjelstvo standartov, Trudy metrologicheskikh institutov,
edit. 172 (232) (1974), (in Russian).
Indirect measurements are characterized by the fact that between the final quantity x and
the measured quantities $x_1....x_m$, there can exist a known relationship x =
$f(x_1....x_m)$. For those cases this paper contains statistical methods for the critical
evaluation of such measurements.

KVB Kassenärztliche Vereinigung Bayerns,
Die Qualitätssicherung der Quantitativen Klinisch-Chemischen Untersuchungen,
München (1980) (in German).
This publication gives rules for the quality control of clinical chemical analyses as well
as methods for reporting accuracy control and precision control in analytical work in
clinical laboratories.

T. S. Kuhn,
The Function of Measurement in Modern Physical Sciences,
ISIS 52, 151-93 (1961).
A broad ranging essay on the nature of the measurement process and its role in the
aquisition of knowledge.

P. D. LaFleur, (ed.),
Accuracy in Trace Analysis,
Nat. Bur. Stand. Spec. Pub. 422, Vol. I and II,
US Govt. Printing Officer, Washington (1976).
This book is the formal report of the proceedings of the 7th Materials Research Symposium
on Accuracy in Trace Analysis. The volume contains papers presented at the Symposium
which discuss sampling, sample handling and analytical methodology. Many important
techniques and methods are described, and extensive references are presented in order to
give deeper insight into the problems of obtaining accurate results in trace analytical
chemistry. Accordingly, this volume should not only stimulate greater interest in
research in these areas but should provide a valuable guide for everyday analytical
problems. These proceedings include the following papers which discuss or treat the
problem of accuracy in chemical analysis: A. C. Kilbye, The Need for Accuracy in a
Regulatory Agency, pp 3-8; R. G. Lewis, Accuracy and Trace Organic Analyses, pp 9-34; J.
H. Boutwell, Accuracy and Quality Control in Trace Element Analysis, pp. 35-40; J. P. Cali
and W. P. Reed, The Role of NBS-SRM in Accurate Trace Analysis, pp. 41-63; G. H. Morrison,
Interpretation of Accuracy of Trace Element Results in Biological Materials, pp. 65-77; F.
P. Byrue, The Analyst and Accuracy, pp 123-26; K. Heydour, Detection of Systematic Errors
by the Analysis of Precision, pp 127-39; D. E. King, Detection of Systematic Error in
Routine Trace Analysis, pp. 141-50; L. Gonski, et al., The Estimation of Accuracy in Trace
Analysis, pp 189-98; R. K. Skogerboe and S. R. Koirtyohann, Accuracy Assurance in the
Analysis of Environmental Samples, pp 199-210. The above papers are those that address
themselves most directly to problems of accuracy, systematic error, and uncertainty.
Numerous other papers touch on these matters. The entire work is highly recommended.

W. H. Lawton, E. A. Sylvestre, and B. J. Young-Ferraro,
Statistical Comparison of Multiple Analytic Procedures:
Application to Clinical Chemistry,
Technometrics 21, 397-409 (1979).
Abstract: "The basic sciences all require an ability to measure the amounts of substance
under study. With new methods of measurement constantly being proposed there is a need for

techniques for comparing these methods in terms of precision and accuracy. Of particular
interest is the case where none of the individual methods are known to measure "truth". A
multiple methods comparison technique for this case is proposed in this paper, and is
illustrated by an example from the field of clinical chemistry. Estimates of the
components of variance for each method are developed and some of their properties
explored."

C. Liteanu and I. Rica,
Statistical Theory and Methodology of Trace Analysis,
Ellis Horwood Ltd, Chichester (1980).
A thorough treatment of the principles of quantitative physico-chemical measurement and
the statistical aspects of both random and systematic error.

P. H. Lloyd,
A Scheme for the Evaluation of Diagnostic Kits,
Ann. Clin. Biochem. 15, 136-45 (1978).
A technical guide to the statistical evaluation of analytical imprecision and accuracy for
the working clinical analyst.

G. E. F. Lundell,
The Chemical Analysis of Things as they Are,
Ind. and Eng. Chem. (Anal. Ed.) 5, 221-25 (1933).
A down-to-earth discussion of chemical analysis in pragmatic situations. One of the first
references as to how the accuracy of standard samples (now called reference materials) is
established. He says this "most probable value is a ...long story that can be summarized
by the statement that it is based on experience, on the work of others in the field, and
usually on determinations made by as many fundamentally different methods as possible."
Lundell also discussed accuracy of results in very pragmatic terms, insisting that the
accuracy of the results of chemical analyses should be in harmony (in terms of effort,
cost, timeliness, etc.) with end-use requirements.

J. Mandel,
The Measuring Process,
Technometrics 1, 251-67 (1959).
Abstract: "This paper deals with the theory of a proposed method for the statistical study
of measuring processes. The practical aspects of the method, including computational
details, are discussed in a companion paper published in the ASTM Bulletin 239 (1959). In
the present article a theoretical framework is proposed for the mathematical expression of
the sources of variation in measuring methods and a suitable method of statistical
analysis is described. Particular attention is given, both here and in the companion
paper, to interlaboratory studies of test methods. An illustration based on data taken
from the chemical literature is appended."

J. Mandel,
Statistical Analysis of Experimental Data,
John Wiley, New York (1964).
Offers experimental scientists an appreciation of the statistical approach to data
analysis. Most of the examples cited are based on actual results drawn from real
experimental and laboratory situations. In Chapter 6 the precision and accuracy of
measurements and experimental error (both random and systematic) are discussed.

J. Mandel,
The Evaluation of Referee Methods in Clinical Chemistry,
Med. Instrum. 8, 26-29 (1974).
Referee methods are now called reference methods and are methods of demonstrated accuracy.
This paper reviews the results of the calcium reference method (the first in clinical
chemistry) suggested by Cali and coworkers from the view point of the statistician. The
importance to health of clinical testing demands that both precision and accuracy be
achieved. Proficiency testing has shown that generally this is not the case. The next
task is to study some individual clinical methods in depth to provide the profession with
well developed referee methods. An interlaboratory study of a test method involves five
elements; the protocol, samples, laboratories, statistical design, and analysis of the
results. These elements are discussed and illustrated in terms of determining calcium in
serum by atomic absorption spectrometry, using isotope dilution mass spectrometry as a

standard of accuracy. This study, consisting of five exercises, shows that only through constant vigilance and an attitude of real concern can acceptable levels of precision and accuracy be achieved.

J. Mandel,
The Analysis of Interlaboratory Test Data,
Stand. News 5, 17-20 (1977).
Precision and accuracy, as parameters for the evaluation of test results, are discussed. The precision is shown to involve both within- and between-laboratory variability. Therefore a proper evaluation of a test method requires that a properly designed interlaboratory study be conducted.
This paper deals with the design and analysis of the results of interlaboratory studies. It is shown that the conventional two-way analysis of variance can lead to erroneous values for the within- and between-laboratory components of variance. It is strongly recommended that the data be analysed level by level, that is, separately for each material or level, by means of a one-way analysis of variance (within-between analysis). This method leads to unbiased estimates for the precision components and allows the latter to be evaluated as functions of the level. A numerical example is included.

J. Mandel and I. Lashof,
The Interlaboratory Evaluation of Testing Methods,
ASTM Bulletin No. 239 (TP133) (1959).
Abstract: "The various sources of variability in test methods are examined and a new general scheme to account for them is proposed. The assumption is made that systematic differences exist between sets of measurements made by the same observer at different times or on different instruments or by different observers in the same or different laboratories and that these systematic differences are linear functions of the magnitude of the measurements. Hence the proposed scheme is called the linear model. The linear model leads to a simple design for the round-robin tests but requires a new method of statistical analysis, geared to the practical objectives of a round-robin. The design, analysis, and interpretation of a round-robin in accordance with the linear model are presented, and the procedure is illustrated in terms of the results obtained in an interlaboratory study of the Bekk smoothness tester for paper. It is believed that this approach will overcome the frustrations that are often associated with the interpretation of round-robin test data."

J. Mandel and F. J. Linniq,
Statistical Methods in Chemistry,
Anal. Chem. 28, 770-77 (1956),
Anal. Chem. 30, 739-47 (1958).
These two comprehensive review articles cover the literature on the subject named over the years roughly 1951 through 1957. The years prior to 1951 are covered in reviews by Wernimont, Anal. Chem. 21, 115-20 (1949) and Hader and Youden, Anal. Chem. 24, 120-24 (1952). In both these reviews a section on accuracy and precision of methods and equipment cites many references, most of which deal with specific analytical situations or procedures.

J. Mandel and F. J. Linniq,
Study of Accuracy in Chemical Analysis using Linear
Calibration Curves,
Anal. Chem. 29, 743-749 (1957).
Abstract: "In situations characterized by linear calibration curves such as the relation between found and added in studies of accuracy in chemical analysis, the usual method for deriving confidence intervals for the slope and the intercept of the fitted straight line may lead to erroneous conclusions. The difficulty results from the interdependence of multiple conclusions drawn from the same data, especially when there is a strong correlation between the parameters involved. The method of joint confidence regions eliminates these difficulties and has the further advantage of allowing for the evaluation of the uncertainty of the calibration line as a whole, as well as of any values or functions of values derived from it."

J. Mandel and L. F. Nanni,
Measurement Evaluation,
Chapter 4 in Quality Assurance Practices for Health Laboratories,
S. L. Inharn, (ed.), Amer. Publ. Health Assoc., Washington (1978).
This chapter deals with basic statistical concepts and their application to the evaluation of health related measurements. It covers, amongst other topics, the evaluation of

precision and accuracy, using point and interval estimates, statistical distribution functions, straight line fitting, the evaluation of diagnostic tests, and control chart techniques.

H.-J. v. Martens and E. Pippig,
Beschreibung des Fehlers eines korrigierten Messergebnisses,
Feingeraetetechnik 28, 359-64 (1979), (in German).

H.-J. v. Martens,
Vergleichende Betrachtungen uber Verfahren zur Beschreibung
der Messfehlers (Teil 1 und 2),
Feingeraetetechnik 30, pp. 443 and 514 (1981), (in German).
This article analyses different methods and procedures for the description of measuring errors. Also included are procedures for the determination of approximate values for the confidence limits of the error and its random and nondetectable systematic part.

G. C. Martorelli and A. Zanini,
Expression for the Combination of Measuring Errors by
Means of a Normal Distribution.
Alta Frequenza 44, 629-35 (1975), (in Italian).
A proposal is presented for expressing the resulting error from the combined effect of various error sources of different types into an easy to calculate parameter that has a precise statistical meaning.

Y. Mashiko and K. Iisuka,
The Precision and Accuracy of Measurements,
Kagaku Sosetsu 10, publ. Todai Shuppan Kai (1976), (in Japanese).
Fundamental concepts of precision and accuracy are explained and defined in an attempt to remove the confusion that surrounds these concepts. Errors in measurements are analysed by means of statistical mathematics, and descriptions of errors are discussed.

E. F. McFarren, J. R. Lishka, and J. H. Parker,
Criterion for Judging Acceptability of Analytical Methods,
Anal. Chem. 42, 358-65 (1970).
A procedure is described for calculating the so-called total error of a chemical analytical method. In order to use the true value in the calculations, the authors make the assumption that the mean value calculated from the results from many laboratories presumably more truly represents only the method bias (accuracy). This process represents an attempt to introduce true values into uncertainty evaluation but is based on a premise that has been shown to be risky at best.

A. G. McNish and J. M. Cameron,
Propagation of Error in a Chain of Standards,
IRE Trans. Instr. I-9, 101-104 (1960).
A discussion on the concept of traceability in the chain of measurements and comparisons whereby assurance is gained that a manufactured measuring device is in consonance with a national standard. The authors demonstrate that the first question to ask in such a proposed traceability chain exercise is "to what end-purpose is the measurement to be used and what accuracy is required for such an end-purpose."
The main body of the paper discusses in detail how error is propagated during the process of comparing secondary standards with primary ones, or in the calibration of an instrument by one or more standardizing laboratories.

A. Michalik,
Instrumental Errors of the UV and Vis Spectrophotometric Measurements
and Methods of their Evaluation. Part VI. Ch. 2 in the Monograph
Complex use of the Molecular Spectroscopy in Chemical Analysis,
(Kompleksowe stosowanie metod spektrometrii czasteczkowej w analizie
chemicznej),
Wydawnictwo Ossolineum, PAN, Warszawa (1977), (in Polish).
The sources and characteristics of instrumental errors caused, for example, by nonlinearity of the photometric scale, stray light, spectral slit width, defects of cuvettes, etc., are described. Methods are recommended for the evaluation of the accuracy of the wavelength and absorbance scales by referring them to selected certified reference materials.

J. W. Mueller,
Les Incertitudes de Mésures,
dans *La Physique*, Vol. 4, pp. 11-17,
    Encyclopédie Scientifique de l'Univers, Gauthier-Villars, Paris (1981), (in French).
Reviews the problems in the evaluation of uncertainty in experimental results as well as
types of error and propagation of errors.

J. W. Mueller,
Un Nouveau Regard sur les probabilitiés à priori,
    Bureau International des Poids et Mesures, Rapport BIPM 80/6 (1980), (in French).
The well-known Bayes theorem which establishes a probability link between observed random
events and their possible causes allows us, for a series of measurements, to dispose at
any time of an updated knowledge of the parameter Θ which characterizes the underlying
distribution law.
Previous discussion has been mostly centered on the assignment of the *a priori* density
$f(\Theta)$, and in particular, for the situation where nothing is known about Θ, it was usually
concluded that there does not exist a well defined form of $f(\Theta)$ corresponding to complete
ignorance. However, a paper by E. T. Jaynes shows that this is not the case. It
describes in some detail Jaynes' analysis which leads, for the two important cases of a
Poisson and a binomial variate, to an unambiguous determination of their respective prior
distributions. They both differ markedly from a constant probability density which is
sometimes assumed to describe ignorance.

J. W. Mueller,
Some Second Thoughts on Error Statements,
Nucl. Instr. Mech. 163, 241-51 (1979).

J. W. Mueller,
The Assignment of Uncertainties to the Results of Experimental Measurements,
Second International Conference on Precision Measurement and
Fundemental Constants, Gaithersburg, (1981).

R. B. Murphy,
On the Meaning of Precision and Accuracy,
Materials Res. and Stand. 4, 264-67 (1961).
A short, but cogent, discussion on the meaning of precision and accuracy. The author
makes clear the differences between the two schools of thought concerning accuracy, and
makes recommendations as to measures of precision and accuracy.

V. V. Nalimov,
The Application of Mathematical Statistics to Chemical Analysis,
    Pergamon, Oxford (1963).
The book, a translation from the Russian original, deals with the application of the
methods of mathematical statistics to chemical analysis. It is intended to be used in
analytical laboratories. The main emphasis is therefore on examples taken from chemical
analysis and the appropriate mathematical statistics are given. Theoretical problems are
considered in so far as they are necessary for an understanding of the metrological
aspect.

National Physical Laboratory of India,
Standards and Industrial Research Institute of Malaysia,
Guidelines for estimation and statement of overall uncertainty
in measurement results,
CSC (80) MS-9, Commonwealth Science Council, London (1980).

M. G. Natrella, (ed.),
Experimental Statistics Nat. Bur. Stand. Handbook 91,
    US Govt Printing Office, Washington (1963).
This handbook of 504 pages is intended for use by scientists with an engineering
background who, "although he has an occasional need for statistical techniques, does not
have the time or inclination to become an expert on statistical theory and methodology."
Particularly recommended for scientists without extensive formal training in statistics,
the handbook provides step-by-step instructions for reaching a certain goal together with
the conditions necessary for the validity of a particular procedure. Particularly
attractive is the manner in which the material is laid out; the left hand pages provide

theory, generalizations, and formulae and the right hand pages follow in parallel but gives actual examples.

G. Olofsson
Assignment of uncertainties
Chapter 6 in Experimental Chemical Thermodynamics, Vol. 1,
Combustion Calorimetry (editors, S.Sunner, M. Månsson)
Pergamon, Oxford (1979).

T. Plebanski (ed.),
Physicochemical Certified Reference Materials,
Wydawnictwa Normalizacyjne, Warszawa (1979).
This book is in two parts: Part 1 (in Polish) is a description of certification procedures used by the Centralny Ośrodek Badawczo-Rozwojowy Wzorców Materiałów (WZORMAT), and Part 2 (in Polish and English) is a listing and description of Reference Materials produced by WZORMAT.
Part I presents, in a systematic way, the various mathematical models used for calculating the accuracy and precision of the reference materials (RM) produced by WZORMAT. The propagation of errors through the calibration chain begins with the base units of the SI and then follows the chain: (1) the basic standards of the derived SI units, including primary RMs; (2) high accuracy of physicochemical RMs and associated instruments; ending at (3) commercially certified RMs of an accuracy sufficient for the intended end-use. Various statistical treatments for combining systematic and random errors, for treating the influence of outside conditions, and for combining the results of two different methods are discussed.

P. E. Pontius,
The Measurement Philosophy of the Pilot Program for Mass Calibration,
Nat. Bur. Stand. (USA) Tech. Note 288, US Govt Printing Office,
Washington (1966, reprinted 1968).
Abstract: "The Pilot Program for mass measurement is the result of a consideration in which the values produced are thought of as the products of a mass measurement process. The collective performance of elements of the mass measurement process results in establishing the process precision which, under certain conditions, can be described quantitatively by pertinent performance parameters. The uncertainty attached to the product of the process, the measured value, is computed from these parameters and reflects the total performance of the process rather than the immediate measurement which might have produced the value. Interpretations of uncertainty and surveillance tests are discussed. The Pilot Program in mass measurement, whereby suitable process performance parameters can be established for precise mass measurement processes in other facilities, is discussed."

P. E. Pontius,
Notes on the Fundamentals of Measurement and Measurement as
a Production Process,
Nat. Bur. Stand. Int. Rep. 74-545 (1974); NTIS as COM 74-11656.

R. Puschel,
Zum Problem der "Genauigkeit" chemischer Analysen,
Mikrochemica Acta, 783-801 (1968), (in German).
Includes a discussion regarding the term accuracy and the difficulties arising from its use. With the help of practical examples he shows that the accuracy of analysis depends on the content of the constituent in a given sample. He indicates ways of overcoming the various problems described.

B. D. Reeves and D. W. Calhoun,
Reliability Criteria for Saturation Analysis of Steroids by
Competitive Protein Binding,
Acta. Endocrin. 64, suppl. 147, 61-78 (1970).
This paper discusses the reliability of competitive protein binding assays in terms of the criterion of accuracy, through a consideration of the properties of the measurement system, the assay design and statistical control.

C. B. Reimer and S. C. Maddison,
Standardisation of Human Immunoglobulin Quantitation:
A Review of Current Status and Problems,
Clin. Chem. 22, 577-82 (1976).
An accessible review of the problem of accuracy in the context of quantifying molecularly heterogeneous analytes.

F. D. Rossini and W. E. Deming,
The Assignment of Uncertainties to the Data of Chemistry and
Physics, with Specific Recommendations for Thermochemistry,
J. Wash. Acad. Sci. 29, 416-41 (1939).

E. B. Sandell,
Errors in Chemical Analysis, in Treatise on Analytical Chemistry,
I. M. Kolthoff and P. J. Elving, (eds.), Part I, Vol. I, Chapter 2, pp. 19,
Wiley, New York (1959).
A clear exposition of the kinds of systematic (called determinate by the author) errors likely to be encountered in chemical analysis. They are: errors inherent in the analytical method itself, operative errors (commonly called personal error), and instrumental errors. A second section deals with indeterminate errors (random) and these are sub-divided into randomly varying fluctuations in systematic errors, and those truly random (i.e., causes not known or uncontrollable). A third section discusses the determination of accuracy by either (1) absolute methods of analysis (each source of known or expected systematic error is investigated and either eliminated or its magnitude determined), or (2) via the comparative method where the method under test is evaluated using materials certified to be of known (true) composition. These materials (certified reference materials) have usually been analyzed via absolute methods.
The chapter concludes with sections on (1) accuracy sought and attainable in chemical analysis; (2) testing the accuracy of analyses; (3) methods for improving the accuracy of analysis.

R. Seward, (ed.),
Standard Reference Materials and Meaningful Measurements,
Nat. Bur. Stand. Spec. Publ. 408, US Govt Printing Office, Washington (1975).

W. A. Shewhart,
The Economic Control of Quality of Manufactured Product,
Van Nostrand, New York (1931).
A key text by the father of statistical quality control concepts. With Skewhart began many of the modern techniques in use today to assure that measurement systems were in a state of statistical quality control. While extremely well written throughout, of special interest to the study of the subject of accuracy is Part VII, Quality Control in Practice.
In the four chapters comprising this part the following are discussed: (1) a summary of fundamental principles, including a discussion of probability, maximum likelihood, and the empirical method; (2) problems of sampling and measurement with a good discussion of various types of error and how each should be handled; (3) sampling; and (4) a resume of the control program - including control of measurements made in research environments.
A bibliography, annotated and extensive and covering much pertinent literature up to 1931, is appended and it is of great value for researching the history of the accuracy problem.

W. A. Shewhart,
Statistical Method from the Viewpoint of Quality Control,
The Graduate School, US Dept. of Agriculture, Washington (1931).
In this book Skewhart, the originator of the concept of statistical control, gives the philosophical basis of how measurement results are to be presented, and the specification of precision and accuracy.

S. S. Stevens,
in Measurement, Definitions and Theories, C. W. Churchman and R. Philburn (eds.),
John Wiley, New York (1959).

P. W. Strike,
Medical Laboratory Statistics,
Wright - PSG, Bristol (1982).
A review of statistical principles and practice for clinical laboratory analysts with particular emphasis on laboratory measurement. The subject of accuracy in measurement is

reviewed in the context of assay comparison studies under the assumptions of structural
error-in -variables model with the support of fully worked examples.

Student,
The Probable Error of a Mean,
Biometrika, 6, 1-25 (1908).

Student,
Errors of Routine Analysis,
Biometrics, 19, 151-64 (1927).

J. W. Tukey,
Conclusions vs. Decisions,
Technometrics 2, 423-33 (1960).
An account of an interesting talk that points out clearly the differences between the
validity of an experimenter's conclusions and those of a statistician, both examining the
same data or set of observations. Systematic errors, or even blunders, and their
correction or removal, must be the responsibility of the experimenter. Errors of this
type are not to be charged against statistical conclusions. The difference between the
statistician's true value and that of the experimenter is clearly distinguished and
results (i.e. conclusions vs decisions) of this understanding made clear.

G. A. Uriano and C. C. Gravatt,
The Role of Reference Materials and Reference Methods in Chemical
Analysis,
Crit. Rev. Anal. Chem. 4, 361-411 (1977).
A comprehensive review of the role reference materials and methods play in assuring
compatible measurement systems. Basic terms are defined and illustrated and arguments
given to show how accuracy in measurement leads to measurement compatibility. There are 6
chapters. Some important discussions treat: (1) transferring accuracy throughout large
measurement networks; (2) definition of terms and concepts; (3) a systems approach to
measurement compatibility; (4) criteria and guidelines for the development of reference
materials and methods; (5) the application of reference materials and methods to large
measurement networks.
A comprehensive bibliography of 169 citations is included.

S. R. Wagner.
On the Quantitative Characterization of the Uncertainty of
Experimental Results in Metrology,
PTB - Mitteilungen, 89, 83-9 (1979).

S. R. Wagner,
Combination of Systematic and Random Uncertainties,
Conference on Precision Electromagnetic Measurements,
Braunschweig, June 1980.

G. Wernimont,
Design and Interpretation of Interlaboratory Studies of Test
Methods,
Anal. Chem. 23, 1572-76 (1951).

A. L. Wilson,
The Performance-characteristics of Analytical Methods,
(Note: in four parts over five years as referenced)
Talanta 17, 21 (1970);
Talanta 17, 31 (1970);
Talanta 20, 725 (1973);
Talanta 21, 1109 (1974).

E. B. Wilson, Jr.,
An Introduction to Scientific Research, Chapters 7,8,9,
McGraw-Hill, New York (1952).
A fine introductory text on the scientific method in all its aspects including the choice

and definition of the problem to be investigated or solved, the planning and execution of the experimental work, and data evaluation and its reporting.

A. G. Worthing and J. Geffner,
Treatment of Experimental Data
Wiley, New York (1943).
A chapter on the representation of data by tables, graphs and equations is followed by a discussion of normal frequency distributions, means, precision indexes and their propagation, least-squares fitting of straight lines and polynomials, and criteria for closeness of fit.

W. J. Youden.
Youden was one of the most prolific, readable, and clearheaded writers on questions of accuracy, uncertainties, and systematic error analysis, especially as applied to chemical analyses. Following this entry are several of his papers which have been annotated. Here we list simply several additional references for the reader interested in pursuing the development of Youden's thinking on these topics.

(1)   Locating Sources of Variability in a Process,
       Ind. Eng. Chem. 43, 2059-62 (1951).

(2)   Statistical Methods for Chemists,
       Wiley, New York (1951).

(3)   Sets of Three Measurements,
       Sci. Monthly 77, 143-47 (1953).

(4)   Measurement made by Matching with Known Standards,
       Technometrics 1, 101-09 (1959).

(5)   Statistical Design,
       Ind. and Eng. Chem. (1954 to 1959 bimonthly articles, reprinted)
       ACS Applied Publ., Washington (1960).

(6)   What is the Best Value?,
       J. Wash. Acad. Sci. 51, 95-97 (1961).

(7)   Systematic Errors in Physical Constants,
       Phys. Today 14, 32-34 (1961); also in
       Technometrics 4, 111-23 (1961).

(8)   Uncertainties in Calibration,
       IRE Trans. Instr. I-11, 133-38 (1962).

(9)   Realistic Estimates of Errors in Measurements,
       ISA Journal 9, 57-58 (1962).

W. J. Youden,
Technique for Testing the Accuracy of Analytical Data,
Anal. Chem. 19, 946-50 (1947).
Abstract: "A new or modified quantitative analysis procedure is usually tested over a range of sample size or amount of substance sought. Frequently such data, when plotted, should give a straight line through the origin. Several examples of published analytical data have been utilized to illustrate a statistical technique for determining whether the line may be considered to pass through the origin and to point out the advantage of using the slope of the line as a check on the accuracy of the analytical method".

W. J. Youden,
The Sample, the Procedure, and the Laboratory,
Anal. Chem. 32, 23A-37A (1960).
A practical, down-to-earth discussion of the various components comprising a measurement laboratory, and what steps need to be taken to ensure reliable analytical results within and between laboratories. The author concludes with: "Very careful efforts on analytical work are associated with atomic weight determinations and with the work on standard samples or reference materials. The approach here is chemical rather than statistical. Using every iota of available chemical information elaborate precautions are taken to eliminate, or correct for, every possible source of systematic error. Comparatively little dependence is placed upon repeat determinations. Here the chemist supplies his own

testimony to support the position taken in this paper. Systematic errors are the real
headache. If enough care is taken, or alternative procedures are employed, the systematic
error can be greatly reduced. By such means atomic weights and standard samples gain
acceptance. In the ordinary work of analytical chemistry, most of these precautions are
not feasible. Nevertheless the goal of general agreement among laboratories, using a
procedure with a very small bias, is the task of the analytical laboratories. To achieve
their goal, the laboratories must get the right kind of data and interpret them properly".

      W. J. Youden,
      Accuracy and Precision: Evaluation and Interpretation of Analytical Data,
      Treatise on Analytical Chemistry, M. I. Kolthoff and P. J. Elving (eds.),
      Part 1, Vol. 1, Chapter 3, 47-66, Wiley, New York (1959).
Youden has very clear ideas as to the differences between accuracy and precision and the
limitations of statistics to systematic error analysis. In this chapter there is first a
discussion of precision and measures for precision and their use in chemical analysis.
Following is a section dealing with the detection and evaluation of the magnitude of
systematic errors through the use of either materials of known composition, or the use of
two or more different analytical procedures. The procedure and the simple statistics used
to demonstrate whether systematic error is present are clearly shown.

      W. J. Youden,
      How to Evaluate Accuracy,
      Mat. Res. and Std. 1, 28-71 (1961).
The errors present in a measurement are broken down into (1) the systematic error inherent
in the procedure; (2) the locally produced systematic error introduced by a laboratory
using the procedure, and (3) the random error. Recognizing these three sources of error
should help attain better measurement accuracy. Several methods whereby sources of error
may be identified are given. Three of these are: measurement (or use of materials of
known properties; comparison with other measurement procedures; and comparison with
modifications of the given procedure. Examples are given.

      W. J. Youden,
      Statistical Technique for Collaborative Tests,
      Association of Official Analytical Chemists, Washington (1967).
This manual presents statistical and other advice and techniques that may be used in
collaborative testing of analytical methods. Youden points out that the proper
interpretation of the results of a collaborative test involves consideration of three
sources of error: random error; inherent systematic error in the procedure; and the
modification of the systematic error by each of the participating laboratories, as a
result of its own equipment, personnel and environment. He shows further how the proper
use of well-characterized reference materials potentially can cancel the second and third
of the above-mentioned errors.