

## TuG1

## KEY COMPARISONS FOR DUMMIES: LESSONS LEARNED

D. G. Jarrett

National Institute of Standards and Technology\*  
Gaithersburg, MD 20899-8980Abstract

This paper will discuss some of the challenges encountered during a recent key comparison. From these experiences, general lessons will be drawn that should prove useful both for the consultative committees and for future pilot laboratories as they consider organization of future key comparisons.

Introduction

In recent years NIST and many other national metrology institutes (NMIs) have been involved in a significant number of key comparisons in support of the Mutual Recognition Arrangement (MRA) and associated needs related to international trade. In past years, international comparisons were generally driven by the science and research needs of the metrology community. The additional needs and constraints brought by the change of focus onto issues directly related to international trade have substantially increased the number of comparisons, significantly altered the direct users of their results, and led to an ongoing evolution of the rules and guidelines that govern the execution of the comparisons.

The changes driven by the new focus on international trade have significant ramifications that have not been fully understood or appreciated by researchers or managers in the NMIs, leading to substantial confusion about the change in focus. Despite the existence of key comparison guidelines, reigning confusion or lack of awareness of such a document has led to the evolution of rules and ad-hoc guidelines while the comparisons have been conducted. These influences have sometimes led to significant changes during and even after completion of the measurement and analysis phases of the comparisons. This creates a substantial and unnecessary burden for key staff members of the pilot laboratories, which is both wasteful of valuable resources and alienating for the affected staff.

The CCEM-K2 [1] comparison, for which we were the pilot laboratory, provides several good examples to illustrate the current problems and potential solutions related to key comparisons. We would first like to state that we believe that overall the Consultative Committee on Electricity and Magnetism (CCEM) is doing a good job in selecting and managing an appropriate collection of key comparisons. However, during CCEM-K2 we directly faced several issues that proved more burdensome than necessary. While our examples are taken specifically from this comparison, we have attempted to draw general lessons that apply broadly across the consultative committees in the hope that we in the metrology community can derive greater benefit with less difficulty in future comparisons. Some of the issues are: the need to clearly define the goals and expectations of a comparison; to develop clear, appropriate, and fixed protocols for the measurements, analysis, and reporting of results; and to continue education of the metrology community concerning these new goals and expectations.

A Clear Protocol

Once a consultative committee has decided to sponsor a key comparison in support of a measurement parameter, a clear consensus must be reached on a broad range of issues before a protocol can be drafted. Decisions regarding issues such as uncertainty budgets, acceptable statistical analysis methods, the reporting of results, the handling of anomalous results, and the determination of reference values need to be made prior to developing a protocol. The protocol defines the key comparison and provides details necessary for an NMI to make decisions regarding participation in a key comparison.

Only after the broad, foundational issues have been settled can the pilot laboratory develop an appropriate and robust protocol for the measurements and analysis phases of the comparison. The measurement protocol in particular must be considered very carefully, because once the comparison has begun, any significant changes regarding the goals and expectations of the comparison will be very disruptive to the comparison and very difficult to accommodate in a satisfactory manner. There should be a clear understanding of how to evaluate and deal with unexpected problems that will occur during a key comparison. Moreover, sound statistical design should be employed to ensure that artifact drift or failure will not result in unusable data.

Key comparisons are intended to support corresponding claims of calibration and measurement capabilities (CMCs) of the participants over a measurand and parameter range wider than the specific comparison protocol. In order for the broader metrology community to benefit from the wider potential applicability of the results of a key comparison, the final report should offer general recommendations, and the scientific basis for those recommendations, concerning the range of applicability. However, it is unlikely that the range of applicability will be significantly greater than the specific measurement of the protocol unless adequate consideration is given to this issue during the design stages of the key comparison.

For many key comparisons, there has been much debate over whether or not it is necessary to define a key comparison reference value (KCRV) for the comparison and if so how it should be defined. In all cases, it is essential that this debate be completed before the protocol is developed and certainly before the measurements have begun. For the CCEM-K2 comparison, decisions were both made and changed during the reporting phase of the comparison, well after measurements were completed and the analysis essentially complete. An example of the disruption caused by retroactive changes to the measurement protocol is the treatment of uncertainty budgets during CCEM-K2. When CCEM-K2 was started in 1996, the participating NMIs were not required to submit uncertainty budgets. After the MRA was signed in 1999, it was decided that uncertainty budgets would be required for the final report.

---

\* Electricity Division, Gaithersburg, MD. NIST is part of the Technology Administration, U.S. Department of Commerce. Official contribution of the National Institute of Standards and Technology; not subject to copyright in the United States.

Collecting uncertainty budgets years after the measurements were concluded was difficult and raised issues of the accuracy and validity of the budgets. While it is now CCEM policy for the protocol to include a template for uncertainty budgets, the general lesson is to minimize the disruption caused by changes to the protocol after the fact.

#### Ignorance of the Measurement Community

The people asked to conduct key comparisons are experts in their field of measurement. However, unless they have gone through the complete process themselves, they may not have complete information on conducting or documenting a key comparison.

Metrology professionals at many NMIs have been involved for most of their careers in a range of very demanding interlaboratory comparisons designed to ensure a sound scientific basis for the international metrology system. The key comparisons of today are designed with trade goals in mind and are expected to serve a different set of needs. Many metrologists, at all levels from the bench to upper management, have not adjusted their understanding of comparisons to accommodate this new reality. Perhaps the most dramatic example of this confusion is related to the issue of whether key comparisons should be performed at the highest possible accuracy and precision achievable by the research staff at the NMI, or whether the key comparison should be performed at the level best achievable in the calibration laboratory by staff that routinely perform calibrations. At many NMIs, these are not the same staff, sometimes not at the same location. Another example relates to the statistical analysis of the data. A full analysis requires a complete understanding of the correlation effects between the measurements. However, often the influences of these effects are significant only for the highest precision measurements and may not be significant for comparisons related to trade issues. The time and effort spent doing such a detailed statistical analysis then would be wasted.

Competent technical experts from NMIs who are charged with advising the pilot lab do not always know what is required for a final report to be included in Appendix B of the key comparison data base (KCDB). These requirements, as they have evolved, have often not been disseminated to the pilot lab staff preparing the reports or to those appointed to help and advise the pilot lab. There are often multiple and conflicting answers to important questions. The development of this system, the length of time it takes for multiple committees and groups to come to a consensus, and the turnover of committee membership make the task frustrating to technical staff.

#### Alienation of Key Staff

Once the measurements are completed, it is often the responsibility of the pilot laboratory to analyze the data and prepare a Draft A report. For any set of data, there are several approaches that can be taken to interpret the results and those involved in the review process do not always agree on the best approach. In the case of CCEM-K2, there were five draft reports produced in response to issues raised by various sequential reviewers. To fully investigate these issues, "leaving no stone unturned", the pilot lab sought guidance from statistical experts. There were a number of consultations between the pilot lab, the subgroup review committee, and the statistical experts. There were significant differences of opinion among these three groups about how to analyze the data and no clear path to meeting all concerns. Perhaps some general guidance from the

consultative committees during the design phase could address the importance of some of these subtle issues and provide some clear and referable guidelines.

Such guidance could be particularly valuable during the review of the final report. Since the constituents of the consultative committees, the working groups on key comparisons, and the appointed subgroups often change from one meeting of these bodies to another, obtaining a consensus is a challenge. Each group reviews the report at a different time with a different perspective, thus producing different, and sometimes conflicting, sets of comments.

It was the experience of pilot lab staff for CCEM-K2 that what was acceptable at one meeting of experts was not acceptable at a later time after the recommended changes had been made. The constant revising of reports is discouraging to staff who are advised to change X to Y and then six months later advised to change Y to X by the same body of experts. The process of documenting a key comparison is such an all-consuming task that staff who have experience in these key comparisons often run and hide from colleagues seeking help as they struggle through the preparation of the report. A rational person would think twice before subjecting himself to this process a second time.

Indeed, the total elapsed time from beginning to end of CCEM-K2 was six-years, extending past the retirement of the key staff member at the pilot lab who had been responsible for the comparison and who had expected to see it through to completion. This example illustrates the point that key comparisons can take too long for achieving maximum benefit.

#### Conclusions / Recommendations

As the system of key comparisons to support the MRA develops, the staff of NMIs responsible for key comparisons are confronted with continually changing requirements and an insufficiently clear set of instructions. Many technical experts have been frustrated with the developing system, changing requirements, and conflicting answers, which puts at risk their willingness to participate in future key comparisons.

Some suggestions for improving the system are included. Develop clear instructions and post them on a web site so that they are accessible, organized, and discernable in a way that staff can be educated, not overwhelmed. Software and templates for preparing the reports would be useful tools. The complete protocol, including such details as a template for reporting uncertainty budgets, needs to be finalized up front before the comparison starts, not years after the data was collected. Adequate statistical support needs to be included during development of the comparison protocol. The review process should be streamlined in such a manner as to allow the concerns of all ultimate reviewers to be heard and addressed in the early, rather than final, stages of review. Finally, we should not expect a key comparison, or a key comparison report, to meet expectations, or comply with requirements, imposed after the fact.

#### References

- [1] R. F. Dziuba and D. G. Jarrett, "CCEM-K2 Key Comparison of Resistance Standards at 10 M $\Omega$  and 1 G $\Omega$ ," *Appendix B of Mutual Recognition Arrangement, Bureau International des Poids et Mesures (BIPM)*, 2002.