Teaching Measurement and Uncertainty the GUM Way

Andy Buffler and Saalih Allie, University of Cape Town, South Africa

Fred Lubben, University of York, United Kingdom

This paper describes a course aimed at developing understanding of measurement and uncertainty in the introductory physics laboratory. The course materials, in the form of a student workbook, are based on the probabilistic framework for measurement as recommended by the International Organization for Standardization in their publication *Guide to the Expression of Uncertainty in Measurement* (GUM).

If you have ever been involved in the design or teaching of introductory physics laboratories, then it is likely that you have been kept awake at night worrying about two things: how to make your laboratories relevant and interesting and how to reasonably deal with measurement errors. The first issue would have stimulated you to stay up late, designing new, exciting experiments. The second issue is likely to have kept you awake when you finally did try to sleep. The nightmarish landscape of error analysis in the introductory physics laboratory is littered with the rocks of well-defined procedures such as calculating standard deviations, and the potholes of rules of thumb. One wonders how students are expected to navigate through the typical laboratory course and emerge with a coherent understanding of the nature of scientific measurement and uncertainty.

What might allow you to rest more easily is the knowledge that similar concerns have been entertaining the experts who represent the national and international metrology bodies, such as the National Institute of Standards and Technology (NIST). It has long been recognized that relying on the traditional frame-work for the statistical analysis of data[1-3] (the so-called conventional or "frequentist" approach) presents both philosophical and technical difficulties. It is not surprising then that physics instructors have found the teaching of error analysis a headache, especially at the introductory level.

The fundamental difficulties intrinsic to the traditional approach to measurement,[4] together with the fragmented way that the formalism and terminology of measurement has been applied across different science disciplines, led the Bureau International des Poids et Mesures to review the situation with regard to calculating and reporting measurements and uncertainties.[5,6] These efforts, which started in the late 1970s, culminated in the 1990s with the issue of a set of recommendations and guidelines contained in the *Guide to the Expression of Uncertainty in Measurement* (GUM),[7] which have now been adopted by all international standards organizations including the International Union of Pure and Applied Physics and the NIST.[8]

A technical and didactical comparison between the traditional frequentist approach and the ISO-recommended probabilistic approach have been made in a companion paper.[4] In brief, one of the key features of the probabilistic framework[6] is that measurement is viewed as a problem of inference using probability theory to model knowledge claims based on the information at hand, in keeping with the Laplace-Bayesian approach to analyzing and interpreting data.[9] A number of excellent summaries and critiques of the ISO-GUM approach are now available.[6,10] The key issue to

Table I. Outline of the content of the interactive student workbook.

Unit	Description
1. Introduction to measurement	– The relationship between science and experiment. – The nature and purpose of measurement.
2. Basic concepts of measurement	– Probability and inference. – Reading digital and analog scales. – The nature of uncertainty. – A probabilistic model of measurement.
3. The single reading	– Probability density functions. – Representing knowledge graphically using a pdf. – Evaluating standard uncertainties for a single reading. – The result of a measurement.
4. Repeated readings that are dispersed	– Dispersion in data sets. – Evaluating standard uncertainties for multiple readings. – Type A and Type B evaluation of uncertainties.
5. Working with uncertainties	– Propagation of uncertainties. – Combined standard uncertainty. – The measurement equation. – The uncertainty budget . – Comparing different results. – Repeatability and reproducibility.
6. Modeling trends in data	– Fitting of straight line functions to data.

Fig. 1. An extract from the workbook dealing with reading analog scales.

consider as educators is whether or not the methods of data analysis used in the physics teaching laboratory should mirror the methods being advocated for use at the research level. In addition, there is a growing body of research[11-14] that shows that even if students can demonstrate an adequate proficiency in carrying out the technical aspects of experimentation and data analysis, they seldom display appropriate understanding of the conceptual framework underpinning these procedures. We have argued[4] that this is, in part, a consequence of using the frequentist framework of measurement with its inherent limitations and logical inconsistencies. One of the reasons we believe that the approach advocated by the GUM has not made its way into the undergraduate curriculum on a large scale to date is that it is written in a technical fashion that does not lend itself to direct use at the freshman level.

To this end we have designed and written a set of materials[15] based on the framework of measurement and uncertainty as specified by the GUM[7] that can be used at the introductory level. The materials combine our research findings relating to students' understanding of measurement[14,16] with the main ideas of the probabilistic framework for measurement in a structured way. The broad content areas in the workbook are listed in Table I. Students can work through the activities in the workbook either alone or in small groups in a tutorial-type environment, and should have additional "hands-on" laboratory activities supporting the ideas about measurement. The course was piloted in the physics department at the University of Cape Town in 2002 and has been running since with minor modifications following various forms of evaluation.

The Course Materials

The workbook[15] introduces the concept of a "measurand" and the idea that a measurement always involves a comparison with a reference standard. Subsequent exercises explore the different purposes of measurement in both everyday and scientific contexts, and deal explicitly with the difference between a reading observed on a measuring instrument and the information that can then be inferred about the value of the measurand. One of the recommendations in
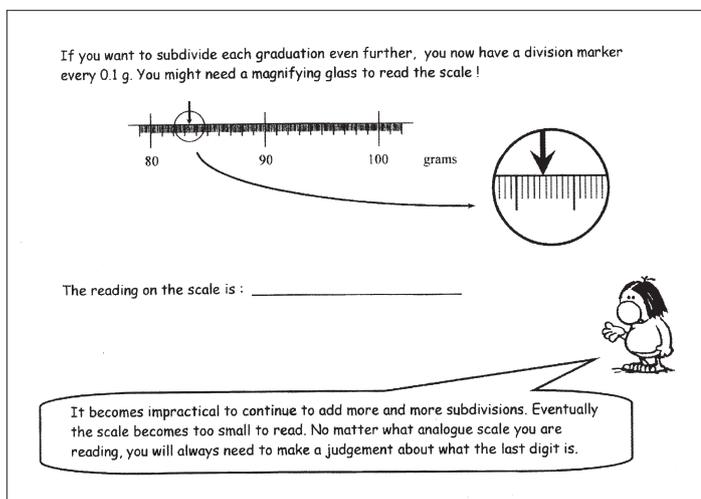
the GUM[7] that lends itself to the teaching situation is that an "uncertainty budget" should be part of all reported experiments. This idea is introduced qualitatively at first by asking students to reflect on experiments and write down all the factors that could have influenced their results and then to judge whether they thought these influences would have a "large" or "small" effect on the results.

The course materials then focus on what information can be inferred about a measurand from a single observation. Students are asked to consider a digital scale and to predict what digit will be displayed if the sensitivity of the instrument is increased by a factor of 10. Most students easily realize that there is an equal probability of the next (unknown) digit being any number between 0 and 9. A particular instrument can never be made "infinitely sensitive," i.e., able to provide a reading with an infinite number of digits. Even in the absence of all other sources of uncertainty, the knowledge about the measurand will always be limited to an interval, the width of which can never be reduced to zero. In this way a student's belief in the possibility of uncovering the "true value" is challenged. The same is shown to be the case with respect to reading an analog scale, which in addition also requires some form of judgment on the part of the observer (Fig. 1). Students are provided with simple apparatus such as an analog voltmeter and a penlight battery and are asked to make measurements and consider the uncertainty associated with reading the scale of the instrument together with all other possible sources of uncertainty in each case.

At this stage the more formal tools for dealing with uncertainty are introduced, including the probability density function (pdf). The idea that the pdf is a tool that models what we know, based on all the available information, is illustrated in the context of reading digital and analog scales (Fig. 2). The best approximation and the standard uncertainty are introduced as the two quantities that may be used to summarize the information associated with a particular pdf. The final stage in the sequence has to do with reporting the result of a measurand as a probabilistic statement. Handling scatter in an ensemble of repeated observations of the same measurand is deliberately delayed until students are adequately able to deal with a single reading. This is necessary to dispel the pervasive belief
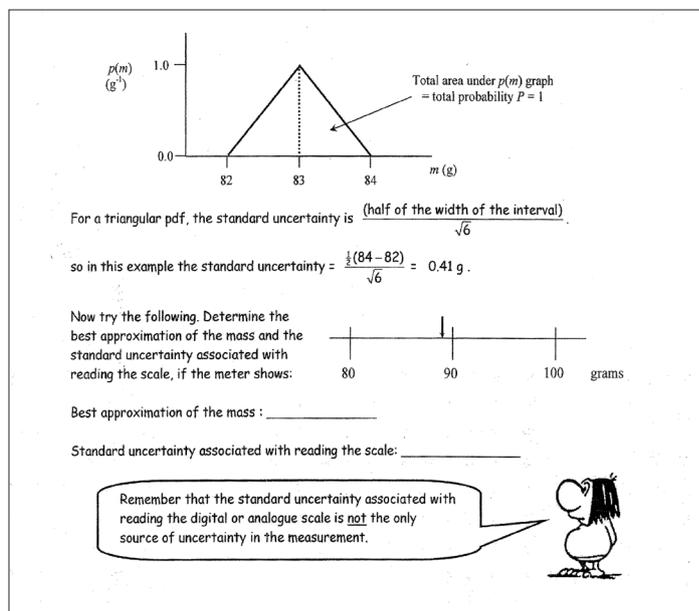
Fig. 2. An extract from the workbook dealing with the uncertainty associated with reading an analog scale.
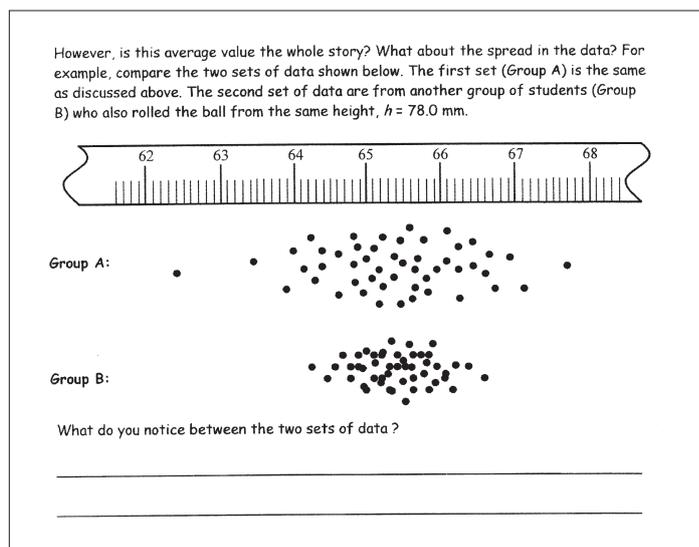

Fig. 3. An extract from the workbook dealing with the uncertainty associated with the spread in repeated observations.

among university entrants that the average value accounts for all experimental errors. By first dealing with the fundamentals of measurement uncertainty in the case of a single reading, dispersion in data may then be introduced as one of many sources of uncertainty and not necessarily the dominant one. We deal with dispersion by using the same experimental context as used for one of their experimental tasks that provides data with scatter (Fig. 3). A plausibility argument is

9.6 Experiment: the falling paper.

You are supplied with a piece of A4 paper and a digital stopwatch.

You are asked to measure the time it takes for the sheet of paper to fall from the top of your desk to the floor, starting with the paper horizontal and level with the top of the table.
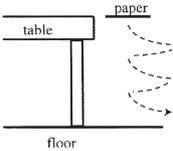
Release the paper a few times and watch what happens.

Now record 20 time readings here:

Complete the uncertainty budget table below. List all the significant sources of uncertainty for *t* and state whether you will calculate each by a Type A or a Type B evaluation of uncertainty (you might <u>not</u> list 4 below).

Uncertainty component (describe carefully).	Type of evaluation
1.	
2.	
3.	
4.	

Fig. 4. An extract from the workbook dealing with drawing up an uncertainty budget.

used to introduce the Gaussian pdf as an appropriate pdf to model the available information from the data. The statistical formula for the standard deviation of the mean is introduced at this stage as a measure of the uncertainty (Type A evaluation of uncertainty).

Once students have been exposed to a range of sources of uncertainty and can undertake both Type A and Type B evaluations of uncertainty, the quantitative version of the uncertainty budget is employed (Fig. 4). This includes both a summary of uncertainties in a measurement and the procedures of combining uncertainties from different sources. The examples in the workbook guide students through a range of measurement contexts, calculating standard uncertainties for a variety of sources of uncertainty and "summing" these to provide a combined standard uncertainty for the measurement. In this way, the theme of considering and evaluating every possible source of uncertainty culminates in the students being able to draw up uncertainty budgets and to estimate a reasonable total uncertainty for their measurements in practical tasks. The idea of assigning "human error" is more easily dispelled as each source of uncertainty has a pdf counterpart and students soon realize that the phrase has no currency.

Conclusion

The ISO-advocated probabilistic framework for metrology[7,8] offers a consistent method for making inferences about a measurand in cases of both single and multiple observations and an unambiguous language for communicating measurement results. Since the new approaches are being implemented at the research level,[6] they ought to be taught to undergraduate physics students.[2,17] Our materials[15] have been successfully used at the freshman level and have been recently evaluated[18] for their effectiveness in improving students' understanding of measurement and uncertainty, with statistically significant improvements being achieved across a range of measurement situations. We welcome physics instructors to engage with our materials and try them in their own teaching contexts.

References
1. C. Eisenhart, "Expression of the uncertainties of final results," *Sci.* **23**, 1201–1204 (June 1968).
2. P.R. Bevington, and D.K. Robinson, *Data Reduction and Error Analysis*, 3rd ed. (McGraw Hill, New York, 2003).
3. J.R. Taylor, *An Introduction to Error Analysis,* 2nd ed. (University Science Books, Sausalito, CA, 1997).
4. S. Allie, A. Buffler, B. Campbell, F. Lubben, D. Evangelinos, D. Psillos, and O. Valassiades, "Teaching measurement in the introductory physics laboratory," *Phys. Teach.* **41**, 394–401 (Oct. 2003).
5. W. Bich, M.G. Cox, and P.M. Harris, "Evolution of the '*Guide to the Expression of Uncertainty in Measurement*,'" *Metrologia* **43**, S161–S166 (Aug. 2006).
6. R. Kacker, K-D. Sommer, and R. Kessel, "Evolution of modern approaches to express uncertainty in measurement," *Metrologia* **44**, 513–529 (Dec. 2007).
7. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML,

Guide to the Expression of Uncertainty in Measurement (GUM) (International Organization for Standardization, Geneva, 1995, ISBN 92-67-10188-9).

8. B.N. Taylor and C.E. Kuyatt, *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results.* NIST Technical Report TN1297 (National Institute of Standards and Technology, Gaithersburg, MD, 1994).

9. G. D'Agostini, "Teaching statistics in the physics curriculum: Unifying and clarifying the role of subjective probability," *Am. J. Phys.* **67**, 1260–1268 (Dec. 1999).

10. L. Kirkup, "A guide to GUM," *Eur. J. Phys.* **23**, 483–487 (Sept. 2002).

11. J. Ryder and J. Leach, "Interpreting experimental data: The views of upper secondary school and university science students," *Int. J. Sci. Educ.* **22**, 1069–1084 (Oct. 2000).

12. M-G. Séré, R. Journeaux, and C. Larcher, "Learning the statistical analysis of measurement error," *Int. J. Sci. Educ.* **15**, 427–438 (1993).

13. R.L. Kung, "Teaching the concept of measurement: An example of a concept-based laboratory course," *Am. J. Phys.* **73**, 771–777 (Aug. 2005).

14. A. Buffler, S. Allie, F. Lubben, and B. Campbell, "The development of first year physics students' ideas about measurement in terms of point and set paradigms," *Int. J. Sci. Educ.* **23,** 1137–1156 (Nov. 2001).

15. A. Buffler, S. Allie, F. Lubben, and B. Campbell, *Introduction to Measurement in the Physics Laboratory. A Probabilistic Approach,* Ed. 3.4 (Department of Physics, University of Cape Town, 2007). Can be downloaded and used by instructors from http://www.phy.uct.ac.za/people/buffler/labmanual.html.

16. F. Lubben, A. Buffler, S. Allie, and B. Campbell, "Point and set reasoning in practical science measurement by entrant university freshmen," *Sci. Educ.* **85**, 311–327 (July 2001).

17. L. Kirkup and R.B. Frenkel, *An Introduction to Uncertainty in Measurement* (Cambridge University Press, Cambridge, 2006).

18. S. Pillay, A. Buffler, S. Allie and F. Lubben, "Effectiveness of a GUM-compliant course for teaching measurement in the introductory physics laboratory," *Eur. J. Phys.* **29**, 647–659 (2008).

PACS codes: 01.55.+b, 06.00.00

Andy Buffler *teaches physics at the University of Cape Town, South Africa, and undertakes research in both applied nuclear physics and physics education, with a special emphasis on physics laboratory work and the role of visualization in physics teaching and learning.*
andy.buffler@uct.ac.za

Saalih Allie *is an associate professor in the Department of Physics and the Coordinator of the Science Academic Development at the University of Cape Town. His Physics Education research interests include student laboratory work, writing in physics, and the relationship between cognitive science and the teaching and learning of physics.*
saalih.allie@uct.ac.za

Fred Lubben *is a researcher in the Centre for Innovation and Research in Science Education (CIRSE) at the University of York, UK. His research interests include the teaching of laboratory work, the nature of science, and the link between science and society, at secondary and tertiary level.*
fel1@york.ac.uk

etcetera...

etcetera... Editor *Albert A. Bartlett*
Department of Physics
University of Colorado
Boulder, CO 80309-0390

In the Checkout Line

"Some checkout aisles in local supermarkets feature grammatically correct '10 Items or Fewer' signs instead of the commonly seen '10 Items or Less' notice. [Local folklore (in Cambridge, Massachusetts) has it that anyone in the 10-item lane carrying 20 items is either a Harvard student who can't count or an M.I.T. student who can't read.]"[1]

1. Steve Mirsky, "A bridge too far," *Sci. Am.* **288**, 105 (Nov. 2003).