Aarne Mämmelä, version 1.1, 27.10.2007

The book is perhaps the only modern extensive book about research methods in physical sciences and engineering. The author is an emeritus professor of artificial intelligence from the George Washington University. The layout of the text looks nice. The illustrations of Bettina Scheibe are excellent and amusing. The first part of the title (“Getting It Right”) may unnecessarily distract some of the serious readers. The front cover could be improved to show that the content is serious. The book is attached with some errata sheets. Some of them are delivered by the author.

There are only a few similar books and they are either quite brief, for example [Goddard04], or outdated, for example [Wilson90]. There are some similar books for social sciences [Singleton99], but the research methods are somewhat different although there is some overlap. Many books claimed to be written about research methods are actually writing instructions. A closely related area is philosophy of science [Rosenberg00], [Niiniluoto02], [Niiniluoto83]. The handbook [Honderich05] also includes several useful articles on philosophy of science.

The preface of the author is missing, therefore for example the target audience and the structure of the book are not presented and they must be deduced from the table of contents. The book includes a bibliography, vocabulary, and summaries of hints. Each chapter could include some bibliographical notes to guide the reader to the literature. One rare reference to the literature in the text is on page 142.

Some important topics have not been given a deep enough discussion. For example brainstorming is not mentioned at all. Other topics that are only briefly covered include causality (pages 8 and 200), creativity (page 64), reductionism (page 64), and deduction (pages 79 and 89). Scientific or incomplete induction and its limitations, called Hume’s problem of induction, are not discussed at all. Mathematical induction, which is a form of complete induction, is mentioned on pages 79 and 165 without separating it from scientific induction. The problem of induction is related to the question of verification and falsification by Karl Popper who is briefly mentioned on page 167. Empirical verification (page 230) is always based on some kind of inductive reasoning and is therefore fallible. Thomas Kuhn’s important concept of paradigm is only briefly mentioned on page 74. Kuhn himself is mentioned on page 167.

The author has in general done good work for unification of terminology and corresponding definitions. Similar unification has been done by various standardization organizations, and they have sometimes different opinions. We should perhaps look at the standards of some international organizations as the International Standards Organization (ISO). It has the authority especially in measurement terminology.
The three terms 1) performance metric (page 215), 2) performance criterion (page 187), and 3) performance measure (page 217) are very important. However, I recommend, respectively, the terms 1) performance metric or, synonymously, performance criterion or performance measure, 2) performance requirement, and 3) performance value or briefly performance. The terms should be defined together in one place. The terminology should be unified. The author is using the terms performance measure (page 217), measure of performance (page 88), performance value (page 169), and value of performance metric (page 114) synonymously although this is not clearly stated anywhere.

In systems engineering the term “requirement” is used instead of “criterion” [Hall62, p. 62], [Ulrich95, p. 55]. In estimation theory and telecommunications the term “criterion” refers to a metric [Kay93, p. 19], [Haykin01, p. 551]. In telecommunications the term “measure” also refers to a metric [Haykin01, p. 614]. In [ISO Guide 99:2004] the term “quantity value” is defined and thus the term “performance value” would follow this convention. The term “value” is also used in [Ulrich95, p. 55].

The selection of some other terms is also unconventional. Such terms include task domain (page 36), task range (page 67), and task unit (page 37). A task is compared to a mathematical function (page 35), which is misleading in the sense that a research task is a creative process that is not at all straightforward mapping from input to output as explained elsewhere in the textbook on page 64.

It is best to make a difference between axiomatic systems in formal sciences as mathematics (page 79) and those in empirical sciences as physics [Niiniluoto02, p. 211]. In formal sciences knowledge is independent of experience, but empirical sciences are based on observation and experiments. In formal sciences the axiomatic systems are not interpreted and are called Hilbertian axiomatic systems. Conjectures (page 47) are hypotheses [Honderich05, p. 411] that are derived or proved deductively from axioms, after which the conjectures become theorems. In empirical sciences the axiomatic systems are interpreted and are called hypothetico-deductive systems. Axioms are hypotheses, which are verified indirectly by deriving testable results in the hypothetico-deductive method (page 167). Models are systems that make the axioms true [Niiniluoto02, p. 211]. In practice most theories in empirical sciences are based on mathematical models [Rosenberg00, p. 96], [Honderich05, p. 617, 914]. If we compare science and mathematics, the use of the term “hypothesis” is upside down: in science the hypothesis is on the top (“model”, “theory”, “solution”, or “axioms” from which results are derived deductively) and in mathematics it is in the bottom (“conjecture” or tentative theorem to be proved or derived deductively from axioms).

The author has used the term “hypothesis” as an expected conclusion (page 205) whereas in science a hypothesis is the possible solution to the problem or an expected theory. The author’s idea of interpreting a hypothesis as a conclusion comes probably from mathematics where hypothesis is a conjecture, which is some kind of conclusion. We should rather use the terminology from science since tasks in engineering resemble science more than mathematics although both science and engineering are heavily based on mathematics. In science, engineering, and mathematics the coherence of the theory is important. However, mathematics is a
formal science where correspondence to reality is not relevant whereas in science and engineering it is crucial.

The hypothetico-deductive method is briefly mentioned on page 167. The author could not explain the term hypothetico-deductive method because in his terminology a hypothesis is a conclusion and the terms do not match in his case. The term hypothetico-deductive method is usually preferred by philosophers [Pagels88] to describe the research method, probably due to the fact that the creative hypothesis construction cannot be easily described. It is not a method of discovering hypotheses, but a method of verifying existing hypotheses. Discovery is not only simple induction but more like abduction, i.e., inference to the best explanation [Honderich05]. Abduction can be approximated by using strong inference based on competing hypotheses [Platt64].

Other unconventional uses of terms in the textbook include goal (page 203) and solution (page 63). The term objective (page 36) is correctly defined as a refined problem statement as in [Hall62, p. 9]. The terms with their conventional definitions are explained in Appendix I. The reading of the textbook is sometimes a challenge because of the unconventional terminology.

Some definitions look too abstract although in general the book is very practical. For example, the definition of the term “null hypothesis” is usually seen as an opposite to an alternative hypothesis. If the null hypothesis cannot be proved, the alternative hypothesis would be true. The null hypothesis is sometimes used to guarantee objectivity. After reading the author’s definition on page 338, I could not understand the meaning of the term. The definition includes the abstract term “decision space” (page 283) that is not defined elsewhere in the glossary. It is not even in the index. The definitions must be concrete enough to be understandable.

Sometimes I felt that the author is too wordy, especially when discussing the terms accuracy (page 112), precision (page 146), and uncertainty (page 119). The definitions and examples could be improved. In the ISO documents the term “trueness” refers to systematic errors, the term “precision” refers to random errors, and the terms “accuracy” and “uncertainty” to both of them [amc03], [ISO Guide 99:2004]. For trueness a reference value is needed. Some authors claim that “accuracy” refers only to systematic errors [Young02]. According to ISO, trueness and accuracy are qualitative concepts and no numerical value should be given to them. It is possible to give a numerical value to imprecision although this is not recommended in [Taylor94].

The most appropriate means of expressing the accuracy is in terms of uncertainty. If the systematic errors are removed, the standard deviation possibly multiplied with a coverage factor can be used to define uncertainty [Taylor94]. Uncertainty can also be defined with coverage interval and coverage probability [Taylor94], [amc03], [ISO Guide 99:2004]. In statistics coverage interval is called “confidence interval” and coverage probability is called “confidence level” [Kreyszig83]. The author is using similar terminology on page 125.

On page 111 the accuracy of 0.001% looks suspicious since one could believe that this is a very good accuracy, but after reading the text this accuracy is extremely poor.
Instead of accuracy, uncertainty should be given. The definition of precision on page 146 seems to be valid for example for the spatial precision of a monitor, but for a voltmeter a unitless precision of 1000 (page 147) looks unconventional.

On page 124 the author is using an arbitrary scaling factor \( s \). I feel that then the result is also arbitrary and the percentage does not have any physical meaning. The uncertainty values in Table 8.3 seem to be meaningless since by using a different scaling factor, we would obtain a completely different set of uncertainties. The percentage does not really say anything although the reader is led to believe so. By suitable scaling you could always have numbers close to 0% or alternatively close to 100%.

Section 10.3 “Investigate related work” on pages 191-193 looks quite brief. The topic is important but the search of existing knowledge is a challenge. The use of bibliographies, citations, and databases should be explained in detail. The outline for a document on page 305 follows the scientific method presented on page 168. However, in present scientific papers the introduction, materials and methods, results, and discussion (IMRAD) structure has become a standard after its invention in the 1850’s [Day98]. I feel that a project document and a scientific paper should be different from a research plan: a document should have a deductive top-down structure; research on the other hand is more like an inductive or rather abductive, bottom-up process. On page 306 the author recommends the bottom-up order for documents. However, in a scientific documents the opposite top-down or deductive order should be used. Bottom-up or inductive order is used in textbooks [Felder88].

The bibliography could be improved. I would expect general books about technical writing [Higham98], philosophy of science [Rosenberg00], and general systems theory [Checkland99]. I would also expect some general books about history of science and technology although it is difficult to say which should be preferred. Many books in the bibliography seem to be quite specific.

To summarize, the author, the illustrator, and the publisher have prepared a book that is tempting to read, and it should be on the bookshelf of all researchers in science and engineering. It is more practical than the corresponding texts about philosophy of science, but some of the terminology is unconventional and some definitions are too abstract. The terminology is further explained in Appendix I and Fig. 1.
Appendix I Explanation of some terminology

The format used in the explanations:

**term**  >  use of the term in the textbook (page number)
   definition elsewhere in the literature [reference]

**accuracy**  >  a term whose definition needs clarification (page 112)
   *accuracy* is the “closeness of agreement between a measurement result and the accepted reference value,” includes both systematic and random errors, no numerical values should be given (ISO term) [amc03], [ISO Guide 99:2004], see precision and trueness

**goal**  >  definition of an experiment (page 203)
   (in systems engineering) *goal* is “some new system matched to a certain environment” [Hall62, p. 74], the goal is the same as hypothesis and solution, see objective

**hypothesis**  >  expected conclusion from an experiment (page 205)
   (in philosophy of science) *hypothesis* is “a possible solution to a problem” [Honderich05, p. 411] or “a provisional theory” [Random House99], ideally a causal relationship but may be also correlation
   (in systems engineering) *hypothetical system* is the same as goal and solution [Hall62, p. 9, p. 74]
   (in mathematics) *hypothesis* is a conjecture or a possible theorem that has not yet been proved [Honderich05, p. 411]

**induction**  >  mathematical induction (pages 79 and 165)
   (in philosophy of science) *induction* or *scientific induction* is “any form of reasoning in which the conclusion, though supported by the premises, does not follow from them necessarily” [Random House99], [Honderich05, p. 432]
   (in mathematics) *mathematical induction* is a form of complete induction where all the special cases of the generalization are enumerated, used in mathematical proofs [Niiniluoto83, p. 25]

**meter-kilogram-second (MKS) system**  >  Systeme International d’Unites (SI)
   *system* (page 110)

**objective**  >  a term whose definition needs clarification (page 36)
   (in systems engineering) selecting *objectives* is “the logical end of problem definition” [Hall62, p. 9], i.e., an objective is a refined problem statement, see goal

**performance criterion**  >  performance requirement (page 187)
   (in estimation theory and telecommunications) *criterion* is the same as metric [Kay93, p. 19], [Haykin02, p. 551]
(in systems engineering) the term \textit{requirement} is used in [Hall62, p. 62], [Ulrich95, p. 55]

\textbf{performance measure} \Rightarrow \textit{performance, performance value} (page 217)

(in telecommunications) \textit{measure} is the same as metric [Haykin01, p. 614]

(in systems engineering) the term \textit{value} is used in [Ulrich95, p. 55]

\textbf{precision} \Rightarrow \textit{a term whose definition needs clarification} (page 146)

\textit{precision} is “the closeness of agreement between independent measurement results obtained under stipulated conditions,” includes random errors, no numerical values should be given (ISO term) [amc03], [ISO Guide 99:2004], see accuracy and trueness

\textbf{task domain} \Rightarrow \textit{materials and resources, inputs of a task}, see task unit (page 36)

(in mathematics) \textit{domain} is “the set of values for which a function is defined” [Weisstein05]

(in philosophy of science) the term \textit{materials} is used [Day98]

\textbf{task range} \Rightarrow \textit{results and conclusions, outputs of a task} (page 67)

(in mathematics) \textit{range} is “the set of all values that a function can take as its argument varies over the domain” [Weisstein05], see task domain

\textbf{task unit} \Rightarrow \textit{materials} (page 37)

(in philosophy of science) \textit{materials} are the set of objects and concepts that form the input for a task, excluding resources [Day98]

\textbf{solution} \Rightarrow \textit{method to find the solution, excluding experiments}, see goal (page 63)

(in systems engineering) \textit{solution} is an answer to a problem, the same as goal and hypothetical system [Li99], [Hall62, p. 9, p. 74]

\textbf{trueness} \Rightarrow \textit{a term not used in the textbook although it should be there}

\textit{trueness} is “the closeness of agreement between the average value obtained from a large series of measurement results and the accepted reference value,” includes systematic errors, no numerical values given (ISO term) [amc03], [ISO Guide 99:2004], see accuracy and precision

\begin{center}
\begin{tikzpicture}
\node (a) at (0,0) {Materials};
\node (b) at (2,0) {Methods};
\node (c) at (0,-1) {Problem};
\node (d) at (2,-1) {Solution};
\node (e) at (1,0) {Research task};
\draw[->] (a) -- (b);
\draw[->] (c) -- (d);
\end{tikzpicture}
\end{center}

Fig. 1 Some common terminology related to research methods.
References
