5 Eignungsprüfungen

Von der esz AG werden Eignungsprüfungen für Kalibrierlabore angeboten gemäß den Anforderungen der DIN EN ISO/ IEC 17043. Eignungsprüfungen werden im nach DIN EN ISO/ IEC 17025 akkreditierten Umfang angeboten gemäß den Vorschriften der DIN EN ISO/ IEC 17043 sowie der DIN EN ISO/ IEC 17025.

5.1. Räumlichkeiten und Umgebungsbedingungen

Die Messungen für Eignungsprüfungen werden in den klimatisierten Räumlichkeiten der esz AG durchgeführt. Detaillierte Informationen dazu sind im QMH Kapiteln 00-3-Unterstützende-QM-Infrastruktur sowie QMH Kapitel 02-I-Umgebungsbedingungen zu finden.

5.2. Verantwortlichkeiten

Die Verantwortlichkeiten sind in QMH Kapitel 1.4.2 festgehalten. Grundsätzlich ist Personal das für die Durchführung von allen Messungen bei Eignungsprüfungen eingesetzt wird identisch mit Personal des Kalibrierlaboratoriums. Die technische Qualifikation für die Durchführung von Messungen ist durch Erfahrung im Kalibrierlabor sichergestellt und in der Kompetenzmatrix des Kalibrierlabors dokumentiert. Es gibt kein dediziertes Personal welches nur für Eignungsprüfungen eingesetzt wird, ebenso kann jede im Labor eingesetzte Person an Eignungsprüfungen beteiligt sein.

Die Koordination und die Überwachung von Eignungsprüfungen erfolgt durch das Qualitätsmanagement-Team der esz AG, insbesondere durch die Eignungsprüfungsmanager*in. Die Aufgaben und Verantwortlichkeiten des Koordinators sind in QMH Kap. 1.4.2. festgelegt. Die Auswertung der Messdaten und das Erstellen des Berichts erfolgt durch das Qualitätsmanagement-Team der esz AG, durch Personen, welche nicht an einer Messung für dieselbe Eignungsprüfung beteiligt waren (z.B. in Durchführung der Kalibrierung oder Freigabe).

Die Aufgaben des Managers für Eignungsprüfungen[1] sind

- Überwachung und Durchführung der Auswahl, Vorbereitung und Verteilung von Prüfgegenstände für die Eignungsprüfung
- Planung des Eignungsprüfungsprogramms , inkl. Wahl des Verfahrens und Überwachung der Erstellung der technischen Beschreibung
- Sicherstellung der Kompetenz der einbezogenen Mitarbeiter
- Dokumentation von Beobachtungen
- Mitwirkung an der Bestimmung der Stabilität und Homogenität des Prüfgegenstandes
- Mitwirkung an der statistischen Analyse der Ergebnisse
- Genehmigung der Herausgabe der Berichte

[1] Siehe auch QMH 1.4.2.6

5.3. Ablauf

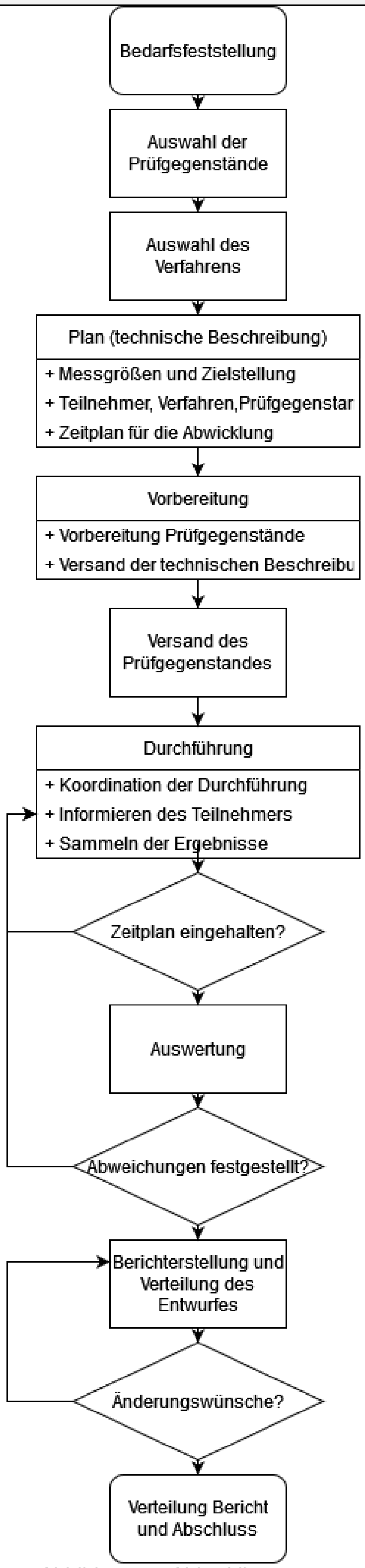
<table>
<tr><th colspan="2">Ablauf</th></tr>
<tr><td rowspan="5">

Abbildung 1 - Ablaufdiagramm Eignungsprüfungen</td>
<td>Bedarfsfeststellung</td></tr>
<tr><td>Plan (Nach DIN EN ISO/ IEC 17043:2023 7.2</td></tr>
<tr><td>Durchführung
• Vorbereiten des Prüfgegenstandes (gemäß Plan bzw. DIN EN ISO/ IEC 17043 4.4.1)
• Versand des Prüfgegenstandes (Entfällt bei Messung an festem Ort durch Teilnehmer)
    ○ Versand und Stand wird kommuniziert an Teilnehmer
    ○ Anforderungen an Versand zwischen Teilnehmern werden im Plan kommuniziert.
• Änderungen im zeitlichen Ablauf müssen zeitnah nach Bekanntwerden betroffenen Teilnehmern kommuniziert werden.
• Sammeln der Ergebnisse</td></tr>
<tr><td>Auswertung
• Anwendung statistischer Verfahren/Modelle. Vgl. 2.1.4.5
• Auswertung so wie im Plan vermerkt. Bei Abweichungen muss die Zustimmung der Teilnehmer eingeholt werden.
• Falls Ausreißer vermutet werden, so müssen diese über nachvollziehbare Verfahren und Annahmen identifiziert werden. Die Verfahren und getroffenen Annahmen müssen den Teilnehmern einer Eignungsprüfung mitgeteilt werden.</td></tr>
<tr><td>Bericht
• Der Bericht wird zunächst in einer Entwurfsfassung allen Teilnehmern zur Verfügung gestellt. Kein einzelner Teilnehmer erhält bevorzugten Zugang zu Ergebnissen.
• Der Bericht muss DIN EN ISO/ IEC 17043:2023 7.4.3.2 berücksichtigen (Erklärung der Unterschiede zwischen Teilnehmern).
• Falls ein Teilnahmezertifikat ausgestellt wird so darf die Leistungsbestätigung nicht widersprüchlich sein.
• Der Bericht muss den Vorgaben der DIN EN ISO/ IEC 17043:2023 Abschnitt 7.4.3 genügen.</td></tr>
</table>

5.3.1.Vertraulichkeit

Die Mitarbeiter des Eignungsprüfungsprogramms sind zur Wahrung der Vertraulichkeit der Identität der Teilnehmer und deren Ergebnisse verpflichtet, wie auch im QMH 00-1 beschrieben. Der Datenschutz ist grundsätzlich einzuhalten. Sollte ein interessierter Kreis oder Behörde an den Ergebnissen der Eignungsprüfung interessiert sein, werden Teilnehmer hierüber in Kenntnis gesetzt.

5.3.2.Kommunikation mit den Teilnehmern

Die Kommunikation mit den Teilnehmern wird durch die Eignungsprüfungsmanagerin koordiniert. Die Auftragsabwicklung erfolgt wie in QMH 00-4 Auftragsbearbeitung beschrieben. Die Teilnehmer werden bei etwaigen Veränderungen in der Programmgestaltung umgehend informiert. Den Teilnehmern wird im Anschluss des Programms Gelegenheit gegeben, Einspruch zu erheben oder Wünsche mitzuteilen. Die Teilnehmer werden schriftlich informiert sofern Produkte und Dienstleistungen extern beschafft werden.

5.3.3.Prüfgegenstand

Die Auswahl der Prüfgegenstände obliegt dem Eignungsprüfungsmanager. Abhängig von der Messaufgabe stellt die esz AG einen Prüfgegenstand oder beschafft einen geeigneten Prüfgegenstand von einem Lieferanten oder Partner. Die Beschaffung und der Umgang mit dem Prüfgegenstand folgen den allgemeinen Vorschriften des akkreditierten Kalibrierlaboratoriums der esz AG. Die Lagerung erfolgt im klimatisierten, zugangsgeschützten Laborbereich der esz AG.

Stabilität: Die Eignung als Prüfgegenstand, insb. im Hinblick auf die Stabilität, wird im Vorfeld untersucht und bestätigt – siehe auch Abschnitt 5.3.1.4. Stabilitätskriterien. Die Freigabe als Prüfgegenstand erfolgt nach bestandener Kalibrierung und Stabilitätsuntersuchung.

Der Prüfgegenstand ist eindeutig mit einem Aufkleber „EIGNUNGSPRÜFUNGSGEGENSTAND“ gekennzeichnet. Die Handhabung ist im technischen Protokoll beschrieben. Bei Bedarf wird die Bedienungsanleitung zur Verfügung gestellt.

Der Umgang mit verloren gegangenen oder beschädigten Prüfgegenständen wird in der Auftragsbestätigung bzw. den AGBs der esz AG geregelt. Anforderungen und Anweisungen für den Transport sind im Bedarfsfall Teil der technischen Beschreibung. Erfolgte Lieferungen werden durch die Teilnehmer an die Eignungsprüfungsmanagerin kommuniziert.

5.3.4.Planung und Programm

Für die Planung eines Eignungsprüfungsprogramms ist das QM-Team der esz AG zuständig. Im Zuge der Planung werden folgende Unterlagen erstellt:

- eine technische Beschreibung (technisches Protokoll), die Informationen zum Anbieter, Koordinator, Messgröße, Verfahren, Prüfgegenstand, Terminplanung und ggf. Teilnahmekriterien enthält (DIN EN ISO/ IEC 17043:2023 7.1.2). Die Vorlage dafür ist Vorlage-EP-Technisches-Protokoll-ABXXXXXX.docx
- ein Zeitplan für den Ablauf inkl. Fristen (auch für Stabilitäts- und ggf. Homogenitätsprüfungen)

Für die Erfassung der Teilnehmerergebnisse dient die Messdatenerfassungsvorlage, auszufüllen durch die Teilnehmer, im Excel-Format.

Wahl des Verfahrens: Bei der esz AG kommen bevorzugt akkreditierte Verfahren zum Einsatz. Das zu verwendende Verfahren wird gemäß den individuellen Anforderungen einer Eignungsprüfung durch den für die Messgröße zuständigen Laborleiter der esz AG oder durch den Eignungsprüfungskoordinator ausgewählt. Sofern Besonderheiten zu beachten sind, werden diese in der technischen Beschreibung den Teilnehmern mitgeteilt. Bei Bedarf wird die Bedienungsanleitung der Geräte zur Verfügung gestellt.

Alle Vorlagen sind im Ordner K:\Intranet\DKD-QS\DKD\Eignungsprüfungsanbieter abgelegt. Die vorbereiteten Unterlagen werden abgelegt und im Vorfeld der Eignungsprüfung an die Teilnehmer verteilt. Dabei wird die Vertraulichkeit jederzeit gewahrt.

EP werden primär auf Kundennachfrage angeboten. Bei konkreten Anlass werden weitere potentielle Teilnehmer kontaktiert. Die Teilnehmerzahl muss so gewählt werden, dass eine Auswertung nach einer der Verfahren in QMH 5.4 möglich ist.
Kriterium für die Zulassung: Teilnehmer sollte ein Kalibrierlabor sein, das sich an der ISO 17025 orientiert (Akkreditierung oder Selbsterklärung).

5.3.5.Durchführung

Wie im Plan festgelegt, wird zunächst eine Erstmessung / Ausgangsmessung bei der esz AG eingeleitet. Verpackung und Versand erfolgt analog zum Versand regulärer Kundengeräte nach der Kalibrierung2. Die Kommunikation mit den Teilnehmern läuft über die Eignungsprüfungsmanagerin, so auch die Benachrichtigungen über Versand und Lieferstatus der Prüfgegenstände, Entgegennahme von Versandbestätigungen, Verteilung der technischen Beschreibung und weitern Informationen an den Teilnehmern. Allfällige Änderungen im zeitlichen Ablauf werden zeitnah nach Bekanntwerden betroffenen Teilnehmern kommuniziert. Die Teilnehmer übermitteln nach erfolgter Messung ihre Daten an pt@esz-ag.de.

[2] Der Umgang mit Kundengeräten ist in der AA0020 beschrieben.

5.3.6.Auswertung und Leistungsbewertung

Die Auswertung der Ergebnisbericht-Vorlagen wird durch die QM der esz AG mit Hilfe einer validierten Auswertungsexcel vorgenommen. Dabei kommen statistische Verfahren zum Einsatz wie im Abschnitt 5.4 dieses Dokuments ausführlich beschrieben. Die Bewertung der Teilnehmer-Ergebnisse sowie der Umgang mit Ausreißern erfolgt ebenfalls wie im Abschnitt 5.4 beschrieben und den Teilnehmern im Plan (technische Beschreibung) kommuniziert. Die Auswertung erzeugt automatisch Informationen und graphische Darstellungen zur Leistungsbewertung (Beispiel).

Sollten bei der Auswertung Auffälligkeiten oder Ausreißer festgestellt werden, werden betroffene Teilnehmer informiert. Der Umgang mit Ausreißern wird den Teilnehmern mitgeteilt. Bei Abweichungen der Auswertung vom Plan wird die Zustimmung der Teilnehmer eingeholt.

Die Messergebnisse der Teilnehmer werden in 3 Schritten auf Auffälligkeiten geprüft.

Schritt 1:
Unmittelbare Sichtung des Messergebnisses nach Abgabe durch den Teilnehmer und Prüfung der Messergebnisse auf Plausibilität. Dies beinhaltet beispielsweise:

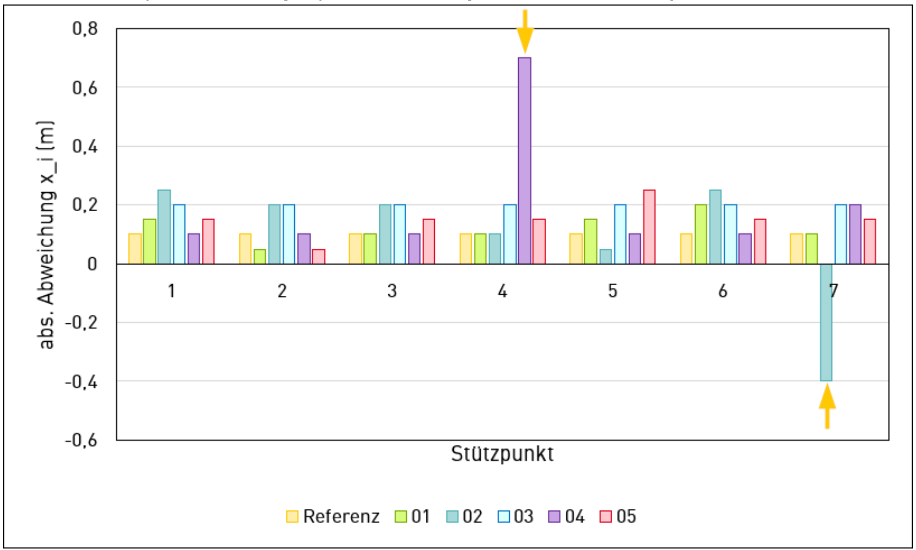
- Ist die Einheit des Messwertes plausibel?
  (Beispiel: statt 32,000 V wurden 32,000 mV abgegeben)
- Ist das Vorzeichen des Messwertes plausibel?
  (Beispiel: statt 32,000 V wurden -32,000 V abgegeben)
- Ist Größenordnung des Messwertes plausibel?
  (Beispiel: statt 32,000 V wurden 320,00 V abgegeben)
- Gibt es einen Zahlendreher im Messwert?
  (Beispiel: statt 32,000 V wurden 23,00 V abgegeben)
- Ist die Größenordnung der Messunsicherheit plausibel?
  (Beispiel: Teilnehmer ist mit 1 mV akkreditiert, es wird aber 0,1 mV oder 1 V als Messunsicherheit abgegeben)
- Fehlen Messwerte bzw. weichen von den Stützpunkten ab?
- Fehlen Messunsicherheiten?

Werden Auffälligkeiten festgestellt, so wird der entsprechende Teilnehmer kontaktiert und gebeten das abgegebene Messergebnis noch einmal zu prüfen und entweder zu bestätigen oder zu korrigieren. Der Teilnehmer erhält dabei keinen Hinweis darauf, wo die Auffälligkeit festgestellt wurde, d.h. an welchem Messpunkt und ob es sich um Messwert oder Messunsicherheit handelt.

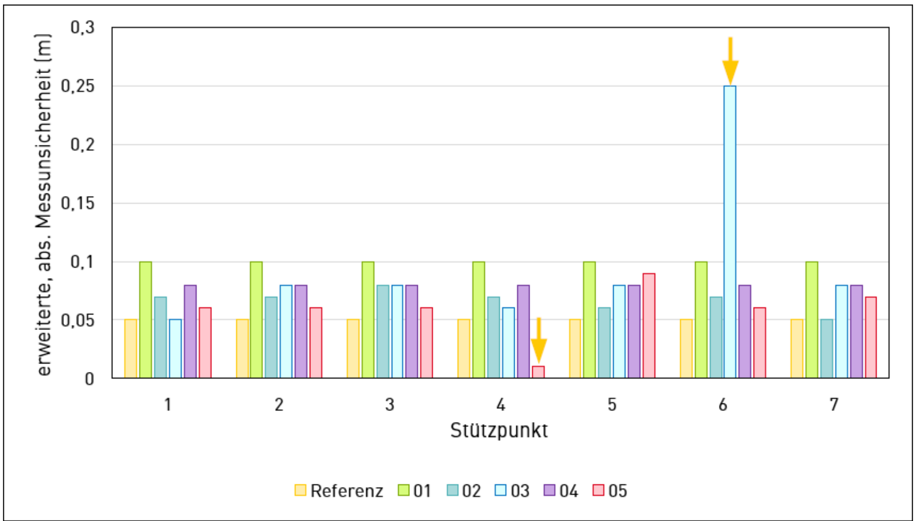
Schritt 2:
Nach Abschluss aller Messungen werden die Messergebnisse aller Teilnehmer in die Auswertungsexcel eingetragen. Die Messwerte und Messergebnisse werden anhand der Histogramme graphisch verglichen. Dabei werden die eingereichten Ergebnisse der Teilnehmer auch im Hinblick auf die Stabilitätsmessungen betrachtet. Dies ist insbesondere

dann relevant, wenn Auffälligkeiten bei zeitlich aufeinander folgenden Teilnehmern vorliegen. Zudem werden die Auffälligkeiten im Messwert auch in Bezug auf die berichtete Messunsicherheit betrachtet. So kann bspw. ein abweichender Messwert eines Teilnehmers plausibel sein, wenn er mit einer hinreichend großen Messunsicherheit berichtet wird. Bei groben Ausreißern in Messwert oder Messunsicherheit wird der entsprechende Teilnehmer unter Anwendung des Vorgehens aus Schritt 1 kontaktiert.

Beispiel für den graphischen Vergleich der Messergebnisse:



Teilnehmer 4 weist eine Auffälligkeit im Messwert beim 4. Stützpunkt auf. Teilnehmer 2 weist eine Auffälligkeit im Messwert am 7. Stützpunkt auf.



Teilnehmer 5 weist eine Auffälligkeit in der Messunsicherheit beim 4. Stützpunkt auf. Teilnehmer 3 weist eine Auffälligkeit in der Messunsicherheit am 6. Stützpunkt auf.

<u>Schritt 3:</u>

Die Auswertung wird durchgeführt. Bei Anwendung von Methode B, C oder D zur Bestimmung der Referenzdaten ist eine Eliminierung von Ausreißern (Methode B und D) bzw. eine Modifizierung von Ausreißern (Methode C) enthalten (vgl. Kapitel 5.4). Wird das Messergebnis eines Teilnehmers in diesem Schritt durch den Algorithmus der jeweiligen Methode als Ausreißer erkannt, so fließt dieses Messergebnis nicht in die Bestimmung der Referenzdaten ein, die Leistung des Teilnehmers wird weiterhin bewertet. Der Teilnehmer wird in diesem Schritt nicht mehr vorab informiert, sondern erhält die Information mit dem Ergebnis der Auswertung im Bericht zu Eignungsprüfung.

Anmerkungen und Erkenntnisse über Verfahren und Einflussfaktoren sowie allfällige weiteren Empfehlungen werden im Ergebnisbericht dokumentiert.

5.3.7. Berichtserstellung

Die Ergebnisberichte werden zunächst als Entwurfsversion basierend auf der Ergebnisbericht-Vorlage erstellt um den Teilnehmern die Gelegenheit zur Stellungnahme zu ermöglichen, dabei erhält kein einzelner Teilnehmer bevorzugten Zugang zu Ergebnissen. Nachfolgend wird der Abschlussbericht finalisiert und den Teilnehmern zugesandt. Ergebnisberichte entsprechen den Anforderungen der DIN EN ISO/ IEC 17043:2023 7.4.3. Sie sind durch eine Berichtsnummer eindeutig gekennzeichnet. Bei nachträglichen Änderungen sind ein Verweis auf den Originalbericht und Informationen zu den Gründen der Neuausgabe enthalten. Die Freigabe der Berichte erfolgt ausschließlich durch Mitarbeiter der Qualitätsabteilung der esz AG.

5.4. Auswertung und statistische Analyse von Eignungsprüfungsdaten

In diesem Abschnitt wird die Vorgehensweise bei der statistischen Auswertung von Eignungsprüfungsdaten beschrieben. Die nachfolgenden Darstellungen orientieren sich dabei an DIN ISO 13528 (Methode C), sowie an den Ausführungen von Maurice Cox in „The evaluation of key comparison data - An introduction" (Metrologia 39, 2002) (Methode B). Des Weiteren finden die Untersuchungen von S. Pommé und J. Keightley zu „Determination of a reference value and its uncertainty through a power-moderated mean" (Metrologia 52, 2015) (Methode D) Anwendung.

5.4.1. Eignungsprüfungsdaten der teilnehmenden Labore

Sei X die zu untersuchende Messgröße und N die Anzahl der teilnehmenden Laboratorien mit $N \geq 1$. Die von Labor i, mit $i \in [1; N]$, übermittelten Daten beinhalten:

- das Verfahren zur Kalibrierung der Messgröße X am Prüfgegenstand, sowie dessen Beschreibung,
- den für dieses Verfahren angesetzten Kalibrierwert $x_{i;\mathrm{cal}}$
- den mit diesem Verfahren ermittelten Messwert $x_{i;\mathrm{meas}}$

Aus diesen Angaben ergibt sich u.a. die Abweichung x_i des Teilnehmers i

$$x_i = x_{i;\mathrm{meas}} - x_{i;\mathrm{cal}}. \tag{1.1}$$

Die Verwendung der Abweichung x_i anstelle von Messwert $x_{i;\mathrm{meas}}$ oder Kalibrierwert $x_{i;\mathrm{cal}}$ bietet sich in der statistischen Auswertung an, da damit auch Messungen abgedeckt sind, für die kein einheitlicher, für alle Teilnehmer gleicher Kalibrierwert vorliegt.

Des Weiteren sind von den Teilnehmern folgende Angaben zu übermitteln:

- die dazugehörige erweiterte Messunsicherheit $U(x_i)$, sodass das berichtete Intervall $x_i \pm U(x_i)$ den wahren Wert der Messgröße mit einer Wahrscheinlichkeit von 95 % enthält,

- sowie falls keine Normalverteilung für die Wahrscheinlichkeitsdichteverteilung $p_i(x)$ angenommen werden kann, eine geeignete Annahme für diese (z.B. Student-t, Rechteck, Dreieck, U/arcsin oder Sonstige),
- eine Angabe darüber, für welche Messpunkte mit dem verwendeten Verfahren eine Akkreditierung vorliegt,
- der CMC für alle akkreditierten Messpunkte.

Nachfolgend wird von einer Auswertung in „absoluten Einheiten“ ausgegangen. Alternativ kann auch eine Auswertung in „relativen Einheiten“ durchgeführt werden. In diesem Fall ist das Messergebnis $x_i \pm U(x_i)$ bestehend aus Abweichung x_i und erweiterter Messunsicherheit $U(x_i)$ durch den Kalibrierwert $x_{i;\mathrm{cal}}$ zu teilen. Für die nachfolgende Diskussion hat die Entscheidung absolut oder relativ keine Auswirkung. Ohne Beschränkung der Allgemeinheit wird daher nachfolgend von absoluten Einheiten ausgegangen.

Aus der Kenntnis der erweiterten Unsicherheit $U(x_i)$ und der angenommenen Wahrscheinlichkeitsdichteverteilung $p_i(x)$ ist $u(x_i)$, d.h. die einfache Standardunsicherheit, bekannt. Kann beispielsweise eine Normalverteilung angenommen werden, so gilt $u(x_i) = U(x_i)/k$, mit $k = 2$. Bei einer Student-t Verteilung ist k entsprechend der Anzahl an Freiheitsgraden anzupassen. Für den Fall, dass eine Rechteck-, Dreieck- oder U/arcsin-Verteilung berichtet wird, wird angenommen, dass die berichtete erweiterte Unsicherheit $U(x_i)$ gerade die Halbbreite dieser Verteilungen definiert, d.h. $u(x_i) = U(x_i)/\sqrt{G}$, wobei $G = 3$ für eine Rechteckverteilung, $G = 6$ für eine Dreieckverteilung und $G = 2$ für eine U/arcsin-Verteilung. Das übermittelte Ergebnis $x_i \pm U(x_i)$ charakterisiert demnach eine Wahrscheinlichkeitsdichteverteilung $p_i(x)$, die für die Werte x der Messgröße X vorliegt, wobei x_i der Erwartungswert ist und $u(x_i)$ die Standardunsicherheit. Aufgrund des zentralen Grenzwertsatzes und der hierfür vorausgegangen Linearitätsannahme der mathematischen Beschreibung der Messung, ist $p_i(x)$ in vielen Fällen näherungsweise durch eine Normalverteilung gegeben. Wird explizit keine andere Wahrscheinlichkeitsdichte genannt (z.B. falls die berichtete Unsicherheit $U(x_i)$ einer Herstellerspezifikation entnommen ist oder ein dominanter, nicht normalverteilter Beitrag in der Unsicherheitsbetrachtung vorliegt), so wird von einer Normalverteilung ausgegangen und

$$p_i(x) = \frac{1}{\sqrt{2\pi}u(x_i)} e^{-\frac{(x-x_i)^2}{2u^2(x_i)}} \tag{1.2}$$

kann angenommen werden. Abbildung 3.1 zeigt Gleichung (1.2) graphisch dargestellt.

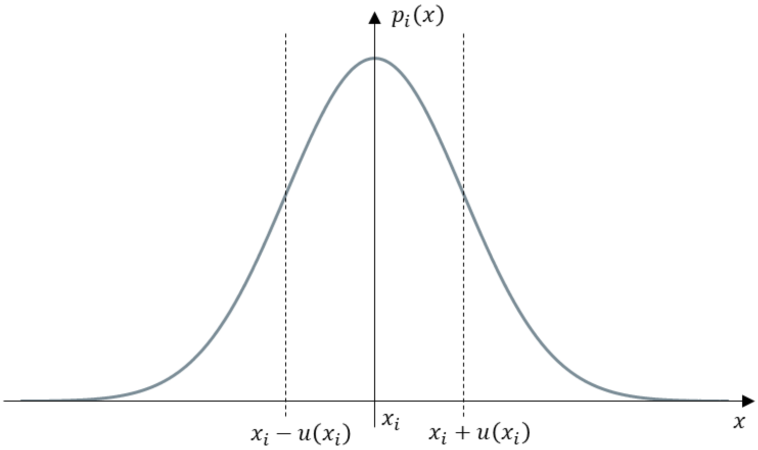


Abbildung 3.1: Darstellung einer Normalverteilung mit Erwartungswert x_i und Standardabweichung $u(x_i)$.

5.4.2. Referenzwert

Zur Bewertung der Leistung eines teilnehmenden Labors i werden ein Referenzwert, sowie eine Referenzunsicherheit benötigt. Für die Bestimmung dieser Referenzdaten wird eine der nachfolgend beschriebenen Methoden A, B, C oder D gewählt. Es werden zunächst keine Stabilitätskriterien oder Stabilitätsbeiträge berücksichtigt. Diese werden in Abschnitt 5.3.4 diskutiert. Insbesondere wird zwischen der Unsicherheit des Referenzwertes und der Stabilitätsunsicherheit unterschieden, sodass Stabilitätsbeiträge stets separat berücksichtigt werden und nicht in der Referenzunsicherheit enthalten sind.

In den Methoden B, C und D werden Teilnehmerdaten in der Bestimmung der Referenzdaten berücksichtigt. Da der Referenzwert eine metrologische Rückführung aufweisen muss, können nur Teilnehmer in die Bestimmung der Referenzdaten einfließen, die am jeweiligen Messpunkt für das verwendete Verfahren akkreditiert sind und eine Unsicherheit größer oder gleich ihrem CMC berichten. Akkreditierte Teilnehmer, die eine Unsicherheit kleiner als den CMC berichten, können in der Bestimmung der Referenzdaten nicht berücksichtigt werden. Zwar könnte in diesem Fall der CMC für die Berechnung der Referenzdaten verwendet werden und die Leistungsbewertung über die berichtete MU erfolgen, doch läuft man damit in das Problem, dass die Korrelation im E_n-Wert nicht eindeutig klar ist und sogar mathematisch nicht definierte E_n-Werte resultieren können (bspw. falls in Methode B die Teilnehmerunsicherheit kleiner ist als die Referenzunsicherheit). Mehr zum Thema Korrelation ist in Abschnitt 5.4.3 diskutiert. Bei einer Eignungsprüfung, an der lediglich nicht akkreditierte Labore teilnehmen und Methode B, C oder D angewandt wird, muss zusätzlich zur Referenzunsicherheit, die aus den Berechnungen der genannten Methoden resultiert noch ein Beitrag für die Rückführung addiert werden. Diese Vorgehensweise wird jedoch vermieden und in einem solchen Fall, falls möglich, Methode A bevorzugt. Berichtet ein Labor für die Messpunkte der Eignungsprüfung mehrere Messergebnisse, d.h. ein Labor stellt mehrere Teilnehmer, so kann dieses Labor mit maximal einem Messergebnis (= ein Teilnehmer) in die Bestimmung der Referenzdaten einfließen. Diese Einschränkung ist nötig um Korrelationen zwischen Messergebnissen, die den Referenzwert beeinflussen zu vermeiden, bzw. zu minimieren. Das Labor kann vorab entscheiden, welches Messergebnis dafür verwendet werden soll. Für alle anderen Messergebnisse kann eine standardmäßige E_n-Auswertung erfolgen. Macht das Labor keine Angabe darüber, welches Messergebnis für die Bestimmung der Referenzdaten herangezogen werden soll, so entscheidet die esz AG darüber.

Falls nicht anders angegeben, fließt zusätzliche zu den Teilnehmerergebnissen die Anfangsmessung in die Bestimmung der Referenzdaten ein. Bei N vorliegenden Teilnehmerergebnissen führt dies zu maximale $N+1$ Datensätzen, die für die Bestimmung der Referenzdaten herangezogen werden können. Die Abschluss- und Zwischenmessungen dienen lediglich der Stabilitätsüberwachung. Durch diese Vorgehensweise wird verhindert, dass Ergebnisse des Labors, das die Stabilitätsmessung durchführt, stärker gewichtet werden als die der anderen Teilnehmer. Zudem werden dadurch, wie bereits oben beschrieben, Korrelationen zwischen Messergebnissen, die den Referenzwert bilden, vermieden. Standardmäßig werden die Anfangs-, Zwischen und Abschlussmessungen durch die esz AG

durchgeführt, dies wird in der nachfolgenden Beschreibung angenommen. Abweichend davon können auch Teilnehmer oder das Referenzlabor die Stabilitätsmessungen durchführen.
Je nach Kundenwunsch, Teilnehmerfeld und/oder berichteter Ergebnisse, kann es zusätzlich zu den oben beschriebenen Kritiereien von Vorteil sein, für die Bestimmung des Referenzwertes nur die Ergebnisse von $M \leq N+1$ ausgewählten Teilnehmern zu berücksichtigen. Mögliche Auswahlkriterien sind:

- Ein oder mehrere Teilnehmer möchten explizit nicht in die Bestimmung des Referenzwertes einfließen
- Die Anfangsmessung (bspw. der esz AG) fließt nicht in die Bestimmung des Referenzwertes ein.

Weitere Auswahlkriterien sind je nach Kundenwunsch oder Notwendigkeit (z.B. bei zeitlich linearer Drift der Daten oder einem Sprung in den Daten, hervorgerufen durch einen Transporteinfluss) möglich. Für $N \leq 2$ Teilnehmer erfolgt die Auswertung ausschließlich über Methode A. Für eine Teilnehmerzahl $N \geq 3$ wird je nach Anforderung an die Referenzunsicherheit standardmäßig Methode A oder D verwendet. Wird die Anfangsmessung nicht in der Bestimmung der Referenzdaten berücksichtigt, so sollten für die Anwendung von Methode D mindestens $N = 4$ Teilnehmerdaten vorliegen. An dieser Stelle sei angemerkt, dass die Anwendung der Methoden B, C und D bei Vorliegen von zwei Datensätzen (bilaterale Eignungsprüfung) nicht zielführend ist, da die resultierende Leistungsbewertung vorhersehbar ist und beide Teilnehmer mit Sicherheit bestehen werden. Bei 3 Datensätzen können die Methoden B und D theoretisch angewandt werden, allerdings besteht das Risiko, dass bei Vorliegen eines Ausreißers sich der effektive Datensatz für die Bestimmung der Referenzdaten wieder auf zwei Datensätze reduziert. Daher sollten für die Anwendung aller statistischen Methoden mindestens 4 Datensätze vorliegen (bspw. $N = 3$ Teilnehmerdaten + Anfangsmessung). Dennoch kann es je nach Datenlage möglich sein, dass bei einer geringen Teilnehmerzahl Methode D nicht anwendbar ist. Dies ist beispielsweise der Fall, wenn das Vorhandensein von Ausreißern die effektive Anzahl der Messergebnisse, die den Referenzwert bestimmen auf weniger als 4 Datensätze reduziert. Darüber hinaus kann Methode D auch nicht sinnvoll angewandt werden, falls es innerhalb der Teilnehmer, die in die Bestimmung der Referenzdaten einfließen, einen großen Anteil gibt, der die gleiche metrologische Rückführung aufweist (bspw. auf das gleiche Nationalinstitut). In diesem Fall ist von einer starken Korrelation der Teilnehmerdaten auszugehen, wodurch die Referenzdaten verfälscht werden können. In solchen oder anderen Fällen kann es notwendig sein nach Sichtung und Analyse aller berichteten Messergebnisse die Auswertung anhand von Methode A anstatt Methode D (oder B/C) durchzuführen. Über diese Möglichkeit werden die Teilnehmer bereits im technischen Protokoll vorab informiert. Eine genaue Analyse zur Bestimmung der Referenzdaten erfolgt im Bericht der Eignungsprüfung.

A) Referenzwert wird durch ein Referenzlabor bestimmt

Methode A entspricht der weit verbreiteten Vorgehensweise, den Referenzwert einer Eignungsprüfung durch ein Referenzlabor festlegen zu lassen. Als Referenzlabor eignen sich typischerweise Nationalinstitute, aber auch andere Labore, die sich für die gesuchte Messgröße auszeichnen und akkreditiert sind. Durch die Wahl des Referenzlabors wird der Referenzwert und folglich auch seine Unsicherheit im Wesentlichen definiert, bzw. die Messung des Referenzlabors legt den Referenzwert fest. Sei x_0 der Abweichungswert (Differenz aus Messwert und Kalibrierwert entsprechend Gleichung

(1.1)) des Referenzlabors für die gesuchte Größe und $u(x_0)$ die einfache Standardunsicherheit des Referenzlabors, so wird definiert, dass

$$x_{\text{ref}} \equiv x_0 \tag{2.1}$$

$$u(x_{\text{ref}}) \equiv u(x_0) \tag{2.2}$$

Insbesondere fließen die Ergebnisse der N Teilnehmer nicht in die Bestimmung des Referenzwertes und dessen Unsicherheit ein. Die Messung des Referenzlabors sollte, falls möglich, in der Mitte der Eignungsprüfung stattfinden.

B) Referenzwert wird als gewichteter Mittelwert berechnet

Methode B orientiert sich u.a. an Procedure A aus „The evaluation of key comparison data - An introduction" (M. Cox, Metrologia 39, 2002) und verwendet den gewichteten Mittelwert und dessen Standardabweichung als Referenzwert und –unsicherheit. Der statistische Hintergrund der Methode ist die Annahme, dass eine normalverteilte Größe vorliegt, wobei jeder Teilnehmer eine entsprechende Stichprobe berichtet. Der gewichtete Mittelwert ist definiert als

$$x_{\text{ref}} = u^2(x_{\text{ref}}) \sum_{i=1}^{M} \frac{x_i}{u^2(x_i)} \tag{2.2}$$

mit

$$u^2(x_{\text{ref}}) = \left(\sum_{i=1}^{M} \frac{1}{u^2(x_i)} \right)^{-1} . \tag{2.3}$$

Dabei ist M die Anzahl der Messergebnisse, die für die Bestimmung des gewichteten Mittelwertes herangezogen wird ($M \leq N+1$, N Teilnehmer, sowie die Anfangsmessung der esz AG). Im Gegensatz zu Methode A können hier Teilnehmerdaten berücksichtigt werden. Die berichteten Werte x_i der Teilnehmer werden entsprechend ihrer dazugehörigen Standardunsicherheit $u(x_i)$ gewichtet, d.h. Abweichungswerte, die mit einer verhältnismäßig kleinen Unsicherheit berichtet werden, werden stärker gewichtet und haben einen größeren Einfluss auf den gewichteten Mittelwert und somit den Referenzwert, als Labore, die eine große Unsicherheit angeben. Charakteristisch für Methode B ist demnach, dass sowohl Teilnehmerwert, als auch Teilnehmerunsicherheit in die Bestimmung des Referenzwertes und dessen Unsicherheit einfließen. Insbesondere ist auch das Zusammenspiel zwischen Abweichungswerten und Unsicherheiten charakteristisch und soll im Nachfolgenden anhand von 3 Spezialfällen veranschaulicht werden.

- Beispiel 1: Alle Teilnehmer berichten eine ähnlich große Unsicherheit
  Der Einfachheit halber können zunächst zwei Teilnehmer betrachtet werden, die folgende Ergebnisse berichten:

Teilnehmer 1: $x_1 = 1$; $u(x_1) = 0{,}5$	(2.4)
Teilnehmer 2: $x_1 = 2$; $u(x_1) = 0{,}5$	(2.5)
Referenzwert: $x_{\text{ref}} = 1{,}5 = \frac{x_1+x_2}{2}$; $u(x_{\text{ref}}) = 0{,}35 = \frac{u(x_1)}{\sqrt{2}}$	(2.6)

  Im gezeigten Beispiel berichten sowohl Teilnehmer 1, als auch Teilnehmer 2 eine ähnlich große Unsicherheit. Konsequenterweise werden beide Teilnehmerergebnisse im Referenzwert daher gleich gewichtet, sodass der Referenzwert im Wesentlichen durch den arithmetischen Mittelwert gegeben ist. Das ist auch in Abbildung 3.2 gut zu erkennen.

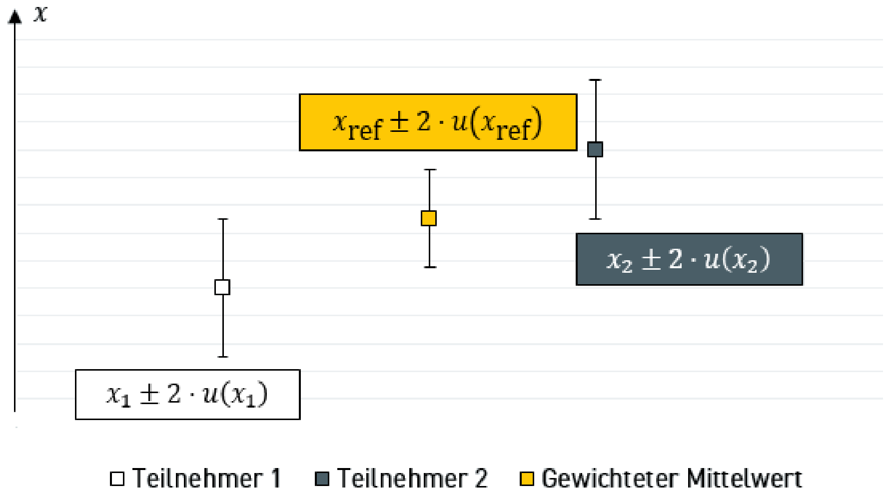


Abbildung 3.2: Im Grenzfall, dass alle Teilnehmer eine ähnlich große Unsicherheit berichten, reproduziert der gewichtete Mittelwert den arithmetischen Mittelwert.

Das Verhalten, das in Abbildung 3.2 und in Gleichung (2.6) gezeigt ist, kann auch auf eine beliebige Anzahl M an Teilnehmern ausgeweitet werden. Sind die Unsicherheiten aller M Teilnehmer ähnlich groß, d.h. $u(x_i) \simeq u(x)$, so gilt

$$x_{\mathrm{ref}} = u^2(x_{\mathrm{ref}}) \sum_{i=1}^{M} \frac{x_i}{u^2(x_i)} \simeq \frac{u^2(x_{\mathrm{ref}})}{u^2(x)} \sum_{i=1}^{M} x_i. \tag{2.7}$$

Weiter gilt für die Referenzunsicherheit

$$u^2(x_{\mathrm{ref}}) = \left(\sum_{i=1}^{M} \frac{1}{u^2(x_i)} \right)^{-1} \simeq \frac{u(x)}{\sqrt{M}}. \tag{2.8}$$

Damit folgt, dass der gewichtete Mittelwert im hier gezeigten Grenzfall den arithmetischen Mittelwert reproduziert.

$$x_{\mathrm{ref}} = \frac{1}{M} \sum_{i=1}^{M} x_i. \tag{2.9}$$

Insbesondere ist hier die Unsicherheit des arithmetischen Mittelwertes nicht durch die empirische Standardabweichung gegeben, d.h. die Streuung der Abweichungswerte x_i wird nicht berücksichtigt. In den Gleichungen (2.6) und (2.8), sowie in Abbildung 3.2 ist zu erkennen, dass die Referenzunsicherheit kleiner ist als die Teilnehmerunsicherheiten. Zudem wird sie umso kleiner, je mehr Teilnehmer in die Bestimmung des Referenzwertes einfließen, d.h. umso größer M ist (vgl. Proportionalität zu $1/\sqrt{M}$). Vereinfacht gesprochen spiegelt das die Annahme wider, dass der Referenzwert umso besser bekannt ist, je mehr Wissen in dessen Bestimmung einfließt.

- <u>Beispiel 2: Ein Teilnehmer berichtet eine deutlich kleinere Unsicherheit als die Anderen</u>
  Zunächst werden auch hier zwei Teilnehmer betrachtet. Dieses Mal berichtet Teilnehmer 1 eine deutlich kleinere Unsicherheit als Teilnehmer 2:

Teilnehmer 1: $x_1 = 1;\ u(x_1) = 0{,}2$	(2.10)
Teilnehmer 2: $x_1 = 2;\ u(x_1) = 0{,}5$	(2.11)
Referenzwert: $x_{\mathrm{ref}} = 1{,}1 \simeq x_1;\ u(x_{\mathrm{ref}}) = 0{,}19 \simeq_{<} u(x_1)$	(2.12)

In Gleichung (2.12) und Abbildung 3.3 sieht man sowohl rechnerisch, als auch graphisch, dass Teilnehmer 1, also der Teilnehmer mit der kleinen Unsicherheit die Bestimmung des Referenzwertes dominiert. Nach wie vor befindet sich der Referenzwert zwischen den Werten von Teilnehmer 1 und Teilnehmer 2. Zudem zeigt sich erneut, dass die Referenzunsicherheit kleiner ist als alle Teilnehmerunsicherheiten. Im Gegensatz zu Beispiel 1, sind nun aber sowohl Referenzwert, als auch Referenzunsicherheit näherungsweise durch die Werte von Teilnehmer 1 gegeben. Dieses Verhalten kann erneut auf M Teilnehmer erweitert werden. Das Ergebnis ist, dass ein Teilnehmer m, der eine signifikant kleinere Unsicherheit berichtet als alle anderen Teilnehmer, zu einer Art Referenzlabor wird, denn

$$x_{\mathrm{ref}} \simeq x_m \tag{2.13}$$

und

$$u(x_{\mathrm{ref}}) \simeq u(x_m) \tag{2.14}$$

Allerdings ist das einzige, das Teilnehmer 1 zu einer Art Referenzlabor entsprechend Methode A macht, die Tatsache, dass er eine kleine Unsicherheit berichtet. Weitere Kriterien, wie sie beispielsweise für die Wahl des Referenzlabors gemäß Methode A angesetzt würden, fließen hier nicht ein. In einem solchen Fall kann zudem die Leistungsbewertung des Teilnehmers m nicht mehr sinnvoll sein, da dieser letztlich mit sich selbst verglichen wird. Diese Tatsache wird auch durch die Berücksichtigung der Korrelation im E_{n}-Wert verdeutlicht, worauf später genauer eingegangen wird.

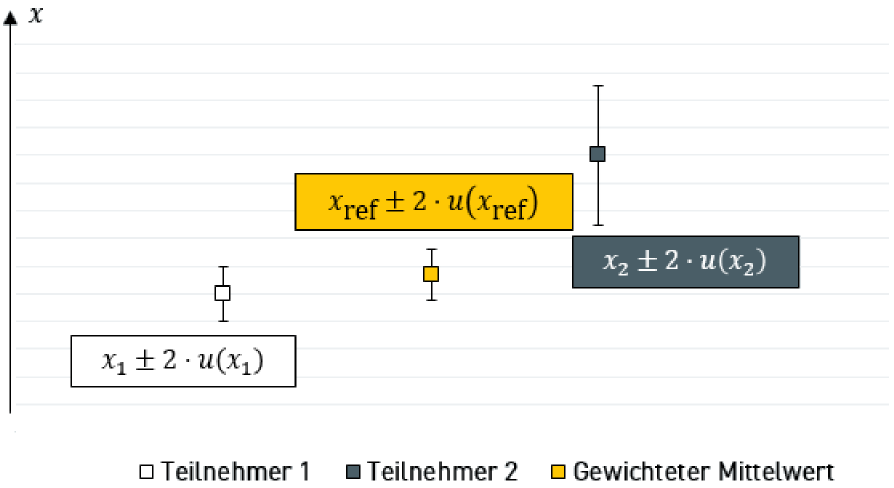


Abbildung 3.3: Berichtet ein Teilnehmer eine signifikant kleinere Unsicherheit als alle anderen Teilnehmer, so wird dieser Teilnehmer im Wesentlichen zu einer Art Referenzlabor, da er die Bestimmung des Referenzwertes und der Referenzunsicherheit dominiert.

- <u>Beispiel 3: Der Teilnehmer mit der kleinen Unsicherheit berichtet einen Ausreißer</u>
  Beispiel 2 hat gezeigt, dass ein Teilnehmer, der eine signifikant kleinere Unsicherheit berichtet als alle anderen Teilnehmer, zu einer Art Referenzlabor wird. Diese ist jedoch dann problematisch, wenn dem Ergebnis, das mit einer kleinen Unsicherheit berichtet wird, kein hohes Vertrauen entgegengebracht wird, beispielsweise, weil es einen Ausreißer darstellt. Nachfolgend seien in Teilnehmergruppe 2 vier Teilnehmer enthalten, die der Einfachheit halber alle den gleichen Abweichungswert mit der gleichen Unsicherheit berichten.

$$\text{Teilnehmer 1: } x_1 = 1;\ u(x_1) = 0{,}2 \tag{2.15}$$

$$\text{Teilnehmergrupp 2: } x_2 = 2;\ u(x_2) = 0{,}5 \tag{2.16}$$

$$\boxed{\text{Referenzwert: } x_{\text{ref}} = 1{,}4;\ u(x_{\text{ref}}) = 0{,}16 < u(x_1)} \tag{2.17}$$

Abbildung 3.4 veranschaulicht, dass Teilnehmer 1 als Ausreißer zu erachten ist, allerdings berichtet Teilnehmer 1 nach wie vor eine signifikant kleinere Unsicherheit als die restlichen Teilnehmer.

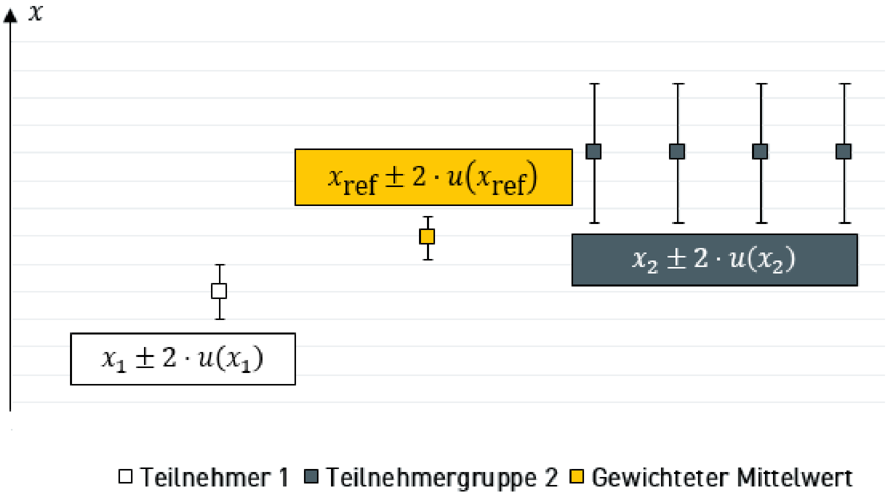


Abbildung 3.4: Liegt ein Ausreißer mit einer signifikant kleinen Unsicherheit vor, so benötigt man ein hinreichend starkes Gegengewicht (Anzahl Teilnehmer in Gruppe 2) um den Einfluss des Ausreißers auf den Referenzwert zu kompensieren.

In Abbildung 3.4 und Gleichung (2.17) ist zu erkennen, dass der Einfluss des Ausreißers mit der kleinen Unsicherheit auf den Referenzwert nach wie vor signifikant ist. Um das starke Gewicht einer kleinen Unsicherheit zu kompensieren, wird ein hinreichend starkes Gegengewicht benötigt. Dieses Gegengewicht wird insbesondere durch die Anzahl der Teilnehmer in Gruppe 2 bestimmt. Je größer die Differenz zwischen der Unsicherheit des Ausreißers und den restlichen Teilnehmerunsicherheiten ist, umso mehr Teilnehmer werden in Gruppe 2 für das „Gegengewicht" benötigt. Es zeigt sich somit, dass Methode B nicht robust ist bzgl. Ausreißern mit kleinen Unsicherheiten. In dieser Situation ist die Wahl von Methode C oder D geeigneter. Hinsichtlich der Referenzunsicherheit zeigt sich erneut, dass diese kleiner ist als alle Teilnehmerunsicherheiten. Im Vergleich zwischen Gleichung (2.12) und (2.17) zeigt sich auch, dass die Referenzunsicherheit kleiner wird, je mehr Teilnehmerdaten in die Bestimmung des Referenzwertes einfließen.

Die Verwendung des gewichteten Mittelwertes als Referenzwert ist nur dann möglich, wenn folgende Bedingungen erfüllt sind:

1. Für die M verwendeten Ergebnisse $x_i \pm U(x_i)$ kann eine Normalverteilung gemäß Gleichung (1.2) angenommen werden.
2. Der χ^2-Test ist positiv, d.h.
$$\Pr(p_\nu(x) > \chi^2_{\text{obs}}) < 1 - \alpha \quad \text{mit} \quad \alpha = 0{,}95, \tag{2.18}$$
wobei $\nu = M - 1$ die Anzahl der Freiheitsgrade der Variable χ^2_{obs} (s. Gleichung (2.19)) ist. Der Term „Pr" ist zu lesen als „Wahrscheinlichkeit, dass". Die Funktion $p_\nu(x)$ stellt die Wahrscheinlichkeitsdichte einer χ^2-Verteilung mit ν Freiheitsgrade dar und ist in Gleichung (2.34) definiert. Die Variable χ^2_{obs} ist definiert als

$$\chi^2_{\text{obs}} = \sum_{i=1}^{M} \frac{(x_i - x_{\text{ref}})^2}{u^2(x_i)}. \tag{2.19}$$

Teilnehmer, die Bedingung 1 erfüllen, jedoch zu einem Fehlschlagen von Bedingung 2 führen, sind sukzessive zu entfernen, solange bis Gleichung (2.18) erfüllt ist. Hierbei werden sukzessive die Teilnehmer mit dem größten Beitrag zu Gleichung (2.19) entfernt.

In den diskutierten Beispiel 1-3 ist der χ^2-Test jeweils bestanden, wenn auch nur knapp in Beispiel 3. In diesem Beispiel wurde festgestellt, dass Methode B kein robustes Verfahren ist, falls Ausreißer mit kleinen Unsicherheiten berichtet werden. Tatsächlich gibt es hier noch die Absicherung des χ^2-Testes, der in einer Situation, wie in Beispiel 3 dargestellt auch fehlschlagen kann, allerdings ist dies nicht immer der Fall.

Zusammenfassend kann festgehalten werden, dass Methode B in jedem Fall zu einer Referenzunsicherheit führt, die kleiner ist als alle Einzelunsicherheiten der Teilnehmer, und daher insbesondere auch kleiner als die kleinste Teilnehmerunsicherheit, die in die Bestimmung des Referenzwertes einfließt. Es kann analytisch gezeigt werden, dass

$$\forall_{i \in [1;M]} \; u(x_{\text{ref}}) < u(x_i). \tag{2.20}$$

Für Teilnehmer, die nicht in die Bestimmung des Referenzwertes einfließen können, da sie beispielsweise Bedingung 1 oder 2 nicht erfüllen können, ist Gleichung (2.20) nicht automatisch gültig. Weiter kann gezeigt werden, dass die Referenzunsicherheit umso kleiner wird, je mehr Teilnehmerdaten in die Bestimmung des Referenzwertes einfließen. Dabei haben insbesondere Teilnehmerergebnisse, die mit einer verhältnismäßig kleinen Unsicherheit berichtet werden, einen dominanten Einfluss auf den Referenzwert. An dieser Stelle sei darauf hingewiesen, dass die Referenzunsicherheit keine Stabilitätsbeiträge enthält und diese gesondert betrachtet werden.

Anmerkungen zu Bedingung 1:

Nachfolgende Diskussion stellt eine Grundlage für die Definition des gewichteten Mittelwertes dar und liefert damit auch eine Begründung dafür, warum die Annahme von Normalverteilungen erforderlich ist. Ohne Beschränkung der Allgemeinheit sei $M = 2$. Es sei angenommen, dass Labor 1 das Ergebnis $x_1 + U(x_1)$ mit $U(x_1) = 2 \cdot u(x_1)$ angibt und Labor 2 ebenso $x_2 + U(x_2)$ mit $U(x_2) = 2 \cdot u(x_2)$ berichtet. In beiden Fällen sei die Wahrscheinlichkeitsdichteverteilung des Werte x der Messgröße X durch eine Normalverteilung gegeben, d.h.

$$p_1(x) = \frac{1}{\sqrt{2\pi}u(x_1)} e^{-\frac{(x-x_1)^2}{2u^2(x_1)}}, \tag{2.21}$$

$$p_2(x) = \frac{1}{\sqrt{2\pi}u(x_2)} e^{-\frac{(x-x_2)^2}{2u^2(x_2)}}. \tag{2.22}$$

Das Produkt der beiden Wahrscheinlichkeitsdichten ist gegeben durch

$$p(x) \equiv p_1(x) \cdot p_2(x) = \frac{1}{2\pi u(x_1)u(x_2)} e^{-\frac{1}{2}\left[\frac{(x-x_1)^2}{u^2(x_1)} + \frac{(x-x_2)^2}{u^2(x_2)}\right]}, \tag{2.23}$$

Der Exponent kann wie folgt umgeformt werden

$$\frac{(x-x_1)^2}{u^2(x_1)} + \frac{(x-x_2)^2}{u^2(x_2)} = x^2\left(\frac{1}{u^2(x_1)} + \frac{1}{u^2(x_2)}\right) - 2x\left(\frac{x_1}{u^2(x_1)} + \frac{x_2}{u^2(x_2)}\right) + \left(\frac{x_1^2}{u^2(x_1)} + \frac{x_2^2}{u^2(x_2)}\right) \tag{2.24}$$

Ziel ist es nun Gleichung (2.24) durch quadratisches Ergänzen in die Form $(ax-b)^2+c$ zu bringen. Dies kann durch folgende Definitionen erreicht werden:

$$a^2=\left(\frac{1}{u^2(x_1)}+\frac{1}{u^2(x_2)}\right) \tag{2.25}$$

$$b=\frac{1}{a}\cdot\left(\frac{x_1}{u^2(x_1)}+\frac{x_2}{u^2(x_2)}\right) \tag{2.26}$$

$$c=\left(\frac{x_1^2}{u^2(x_1)}+\frac{x_2^2}{u^2(x_2)}\right)-b^2 \tag{2.27}$$

Damit erhält man

$$p(x)=\frac{e^{-\frac{c}{2}}}{2\pi u(x_1)u(x_2)}e^{-\frac{1}{2}(a\cdot x-b)^2}=\frac{e^{-\frac{c}{2}}}{2\pi u(x_1)u(x_2)}e^{-\frac{1}{2\cdot 1/a^2}\left(x-\frac{b}{a}\right)^2}. \tag{2.28}$$

Die Normierungsbedingung lautet

$$\int_{-\infty}^{\infty}\mathrm{d}x\; p_1(x)\int_{-\infty}^{\infty}\mathrm{d}x'\; p_2(x')=1. \tag{2.29}$$

Die Betrachtung von $p(x)$ ist ein Spezialfall, für den Fall, dass $x=x'$, d.h. unter der Nebenbedingung, dass Labor 1 und Labor 2 exakt den gleichen Wert messen, d.h. in Gleichung (2.29) wäre die Dirac-Delta Distribution $\delta(x-x')$ zu ergänzen (zur Definition s. Gleichung (5.2)). Eine Normierung der Funktion $p(x)$ selbst ist daher mathematisch nicht sinnvoll. Gleichung (2.29) ist bereits normiert, da $p_1(x)$ und $p_2(x)$ es sind. Gleichung (2.28) kann somit umgeschrieben werden zu

$$p(x)=\frac{e^{-\frac{c}{2}}}{2\pi u(x_1)u(x_2)}e^{-\frac{(x-x_{\mathrm{ref}})^2}{2u^2(x_{\mathrm{ref}})}} \tag{2.30}$$

mit

$$x_{\mathrm{ref}}=\frac{b}{a}=u^2(x_{\mathrm{ref}})\cdot\left(\frac{x_1}{u^2(x_1)}+\frac{x_2}{u^2(x_2)}\right) \tag{2.31}$$

und

$$u^2(x_{\mathrm{ref}})=\frac{1}{a^2}=\left(\frac{1}{u^2(x_1)}+\frac{1}{u^2(x_2)}\right)^{-1}. \tag{2.32}$$

Der Vorfaktor in Gleichung (2.30) weicht von einer üblichen Normalverteilung ab, spielt für die Definition der Standardabweichung und des Erwartungswertes jedoch keine Rolle. Er steuert lediglich die konkrete Wahrscheinlichkeit für einen gegebenen x-Wert. Eine beispielhafte Darstellung von $p(x)$ ist in Abbildung 3.5 gezeigt. Damit ist allgemein gezeigt, dass der gewichtete Mittelwert x_{ref} der Erwartungswert der Wahrscheinlichkeitsdichteverteilung

$$p(x)=\prod_{i=1}^{M}p_i(x) \tag{2.33}$$

ist und $u(x_{\mathrm{ref}})$ die dazugehörige Standardabweichung. Insbesondere wurde zur Herleitung die Annahme gemacht, dass $p_i(x)$ durch eine Normalverteilung gegeben ist, wodurch Bedingung 1 begründet ist.

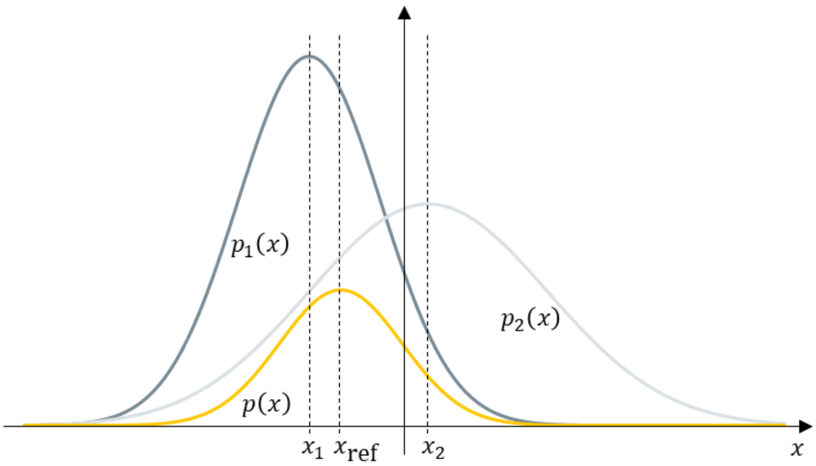


Abbildung 3.5: Darstellung der Wahrscheinlichkeitsdichteverteilungen $p_1(x)$, $p_2(x)$ und $p(x) = p_1(x) \cdot p_2(x)$. Der Erwartungswert von $p(x)$ ist der gewichtete Mittelwert x_{ref} und dessen Unsicherheit ist die dazugehörige Standardabweichung. Gezeigt ist der Fall $u(x_1) < u(x_2)$. Es zeigt sich, dass durch die Gewichtung der Referenzwert zum Erwartungswert mit der kleineren Unsicherheit (hier x_1) geschoben wird.

Anmerkungen zu Bedingung 2:

Sei y' eine normalverteilte Variable mit Erwartungswert $\bar{y}$ und Standardabweichung σ_y, so ist $y \equiv (y' - \bar{y})/\sigma_y$, eine standardnormalverteilte Variable, d.h. der Erwartungswert ist 0 und die Standardabweichung gleich 1, sodass $p(y) = e^{-y^2}/\sqrt{2\pi}$. Unter dieser Voraussetzung ist y^2 eine χ^2-verteilte Variable. Die χ^2-Verteilung ist eine Verteilungsfunktion, deren charakteristischer Parameter nur der Freiheitsgrad ν ist. Erwartungswert (= ν) und Standardabweichungen (= $\sqrt{2\nu}$) sind somit keine Parameter, sondern direkt durch den Freiheitgrad gegeben. Dies ist eine direkte Folge aus der Annahme der Standardnormalverteilung. Die Wahrscheinlichkeitsdichte lautet

$$p_\nu(y) = \left(2^{\nu/2} \cdot \Gamma\left(\frac{\nu}{2}\right)\right)^{-1} \cdot y^{\frac{\nu}{2}-1} \cdot e^{-y/2} \quad \text{für} \quad y > 0 \tag{2.34}$$

mit $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$ und $\Gamma(m+1) = m \cdot \Gamma(m)$. Abbildung 3.6 zeigt beispielhaft die χ^2-Verteilung für Freiheitsgraden $\nu \in \{1; 3; 5\}$. Weiterführend gilt, dass eine Summe von standardnormalverteilten Zufallsgrößen $\sum_i y_i$ ebenfalls standardnormalverteilt ist. Die Quadratesumme $\sum_i y_i^2$ ist dann wiederum χ^2-verteilt, wenn alle y_i standardnormalverteilt sind.

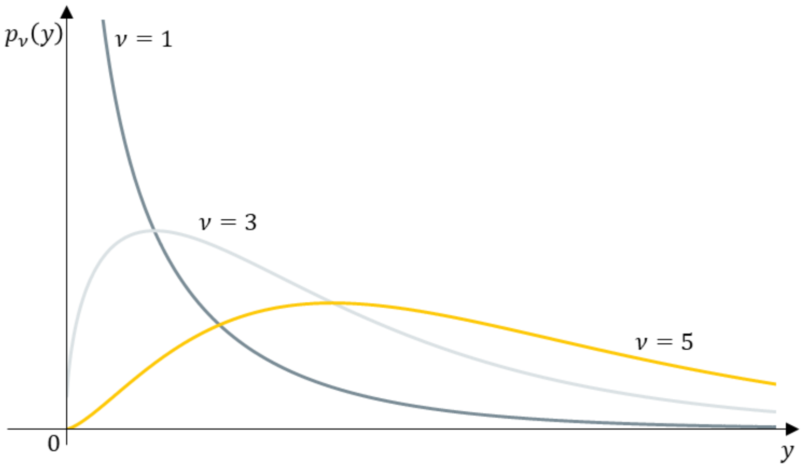


Abbildung 3.6: Darstellung der χ^2-Verteilung für ausgewählte Freiheitsgrade.

Sei nun χ^2_{obs} wie in Gleichung (2.4) definiert. Wenn der gewichtete Mittelwert ein sinnvoller Referenzwert ist, so sollten die einzelnen Ergebnisse der Teilnehmer (zumindest jene, die zur Berechnung des gewichteten Mittelwertes berücksichtigt werden), normalverteilt um diesen liegen, d.h. jeder Term $(x_i - x_{\text{ref}})/u(x_i)$ sollte standardnormalverteilt sein, sodass $(x_i - x_{\text{ref}})^2/u^2(x_i)$ und damit auch χ^2_{obs} jeweils χ^2-verteilte Größen darstellen. An dieser Stelle wird die Erfüllung von Bedingung 1 vorausgesetzt. Der in Gleichung (2.19) definierten Variable χ^2_{obs} wird der Freiheitsgrad $\nu = M - 1$ zugeordnet (M Summanden, und eine Nebenbedingung durch die Definition des gewichtete Mittelwertes in Gleichung (2.1)). Das wiederum bedeut, dass χ^2_{obs} eine χ^2-verteilte Variable ist, die durch die Wahrscheinlichkeitsdichteverteilung $p_\nu(x = \chi^2_{\text{obs}})$ (s. Gleichung (2.34)) beschrieben werden kann. Um diese Aussagen zu testen, wird festgelegt, dass

$$\int_0^{\chi^2_{\text{obs}}} \mathrm{d}x\, p_\nu(x) < \alpha = 0{,}95 \tag{2.35}$$

gelten muss. Gleichung (2.35) liefert damit eine alternative Formulierung zu Gleichung (2.18). Der χ^2-Test prüft somit, ob die definierte Größe χ^2_{obs} mit einer 95 %-igen Wahrscheinlichkeit durch eine χ^2-Verteilung beschrieben werden kann. Ist dies der Fall, so können die obigen Annahmen als korrekt erachtete werden und der gewichtet Mittelwert eignet sich als Referenzwert.

C) Referenzwert wird mit Algorithmus A berechnet

Die hier dargestellte Vorgehensweise bezieht zur Bestimmung der Referenzdaten ebenfalls Teilnehmerdaten mit ein. Die hier gezeigte Berechnung ist zum Großteil aus DIN ISO 13528, Anhang C.3 entnommen. Der dort vorgestellte Algorithmus A stellt ein robustes Verfahren dar, d.h. der Einfluss von Ausreißern auf den Referenzwert wird reduziert. Die Idee hinter dem Algorithmus kann Peter J. Huber zugeordnet werden und findet sich in Ansätzen bspw. in seinem Buch "Robust Statistics" wieder. Das Verfahren ist bis zu einem Ausreißeranteil von 25 % geeignet, sollte jedoch erst ab einer Teilnehmerzahl $N \geq 3$ eingesetzt werden (bei einer Teilnehmerzahl N, ergeben sich durch die zusätzliche Anfangsmessung der esz AG insgesamt $N + 1$ Datensätze, die zur Bestimmung des Referenzwertes herangezogen werden können). Durch eine graphische Analyse der erhaltenen Teilnehmerdaten kann festgestellt werden, ob Ausreißer in einer Eignungsprüfung vorliegen oder nicht.

Sei $\boldsymbol{x} \equiv \{x_1; x_2; \dots; x_{N+1}\}$ der Datensatz, der die übermittelten Abweichungswerte der N Teilnehmer und die Anfangsmessung der esz AG enthält und aufsteigend sortiert ist, d.h. $x_1 \leq x_2 \leq \dots \leq x_{N+1}$. Nachfolgend sei $x_{\text{ref};n}$ der robuste Mittelwert der n-ten Iteration und $u_n(x_i)$ die robuste Standardabweichung. Folgende Anfangswerte werden verwendet

$$x_{\text{ref};0} = \text{med}(\boldsymbol{x}) = \begin{cases} x_{\frac{(N+3)}{2}}, & N \text{ ungerade} \\ \dfrac{x_{\frac{(N+2)}{2}} + x_{2+\frac{N}{2}}}{2}, & N \text{ gerade} \end{cases} \tag{2.36}$$

$$u_0(x_i) = 1{,}483 \cdot \text{med}\left(\left|\boldsymbol{x} - x_{\text{ref};0}\right|\right) \tag{2.37}$$

Mit „med" ist dabei der Median bezeichnet, dessen Definition in Gleichung (2.36) gegeben ist. Gleichung (2.37) berechnet die sogenannte „median absolute deviation" (MAD). Der Faktor 1,483 ist nötig, um die MAD an die Größenordnung einer Standardabweichung einer entsprechenden Normalverteilung anzugleichen. Die MAD ist so definiert, dass 50 % der Werte in $\text{med}(\boldsymbol{x}) \pm MAD$ liegen. Gesucht ist jedoch ein u_0 in der Größenordnung einer Standardabweichung, d.h.

$$\int_{-\sigma}^{\sigma} \mathrm{d}x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} = p = \mathrm{erf}\left(\frac{1}{\sqrt{2}}\right) \tag{2.38}$$

$$\int_{MAD}^{MAD} \mathrm{d}x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} = \mathrm{erf}\left(\frac{MAD}{\sqrt{2}\sigma}\right) = \frac{1}{2} \tag{2.39}$$

Damit erhält man

$$MAD = \sqrt{2}\sigma \cdot \mathrm{erf}^{-1}\left(\frac{1}{2}\right) \tag{2.40}$$

bzw.

$$\sigma = \frac{1}{\sqrt{2} \cdot \mathrm{erf}^{-1}\left(\frac{1}{2}\right)} \cdot MAD \simeq 1{,}483 \cdot MAD \tag{2.41}$$

Hierbei bezeichnet erf die Gauß'sche Fehlerfunktion, bzw. deren Inverse mit erf^{-1}. Für den Fall, dass $u_0(x_i)$ gleich Null ist, kann der vorgestellte Algorithmus nicht los laufen. In diesem Fall ist statt Gleichung (2.37) die Stichprobenstandardabweichung für $u_0(x_i)$ zu verwenden, wobei Ausreißer bereits entfernt werden müssen. Diese Ersetzung ist nur für den Anfangswert $u_0(x_i)$ zu verwenden, der restliche Ablauf bleibt davon unberührt. Um die Funktionsweise des vorgestellten Algorithmus zu veranschaulichen, werden die nachfolgenden Schritte anhand eines Beispieldatensatzes visualisiert. Verwendet wird hierfür ein Datensatz bestehend aus 10 Werten und einem Ausreißeranteil von 20 %, sodass die eingangs genannten Voraussetzungen erfüllt sind:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1,0	1,0	1,1	1,2	1,2	1,2	1,3	1,5	2,5	4,5

Abbildung 3.7 zeigt die graphische Darstellung des Datensatzes.

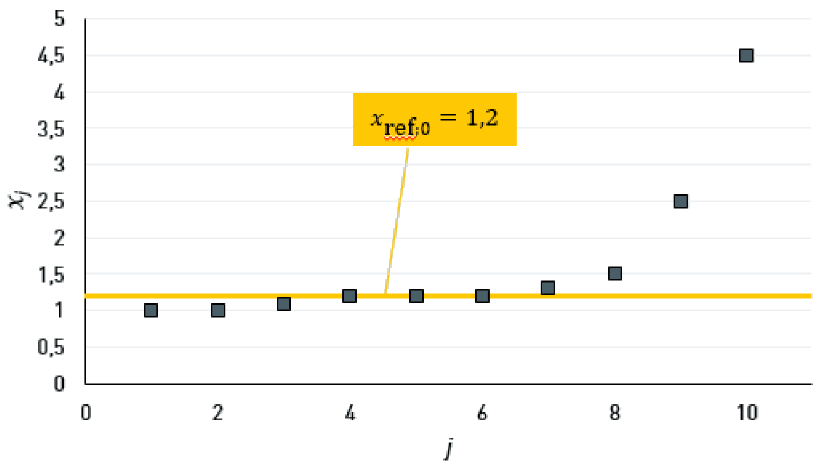


Abbildung 3.7: Darstellung des Beispieldatensatzes und des resultierenden Anfangswertes.

Nachdem die Anfangswerte definiert sind, folgt der iterative Teil. Für jeden Iterationsschritt n wird

$$\delta_n = l \cdot u_{n-1}(x_i) \quad \text{mit} \quad l = \frac{3}{2} \tag{2.42}$$

definiert. Die Größe δ_n definiert den Bereich um $x_{\mathrm{ref};n}$ der bei der n-ten Iteration abgeschnitten wird, denn für jedes $x_{i;n}$ mit $i \in \{1; \dots; N+2\}$ wird folgende Modifikation vorgenommen

$$x_{i;n} = \begin{cases} x_{\mathrm{ref};n-1} - \delta_n, & \text{wenn } x_i < x_{\mathrm{ref};n-1} - \delta_n \\ x_{\mathrm{ref};n-1} + \delta_n, & \text{wenn } x_i > x_{\mathrm{ref};n-1} + \delta_n. \\ x_{i;n-1}, & \text{sonst} \end{cases} \tag{2.43}$$

Abbildung 3.8 visualisiert Gleichungen (2.42) und (2.43). Werte, die bereits in der δ_n-Umgebung liegen, bleiben von der Modifikation in Gleichung (2.43) unberührt. Im Gegensatz dazu, werden Werte die außerhalb der δ_n-Umgebung liegen auf die Grenzen dieser Umgebung verschoben. Dies betrifft insbesondere Werte, die als Ausreißer bezeichnet werden.

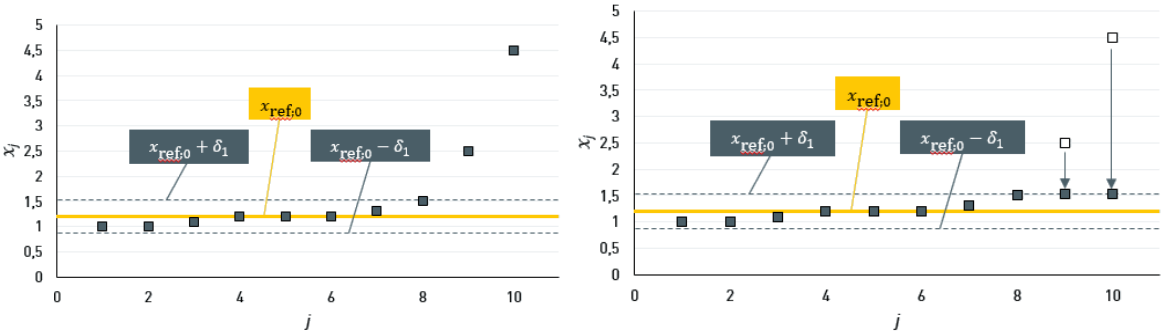


Abbildung 3.8: Werte, die außerhalb der δ-Umgebung liegen, werden modifiziert und auf die Grenzen der genannten Umgebung verschoben.

Durch die Modifikationen in Gleichung (2.43) resultiert ein neuer Datensatz. Im gezeigten Beispiel ändert sich dieser zu

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1,0	1,0	1,1	1,2	1,2	1,2	1,3	1,5	1,6	1,6

Der robuste Mittelwert des n-ten Iterationsschrittes ist dann

$$x_{\mathrm{ref};n} = \sum_{i=1}^{N+1} \frac{x_{i;n}}{N+1} \tag{2.44}$$

und die dazugehörige Unsicherheit ist

$$u_n(x_i) = 1{,}134 \cdot \sqrt{\sum_{i=1}^{N+1} \frac{\left(x_{i;n} - x_{\mathrm{ref};n}\right)^2}{N}}. \tag{2.45}$$

Die Berechnungen gemäß Gleichungen (2.42) bis (2.45) werden solange iterativ wiederholt, bis sich die Ergebnisse in der dritten signifikanten Stelle nicht mehr ändern. In diesem Fall kann von einer Konvergenz bzgl. dieses Wertes ausgegangen werden.

Sei n_c der Laufindex, bei dem Konvergenz eintritt, so ist am Ende die Unsicherheit des Referenzwertes

$$x_{\mathrm{ref}} = x_{\mathrm{ref};n_c} \tag{2.46}$$

gegeben durch

$$u(x_{\mathrm{ref}}) = \frac{u_{n_c}(x_i)}{\sqrt{N+1}} \tag{2.47}$$

An dieser Stelle sei darauf hingewiesen, dass die Referenzunsicherheit nach Methode C in Gleichung (2.47) zunächst keinen Beitrag für die Rückführung aufweist und dieser zusätzlich berücksichtigt werden muss, insbesondere wenn aufgrund kleiner Streuung Gleichung (2.47) zu kleine Werte liefert. Alternativ kann die Referenzunsicherheit auch durch $u(x_{\mathrm{ref}}) = \frac{1}{N+1}\sqrt{\sum_{i=1}^{N} u^2(x_i)}$ berechnet werden, d.h. über die direkte Fehlerfortpflanzung des empirischen Mittelwertes.

Anhand Gleichung (2.47) ist klar, dass, wie in Methode B, die Referenzunsicherheit umso kleiner wird, je mehr Teilnehmerdaten vorliegen. Als Beispiel sei hier das Konvergenzverhalten für die folgenden zwei Datensätze gezeigt.

Datensatz 1:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1,0	1,0	1,1	1,2	1,2	1,2	1,3	1,5	2,5	4,5

Datensatz 2:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1,0	1,0	1,1	1,1	1,2	1,2	1,2	1,3	1,4	1,5

Datensatz 1 enthält zwei Ausreißer x_9 und x_{10}, wohingegen in Datensatz 2 keine Ausreißer vorliegen, ansonsten sind die Daten gleich. In Abhängigkeit von der Iteration n erhält man die in Abbildung 3.9 dargestellten Ergebnisse für den Referenzwert und dessen Unsicherheit. Bei 50 Iterationen erkennt man eine klare Konvergenz. Man sieht zudem, dass Methode C annähernd den arithmetischen Mittelwert (durchgezogene Linie) und die Stichprobenstandardabweichung reproduziert, falls wie in Datensatz 2 keine starken Ausreißer vorhanden sind. Liegen Ausreißer vor, sowie in Datensatz 1, so zeigt sich, dass Methode C einen abweichenden Referenzwert liefert, dieser ist in etwa zwischen arithmetischen Mittelwert (durchgezogene Linie) und dem Ergebnis für Datensatz 2, d.h. für den Fall, dass keine Ausreißer vorliegen. Das Gleiche gilt für die Unsicherheit. Somit erweist sich Methode C als robustes Verfahren, da der Einfluss der Ausreißer auf den Referenzwert deutlich verringert wird.

Im Gegensatz zu Methode B wird im vorliegenden Fall die Unsicherheit der Teilnehmer nicht berücksichtigt, sondern lediglich deren berichteter Abweichungswert. Insbesondere ist auch die Unsicherheit des Referenzwertes letztlich durch die Streuung der Abweichungswerte charakterisiert. Bei Methode B hingegen ist gerade das Zusammenspiel aus Abweichungswert und Unsicherheit charakteristisch.

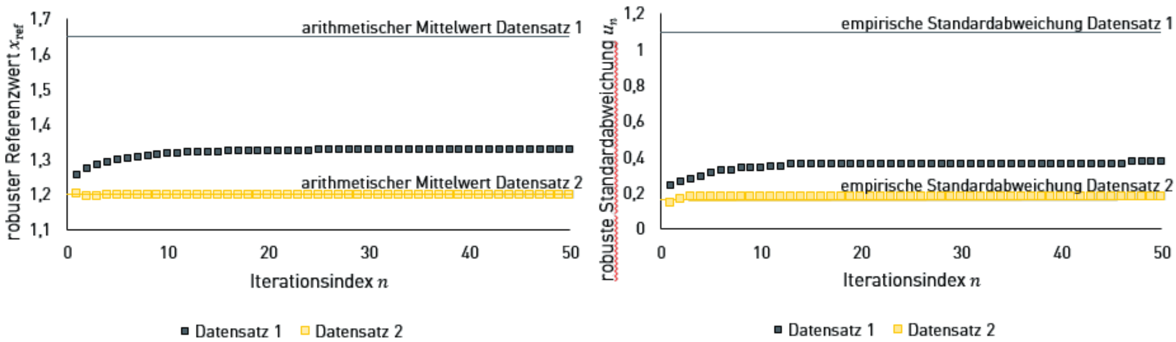


Abbildung 3.9: Die durchgezogene Linie zeigt jeweils den arithmetischen Mittelwert, bzw. die Stichprobenstandardabweichung. Im Falle von Datensatz 1 ist diesbezüglich ein deutlicher Unterschied zu erkennen. In Datensatz 2 liefert Methode C ähnliche Ergebnisse wie der arithmetische Mittelwert und die Stichprobenstandardabweichung. Dies ist klar, da keine starken Ausreißer im Datensatz enthalten sind. In Datensatz 1 ermöglicht die Verwendung von Methode C eine Verringerung des Einflusses der Ausreißer.

<u>Anmerkungen zu Gleichung (2.45):</u>

In Gleichung (2.45) fällt auf, dass die angegebene Standardabweichung im Vergleich zur empirischen Standardabweichung mit einem Faktor 1,134 multipliziert wird. Im nachfolgenden soll dieser Faktor hergeleitet werden. Vorab kann bereits angemerkt werden, dass der Faktor 1,134 den Fehler korrigiert, der dadurch zustande kommt, dass alle Werte außerhalb der δ_n-Umgebung auf deren Grenzen verschoben werden. Durch diese Modifikation der Werte, wird die empirische Standardabweichung unterschätzt und fällt kleiner aus, als die gemäß einer Normalverteilung zu erwarten wäre. Dieser Fehler wird durch die Multiplikation mit dem Faktor 1,134 korrigiert. Insbesondere fließt hier also die gleiche statistische Annahme ein, die bereits Methode B zugrunde liegt. Auch für Methode B wird angenommen, dass die gesuchte Größe durch eine Normalverteilung beschrieben wird und jeder Teilnehmer eine entsprechende Stichprobe berichtet.

Zwar setzt sich die empirische Standardabweichung aus diskreten Punkten zusammen, allerdings soll der Einfachheit halber im Nachfolgenden eine kontinuierliche Berechnung durchgeführt werden. Zur Herleitung des Faktor sei oBdA eine Normalverteilung mit Standardabweichung σ und Erwartungswert $E(x) \equiv 0$ angenommen, d.h.

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}. \tag{2.48}$$

Entsprechend der Definition der Standardabweichung folgt für $E(x) = 0$:

$$\sigma^2 = E(x^2) - E(x)^2 = \int_{-\infty}^{\infty} \mathrm{d}x \; p(x) x^2 \tag{2.49}$$

Zu Beginn, d.h. bevor Gleichungen (2.42) und (2.43) angewandt werden, liegt den Daten eine Normalverteilung, wie beispielsweise in Gleichung (2.48) gezeigt zugrunde. Nach Anwendung der Gleichungen (2.42) und (2.43), wird dies Normalverteilung modifiziert. Da alle Punkte außerhalb der δ_n-Umgebung auf deren Grenzen verschoben wird, werden die Bereich der Normalverteilung, die außerhalb $\pm\delta_n = \pm l \cdot \sigma$ (mit $l = 3/2$) auf exakt $\pm l \cdot \sigma$ verschoben, d.h. die Wahrscheinlichkeitsdichte an diesem Punkt erhöht sich und eine neue Wahrscheinlichkeitsverteilung resultiert. Eine Veranschaulichung der modifizierten Wahrscheinlichkeitsdichte ist in Abbildung 3.10 gezeigt.

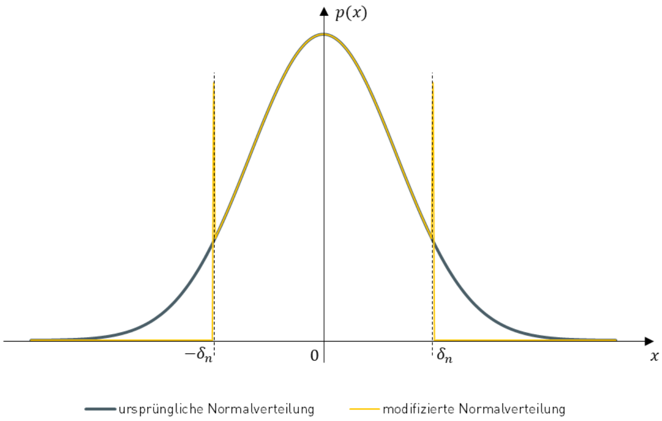


Abbildung 3.10: Darstellung der modifizierten Normalverteilung, nachdem Gleichungen (2.42) und (2.43) angewandt wurden. Alle Werte außerhalb der δ_n-Umgebung werden auf deren Grenzen verschoben, sodass die Wahrscheinlichkeit an diesen Grenzen im Vergleich zu einer Normalverteilung steigt.

Gleichung (2.45) berechnet nur die empirische Standardabweichung für die modifizierte Normalverteilung, d.h. mit (unter der Annahme, dass $E(x) = 0$):

$$p_{\mathrm{mod}}(x) = \Theta(\delta_n - |x|) \cdot \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} + \big(\delta(x - \delta_n) + \delta(x + \delta_n)\big) \cdot \frac{1}{2} \cdot \left(1 - \mathrm{erf}\left(\frac{l}{\sqrt{2}}\right)\right) \tag{2.50}$$

Hierbei ist $\Theta(x)$ die Stufenfunktion und $\delta(x)$ die Dirac-Delta Distribution. Die Standardabweichung σ ist nach wie vor über $p(x)$ definiert, d.h. über Gleichung (2.49). Mit der modifizierten Normalverteilung in Gleichung (2.50) kann nun eine ebenfalls modifizierte Standardabweichung berechnet werden. Das Ergebnis der Integration liefert

$$\sigma_{\text{mod}}^2 = \left[\operatorname{erf}\left(\frac{l}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}} \cdot l \cdot e^{-\frac{l^2}{2}} + l^2 \left(1 - \operatorname{erf}\left(\frac{l}{\sqrt{2}}\right)\right) \right] \cdot \sigma^2 \tag{2.51}$$

Damit erhält man $\sigma = m \cdot \sigma_{\text{mod}}$, mit

$$m = \frac{1}{\sqrt{\operatorname{erf}\left(\frac{l}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}} \cdot l \cdot e^{-\frac{l^2}{2}} + l^2 \left(1 - \operatorname{erf}\left(\frac{l}{\sqrt{2}}\right)\right)}}. \tag{2.52}$$

Für $l = 1{,}5$ aus Gleichung (2.42) ergibt sich dann $m = 1{,}134 \simeq \frac{1}{2} \cdot \left(1 - \operatorname{erf}\left(\frac{l}{\sqrt{2}}\right)\right) + 1$, was gerade dem Korrekturfaktor aus Gleichung (2.45) entspricht. Damit ist klar, dass die Wahl des Korrekturfaktors der Standardabweichung unmittelbar mit der Wahl der δ_n-Umgebung zusammenhängt.

D) Referenzwert wird als power-moderated mean berechnet

Die nachfolgenden Darstellungen orientieren sich an der Veröffentlichung von S. Pommé und J. Keightley zu „Determination of a reference value and its uncertainty through a power-moderated mean" (Metrologia 52, 2015). Eine detaillierte Analyse der Methode ist dort zu finden. Im nachfolgenden werden die Darstellungen aus der genannten Veröffentlichung unter Methode D zusammengefasst. Methode D kann als Symbiose der positiven Eigenschaften von Methode B und C erachtet werden und ist aus diesem Grund nach Möglichkeit anzuwenden.

Ähnlich wie bei Methode B, fließen auch hier sowohl Teilnehmerwert, als auch deren dazugehörige Unsicherheit in die Bestimmung des Referenzwertes ein. Ausgangspunkt ist der sogenannte Mandel-Paule Mittelwert x_{mp}, der im Gegensatz zum herkömmlichen gewichteten Mittelwert aus Methode B, einen Parameter s quadratisch zur Teilnehmerunsicherheit addiert

$$x_{\text{mp}} = u^2(x_{\text{mp}}) \sum_{i=1}^{M} \frac{x_i}{u^2(x_i) + s^2} \tag{2.53}$$

mit

$$u^2(x_{\text{mp}}) = \left(\sum_{i=1}^{M} \frac{1}{u^2(x_i) + s^2} \right)^{-1}. \tag{2.54}$$

Für $s = 0$, reproduzieren die Gleichungen (2.53) und (2.54) den gewichteten Mittelwert, wie in den Gleichungen (2.2) und (2.3) beschrieben. Zur Bestimmung des Parameters s wird zunächst die Hilfsgröße $\tilde{\chi}_{\text{obs}}^2$ definiert. Die Definition orientiert sich an Gleichung (2.19) und ist wie der Mandel-Paule Mittelwert um den Parameter $s \geq 0$ erweitert

$$\tilde{\chi}_{\text{obs}}^2 = \frac{1}{M-1} \sum_{i=1}^{N} \frac{\left(x_i - x_{\text{mp}}\right)^2}{u^2(x_i) + s^2}. \tag{2.55}$$

Der Parameter s ist nun so zu bestimmen, dass

$$\tilde{\chi}_{\text{obs}}^2(s) \leq 1, \tag{2.56}$$

wobei für s der minimale Wert anzusetzen ist, für den Gleichung (2.56) erfüllt ist. Ist Gleichung (2.56) bereits für $s = 0$ erfüllt, so ist keine weitere Anpassung nötig. In allen anderen Fällen, wird s solange sukzessiv erhöht, bis Gleichung (2.56) erfüllt ist. Für die nachfolgenden Rechnungen wird dann der so ermittelten s-Parameter angewandt.

Anhand Gleichung (2.55) wird klar, dass eine Anpassung des s-Parameters insbesondere dann resultiert, wenn Unsicherheiten berichtet werden, die im Vergleich zur Differenz zwischen Teilnehmerwert und Mandel-Paule Mittelwert klein sind. In diesem Fall wird das starke Gewicht einer kleinen Unsicherheit reduziert, da stets

$$\frac{1}{u^2(x_i)} \geq \frac{1}{u^2(x_i)+s^2} \quad \text{für} \quad s \geq 0 \tag{2.57}$$

Insbesondere wird Gleichung (2.56) nicht erfüllt sein, wenn Ausreißer vorliegen, die mit einer kleinen Unsicherheit berichtet werden. In diesem Fall wird das unerwünschte Verhalten aus Methode B durch das Einfügen des Parameters s vermieden. Im Vergleich zu Methode B wird demnach, falls nötig, der Einfluss der Unsicherheiten auf die Gewichtung der Abweichungswerte reduziert. Ist $s \gg u(x_i)\ \forall_{i\in[1;M]}$, so werden im wesentlichen alle Abweichungswerte gleich gewichtet und Gleichung (2.53) reproduziert den arithmetischen Mittelwert. Je nach Grenzfall vermittelt der Mandel-Paule Mittelwert demnach zwischen gewichteten und arithmetischen Mittelwert

$$x_{\text{mp}} = u^2(x_{\text{mp}}) \sum_{i=1}^{M} \frac{x_i}{u^2(x_i)+s^2} \rightarrow \begin{cases} \frac{1}{M}\sum_{i=1}^{M} x_i\,, & s \gg u(x_i)\ \forall_{i\in[1;M]} \\ \left(\sum_{i=1}^{M} \frac{1}{u^2(x_i)}\right)^{-1} \cdot \sum_{i=1}^{M} \frac{x_i}{u^2(x_i)}, & s \ll u(x_i)\ \forall_{i\in[1;M]} \end{cases} \tag{2.58}$$

Für die Unsicherheit des Mandel-Paule Mittelwertes gilt entsprechend

$$u^2(x_{\text{mp}}) = \left(\sum_{i=1}^{M} \frac{1}{u^2(x_i)+s^2}\right)^{-1} \rightarrow \begin{cases} \frac{s^2}{M}, & s \gg u(x_i)\ \forall_{i\in[1;M]} \\ \left(\sum_{i=1}^{M} \frac{1}{u^2(x_i)}\right)^{-1}, & s \ll u(x_i)\ \forall_{i\in[1;M]} \end{cases} \tag{2.59}$$

Als Beispieldatensatz sei hier

x_1	x_2	x_3	x_4	x_5
10	8	11	13	12
$u(x_1)$	$u(x_2)$	$u(x_3)$	$u(x_4)$	$u(x_5)$
1,5	0,25	2,0	2,5	1,5

Angenommen, wobei Teilnehmer 2 offensichtlich einen Ausreißer mit kleiner Unsicherheit berichtet. Nachfolgende Abbildung 3.11 zeigt die Ergebnisse gemäß Gleichungen (2.58) und (2.59)

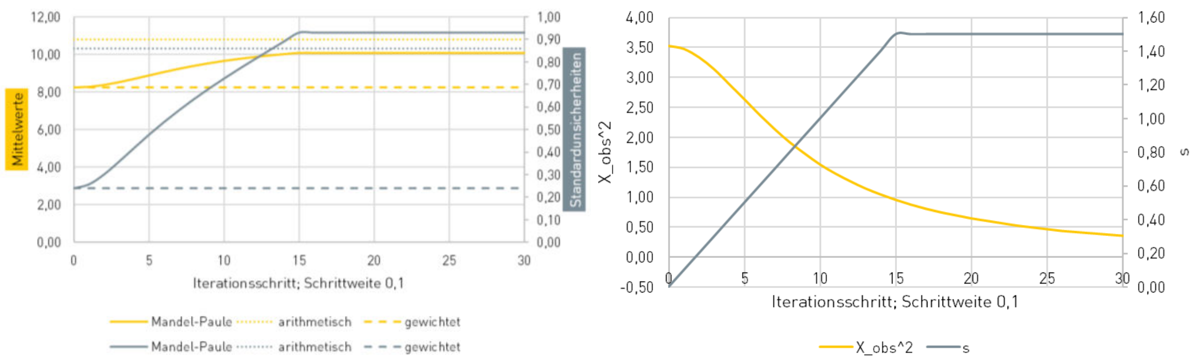


Abbildung 3.11: Links ist ein Vergleich der verschiedenen Mittelwerte dargestellt. Rechts ist die iterative Bestimmung des Parameters s gezeigt. Zur Berechnung der Standardunsicherheit des arithmetischen Mittelwertes wurde hier die empirische Standardabweichung angesetzt.

Es ist gut zu erkennen, dass der gewichtete Mittelwert von Teilnehmer 2 dominiert ist. Der Mandel-Paule Mittelwert hingegen liefert einen Wert, der zwischen gewichteten und arithmetischen Mittelwert liegt. Die Unsicherheit des gewichteten Mittelwertes nach Methode B ist immer kleiner als die kleinste Teilnehmerunsicherheit. Dies ist besonders dann ein Defizit, wenn ein Teilnehmer eine nicht plausible, kleine Unsicherheit berichtet und zudem einen Ausreißer darstellt. Methode B impliziert dann eine Kenntnis des Referenzwertes, die nicht realistisch ist. Der Mandel-Paule Mittelwert kompensiert das durch den Parameter s und liefert so eine Standardunsicherheit, die größer ist als die des gewichteten Mittelwertes. Je größer der Parameter s, desto größer wird die Unsicherheit des Mandel-Paule Mittelwertes. Das bedeutet, dass insbesondere bei diskrepanten Abweichungswerten und kleinen Unsicherheiten, die Unsicherheit des Mandel-Paule Mittelwertes erhöht wird und eine realistischere Größenordnung liefert als der gewichtete Mittelwert gemäß Methode B.

Nachdem der Parameter s bestimmt wurde, wird eine zusätzlicher Parameter $\alpha \in [0; 2]$ definiert. Dieser wird als Exponent in der finalen Gewichtung angesetzt. Je nach Datenlage ist die Wahl des Parameters α durch die nachfolgende Tabelle gegeben:

$\alpha =$	Verlässlichkeit der Unsicherheiten
0	Die Unsicherheiten $u(x_i)$ sind nicht informativ und sind nicht in Einklang mit den Differenzen der Abweichungswerte x_i (Grenzfall: arithmetischer Mittelwert)
$2 - 3/M$	Informative Unsicherheiten $u(x_i)$, die jedoch tendenziell unterschätzt sind.
2	Realistische Unsicherheiten $u(x_i)$, die mit den berichteten Abweichungen zusammen passen.

Standardmäßig wird

$$\alpha = 2 - \frac{3}{M} \tag{2.60}$$

angesetzt, wobei M die Anzahl an Daten ist, die in die Berechnung des Referenzwertes einfließen. Insbesondere ist daher $\alpha \simeq 2$, falls eine Vielzahl an Teilnehmerdaten in die Bestimmung des Referenzwertes einfließen.

Als dritter Parameter wird noch

$$S = \sqrt{M \cdot \max\left(u^2(\bar{x}); u^2(x_{\mathrm{mp}})\right)} \tag{2.61}$$

mit

$$u^2(\bar{x}) = \frac{1}{M(M-1)} \sum_{i=1}^{M} (x_i - \bar{x})^2 \quad \text{und} \quad \bar{x} = \frac{1}{M} \sum_{i=1}^{M} x_i \tag{2.62}$$

Definiert. Insbesondere berücksichtigt dieser S-Parameter die Streuung der berichteten Abweichungswerte, falls diese groß ist. Im Wesentlichen ist Methode D somit sensitiv gegenüber der Größenordnung der berichteten Messunsicherheiten, sowie der Streuung der Abweichungswerte.

Unter Berücksichtigung der Parameter s, α und S, wird der Referenzwert als power-moderated mean berechnet

$$x_{\text{ref}} = \sum_{i=1}^{M} w_i x_i \tag{2.63}$$

mit

$$w_i = \frac{u^2(x_{\text{ref}})}{\left(\sqrt{u^2(x_i) + s^2}\right)^{\alpha} \cdot s^{2-\alpha}} \tag{2.64}$$

und

$$u^2(x_{\text{ref}}) = \left(\sum_{i=1}^{M} \frac{1}{\left(\sqrt{u^2(x_i) + s^2}\right)^{\alpha} \cdot s^{2-\alpha}} \right)^{-1}. \tag{2.65}$$

Für $s = 0$ und $\alpha = 2$ wird der Referenzwert aus Methode B reproduziert. In Abbildung 3.12 ist ein Vergleich der Ergebnisse aus Methode B, C und D gezeigt, für den Datensatz

x_{N}	x_1	$u(x_1)$	x_2	$u(x_2)$	x_3	$u(x_3)$	x_4	$u(x_4)$	x_5	$u(x_5)$
10	10	1,5	8	0,25	11	2,0	13	2,5	12	1,5
20	20	1,5	18	0,25	21	1,5	21	3,5	21	1,0
30	30	1,5	28	0,25	32	2,5	32	3,5	34	2,0
40	40	1,5	38	0,25	43	3,5	44	4,0	43	3,0

gezeigt. Im vorliegenden Beispieldatensatz berichtet Teilnehmer 2 jeweils einen Ausreißer mit kleiner Unsicherheit. Methode B liefert somit einen Referenzwert, der im Wesentlichen durch die Ergebnisse von Teilnehmer 2 gegeben ist. Methode D hingegen liefert einen Referenzwert, der deutlich weniger stark von Teilnehmer 2 beeinflusst ist. Entsprechend ist auch die Unsicherheit nach Methode D größer und liefert eine realistischere Abschätzung für die Unsicherheit des Referenzwertes, welche der vorliegenden Daten gerecht wird. Im gezeigten Beispiel ist der χ^2-Test in Methode B bestanden.

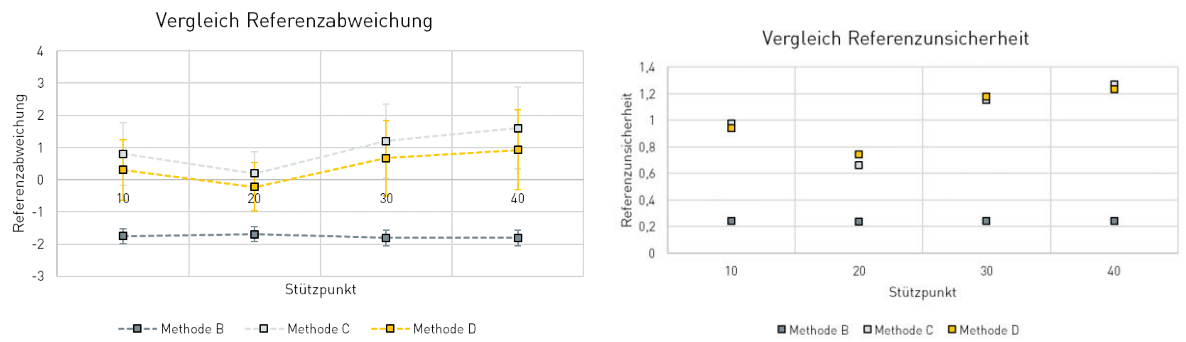


Abbildung 3.12: Vergleich der Ergebnisse nach Methode B, C und D für den obigen Beispieldatensatz.

Wie Methode C, stellt auch Methode D ein robustes Verfahren dar. Weitere Eigenschaften der Methode D sind in der eingangs genannten Veröffentlichung diskutiert. Im Gegensatz zu Methode C hat Methode D den Vorteil, dass die Unsicherheiten der Teilnehmer berücksichtigt werden und keine Mindestanzahl an Daten für die Verwendung notwendig ist. Des Weiteren liefert Methode D eine Vorgehensweise, wie Ausreißer eliminiert werden. Nachdem in einem ersten Durchlauf die Referenzwerte gemäß Gleichungen (2.63) bis (2.65) bestimmt wurden, wird getestet, ob

$$|e_i| \leq k \cdot u(e_i), \tag{2.66}$$

wobei $k = 2{,}5$ und

$$e_i = x_i - x_{\text{ref}}, \tag{2.67}$$

sodass unter der Berücksichtigung der Korrelation zwischen x_i und x_{ref}

$$u(e_i) = \sqrt{(1 - 2w_i)u^2(x_i) + u^2(x_{\text{ref}})} \tag{2.68}$$

folgt (mit w_i aus Gleichung (2.64)). Liegt keine Korrelation vor, d.h. fließt x_i nicht in die Bestimmung des Referenzwertes ein, so ist

$$u(e_i) = u^2(x_i) + u^2(x_{\text{ref}}). \tag{2.69}$$

Teilnehmer, für die Gleichung (2.66) nicht erfüllt ist, sind aus der Bestimmung des Referenzwertes auszuschließen. Mit den verbleibenden Teilnehmern ist die Berechnung gemäß Gleichungen (2.53) bis (2.65) zu wiederholen. Dies geschieht solange, bis Gleichung (2.66) mit einen bestimmten Referenzwert für alle Teilnehmer erfüllt ist. An dieser Stelle sei angemerkt, dass Gleichungen (2.68) und (2.69) im Gegensatz zur eingangs genannten Veröffentlichung modifiziert sind. Der Grund ist, dass die vereinfachten Gleichungen der Publikation ((C.1) und (C.2)) nicht reproduziert werden können und zu einem nicht zielführenden Kriterium in Gleichung (2.66) führen. Für den resultierenden Referenzwert und dessen Unsicherheit wird angenommen, dass eine Normalverteilung zugrunde liegt.

Abschließend sei noch angemerkt, dass die zitierte Publikation keine Vorgabe hinsichtlich einer Mindestanzahl an Teilnehmerdaten macht. Daher wird Methode D ab einer Teilnehmerzahl von $N \geq 3$ verwendet, sodass einschließlich der Anfangsmessung der esz AG insgesamt 4 Datensätze vorliegen.

5.4.3.Kriterium zur Leistungsbewertung

Als Kriterium zur Leistungsbewertung wird der für Kalibrierlabore übliche E_{n}-Wert („normalized Error") verwendet. Dessen Definition ist für das i-te Labor gegeben durch

$$E_{\text{n};i} = \frac{d_i}{U(d_i)}. \tag{3.1}$$

mit

$$d_i = x_i - x_{\text{ref}} \tag{3.2}$$

und

$$U(d_i) = 2 \cdot \sqrt{u^2(x_i) + u^2(x_{\text{ref}}) + 2\frac{\partial d_i}{\partial x_i}\frac{\partial d_i}{\partial x_{\text{ref}}}u(x_i; x_{\text{ref}})}. \tag{3.3}$$

Hierbei ist $u(x_i; x_{\text{ref}})$ die Kovarianz, welche Korrelationen zwischen x_i und x_{ref} berücksichtigt. Die Leistung des i-ten Labors wird als positiv bewertet, falls

$$-1 < E_{\text{n};i} < 1. \tag{3.4}$$

Um die Information, $x_i > x_{\text{ref}}$ oder $x_i < x_{\text{ref}}$ nicht zu verlieren ist es vorteilhaft, das Vorzeichen des $E_{\text{n};i}$-Wertes nicht zu vernachlässigen. Das E_{n}-Kriterium prüft im Wesentlichen, ob die Differenz d_i von Labor i zum Referenzwert innerhalb der erweiterten Unsicherheit $U(d_i)$ dieser Differenz liegt. Es sei an dieser Stelle auch darauf hingewiesen, dass für $u(x_i; x_{\text{ref}}) = 0$ gilt $U(d_i) = \sqrt{U^2(x_i) + U^2(x_{\text{ref}})} < U(x_i) + U(x_{\text{ref}})$. Das bedeutet, dass selbst wenn sich die Messunsicherheitsbalken des Ergebnisses $x_i \pm U(x_i)$ mit denen des Referenzwertes $x_{\text{ref}} \pm U(x_{\text{ref}})$ noch überlappen, ein $E_{\text{n};i} \geq 1$ resultieren kann. Für das positive Bestehen ist das Überlappen der Messunsicherheitsbalken somit ein notwendiges, aber kein hinreichendes Kriterium. Dieser Fall ist beispielhaft in Abbildung 3.13 dargestellt.

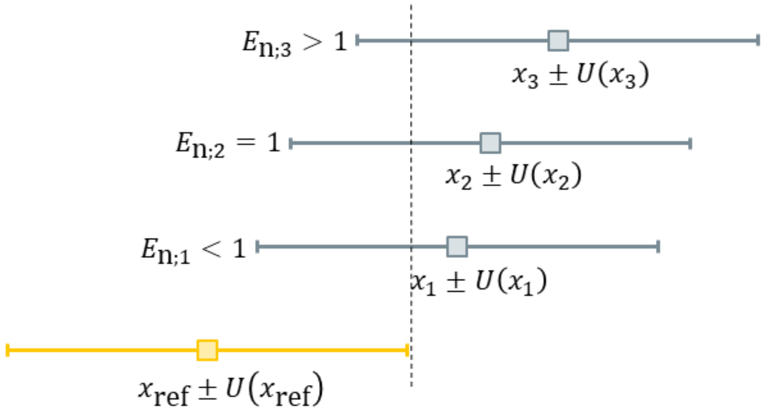


Abbildung 3.13: Im Beispiel hier ist $U(x_{ref}) = U(x_1) = U(x_2) = U(x_3)$ und $x_2 - x_{ref} = \sqrt{U^2(x_2) + U^2(x_{ref})}$. Es zeigt sich, dass ein Überlappen der Unsicherheitsbalken selbst für $E_{n;i} > 1$ möglich ist. Das Überlappen der Messunsicherheitsbalken ist somit ein notwendiges, aber kein hinreichendes Kriterium.

Für den Fall, dass der Referenzwert x_{ref} als gewichteter Mittelwert durch Methode B oder als power-moderated mean gemäß Methode D bestimmt wurde, so ist zu berücksichtigen, dass x_i und x_{ref} korreliert sind, d.h. $u(x_i; x_{ref}) \neq 0$, falls die Ergebnisse des i-ten Labors in die Bestimmung des Referenzwertes eingeflossen sind. Ist dies der Fall, so ist

$$u(x_i; x_{ref}) = w_i \cdot u^2(x_i) \tag{3.5}$$

mit

$$w_i = \begin{cases} \dfrac{u^2(x_{ref})}{u^2(x_i)}, & \text{Methode B} \\ \dfrac{u^2(x_{ref})}{\left(\sqrt{u^2(x_i) + s^2}\right)^{\alpha} \cdot S^{2-\alpha}}, & \text{Methode D} \end{cases} \tag{3.6}$$

und wegen $\partial d_i / \partial x_{ref} = -1$ ergibt sich der komplette Korrelationsterm in Gleichung (3.3) zu

$$2 \frac{\partial d_i}{\partial x_i} \frac{\partial d_i}{\partial x_{ref}} u(x_i; x_{ref}) = -2 w_i \cdot u^2(x_i). \tag{3.7}$$

Gleichung (3.5) kann über die Definition des Korrelationsterms hergeleitet werden. Hierfür sei x die Variable, die das Ergebnis von Labor i parametrisiert, d.h. $E(x) = x_i$ und $E(x^2) - E(x)^2 = u^2(x_i)$, wobei $E(x^n)$ den Erwartungswert der Größe x^n repräsentiert, dessen Definition gegeben ist durch

$$E(x^n) = \int_{-\infty}^{\infty} \mathrm{d}x \; p_i(x) x^n. \tag{3.8}$$

Die Kovarianz zwischen x_i und x_{ref} ist definiert als

$$u(x_i; x_{ref}) = E\big(x \cdot x_{ref}(x)\big) - E(x) \cdot E\big(x_{ref}(x)\big) = E(x \cdot x_{ref}) - x_i x_{ref} \tag{3.9}$$

mit $E\big(x_{ref}(x)\big) = x_{ref}$ und

$$x_{ref}(x) = \left(w_i \cdot x + \sum_{\substack{j=1 \\ j \neq i}}^{M} w_j \cdot x_j \right) = w_i(x - x_i) + x_{ref} \quad \text{und} \quad i \in [1; M], \tag{3.10}$$

d.h. das von Labor i berichtete Ergebnis fließt in die Berechnung des Referenzwertes ein und wird parametrisiert durch die Variable x. Damit folgt

$$E(x_i \cdot x_{ref}) = \int_{-\infty}^{\infty} \mathrm{d}x \; p_i(x) x \cdot [w_i(x - x_i) + x_{ref}] = w_i(E(x^2) - E(x)x_i) + E(x)x_{ref} \tag{3.11}$$

Mit $E(x) = x_i$ folgt

$$E(x_i \cdot x_{\mathrm{ref}}) = w_i(E(x^2) - x_i^2) + x_i x_{\mathrm{ref}}. \tag{3.12}$$

Die Kombination von Gleichung (3.10) und (3.7) liefert dann

$$u(x_i; x_{\mathrm{ref}}) = w_i(E(x^2) - x_i^2) = w_i(E(x^2) - E(x)^2) = w_i \cdot u^2(x_i), \tag{3.13}$$

womit das Ergebnis aus Gleichung (3.5) hergeleitet ist. Allgemein gilt daher

$$u(d_i) = (1 - 2w_i)u^2(x_i) + u^2(x_{\mathrm{ref}}), \tag{3.14}$$

Wobei w_i die entsprechende Gewichtung ist. Für den arithmetischen Mittelwert ist die Gewichtung beispielsweise durch $w_i = 1/M$ gegeben.

Die Berechnung des E_{n}-Wertes kann damit zusammengefasst werden als

$$E_{\mathrm{n};i} = \begin{cases} \dfrac{x_i - x_{\mathrm{ref}}}{\sqrt{U^2(x_i) + U^2(x_{\mathrm{ref}})}}, & \text{Methode A, B, C, D wenn Teilnehmer } i \text{ nicht Teil von } x_{\mathrm{ref}} \\ \dfrac{x_i - x_{\mathrm{ref}}}{2\sqrt{(1 - 2w_i)u^2(x_i) + u^2(x_{\mathrm{ref}})}}, & \text{Methode B, C, D wenn Teilnehmer } i \text{ Teil von } x_{\mathrm{ref}} \end{cases} \tag{3.15}$$

Die Bestimmung des Referenzwertes nach Methode C ist ebenfalls mit den Teilnehmerergebnissen korreliert, d.h. es besteht ein funktionaler Zusammenhang zwischen x_i und x_{ref}. Eine analytische Bestimmung der Kovarianz $u(x_i; x_{\mathrm{ref}})$ ist jedoch nicht möglich, bzw. aufgrund des iterativen Algorithmus nicht intuitiv exakt zugänglich. In DIN ISO 13528 wird kein Hinweis auf die Berechnung der Korrelation, bzw. die Notwendigkeit einer Korrelationsberücksichtigung beschrieben.

An dieser Stelle sei angemerkt, dass der $E_{\mathrm{n};i}$-Wert bei Vorliegen einer Korrelation stets definiert ist, falls

$$w_i \leq \frac{1}{2} \cdot \left(\frac{u^2(x_{\mathrm{ref}})}{u^2(x_i)} + 1\right) \tag{3.16}$$

Sowohl für den arithmetischen als auch für den gewichteten und den power-moderated Mittelwert ist dies stets erfüllt. Da $w_i > 0$ per Definition, ist

$$(1 - 2w_i)u^2(x_i) + u^2(x_{\mathrm{ref}}) < u^2(x_i) + u^2(x_{\mathrm{ref}}), \tag{3.17}$$

sodass Teilnehmer, die in die Bestimmung des Referenzwertes einfließen, stets ein schärferes E_{n}-Kriterium erfüllen müssen, als Teilnehmer, die nicht in die Berechnung einfließen und somit mit dem Referenzwert nicht korreliert sind.

Aus dieser Diskussion wird klar, dass eine Nicht-Berücksichtigung der Korrelation zwischen Teilnehmer und Referenz in Methode C zu einer Unterschätzung des E_{n}-Wertes führt, d.h. im schlimmsten Fall zu einer falsch-positiven Leistungsbewertung. Aus diesem Grund wird folgender Gewichtungsfaktor w_i in Gleichung (3.15) für Methode C (in Anlehnung an den arithmetischen Mittelwert) definiert und berücksichtigt:

$$w_i = \frac{w_{i;\mathrm{aux}}}{\sum_{i=1}^{N+1} w_{i;\mathrm{aux}}} \quad \text{mit} \quad w_{i;\mathrm{aux}} = \begin{cases} \dfrac{1}{N+1}, & x_{i;n_{\mathrm{c}}} = x_i \\ 0, & x_{i;n_{\mathrm{c}}} \neq x_i \end{cases} \tag{3.18}$$

Wie in Abbildung 3.14 gezeigt, besteht zwischen Referenz- und Teilnehmerwert keine Abhängigkeit mehr, sobald der Teilnehmerwert modifiziert wird, d.h. wenn $x_{i;n_{\mathrm{c}}} \neq x_i$. Wie in Abbildung 3.14 dargestellt, ist die Steigung in diesem Fall gleich null. Da der Gewichtungskoeffizient der Korrelation w_i gerade durch die Steigung gegeben ist (vgl. auch Gleichungen (3.8) bis (3.13)), gilt $w_i = 0$ falls $x_{i;n_{\mathrm{c}}} \neq x_i$, wie in Gleichung (3.18) definiert. Im Bereich $x_{i;n_{\mathrm{c}}} = x_i$ kann die Abhängigkeit von x_i durchaus auch nicht-linear sein, da zur Bestimmung des Referenzwertes auch die Standardabweichung einfließt und diese quadratisch von x_i abhängt. Allerdings werden die höheren Momente hier nicht berücksichtigt und die Abhängigkeit des Referenzwertes vom Teilnehmerwert x_i wird linear approximiert, wobei die Steigung durch Gleichung (3.18) gegeben ist, da diese mit dem Gewichtungskoeffizienten w_i übereinstimmt. Diese lineare Approximation deckt den Grenzfall des arithmetischen Mittelwertes ab.

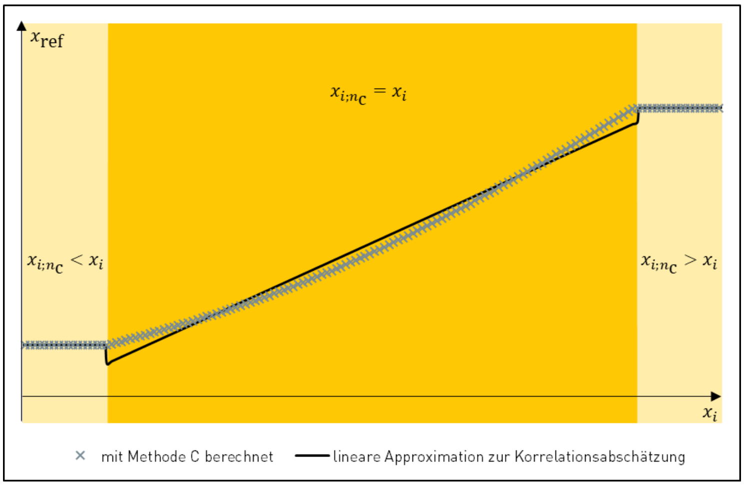


Abbildung 3.14: Eine exemplarische Abhängigkeit zwischen Referenzwert und Teilnehmerwert ist gezeigt. Wird der Teilnehmerwert modifiziert, besteht kein Zusammenhang mehr zwischen Teilnehmer- und Referenzwert, d.h. die Steigung ist null und somit auch die Korrelation. Wird der Teilnehmerwert durch den Algorithmus nicht verändert, so kann die Abhängigkeit linear approximiert werden. Der in Gleichung (3.18) angegebene Wert entspricht der lineare Approximation der Steigung.

Durch die Normierung in Gleichung (3.18) wird sichergestellt, dass

$$w_i \in [0; 1] \quad \text{und} \quad \sum_{i=1}^{N+1} w_i = 1. \tag{3.19}$$

Sei N_{mod} die Anzahl der Teilnehmer, für die gilt $x_{i;n_c} = x_i$, dann gilt

$$w_i = \begin{cases} \dfrac{1}{N_{\mathrm{mod}}}, & x_{i;n_c} = x_i \\ 0, & x_{i;n_c} \neq x_i \end{cases}. \tag{3.20}$$

Da maximal ein Ausreißeranteil von 25 % erlaubt ist, gilt $\frac{3}{4}(N+1) \leq N_{mod} \leq N+1$. Zudem erfordert der Algorithmus mindestens 4 Datensätze, d.h. $N \geq 3$, sodass $N_{\mathrm{mod}} \geq 3$. Damit folgt insgesamt

$$w_i \leq \frac{1}{3}. \tag{3.21}$$

Damit kann gezeigt werden, dass Gleichung (3.16) gilt, denn

$$\frac{1}{2} \cdot \left(\frac{u^2(x_{\mathrm{ref}})}{u^2(x_i)} + 1 \right) \geq \frac{1}{2} > \frac{1}{3} \geq w_i, \tag{3.22}$$

d.h. unter der in Gleichung (3.18) angegebenen Berechnung der Koeffizienten w_i, ist der E_{n}-Wert stets definiert.

Abweichungen zur Definition des E_{n}-Wertes in Gleichung (3.15) sind dann möglich, wenn festgestellt wird, dass die Stabilitätskriterien des Prüfgegenstandes nicht erfüllt sind. Das Vorgehen in diesem Fall ist im nachfolgenden Abschnitt beschrieben.

5.4.4. Stabilität

Um die Stabilität des Prüfgegenstandes zu überprüfen, führt die esz AG calibration & metrology Anfangs-, Zwischen und Abschlussmessungen durch. Abweichend davon können die Stabilitätsmessungen auch von einem teilnehmenden Labor oder dem Referenzlabor durchgeführt werden. Die nachfolgenden Ausführungen betrachten den Fall, dass die esz AG die Stabilitätsmessungen durchführt. Insbesondere dienen die Zwischen- und Abschlussmessungen lediglich der Überprüfung der Stabilität und fließen nicht in die Bestimmung des Referenzwertes ein (Methode B, C, D). Dadurch wird insbesondere vermieden, dass die Messergebnisse der esz AG im Vergleich zu den anderen Teilnehmern stärker gewichtet werden und Korrelationen den Referenzwert beeinflussen. Bei Methode A findet zudem eine Kalibrierung durch ein Referenzlabor etwa in der Mitte der Vergleichsrunde statt. Diese Messungen können ebenfalls zur Bewertung der Stabilität herangezogen werden. Für die Bewertung der Stabilität genügt die Durchführung einer Anfangs- und Abschlussmessung. Die Durchführung von Zwischenmessung dient der Risikominimierung und der präziseren Bewertung der Stabilität, sowie der Bestimmung der Stabilitätsbeiträge. Die Anzahl der Zwischenmessung kann je nach Eignungsprüfung (u.a. aufgrund der Teilnehmerzahl) variieren und wird vorab im technischen Protokoll kommuniziert. Abbildung 3.16 zeigt einen möglichen schematischen Ablauf der Eignungsprüfung.

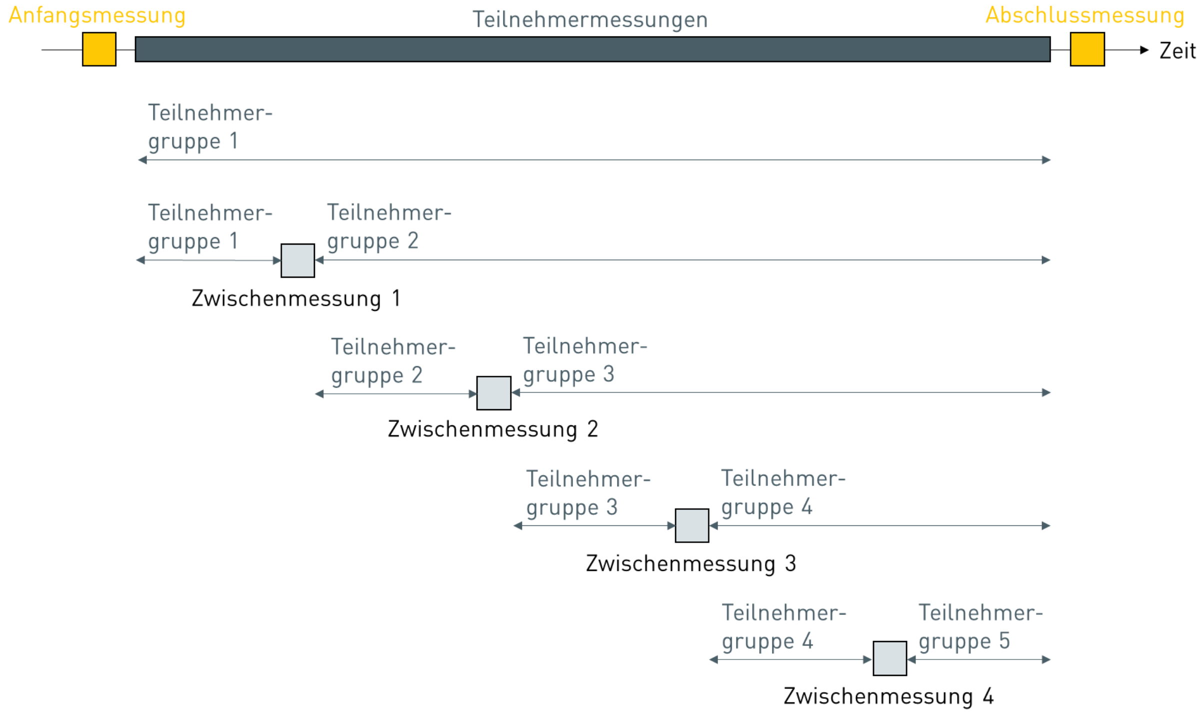


Abbildung 3.16: Möglicher zeitlicher Verlauf der in einer Eignungsprüfung durchgeführten Messungen. Bei Methode A findet die Referenzmessung in etwa in der Mitte der Eignungsprüfung statt und kann ggf. eine Zwischenmessung ersetzen. Die konkrete Anzahl der Zwischenmessungen kann je nach Teilnehmerzahl variieren. Gezeigt ist zudem die Gruppierung der Teilnehmermessungen in Abhängigkeit von der Anzahl der Zwischenmessungen

Die Gruppierung der Teilnehmermessungen erfolgt entsprechend Abbildung 3.16 und hängt von der konkreten Anzahl der Zwischenmessungen ab. Sei $Z \geq 0 \in \mathbb{N}$ die Gesamtzahl der Zwischenmessungen, dann gehören alle Teilnehmer, deren Messungen zwischen Anfangsmessung und der ersten Zwischenmessung, bzw. der Abschlussmessung für $Z = 0$ zur Teilnehmergruppe 1. Alle Messungen von Teilnehmern zwischen den Zwischenmessung m und $m+1$ gehören zur Teilnehmergruppe $m+1$. Dies gilt auch, wenn $m=Z$ ist und somit die Zwischenmessung $m+1$ durch die Abschlussmessung gegeben ist. Bei Z Zwischenmessungen liegen somit $Z+1$ Teilnehmergruppen vor.

Für die Anfangsmessung wird der Index „stab; 0" verwendet, für die Abschlussmessung der Index „stab;$Z+1$". Die m-te Zwischenmessung trägt den Index „stab;m".

Die Stabilitätskriterien 1 bis $Z+1$ werden zwischen zwei aufeinanderfolgenden Stabilitätsmessungen bewertet und sind wie folgt definiert:

$$\frac{\left|x_{\mathrm{stab};m}-x_{\mathrm{stab};m+1}\right|}{\sqrt{U^2\left(x_{\mathrm{stab};m}\right)+U^2\left(x_{\mathrm{stab};m+1}\right)}}<\frac{1}{2} \quad \text{mit} \quad m \in [0;Z] \tag{4.1}$$

Zusätzlich wird das Stabilitätskriterium 0 in Gleichung (4.2) definiert um die Gesamtstabilität durch den Vergleich von Anfangs- und Abschlussmessung zu bewerten.

$$\frac{\left|x_{\mathrm{stab};0}-x_{\mathrm{stab};Z+1}\right|}{\sqrt{U^2\left(x_{\mathrm{stab};0}\right)+U^2\left(x_{\mathrm{stab};Z+1}\right)}}<\frac{1}{2} \tag{4.2}$$

Entsprechend liegen bei Z durchgeführten Eignungsprüfungen $Z+2$ Stabilitätskriterien vor. Nachfolgend soll die Wahl der oberen Grenze (linke Seite in Gleichung (4.1) bzw. (4.2) gleich 1/2) erläutert werden. Die Standardabweichung aus zwei Messwerten kann vereinfacht werden zu

$$\frac{1}{\sqrt{2-1}}\sum_{i=1}^{2}(x_1-\bar{x})^2=\frac{|x_1-x_2|}{\sqrt{2}}, \tag{4.3}$$

mit $\bar{x}=(x_1+x_2)/2$ als arithmetischer Mittelwert. Sei nun

$$u=\frac{1}{2}\cdot\max\left(U(x_1);U(x_2)\right), \tag{4.4}$$

dann gilt nach Gleichung (4.1) bzw. (4.2)

$$\frac{1}{2}>\frac{|x_1-x_2|}{\sqrt{U^2(x_1)+U^2(x_2)}}\geq\frac{|x_1-x_2|}{2\sqrt{2}\cdot u}, \tag{4.5}$$

sodass folgt

$$\frac{1}{2}>\frac{|x_1-x_2|}{2\sqrt{2}\cdot u} \quad\Leftrightarrow\quad \frac{|x_1-x_2|}{\sqrt{2}}<u. \tag{4.6}$$

Mit Gleichung (4.6) wird klar, dass das Stabilitätskriterium in Gleichung (4.1) bzw. (4.2) durch die Wahl der oberen Grenze (= 1/2) eine Streuung (vgl. einfache ($k=1$), empirische Standardabweichung zweier Messwerte, Gleichung (4.3)) zwischen Referenzlabor und Anfangs-/Abschlussmessung erlaubt, die in der Größenordnung der maximalen, einfachen ($k=1$) Messunsicherheit liegt. Eine eventuelle Korrelation zwischen den Stabilitätsmessungen findet keine Berücksichtigung im definierten Stabilitätskriterium. Angenommen $x_{\mathrm{stab};i}$ und $x_{\mathrm{stab};j}$ sind stark positiv korreliert, dann entspräche das einem Gewichtungsfaktor $w_i \simeq 1$ in Gleichung (3.5) und Gleichung (3.7). Die resultierende Unsicherheit im Nenner würde dann sehr klein werden und das Stabilitätskriterium wäre kaum noch zu erfüllen. Um das zu vermeiden, wird die Korrelation in der Definition des Stabilitätskriteriums nicht berücksichtigt.
Existiert ein m, für das Gleichung (4.1) nicht erfüllt ist, d.h. eine nicht zufällige Differenz zwischen zwei zeitlich aufeinander folgenden Stabilitätsmessungen vorliegt, so ist ggf. in der E_{n}-Auswertung der Teilnehmergruppe m ein Stabilitätsbeitrag u_{stab} quadratisch zur Referenzunsicherheit und

Teilnehmerunsicherheit zu addieren, vgl. Gleichung (4.7). Die Bestimmung des Stabilitätsbeitrages der Teilnehmergruppe m kann erst im konkreten Anwendungsfall durchgeführt werden.

$$E_{\mathrm{n};i} = \frac{x_i - x_{\mathrm{ref}}}{\sqrt{u^2(x_i) + u^2(x_{\mathrm{ref}}) - 2u(x_i; x_{\mathrm{ref}}) + u^2_{\mathrm{stab};m}}}, \tag{4.7}$$

wobei Labor i Teil der Teilnehmergruppe m sein muss. Eine mögliche Berechnung des Stabilitätsbeitrages ist durch

$$u_{\mathrm{stab};m} = \frac{|x_{\mathrm{stab};m} - x_{\mathrm{stab};m+1}|}{2 \cdot \sqrt{3}} \quad \text{oder} \quad u_{\mathrm{stab};m} = \frac{|x_{\mathrm{stab};m} - x_{\mathrm{stab};m+1}|}{\sqrt{3}} \tag{4.8}$$

gegeben. Alternativ kann auch die empirische Standardabweichung verwendet werden. Die Größenordnung des Stabilitätsbeitrags ist im Vergleich zu den Teilnehmerunsicherheiten, bzw. im Vergleich zur Referenzunsicherheit zu analysieren. Bei Fehlschlagen des Stabilitätskriteriums aus Gleichung (4.1) ist es auch möglich, dass der betreffende Stützpunkt aus der Auswertung genommen wird. Gleiches gilt, falls das Stabilitätskriterium aus Gleichung (4.2) fehlschlägt. Prinzipiell sei an dieser Stelle aber darauf hingewiesen, dass ein fehlschlagen des Stabilitätskriteriums nicht zwangsläufig schlecht ist. Das Stabilitätskriterium überprüft lediglich, ob die beobachtete Streuung innerhalb der für das Verfahren und das verwendete Prüfmittel erwarteten Unsicherheit liegt (vgl. Unsicherheit der Kalibrierung). Da Eigenschaften des Eignungsprüfungsobjektes, wie Lang- oder Kurzzeitstabilität typischerweise nicht in die Messunsicherheit einer Kalibrierung einfließen, kann es durchaus möglich sein, dass die beobachtete Streuung der Messwerte, die auf das Eignungsprüfungsobjekt zurück zu führen ist, größer ist als die Messunsicherheit, die in das Stabilitätskriterium einfließt. Ein nicht bestandenes Stabilitätskriterium ist daher nicht grundsätzlich bedenklich, jedoch ist es generell zu diskutieren und eventuell ergänzte Stabilitätsbeiträge sollten mit der Größenordnung der erwarteten Stabilität des Eignungsprüfungsobjektes verglichen werden. Informativ wird im Ergebnisbericht angegeben, ob die Unsicherheit eines Teilnehmers größer ist als die Kombination aus Referenzunsicherheit und Teilnehmerunsicherheit. Diese Auswertung fließt jedoch nicht in die Leistungsbewertung ein, die Interpretation bleibt dem jeweiligen Teilnehmer überlassen. Intern wird zudem die Größenordnung des Stabilitätsbeitrags mit einer für das Eignungsprüfungsobjekt bekannten Stabilität verglichen (bspw. Spezifikation, MU des Eignungsprüfungsobjektes, bekannte Drift, etc.). Sollte die beobachtete Instabilität während der Eignungsprüfung größer sein als die erwartete Stabilität, sind entsprechende Konsequenzen zu ziehen. Die nachfolgende Tabelle zeigt eine Übersicht möglicher Instabilitäten und der entsprechenden Konsequenzen. Die Übersicht ist nicht abschließend.

Beispiel der aufgetretenen Instabilität	**Mögliche Konsequenz**
1. Stabilität während der Eignungsprüfung deutlich größer als typische Stabilität des Eignungsprüfungsgegenstandes	▪ Ursachenanalyse (wird im Bericht der Eignungsprüfung kommuniziert) ▪ Analyse, ob Berücksichtigung eines Stabilitätsbeitrages im Vergleich zu den vorliegenden Teilnehmerunsicherheiten sinnvoll ist ▪ Ggf. Ausschluss der betreffenden Messpunkte von der Aus- und Bewertung der Eignungsprüfung ▪ Ggf. Ausschluss des Prüflings von zukünftigen Eignungsprüfungen ▪ Sonstige

2. Stabilitätskriterium aus Gleichung (4.2) schlägt fehl, aber alle Stabilitätskriterien aus Gleichung (4.1) fallen positiv aus	▪ Ursachenanalyse (wird im Bericht der Eignungsprüfung kommuniziert), möglicherweise liegt eine lineare Drift des Eignungsprüfungsgegenstandes während der Eignungsprüfung vor ▪ Detaillierte Analyse aller Stabilitätsmessung vor dem Hintergrund der Referenz- und Teilnehmermessungen (Methode A: Wann fand die Referenzmessung statt? Methode D: Ist die Instabilität hinreichend über die Parameter s, S und α abgedeck?) ▪ Ggf. Berücksichtigung von Stabilitätsbeiträgen Ggf. Ausschluss der betreffenden Messpunkte von der Aus- und Bewertung der Eignungsprüfung ▪ Ggf. Ausschluss des Prüflings von zukünftigen Eignungsprüfungen ▪ Sonstige
3. Ein oder mehrere Stabilitätskriterien aus Gleichung (4.2) schlagen fehl	▪ Ursachenanalyse (wird im Bericht der Eignungsprüfung kommuniziert) ▪ Berücksichtigung von Stabilitätsbeiträgen für die entsprechenden Teilnehmergruppen ▪ Ggf. Ausschluss der betreffenden Messpunkte von der Aus- und Bewertung der Eignungsprüfung ▪ Ggf. Ausschluss des Prüflings von zukünftigen Eignungsprüfungen ▪ Sonstige

An dieser Stelle sei auch angemerkt, dass der Stabilitätsbeitrag als zusätzlicher Beitrag zu erachten ist und nicht allein dem Referenzwert zugeordnet wird. Die Referenzunsicherheit ist lediglich die Unsicherheit, die der Bestimmung des Referenzwertes zuzuordnen ist. Der Stabilitätsbeitrag hingegen berücksichtigt zusätzliche Instabilitäten, die nicht zwangsläufig in der Referenzunsicherheit enthalten sind. Insbesondere kann eine Eignungsprüfung daher auch für Teilnehmer nützlich sein, die eine kleinere Unsicherheit als $\sqrt{u^2(x_{\text{ref}}) + u_{\text{stab}}^2}$ berichten. Der Grund hierfür ist, dass Teilnehmer prinzipiell keine Möglichkeit zur Abschätzung eines Stabilitätsbeitrages in ihrer Unsicherheitsbetrachtung haben, sodass durchaus kleine Unsicherheiten seitens der Teilnehmer berichtet werden können. Idealerweise sind diese jedoch größer als die Referenzunsicherheit $u(x_{\text{ref}})$.

Im Fall, dass der Referenzwert gemäß Methode B, C oder D bestimmt wird, fließen Teilnehmerergebnisse in den Referenzwert mit ein. Idealerweise werden die Teilnehmer, die voraussichtlich in die Bestimmung des Referenzwertes einfließen (z.B. akkreditierte Labore), sinnvoll in die Eignungsprüfungsrunde verteilt. Methoden, die einen zeitlich driftenden Referenzwert berechnen, werden aufgrund ihrer Komplexität und der nicht ausreichenden Datenkenntnis, nicht angewandt. An dieser Stelle sei erneut darauf hingewiesen, dass die Referenzunsicherheit lediglich die

Unsicherheit der Bestimmung des Referenzwertes darstellt. Auch wenn Teilnehmerdaten in die Bestimmung des Referenzwertes einfließen, ist damit nicht unmittelbar klar, dass eventuelle Instabilitäten des Prüfgegenstands berücksichtigt sind. Methode B liefert beispielsweise stets eine kleinere Unsicherheit als alle Teilnehmerunsicherheiten und berücksichtigt Streuungen der Teilnehmerabweichungen x_i nicht. Methode C und D hingegen berücksichtigen die Streuung der Teilnehmerabweichungen, allerdings werden in beiden Methoden Ausreißer eliminiert, sodass auch eventuelle Instabilitäten ausgeschnitten werden können. Zudem ist sowohl für alle drei vorgestellten Berechnungsmethoden die Referenzunsicherheit umso kleiner, je mehr Teilnehmer in die Berechnung einfließen. Auch das kann dazu führen, dass eventuelle Instabilitäten des Prüfgegenstands nicht korrekt berücksichtigt sind. Aus diesem Grund werden, falls nötig, Stabilitätsbeiträge stets separat und zusätzlich zur Referenzunsicherheit angesetzt.

5.4.5. Wahrscheinlichkeiten als Leistungshinweis

Neben der Auswertung des E_n-Wertes, was als einziges Kriterium zur Leistungsbewertung dient, kann es hilfreich sein, die Wahrscheinlichkeit anzugeben, mit der ein Labor i mit dem Referenzwert übereinstimmt. Hierfür kann die Angabe der Wahrscheinlichkeit $P_\text{Ü}$, dass Referenzergebnis und Labor i einen Wert im Bereich $\left[\max\left(x_i - U(x_i); x_\mathrm{ref} - U(x_\mathrm{ref})\right); \min\left(x_i + U(x_i); x_\mathrm{ref} + U(x_\mathrm{ref})\right)\right]$ liefern, hilfreich sein. Es sei darauf hingewiesen, dass die Wahrscheinlichkeiten $P_\text{Ü}$ nur dann berechnet werden kann, wenn sowohl für den Referenzwert, als auch für das von Labor i mitgeteilte Ergebnis die Wahrscheinlichkeitsdichten $p_i(x)$ und $p_\mathrm{ref}(x)$ bekannt sind. Wie eingangs beschrieben, wird falls vom teilnehmenden Labor keine weitere Annahme gemacht wird, davon ausgegangen, dass $p_i(x)$ durch eine Normalverteilung gegeben ist. Wird hingegen explizit eine andere Verteilung genannt, so kann die Auswertung der Wahrscheinlichkeit nur für eine Student-t, Rechteck-, Dreieck- oder U/arcsin-Verteilung berechnet werden. Werden andere Wahrscheinlichkeitsdichten angegeben, so wird die Wahrscheinlichkeit $P_\text{Ü}$ nicht berechnet und erscheint nicht im Ergebnisbericht.

Wahrscheinlichkeit $P_\text{Ü}$ ist definiert als

$$P_\text{Ü} = \int_{\max(x_i - U(x_i); x_\mathrm{ref} - U(x_\mathrm{ref}))}^{\min(x_i + U(x_i); x_\mathrm{ref} + U(x_\mathrm{ref}))} \mathrm{d}x\, p_i(x) \int_{\max(x_i - U(x_i); x_\mathrm{ref} - U(x_\mathrm{ref}))}^{\min(x_i + U(x_i); x_\mathrm{ref} + U(x_\mathrm{ref}))} \mathrm{d}x'\, p_\mathrm{ref}(x') = P_{\text{Ü};i} \cdot P_{\text{Ü};\mathrm{ref}}. \tag{5.1}$$

Im Wesentlichen ist $P_\text{Ü}$ somit definiert als die Wahrscheinlichkeit, dass sowohl das Referenzergebnis, als auch das von Labor i berichtete Ergebnis im Überschneidungsbereich der Unsicherheitsbalken liegen. Je größer der Wert für $P_\text{Ü}$ ist, desto besser ist die Übereinstimmung von Teilnehmer- und Referenzdaten. Wichtig ist jedoch, dass die Wahrscheinlichkeit $P_\text{Ü}$ nicht normiert ist, d.h. selbst wenn Referenz- und Teilnehmerergebnis exakt übereinstimmen resultiert bspw. bei einer Normalverteilung ein Wert $P_\text{Ü} \simeq 0{,}9$. Abbildung 3.17 zeigt die graphische Interpretation von Gleichung (5.1), für den Fall, dass $p_i(x)$ und $p_\mathrm{ref}(x)$ durch Normalverteilungen gegeben sind. Die Berechnung der Wahrscheinlichkeit $P_\text{Ü}$ ist jedoch nicht auf diese Annahme reduziert.

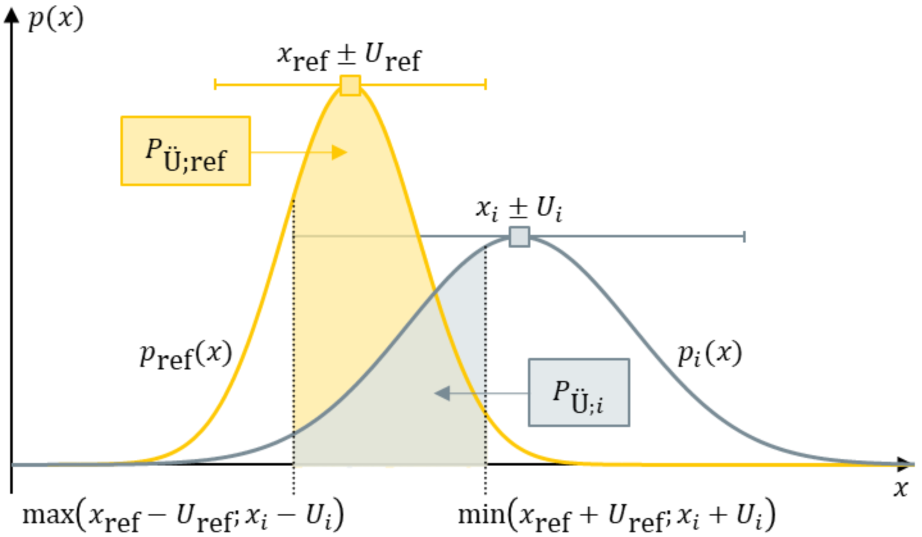


Abbildung 3.17: Die Abbildung zeigt die graphische Darstellung der zu Berechnung von $P_{\text{Ü}}$ nötigen Wahrscheinlichkeiten $P_{\text{Ü};i}$ und $P_{\text{Ü;ref}}$. Im gezeigten Beispiel wurde sowohl für $p_i(x)$ als auch für $p_{\mathrm{ref}}(x)$ eine Normalverteilung angenommen. Die Berechnungen sind aber nicht auf Normalverteilungen beschränkt.

Ergänzend zum E_{n}-Kriterium kann die Wahrscheinlichkeiten $P_{\text{Ü}}$ einen Hinweise auf die Leistung eines teilnehmenden Labors liefern, wenngleich diese kein Kriterium zur positiven oder negativen Bewertung der Eignungsprüfung darstellt. Die Wahrscheinlichkeit $P_{\text{Ü}}$ wird standardmäßig berechnet, unabhängig davon, ob das Stabilitätskriterium der entsprechenden Teilnehmergruppe erfüllt ist oder nicht und, wenn entweder von einer Normalverteilung, einer Student-t, Rechteck-, Dreieck- oder U/arcsin-Verteilung ausgegangen werden kann. Korrelationen werden in der Wahrscheinlichkeit $P_{\text{Ü}}$ nicht berücksichtigt. Auf Kundenwunsch kann Angabe der Wahrscheinlichkeit $P_{\text{Ü}}$ auch entfallen. Eine Bewertung der resultierenden Wahrscheinlichkeiten bleibt dem teilnehmenden Labor überlassen.

5.5. Graphische Darstellung von Eignungsprüfungsdaten

Um die statistische Auswertung der Eignungsprüfungsdaten anschaulich darzustellen, werden den teilnehmenden Laboren verschiedene Plots mitgeliefert.

5.5.1. Graphische Darstellung des Stabilitätskriteriums

Zur Darstellung der Anfangs- und Abschlussmessungen mit dem Referenzwert verglichen. Zur Veranschaulichung wird die nachfolgende Grafik verwendet. Dafür sind beispielhafte Werte hinterlegt.

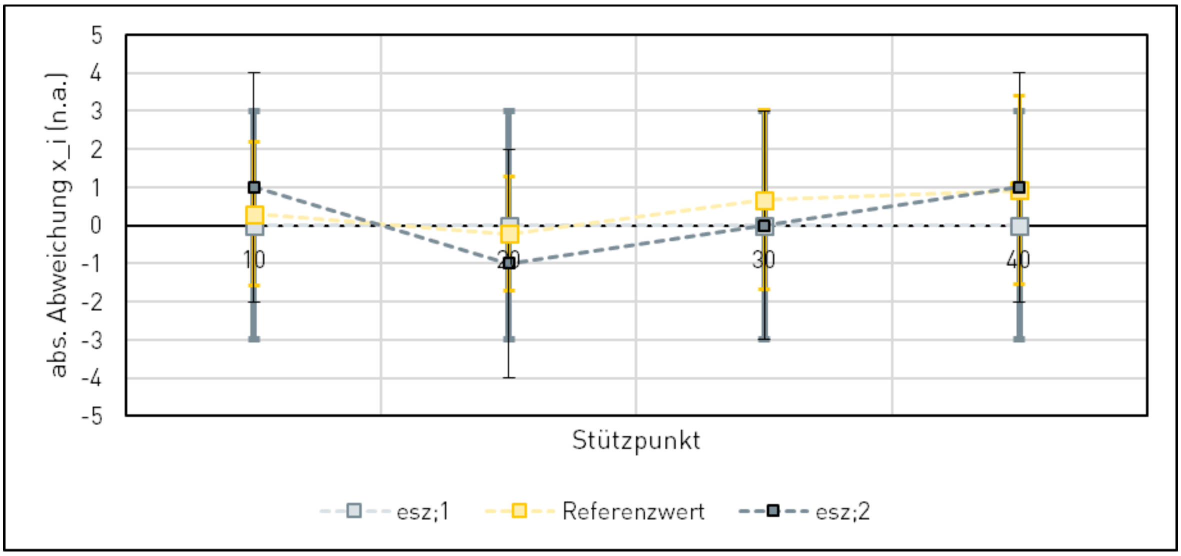
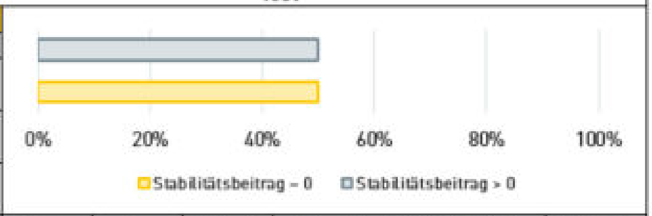


Abbildung 4.1: Vergleich der Anfangs- und Abschlussmessungen durch die esz AG mit dem Referenzwert und dessen Unsicherheit (ohne Berücksichtigung eines eventuellen Stabilitätsbeitrags).

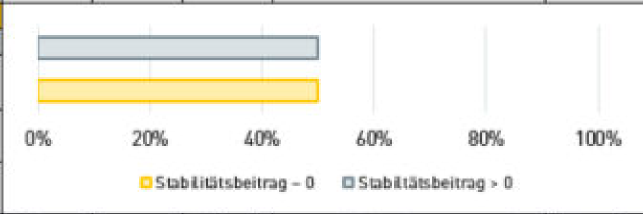
Außerdem werden die Stabilitätsbeiträge berichtet, wie in Abbildung 4.2 anhand von Beispielwerten gezeigt.

Messgröße	Test
Labor	esz;1
Anfangsmessung	
Anzahl gemessener Stützpunkte	8
Anzahl Stabilitätsbeitrag = 0	50%
Anzahl Stabilitätsbeitrag > 0	50%



Stützpunkt	$x_{esz;1}$ a.u.	$U(x_{esz;1})$ a.u.	x_{ref} a.u.	$U(x_{ref})$ a.u.	u_{Stab} Teilnehmergruppe 1 a.u.	In x_{ref} & $U(x_{ref})$?
1	1	2	1	1	0	JA
2	2	2	1	1	0	JA
3	4	2	1	1	1	JA
4	3	2	1	1	1	JA
5	7	2	1	1	1	JA
6	2	2	2	2	0	JA
7	4	2	2	2	1	JA
8	1	2	2	2	0	JA

Labor	esz;2
Abschlussmessung	
Anzahl gemessener Stützpunkte	8
Anzahl Stabilitätsbeitrag = 0	50%
Anzahl Stabilitätsbeitrag > 0	50%



Stützpunkt	$x_{esz;2}$ 0	$U(x_{esz;2})$ 0	x_{ref} a.u.	$U(x_{ref})$ a.u.	u_{Stab} Teilnehmergruppe 2 0	In x_{ref} & $U(x_{ref})$?
1	1	2	1	1	0	NEIN
2	2	2	1	1	0	NEIN
3	4	2	1	1	1	NEIN
4	3	2	1	1	1	NEIN
5	7	2	1	1	1	NEIN
6	2	2	2	2	0	NEIN
7	4	2	2	2	1	NEIN
8	1	2	2	2	0	NEIN

Abbildung 4.2: Darstellung der Stabilitätsbeiträge, neben den Werten der Anfangs- und Abschlussmessung.

5.5.2.Graphische Darstellung der Ergebnisse eines Teilnehmers

Die Teilnehmer werden mit einem Code anonymisiert. Im Beispiel ist dieser „01". Die graphische Darstellung der Ergebnisse eines Teilnehmers ist ähnlich zu den Plots aus Abschnitt 5.4.1. Allerdings wird hier nun die Unsicherheit des Referenzwertes erhöht, falls das Stabilitätskriterium für Gruppe 1 und/oder 2 nicht erfüllt ist und sich der betreffende Teilnehmer in dieser Gruppe befindet. Der zusätzliche Stabilitätsbeitrag ist unter 5.3.4 zu finden.

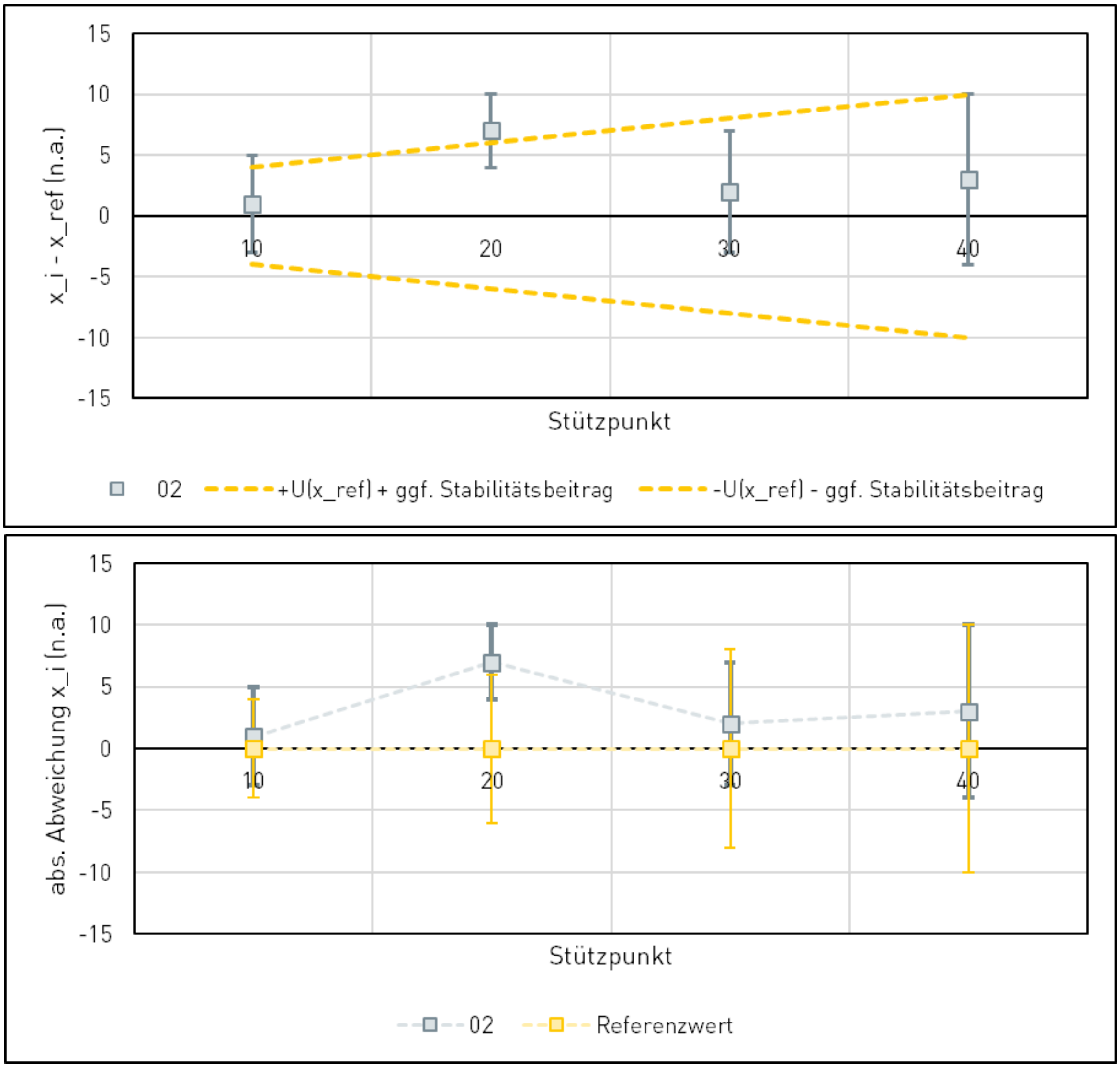


Abbildung 4.1: Die Ergebnisse eines Teilnehmers werden mit dem Referenzwert und dessen Unsicherheit verglichen.

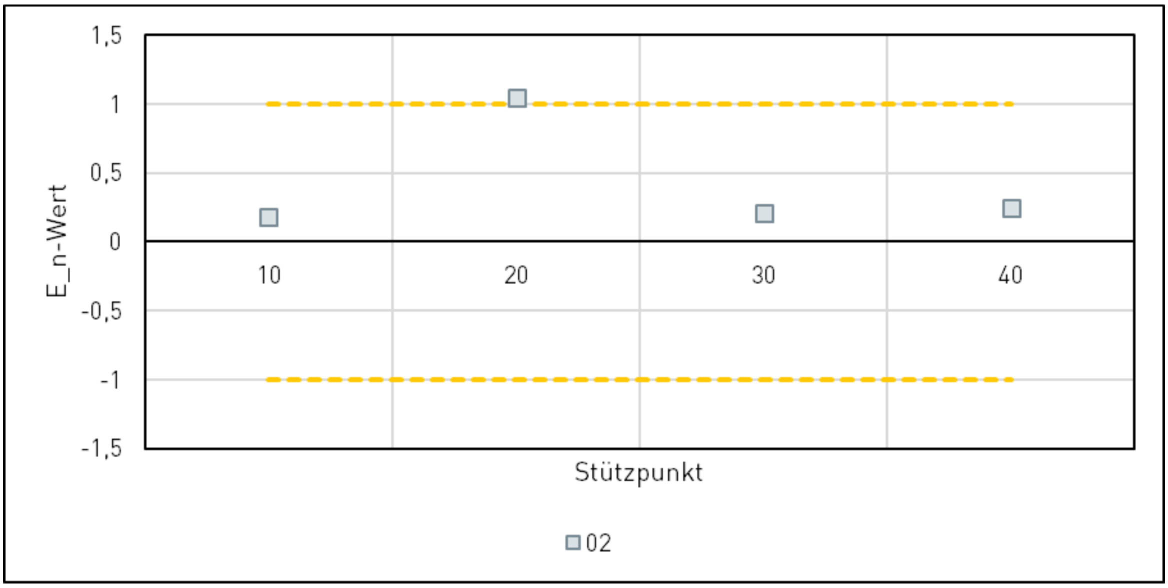
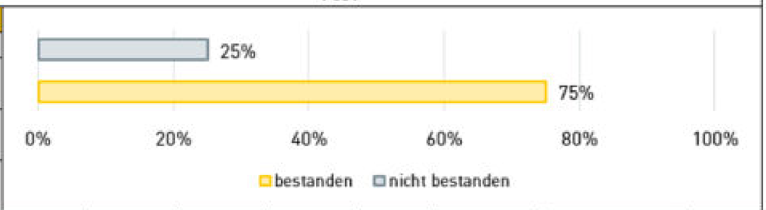


Abbildung 4.4: Darstellung des Leistungskriteriums. Für ein positives bestehen ist ein E_n-Wert zwischen -1 und +1 nötig.

Als Hinweis auf die Gesamtleistung eines Teilnehmers, wird außerdem das Verhältnis von bestandenen und nicht bestandenen Stützpunkten einer Messgröße dargestellt, siehe Abbildung 4.5. Dabei wird die Anzahl der abgegebenen Ergebnisse eines Teilnehmers berücksichtigt, falls diese kleiner ist als die Gesamtzahl der Stützpunkte. Abbildung 4.5 zeigt auch die tabellarische Übersicht, die jedem - Eignungsprüfungsteilnehmer zur Verfügung gestellt wird. Da bei Verwendung von Methode B, C und D Teilnehmer in die Bestimmung des Referenzwertes einfließen, bzw. auch ausgeschlossen werden können, wird für jeden Teilnehmer angegeben, ob dessen Messergebnis letztendlich in die Bestimmung des Referenzwertes eingeflossen ist oder nicht.

Messgröße	Test
Labor	01
Teilnehmergruppe	1
Anzahl eingereichte Ergebnisse	8
Anteil bestandene Stützpunkte	75%
Anteil nicht bestandene Stützpunkte	25%



Stützpunkt	x_i a.u.	$U(x_i)$ a.u.	x_{ref} a.u.	$U(x_{ref})$ a.u.	u_{stab} a.u.	$E_{n,i}$	Bestanden?	$u^2(x_i) > u^2(x_{ref})+u^2_{stab}$?	In x_{ref} & $U(x_{ref})$?
1	1	1	1	1	0	0,00	JA	NEIN	NEIN
2	0	1	1	1	0	-0,71	JA	NEIN	JA
3	3	1	1	1	0	1,41	NEIN	NEIN	NEIN
4	4	2	1	1	0	1,34	NEIN	JA	JA
5	0	5	1	1	0	-0,20	JA	JA	JA
6	1	5	2	2	0	-0,19	JA	JA	JA
7	0	2	2	2	0	-0,71	JA	NEIN	JA
8	1	2	2	2	0	-0,35	JA	NEIN	JA

Abbildung 4.5: Für jeden Teilnehmer wird für eine gegebene Messgröße das Verhältnis aus bestandenen und nicht bestandenen Stützpunkten graphisch dargestellt. Diese wird in die tabellarische Darstellung des Ergebnisses, des Teilnehmers mit dem Code „01" eingebettet.

5.5.3. Graphische Darstellung der Ergebnisse aller Teilnehmer

Um einen Überblick über die Ergebnisse aller Teilnehmer zu bekommen, werden die berichteten Abweichungen, sowie die dazugehörigen erweiterten Messunsicherheiten in Form von Balkendiagrammen gegenübergestellt.

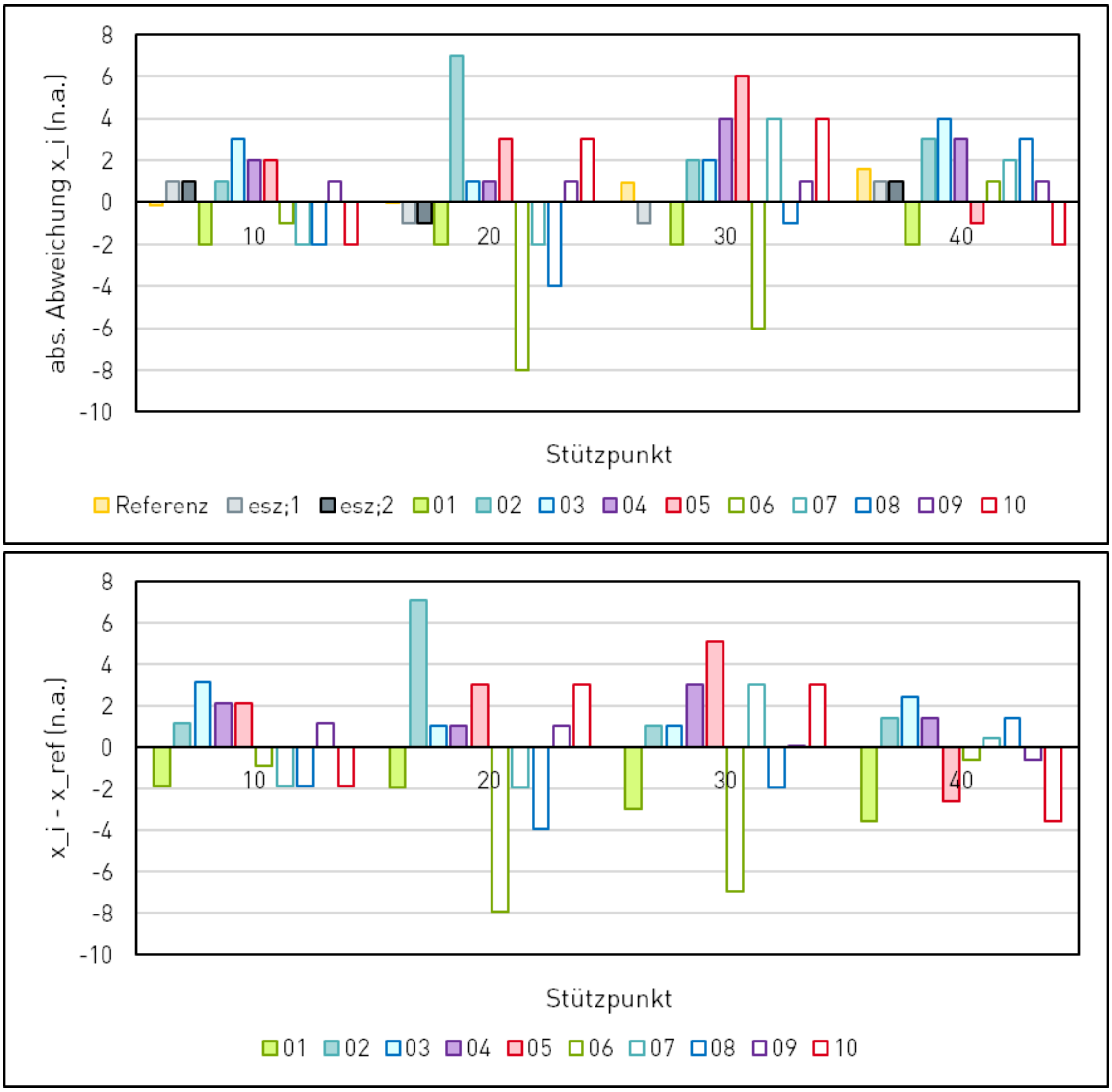


Abbildung 4.6: Darstellung der berichteten Abweichung (oben), sowie die Differenz zum Referenzwert (unten) aller Teilnehmer.

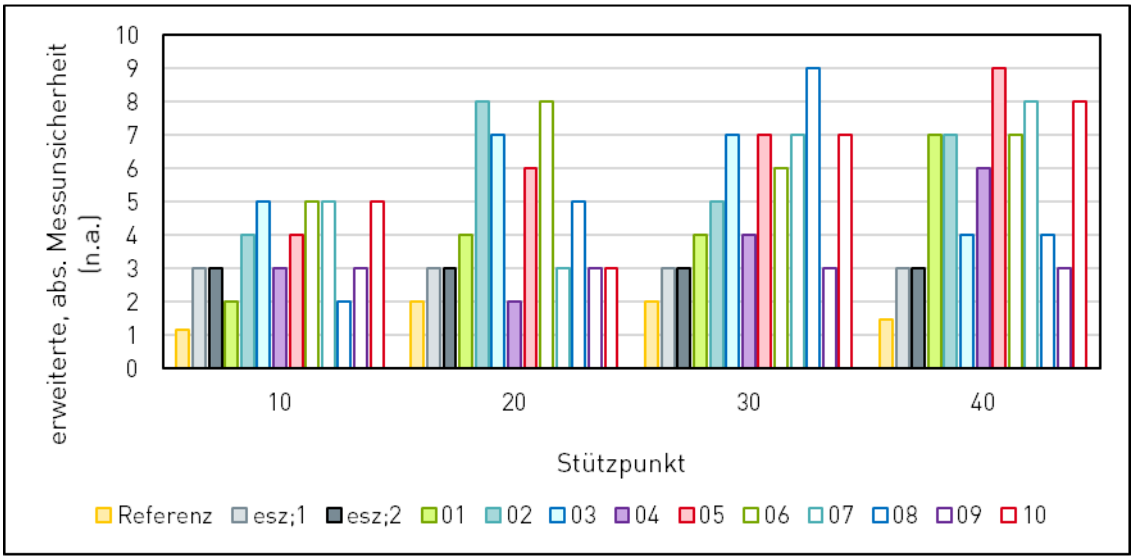


Abbildung 4.7: Darstellung der berichteten erweiterten Messunsicherheit aller Teilnehmer.

Neben dem Vergleich der von den Teilnehmern eingereichten Ergebnisse wird, wie in Abbildung 4.8 auch der E_n-Wert aller Teilnehmer in einer Übersicht dargestellt.

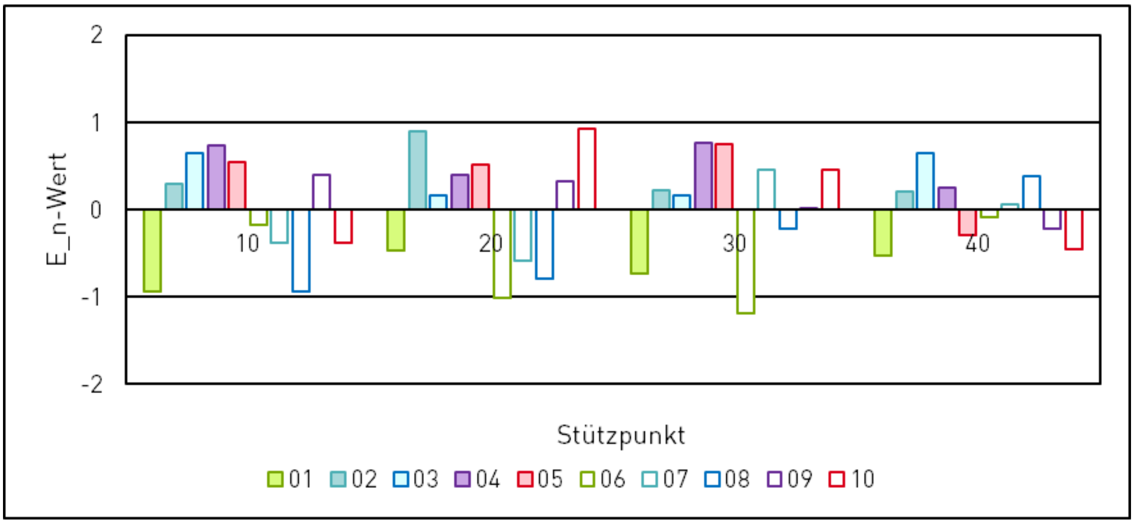


Abbildung 4.8: Darstellung aller E_n-Werte. Die Grenzen ± 1 legen das Leistungskriterium fest.

Zusätzlich wird das Verhältnis der insgesamt bestandenen und nicht bestandenen Ergebnisse graphisch dargestellt.

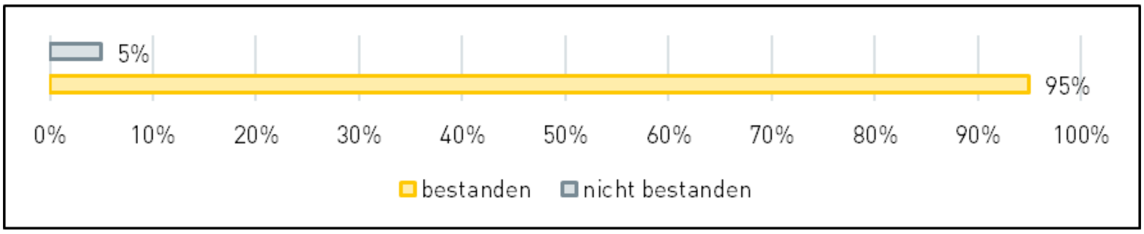


Abbildung 4.9: Darstellung des Anteils der insgesamt bestandenen und nicht bestandenen Ergebnisse.

Die in Abschnitt 5.3.5 beschriebene Wahrscheinlichkeit P_l, wird ebenfalls graphisch dargestellt, wie in Abbildung 4.10 gezeigt.

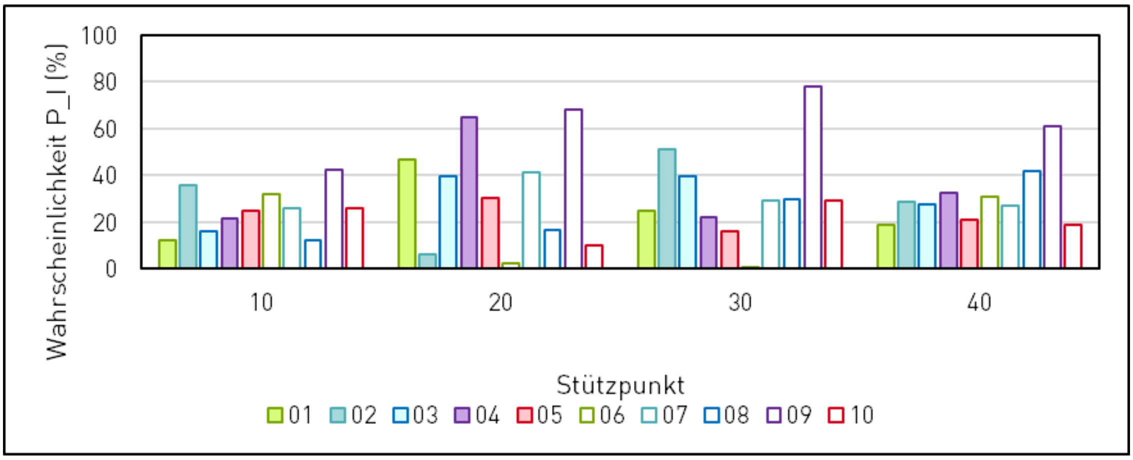


Abbildung 4.10: Darstellung der Wahrscheinlichkeit P_l als Leistungshinweis.