Working Group 7: Key Comparisons
Report to the CCT
May 2014

WG7 Members:	Ken Hill (chair)	NRC (Canada)
	Mark Ballico	NMIA (Australia)
	Stephanie Bell	NPL (United Kingdom)
	Robert Benyon	INTA (Spain)
	Andrea Peruzzi	VSL (Netherlands)
	Steffen Rudtsch	PTB (Germany)
	Richard Rusby	NPL (United Kingdom)
	Greg Strouse	NIST (USA)
	Andrew Todd	NRC (Canada)
	Rod White	MSL (New Zealand)

Terms of reference

The terms of reference of CCT-WG7 are to oversee all aspects of key comparison documentation, starting with the protocol and ending with the Draft B Report and the KCDB entry, including the provision of advice to Pilots on the calculation of degrees of equivalence, key comparison reference values and linkage between RMO and CIPM key comparisons.

Working Group 7 is tasked with:

- examining all relevant documents for each key comparison starting with the protocol and ending with the Draft B Report,
- advising the pilot laboratory in preparing the text of the entry to Appendix B of the CIPM MRA as required, and to approve the Draft B Report on behalf of the CCT for inclusion into the BIPM key comparison database (KCDB),
- advising the pilot laboratory about the preparation of a comparison status document,
- reviewing and commenting on supplementary comparison Draft B reports.

WG7 is a service-oriented Working Group, tasked with reviewing comparison reports to fulfill the requirements of the MRA on behalf of the CCT as detailed in CIPM MRA-D-05. The goal of WG7 is to provide a credible review process that adds confidence to the comparison results, much as an on-site peer review provides confidence that the procedures and quality system of the NMI are appropriate to its tasks.

The Key Comparison Database – Thermometry Metrology Area

A "snapshot" of the BIPM KCDB taken on April 22, 2014 lists a total of **93 comparisons** within the Thermometry Metrology Area. This consists of:

- **28** comparisons **approved** prior to (or during) the May 2012 meeting of the CCT
- **13** comparisons **approved** since the May 2012 meeting of the CCT
- **27** comparisons (in addition to the 13 approvals) required WG7 actions (e.g. **protocol review, report review**) since May 2012
- **25** comparisons with **no WG7 actions** since May 2012 (i.e. waiting for the pilot)

Comparisons approved since May 2012

Comparison ID	Pilot	Date	Description
APMP.T- K4	KRISS	2003 - 2005	Aluminium freezing-point temperatures
APMP.T- S5	NMIA	2008 - 2009	Thermocouple calibration
COOMET.T- K3.1	VNIIM	2008 - 2009	Realizations of the ITS-90 from 302.9 K to 692.7 K
COOMET.T- K3.2	VNIIM	2010 - 2012	Realizations of the ITS-90 from 302.9 K to 692.7 K
EURAMET.T- K3.1	BIM	2008 - 2009	Realizations of the ITS-90 from 234.3 K to 692.7 K
EURAMET.T- K3.3	CEM	2009	Realizations of the ITS-90 from 83.8 K to 933 K
EUROMET.T- K5	VSL	1999 - 2001	Realizations of the ITS-90 up to 1700 °C
EURAMET.T- K6.1	MIKES	2009 - 2010	Comparison of dew-point temperatures
EURAMET.T- K7.1	SMU	2008 - 2009	Comparison of water triple point cells
EURAMET.T- K7.2	VSL	2010	Comparison of water triple point cells
EURAMET.T- K7.3	VSL	2011 - 2012	Comparison of water triple point cells
SIM.T- K6.4	INMETRO	2010 - 2011	Comparison of dew-point temperatures
SIM.T- S2	CEM	2004 - 2005	Comparison of platinum resistance thermometers

Protocol reviews / report reviews since May 2012

Comparison ID	Pilot	Date	Status
CCT- K3.2	NIM	2010	comments on report sent July 25, 2013
CCT- K4.1	NMIA	2012 - 2014	WG7 approves protocol Dec 12, 2012
AFRIMETS.T- S1	NMISA	2009 - 2010	comments on report sent Dec 11, 2012
AFRIMETS.T- S2	NMISA	2012	protocol reviews sent June 15, 2012
AFRIMETS.T- S4	MIRS/UL-FE/LMK	2012 - 2013	protocol reviews sent March 11, 2013
APMP.T- K3.6	NIM	2013 - 2014	WG7 approves protocol Dec 13, 2013
APMP.T- K4.1	NIM	2013 - 2014	WG7 approves protocol Dec 13, 2013
APMP.T- K6.1	NIMT	2013	comments on the protocol sent Dec 19, 2012
APMP.T- K7	CMS/ITRI	2007 - 2009	comments on the report sent Jan 15, 2014
APMP.T- K8	NMIJ	2011 - 2013	protocol reviews sent Aug 21, 2012
APMP.T- S8	NMLPHIL	2011 - 2015	protocol reviews sent July 2, 2013
APMP.T- S10	KRISS	2013	protocol reviews sent July 25, 2013
APMP.T- S11	NMIJ	2013 - 2015	WG7 approves protocol Nov 7, 2013
APMP.T- S12	NMIJ	2013 - 2015	WG7 approves protocol Nov 7, 2013

COOMET.T- K3.3	VNIIM	2014 - 2015	WG7 approves protocol April 14, 2014
COOMET.T- K5	NSC IM	2014 - 2015	protocol under review
COOMET.T- S1	VNIIM	2014 - 2015	WG7 approves protocol April 22, 2014
EURAMET.T- K3.4	MIRS/UL-FE/LMK	2010 - 2011	protocol reviews sent May 2, 2013
EURAMET.T- K3.5	VSL	2013	protocol reviews sent April 8, 2013
EUROMET.T- K8	PTB	2008 - 2012	protocol reviews sent May 13, 2013
EURAMET.T- S3	CEM	2013 - 2015	WG7 approves protocol Jan 24, 2014
SIM.T- K6.2	NIST	2010	report under review by WG7
SIM.T- K6.3	NIST	2010	comments on the report sent Feb 24, 2014
SIM.T- K9.1	NRC	2012	comments on the report sent June 17, 2013
SIM.T- S3	CESMEC, INEN	2007 - 2008	comments on the report sent Nov 20, 2012
SIM.T- S4	PTB	2008	comments on the report sent Aug 2, 2012
SIM.T- S5	CEM	2013 - 2014	WG7 approves protocol Dec 6, 2013

No WG7 actions since May 2012

Comparison ID	Pilot	Date	Description
CCT- K1.1	NIST	2006 - 2014	ITS-90 from 0.65 K to 24.6 K
CCT- K2.2	INRIM		ITS-90 from 24.5 K to 273.16 K
CCT- K2.5	NRC	2006 - 2014	ITS-90 from 13.8 K to 273.16 K
CCT- K6	NPL	2003 - 2010	dew and frost point temperatures
CCT- K6.1	NPL	2008 - 2010	dew and frost point temperatures
CCT- K8	INTA	2012 - 2015	dew-point temperature of humid gas
CCT- K9	NIST	2011 - 2012	ITS-90 from 83.8 K to 692.7 K
CCT- S1	NIST	2006 - 2009	Infrared spectral normal emissivity
CCT- S2	LNE	2007 - 2010	thermal conductivity by GHP
CCT- S3	NMIJ	2007 - 2008	Thermal diffusivity
AFRIMETS.T- S3	NMISA	2012	Calibration of industrial PRTs
APMP.T- K3.4	KRISS	2011 - 2013	SPRT calibration from Hg to Zn
APMP.T- K3.5	KRISS	2011 - 2014	SPRT calibration at the Zn f.p.
APMP.T- K3.7	NIM	2014	SPRT calibration at Argon t.p.
APMP.T- K6.2013	NMC, A*STAR	2013 - 2014	dew point temperatures
APMP.T- K7.1	NMIJ	2011 - 2012	Comparison of water triple point cells
APMP.T- S6	KRISS, NMIA, NML-SIRIM	2009	Comparison of IPRTs
APMP.T- S7	KRISS	2010 - 2013	melting temperature of Co-C eutectic
APMP.T- S9	NMIJ	2013	Thermal diffusivity
APMP.T- S13	NMC, A*STAR and NMIJ	2014 - 2016	Low-frost-point temperature
EURAMET.T- K1	PTB	2008 - 2012	ITS-90 from 2.4 K to 24.6 K
EURAMET.T- K3.2	UME	2009 - 2010	ITS-90 from 83.8 K to 692.7 K
EURAMET.T- S4	NPL	20007 - 2009	parameters for RT medium range
SIM.T- K6.1	NIST	2011	dew/frost-point temperatures
SIM.T- S6	NIST	2012 - 2014	Comparison of Type S thermocouples

WG7 membership

Referring to the minutes of the *20th Meeting of the CCT* (April 2000):

"Members of the working group are the pilot laboratories of ongoing key comparisons. Whenever a new key comparison is started, the pilot laboratory of that key comparison will be a new member of this working group until the key comparison is completed and its report accepted by the CCT. From this rule it follows that the NIST, NMi-VSL, NPL, NRC and PTB are members of this working group. The chairman of Working Group 3 is a personal [*ex officio*] member. The IMGC-CNR and the KRISS were accepted by the CCT as regular members." NML-CSIRO (currently NMIA) was added at the 21st Meeting of the CCT in 2001.

The principal when WG7 was formed was to have members experienced in the running of the previous comparison so that they can comment on the linkage for the next set of comparisons. Over time, the composition of WG7 has become weak. Many of the past comparison pilots have left the CCT. While it is desirable that the members include past pilots, or persons who have written reports and are familiar with the requirements of the CIPM MRA, it is not a matter of having laboratory representatives but members with expertise suited to the work.

With that in mind, Andrea Peruzzi (VSL), Steffen Rudtsch (PTB), and Andrew Todd (NRC) were recruited to bolster the ranks. However, with Mark Ballico's recent resignation from WG7, recruitment of additional members remains a high priority.

The Review Process

The role of WG7 chair is similar to that of a journal or conference proceedings editor. Three WG7 members are assigned to each protocol or report that is submitted for review, and they are asked to deliver their reviews via the WG7 Discussion Forum within 2 weeks. The deadlines are exceeded more frequently than we would like, but that is the challenge of an activity that relies on the good will of volunteers. Increasing the number of WG7 reviewers may alleviate some of the delays, so this is another driver for recruitment.

The "normal" WG7 process is to review the initial protocol and subsequent iterations until WG7 approval is obtained. The comparison is initiated and carried on without WG7 intervention until the Draft B report is agreed by the participants. WG7 then reviews the report and its revisions until the three reviewers approve its publication.

In many cases, there are long delays between subsequent versions of the report, and the responsibility for these delays must be borne by the pilot and the participants. Measurements comparisons are typically carried out over many years, even though it

would be in the best interests of all involved to ensure they conclude as quickly as possible to maximize their relevance.

RMO coordination

Would it be helpful to make available to WG7 the deadlines for CMC reviews by RMO to help establish priorities within WG7? If RMO MWG chairs have specific deadlines to meet (e.g. targets for reporting in advance of annual meetings), then they are encouraged to share this information with WG7.

Within EURAMET, the TC-T deadlines for CMC reviews are: end of February for the contact persons to submit new/revised CMCs; end of May for the EURAMET CMC review group to review the submitted CMCs and, if required, request additional information; end of June for the contact persons to provide the requested additional information.

Should WG7 seek to "improve" comparison reports by correcting errors?

We have had some cases where reports have not really been satisfactory, but we have been reluctant to correct them (or felt inhibited from doing so). If the proficiency-testing element of comparisons is to be preserved, then any changes made to the report based on feedback from WG7 needs to be explicitly documented as such and the original information retained.

As a minimum, we would like to see some sort of note detailing the concerns of WG7 attached to the Final Report that appears in the KCDB. One solution might be to institute a signing-off statement to be added to the report, stating that 'This report has been approved by CCT WG7 with no material changes' (which could allow for some revision), or 'with changes as indicated in the text' (otherwise).

Clearly, the defense that subsequent comparisons are doomed to follow the protocol of preceding ones should not be a sufficient argument to condone the promulgation of poor practice when deficiencies have been identified.

WG7 expectations for protocols and reports

The document "*Measurement comparisons in the CIPM MRA*" (CIPM MRA-D-05) provides a checklist of items that should be included in the protocol:

1. Detailed description of the devices: make, type, serial number, size, weight, packaging, etc., and technical data needed for their operation.

2. Advice on handling the travelling standards, including unpacking and subsequent packing and shipping to the next participant. This should include a complete list of the content of the package including handbooks, etc., and the weight and size of the whole package.
3. Action to be taken on receipt of the standards in a participating institute.
4. Any tests to be carried out before measurement.
5. Conditions of use of travelling standards during measurement.
6. Instructions for reporting the results.
7. Proposal for the method of determination of the key comparison reference value.
8. List of the principal components of the uncertainty budget to be evaluated by each participant, and any necessary advice on how uncertainties are estimated (this is based on the principles laid out in the ISO Guide to the Expression of uncertainty in Measurement). In addition to the principal components of the uncertainty, common to all participants, individual institutes may add any others that they consider appropriate. Uncertainties are evaluated at a level of one standard uncertainty and information must be given on the number of effective degrees of freedom required for a proper estimation of the level of confidence.
9. Timetable for communicating the results to the pilot institute. Early communication helps to reveal problems with the travelling standard during the comparison.
10. Financial aspects of the comparison, noting that in general each participating institute is responsible for its own costs for the measurements, transport and any customs charges as well as any damage that may occur within its country. Overall costs of the organization of the comparison, including the supply of the transfer devices, are normally borne by the pilot institute.
11. Insurance of transfer devices is decided by agreement among the participants taking account of the responsibility of each participant for any damage within its country.

The document is less prescriptive regarding the content of the comparison report. To supplement the content of CIPM MRA-D-05, Mark Ballico prepared supplementary checklists (published within CCT/12-25) as a guideline for WG7 reviewers.

Review criteria for protocols:
The intention of review of protocols is to provide advice to the pilot to ensure that the comparison is successful and fit for purpose. Suggested criteria are:

1. An assurance from the pilot that the protocol has been formally approved by all of the participants.
2. Suitable for the purposes of the MRA (e.g. blindness, etc.)
3. For linking comparisons, the protocol and comparison are substantially equivalent to the relevant CCT key comparison (i.e. a similar measurand and similar experimental techniques being “proficiency-tested”).
4. The linkage mechanism is satisfactory and explicit (uncertainty, reliability and suitability of the link lab) and link laboratory to the CIPM KC results has a sufficiently low uncertainty to support CMC claims of the participants.

5. Consideration of the comparison topology and artefacts.
6. Will have high probability of likely success using the proposed methodology.
7. Will not face any foreseeable problems in draft-B review stage.
8. The protocol should contain a suggested uncertainty template addressing the known components.

Review criteria for Draft-B reports:

1. An assurance from the pilot that the draft-B has been formally approved by all of the participants.
2. Basic editorial comments: Ensure that the text is unambiguous and clear (not worrying too much about English, grammar and formatting).
3. Completeness: the draft-B report should contain enough technical detail of the methodology to allow future use and re-evaluation of the data, potentially for 15 years or more.
4. The comparison satisfies the requirements of the MRA (i.e. blindness, no modification of the uncertainties or measurement values without comments in the report, etc.).
5. For linked comparisons, the protocol and comparison are substantially equivalent to the relevant CCT key comparison (i.e. a similar measurand and similar experimental techniques being “proficiency-tested”).
6. The linkage mechanism is satisfactory and explicit (uncertainty, reliability and suitability of the link lab).
7. Sufficient detail of link-lab standards to allow linking to the KCRV (of same family), and to facilitate linking to subsequent KCs. For example in the K7 and K3 comparisons we need to pay attention to the possibility that labs have made step-changes in their national definition of the TPW between the CCT-KC and RMO-KC. In K3 and K4, perhaps cells were different between CCT and RMO KCs.
8. The draft-B acknowledges and discusses any significant-unresolved-differences (SUDs) (e.g. by discussion, extra-analysis, comments from the participants, etc.) to facilitate later CMC review.
9. The mathematical analysis of both the lab-lab and lab-KCRV differences and their uncertainties are correct and makes sense, particularly with respect to the uncertainties of linkage and artefact stability, etc. A large linkage uncertainty may mask potentially scientifically important SUDs.
10. Bilateral DOE tables are optional for KCs, but it is required to provide the equations to calculate it from the tabulated data in the report.

We encourage pilots to ask critical questions about the inputs received from the participants, rather than simply collating the results and uncertainties. The pilot has a unique view of how the comparison is turning out, and intervention at an early stage may avert errors and inconsistencies that are difficult or impossible to correct at a later stage without compromising the comparison. We suggest that more rigorous evaluation by the pilots will improve the quality of the comparisons.

Linking

Implementing the linkage mechanism and its uncertainty can often prove difficult for the authors of key comparison reports. Some comparison coordinators simply do not know how to perform the linkage, and sometimes the required information can be difficult to locate. For example, in the case of a K3-style comparison, only a few persons have access to the Average Reference Value needed to perform the linkage. Perhaps a short guidelines document would be of help.

Bayesian vs frequentist GUM

WG7 continues to struggle with concerns regarding the application of Bayesian uncertainty analysis to comparison reports. Fortunately, the most recent revision of a report under WG7 review has forsaken the Bayesian approach for the more familiar frequentist-style analysis. While this has solved the problem in the short term, future comparison reports may well revive the debate.

Uncertainty guides

The various WGs of the CCT have devoted considerable time to the preparation of uncertainty guides with the aim of harmonizing uncertainty analysis – but there is little evidence of them being used for comparisons. We encourage their incorporation into the protocols at an early stage, and for the pilots to seek WG7 approval of the protocols (and not merely the final report) well in advance of the measurement schedule.

We note that very few reports or protocols reference the SPRT uncertainty guide CCT/08-19rev. We feel that improvements in practice can be realized by raising expectations, and so we encourage those engaged in the writing of comparison protocols to incorporate current best practices.

Meetings

Most of the members present at TEMPMEKO 2013 attended a brief meeting of WG7. Similar “meetings of opportunity” will be planned subject to need and interest, the next most likely event being TEMPMEKO 2016. Otherwise, WG7 will continue to carry out its tasks by correspondence with plans to meet in advance of the next CCT.