



Methoden zur Bewertung von Vergleichsmessungen



D-K-15186-01

Olaf Schnelle-Werner, Nadine Schiering, G. Wübbeler (PTB)
Zentrum für Messen und Kalibrieren & ANALYTIK GmbH
Bitterfeld-Wolfen



Agenda

- Normative Basis / Literatur
- Stufen und Bedeutung von Vergleichsmessungen
- Vorbereitung einer Vergleichsmessung
- Das Pilotlabor
- Das technische Protokoll
- Möglichkeiten der Durchführung
- Bester Schätzwert / Referenzwert
- Stabilität des Transferstandards
- Bewertung der Messergebnisse
- Der Abschlussbericht



Hintergrund von Vergleichsmessungen

Normative Basis / Literatur:

- ISO/IEC 17025:2005 bzw. ISO/IEC 17025:2017; Allgemeine Anforderungen an die Kompetenz von Prüf- und Kalibrierlaboratorien
- ISO/IEC 17043:2010; Konformitätsbewertung - Allgemeine Anforderungen an Eignungsprüfungen
- Guidelines for CIPM key comparisons; BIPM 1999
- Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes, Paris, 14 October 1999 (CIPM MRA)



Hintergrund von Vergleichsmessungen

Normative Basis / Literatur:

- COOMET Recommendation R/GM/11:2007; REGULATIONS for comparison measurement standard from the national metrological institutes of COOMET
- EURAMET Guide on Comparisons EURAMET Guide No. 4 Version 1.0 (05/2016)



Stufen von Vergleichsmessungen





Vorbereitung einer Vergleichsmessung

Was muss vor Beginn abgestimmt werden?

- **Pilotlabor**
- **Kalibrierverfahren, die von den Laboren benutzt werden**
- **Anwendbare Normen und Richtlinien**
Beispiel Masse: Internationale Richtlinie OIML R111
Beispiel Temperatur: Richtlinie DAkkS-DKD-R 5-7 (Klimakammern)
- **Umfang der Vergleichsmessung**
Beispiel: Vereinbarung, dass nur die Endmaße 0,5 mm und 100 mm aus einem 122 teiligem Satz von Parallelendmaßen kalibriert werden sollen



Vorbereitung einer Vergleichsmessung


Was muss vor Beginn abgestimmt werden?

- Messunsicherheit
 - ✓ CMC-Einträge
 - ✓ Akkreditierte Messunsicherheit
 - ✓ **Angestrebte Messunsicherheit**

- Zeitplan mit Terminen für Messung und Berichterstattung

- Transport der Kalibriergegenstände

- Kosten

Eine Messunsicherheitsbilanz muss bei den teilnehmenden Laboratorien verfügbar sein. 



Das Pilotlabor

Das Pilotlabor ist verantwortlich für die Organisation und Durchführung der Vergleichsmessung sowie für die Auswertung der Ergebnisse.

Qualifikation als Pilotlabor:

- Langjährige Erfahrungen und Akkreditierung in der Messgröße, die zum **Kalibriergegenstand / Transferstandard** gehört
- Akkreditierung (als Kalibrierlaboratorium) oder CMC-Einträge (als NMI) für den entsprechenden Kalibriergegenstand, das Verfahren und den beabsichtigten Umfang

Verantwortlichkeiten des Pilotlabors:

- Vorschlag für eine Aufgabenbeschreibung einschließlich des metrologischen Umfangs
- Diskussion und Abstimmung mit den Teilnehmern
- Formulierung eines technischen Protokolls (alle Anforderungen der Aufgabe müssen dokumentiert werden)



Das technische Protokoll

Das technische Protokoll muss folgendes beinhalten:

- Detaillierte Beschreibung der Kalibriergegenstände: *Messgeräte bzw. -systeme* (Modell, Typ, Seriennummer...), *Referenzlösungen und Referenzmaterialien*
- Bewertung der Stabilität des Transferstandards
- Vorgaben zum Handling einschließlich Aus- und Verpacken, Sendung zum nächsten Teilnehmer
- Durchzuführende Tests vor Beginn der Messungen
- Einsatzbedingungen des Kalibriergegenstandes während den Messungen
- Vorgaben zur Darstellung der Messergebnisse
- Messverfahren
- Wichtigste Komponenten der Messunsicherheit
- Termine für die Übermittlung der Messergebnisse an das Pilotlabor



Bester Schätzwert

Es ist nicht möglich, einen “wahren Wert” zu ermitteln. Aber es ist möglich, unter gegebenen Bedingungen den besten Schätzwert zu bestimmen.

Es gibt verschiedene Möglichkeiten zur Berechnung des besten Schätzwertes. Die geeignete Möglichkeit muss vor Beginn der Vergleichsmessung abgestimmt werden.

Beispiel 1: Messergebnis eines Labors als Referenzwert

Der Referenzwert können die Messergebnisse des Pilotlabors sein.

Das Pilotlabor sollte die besten messtechnischen Möglichkeiten unter den Teilnehmern haben.

In diesem Falle muss das Pilotlabor ein akkreditiertes Kalibrierlaboratorium oder ein NMI mit CMC-Einträgen und langjähriger Erfahrung auf dem Gebiet der entsprechenden Messgröße sein.



Bester Schätzwert

Beispiel 2: NMI's oder Kalibrierlaboratorien mit äquivalenten messtechnischen Möglichkeiten

In diesem Falle kann der Referenzwert der arithmetische Mittelwert sein.

$$x_R = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

x_R ... arithmetischer Mittelwert

x_i ... Messergebnis eines Teilnehmers

n ... Anzahl der Teilnehmer

i ... Zählindex

Auf Grundlage des GUM, Thema auch diskutiert auf der VDI/VDE-Konferenz BEV "Neue Entwicklungen in der Messdatenauswertung und Messunsicherheitsbestimmung",

Wien, 10. bis 11. März 2010



Bester Schätzwert

Beispiel 2: NMI's oder Kalibrierlaboratorien mit äquivalenten messtechnischen Möglichkeiten

Als Referenzwert kann der arithmetische Mittelwert verwendet werden.

Die erweiterte Messunsicherheit ($k=2$) wird wie folgt berechnet:

$$U = 2 \cdot \sqrt{\frac{1}{n(n-1)}} \cdot \sqrt{\sum_{i=1}^n (x_i - x_R)^2} \quad (2)$$

x_R ... arithmetischer Mittelwert

x_i ... Messergebnis eines Teilnehmers

n ... Anzahl der Teilnehmer

i ... Zählindex

U ... Erweiterte Messunsicherheit ($k=2$)

Quelle: H. Bauer,
G. Klingenberg, B. Werner;
*Viscosity comparison
measurements between PTB
and the ZMK-Wolfen in
connection with the CCM.V-K1
key comparison; PTB report
PTB-CP-1 (2002)*



Bester Schätzwert

Beispiel 3: Verwendung eines gewichteten Mittelwertes nach Cox

$$y = \frac{\frac{x_1}{u^2(x_1)} + \dots + \frac{x_N}{u^2(x_N)}}{\frac{1}{u^2(x_1)} + \dots + \frac{1}{u^2(x_N)}} \quad (3)$$

Quelle: M.G. Cox, *The evaluation of key comparison data*, *Metrologia*, 2002, **39**, 589-595

y ... gewichteter Mittelwert

x_1 bis x_N ... Messergebnisse von Teilnehmern

u ... Standardmessunsicherheiten



Bester Schätzwert

Beispiel 3: Verwendung eines gewichteten Mittelwertes nach Cox

Ermittlung der kombinierten Messunsicherheit $u(y)$ des Referenzwertes wie folgt:

$$\frac{1}{u^2(y)} = \frac{1}{u^2(x_1)} + \dots + \frac{1}{u^2(x_N)} \quad (4)$$

Quelle: M.G. Cox, *The evaluation of key comparison data*, *Metrologia*, 2002, **39**, 589-595

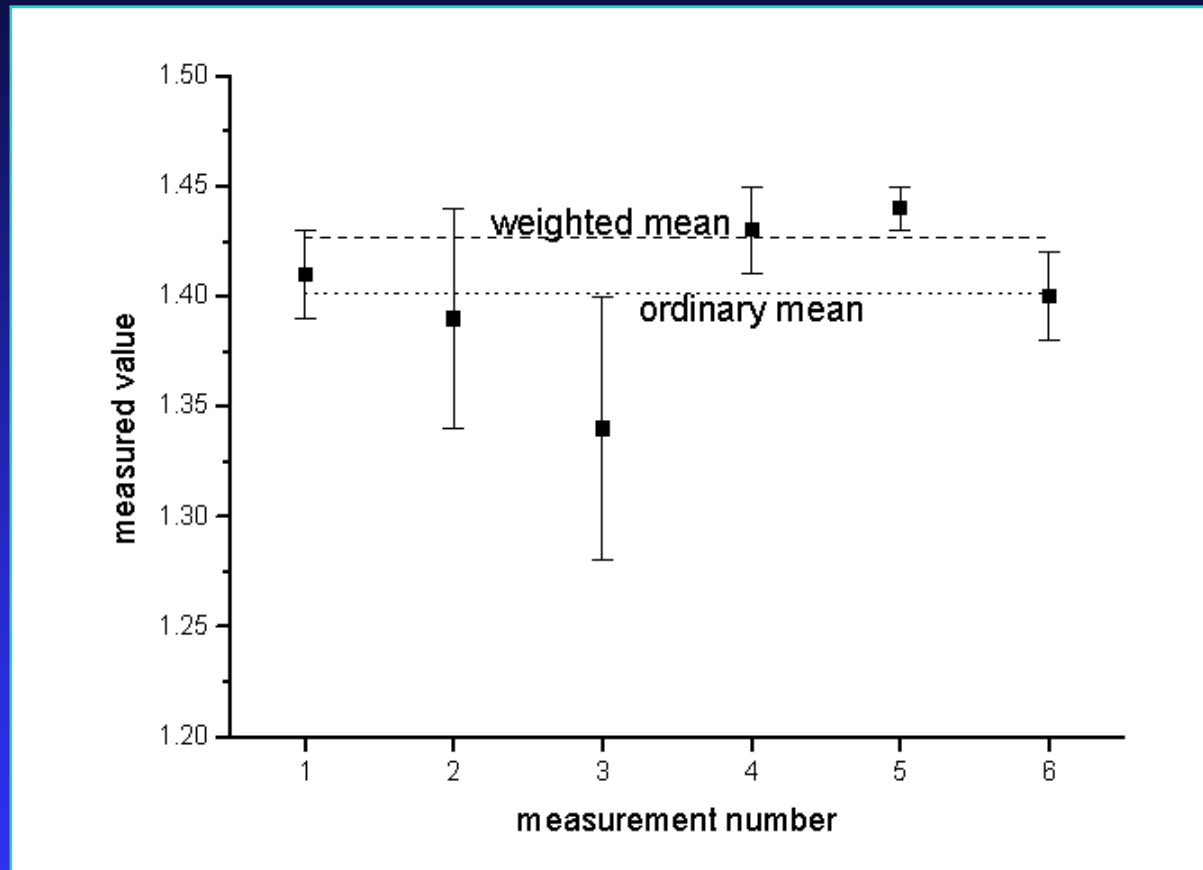
y ... gewichteter Mittelwert

x_1 bis x_N ... Messergebnisse von Teilnehmern

u ... kombinierte Messunsicherheiten der Teilnehmer



Bester Schätzwert



Quelle:

Homepage der
Cambridge
University Press

[www.cambridge.org/
resources/](http://www.cambridge.org/resources/)

Vergleich arithmetischer und gewichteter Mittelwert

**308. PTB-Seminar Berechnung der Messunsicherheit – Empfehlungen für die Praxis,
15. und 16. März 2018, PTB Berlin**

Seite 16



Bester Schätzwert

Resümee:

Die Herangehensweise entsprechend Beispiel 3 wird zur Bestätigung kleiner Messunsicherheiten empfohlen, z.B. in Supplementary Key Comparisons.



Bester Schätzwert

Ermittlung von Ausreißern durch statistische Methoden:

Im Folgenden eine Liste von Möglichkeiten:

- Der Grubbs-Test ist ein statistischer Test zur Ermittlung von Ausreißern in einer gegebenen Probe. Diese werden entfernt und die reduzierte Probe wird durch Iteration optimiert.
- Der Nalimov-Test ist eine Variante von Grubbs.
- Unter dem Chi-Quadrat-Test verstehen wir eine Gruppe von Hypothese-Tests mit einer Chi-Quadrat-Verteilungsfunktion für die Probe:
 1. Test der Verteilungsfunktion (Adaptationstest)
 2. Unabhängigkeitstest
 3. Homogenitätstest



Bester Schätzwert

Ermittlung von Ausreißern durch statistische Methoden:

Im Folgenden ein Beispiel für einen beidseitigen Grubbs-Test:

Bestimmung von Outliner-Daten durch einen beidseitigen Grubbs Test							
N°	Nominalwert in mg	Messergebnis als Abweichung vom Nominalwert in mg	Abweichung des Messergebnisses zum Mittelwert d_i in mg	$g_i = d_i / s$	$g_i > g_{crit}$		
1		95,6	0,2	0,421	ok	Mittelwert m:	95,367
2		95,2	0,2	0,301	ok	Standardabw.s:	0,555
3		95,8	0,4	0,781	ok	max. Abw.:	1,367
4		94	1,4	2,465	outliner	g_{max} :	2,465
5		95,4	0,0	0,060	ok	g_{crit} :	2,215
6		95,3	0,1	0,120	ok	outliner:	Yes
7		95,5	0,1	0,240	ok		
8		95,8	0,4	0,781	ok		
9		95,7	0,3	0,601	ok		
10							
11							
12							
13							
14							
15							
16							



Bester Schätzwert

Resümee:

Die Anwendung von Ausreißertests ist zwingend notwendig für die Herangehensweise entsprechend der Beispiele 2 und 3.



Die Stabilität des Transferstandards – Einfluss auf die Messunsicherheit des Referenzwertes

Ein stabiler Transferstandard ist der **Sollzustand**

Vorgehen bei einem nichtstabilen Transferstandard:

- Ermittlung der Abweichung der Eingangsmessungen zu den Kontroll- bzw. Rückmessungen
- Zuordnung der Abweichung zu der zeitlichen Reihenfolge der Vergleichsmessung
- Ermittlung eines Zusatzbeitrages für die Messunsicherheit des Referenzwertes unter Berücksichtigung der chronologischen Reihenfolge



Die Stabilität des Transferstandards – Einfluss auf die Messunsicherheit des Referenzwertes

Transferstandard:	100 mg E1 – Gewicht
Eingangsmessung:	Abweichung vom Nominalwert (Δm_c) 0,0016 mg
Rückmessung:	Abweichung vom Nominalwert (Δm_c) 0,0008 mg
Erweiterte Messunsicherheit für beide Messungen:	$U (k = 2) =$ 0,0010 mg
Erweiterte Messunsicherheit des Referenzwertes nach Cox:	$U (k = 2) =$ 0,0008 mg
Erweiterte Messunsicherheit des Referenzwertes unter Berücksichtigung der Drift des Transferstandards:	$U (k = 2) =$ 0,0012 mg Driftbeitrag: 0,00046 mg

Für die Drift wurde eine Rechteckverteilung angesetzt, da keine weiteren Kontrollmessungen zwischen den Teilnehmern der Vergleichsmessung vorliegen.

Information zum Zeitpunkt der Veränderung liegen nicht vor.



Bewertung von Messergebnissen

Es existieren verschiedene Möglichkeiten zur Bewertung von Messergebnissen.

- Berechnung des E_n -Wertes
- Degree of equivalence
- Berechnung des Zeta-Scores
- Berechnung des z-Scores (Prüflabore)

- Der Zeta-Score wird von Messunsicherheiten beeinflusst.
Der z-Score wird von einer Standardabweichung beeinflusst.

- Der Zeta-Score erlaubt eine ähnliche Interpretation wie der z-Score.



Bewertung der Messergebnisse – der E_n -Wert

Es muss ein Akzeptanzkriterium für die Messergebnisse festgelegt werden.

Das Ergebnis der Kalibrierung des teilnehmenden Laboratoriums wird dem Referenzwert verglichen (entsprechend **DIN EN ISO/IEC 17043:2010**).

$$E_n = \frac{x_{lab} - x_R}{\sqrt{U_{lab}^2 + U_R^2}} \quad (5)$$

x_R ... Referenzwert

x_{lab} ... Messergebnis eines teilnehmenden Laboratoriums

U_R ... erweiterte Messunsicherheit ($k = 2$) des Referenzwertes

U_{lab} ... erweiterte Messunsicherheit ($k = 2$) eines teilnehmenden Laboratoriums

-
- $|E_n| < 1$ Messergebnis „befriedigend“
 - $|E_n| \geq 1$ Messergebnis auffällig oder „nicht befriedigend“
-



Bewertung der Messergebnisse – Degree of equivalence

Ermittlung des Degree of equivalence und seiner Messunsicherheit nach Cox:

$$d_i = x_i - x_r \quad (6)$$

$$u^2(d_i) = u^2(x_i) - u^2(x_r) \quad (7)$$

Quelle: M.G. Cox, *The evaluation of key comparison data*, *Metrologia*, 2002, **39**, 589-595

d_i ... Degree of equivalence

x_1 ... Messergebnis eines Teilnehmers

x_r ... Referenzwert

u ... kombinierte Messunsicherheiten



Bewertung der Messergebnisse – der Zeta-Score

Der Zeta-Score wird nach der folgenden Gleichung berechnet:

$$\zeta = \frac{x_{lab} - x_R}{\sqrt{u_{lab}^2 + u_R^2}} \quad (8)$$

... Zeta-Score

x_R ... Referenzwert

x_{lab} ... Messergebnis eines teilnehmenden Laboratoriums

u_R ... Standardmessunsicherheit des Referenzwertes

u_{lab} ... Standardmessunsicherheit eines Laboratoriums

Die Berechnung ist ähnlich wie die Berechnung des E_n -Wertes.

Quelle: ISO/IEC 17043:2010 / BS ISO 13528:2015

- $|\zeta| \leq 2.0$ zeigt eine “zufrieden stellende” Leistung an und generiert kein Signal
- $2.0 < |\zeta| \leq 3.0$ zeigt eine “fragwürdige” Leistung an und generiert ein Warnsignal
- $|\zeta| \geq 3.0$ zeigt eine “unbefriedigende” Leistung an und generiert ein Handlungssignal



Die Stabilität des Transferstandards – Einfluss auf die Messunsicherheit des Referenzwertes

Iststand:

- Der absolute E_n Wert ist kleiner gleich 1
- Der degree of equivalence ist kleiner 1,96
- Der Transferstandard hat sich in der Größenordnung deutlich über der erweiterten Messunsicherheit seiner Kalibrierung verändert
- **Wir können die neuen kleinsten Messunsicherheiten CMC bestätigen ?**
- **Wichtig ist nur, dass der E_n Wert kleiner gleich 1 ist !**

Zukunft:

Die Einführung eines zusätzlichen Kriteriums zur Bewertung des Transferstandards z.B. „Loss of Power“ (PTB ..) für die Bestätigung kleinster Messunsicherheiten CMC's wird vorgeschlagen.



Bewertung der Messergebnisse – Power Analyse

Einführung eines Powerfaktors G:

Zur Feststellung des Verlustes an Aussagekraft des Bewertungskriteriums z.B. „degree of equivalence“ werden u.a. die folgenden Größen untersucht:

- u_{ts} - Messunsicherheit des Transferstandards
- d_i - Degree of equivalence
-

Der Powerfaktor G wird für $u_{ts} = 0$; „ G_0 “ und die ermittelte Messunsicherheit des Transferstandards u_{ts} ; „G“ bestimmt.

Der Verlust an Aussagekraft ΔG wird wie folgt bestimmt:

- $\Delta G = G_0 - G$ - loss of power

Quelle: G. Wübbeler, O. Bodnar, B. Mickan, C. Elster, Explanatory power of degrees of equivalence in the presence of a random instability of the common measurand, *Metrologia*, 2015, **50**, 400-405



Bewertung der Messergebnisse – Power Analyse

Beispiel für die Anwendung des Powerfaktors G:

Calculation of the loss of power in comparison measurements

variables

unilateral degrees of equivalence	$\delta_i = x_i - \mu, u(\delta_i)$
uncertainty of the unilateral degrees of equivalence	$u(\delta_i)$
value of the participant	x_i
uncertainty value of the participant	u_i
loss of power	ΔG
uncertainty of the transfer standard	u_{TS}
uncertainty of the reference value	u_R

input data (marked yellow)

u_{TS}	0,08
u_{TS0}	0,00
u_R	0,03
x_R	0,00
<i>all units in</i>	mg
<i>nominal value</i>	1 kg



Bewertung der Messergebnisse – Power Analyse

Beispiel für die Anwendung des Powerfaktors G:

unilateral degrees of equivalence

N°	x_i	u_i	δ_i	$u(\delta_i)$	$abs(\delta_i) / u(\delta_i)$	<i>satisfied</i> $\leq 1,96$	G_0	G	ΔG
1	-0,02	0,15	-0,27	0,15	1,85	Yes	0,07	0,47	-0,40
2	0,02	0,15	-0,23	0,15	1,58	Yes	0,05	0,37	-0,31
3	0,09	0,15	-0,16	0,15	1,11	Yes	0,06	0,20	-0,14

bilateral degrees of equivalence

N°	x_i	u_i	δ_i	$u(\delta_i)$	d_i degree of equiv.	<i>satisfied</i> $\leq 1,96$
1	-0,02	0,15	-0,27	0,15	0,11	Yes
2	-0,01	0,15	-0,26	0,15	0,05	Yes
3	0,09	0,15	-0,16	0,15	0,47	Yes



Der Abschlussbericht

Verantwortlichkeiten des Pilotlabors:

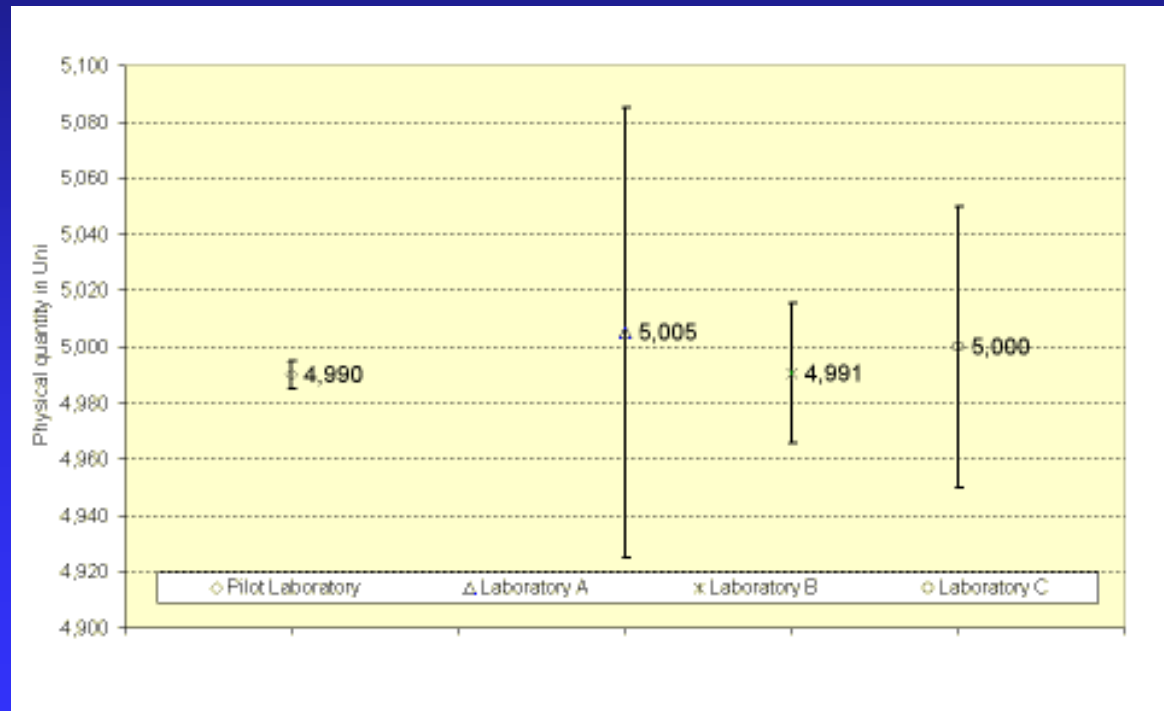
- Erstellung eines Berichtes als Entwurf und Versendung zu den Teilnehmern (zur Überprüfung der Resultate und zur Ermöglichung von Kommentaren und Vorschlägen) – dieser Bericht ist für die Teilnehmer der Vergleichsmessung vertraulich.
- Erstellung der finalen Version unter Berücksichtigung der Kommentare and Vorschläge der Teilnehmer
- **Neben weiteren Anforderungen muss der Abschlussbericht zwingend auch Aussagen zur Stabilität des Transferstandards enthalten.**



Der Abschlussbericht

Graphische Darstellung der Messergebnisse:

- Auch für die graphische Darstellung gilt:
Messergebnisse müssen mit ihrer Messunsicherheit dargestellt werden.





Zusammenfassung

- Eine alleinige Bewertung des Auswertekriteriums ist nicht in jedem Fall ausreichend.
- Für die Bestätigung kleinster CMC ist die Bewertung der Stabilität des Transferstandards essentiell.
- Ggf. muss ein zusätzliches Kriterium wie „loss of power“ * ΔG der PTB eingeführt und bewertet werden.

Quelle: G. Wübbeler, O. Bodnar, B. Mickan, C. Elster, Explanatory power of degrees of equivalence in the presence of a random instability of the common measurand, Metrologia, 2015, 50, 400-405



**Danke für Ihre
Aufmerksamkeit !**

ZMK & ANALYTIK GmbH

www.zmk-wolfen.de

