

Die Eignungsprüfung als Instrument der Qualitätssicherung und Kompetenzbewertung von Prüflaboratorien für mechanisch-technologische Werkstoffprüfungen

Vom Fachbereich Sicherheitstechnik der Bergischen Universität Wuppertal zur Erlangung des akademischen Grades eines Doktor-Ingenieurs genehmigte Dissertation

vorgelegt von

Diplom-Ingenieur Cord-Christoph Vogt aus Waiblingen

Promotionsausschuss:

Vorsitzender: Univ.-Prof. Dr.-Ing. A. Meyna

1. Gutachter: Univ.-Prof. Dr.-Ing. F. Deuerler

2. Gutachter: Univ.-Prof. Dr.-Ing. habil. P. Winzer

3. Gutachter: Univ.-Prof. Dr.-Ing. Dr.-Ing. e.h. B.-J. Vorath

Tag der mündlichen Prüfung: 09. Mai 2003

Wuppertal 2003

D 468

Zusammenfassung

Im Rahmen dieser Arbeit wurden erstmalig Eignungsprüfungen speziell für verschiedene mechanisch-technologische Werkstoffprüfverfahren konzipiert, mit insgesamt 600 Teilnehmern aus 25 Nationen durchgeführt und systematisch ausgewertet. Ziel der Untersuchung war die Entwicklung von Methoden zur Durchführung und Auswertung von Eignungsprüfungen, die einen prüftechnisch angemessenen und wissenschaftlich begründbaren Kompetenznachweis ermöglichen.

Die aus den Ergebnissen dieser Untersuchung resultierenden Eignungsprüfungen lassen sich für folgende Aufgaben verwenden:

Akkreditierte Prüflaboratorien können bei erfolgreicher Teilnahme einen belastbaren Kompetenznachweis erbringen und die in ISO/IEC 17025 gestellte Anforderung erfüllen, an „einem geeigneten Programm für Vergleichsmessungen zwischen Laboratorien“ teilzunehmen.

Nicht akkreditierten Laboratorien ermöglichen diese Eignungsprüfungen erstmalig die Beurteilung ihrer Prüfleistungen im Vergleich mit führenden europäischen Laboratorien auf Grundlage belastbarer Kriterien. Die aus den Daten der Eignungsprüfungen gewonnenen Ergebnisse können zusätzlich zur Weiterentwicklung und Normung der mechanisch-technologischen Prüfverfahren, z. B. zur Bestimmung der Messunsicherheit von Härtemessungen, verwendet werden.

Abstract

Within the scope of this thesis proficiency tests in particular for mechanical materials testing laboratories were designed, conducted with 600 participants from 25 countries and systematically analysed. Main objective of the analysis was the development of methods for execution and evaluation of proficiency tests, that will make possible an appropriate proof of competence on a scientific basis.

The resulting proficiency tests based on this analysis can be used for the following:

Provided a successful participation accredited laboratories can sustainable proof evidence of their competence and fulfil the requirement of ISO/IEC 17025 “to participate in a suitable programme of interlaboratory comparisons”.

The proficiency tests enable non-accredited laboratories for the first time to compare their testing results with European leading laboratories based on sustainable criteria.

The data derived from the proficiency tests can additionally be used for further development and standardisation of mechanical testing methods, e. g. for the determination of measurement uncertainty in hardness testing.

Résumé

Dans le cadre de cette étude on a élaboré pour la première fois des tests de qualification s'appliquant spécifiquement à différentes méthodes d'essais mécaniques de matériau. 600 représentants de 25 nations ont participé à la réalisation et à l'évaluation de ces tests. L'objectif de l'étude était le développement de méthodes destinées à réaliser et à évaluer des tests de qualification permettant de démontrer la compétence sur la base de critères acceptables des points de vue technique et scientifique.

Les tests de qualification résultant de cette étude peuvent être utilisés pour les applications suivantes:

En participant avec succès à ces tests, les laboratoires accrédités peuvent démontrer de manière sûre leur compétence; en même temps, ils satisfont la partie de la spécification ISO/IEC 17025 qui demande „La participation à un programme approprié de comparaisons interlaboratoires“.

Quant aux laboratoires non accrédités ils ont la possibilité, en participant aux tests, de comparer pour la première fois leurs performances en matière d'essais avec les performances des laboratoires européens les plus renommés et ce sur la base de critères sûrs.

De plus, les résultats des tests de qualification pourront être utilisés pour le perfectionnement et la standardisation des méthodes d'essai mécaniques, un exemple de cette application étant la détermination de l'incertitude de mesure dans les essais de dureté.

*Also lautet ein Beschluss:
Daß der Mensch was lernen muß. -
- Nicht allein das ABC
Bringt den Menschen in die Höh';
Nicht allein im Schreiben, Lesen
Übt sich ein vernünftig Wesen;
Nicht allein in Rechnungssachen
Soll der Mensch sich Mühe machen;
Sondern auch der Weisheit Lehren
Muß man mit Vergnügen hören.*
Wilhelm Busch

*Wenn uns im Übrigen eine Reihe gigantischer Irrtümer unterlaufen wird, werden wir
uns in bester Gesellschaft befinden.*
Stanislaw Lem

Danksagung

Ich möchte mich hier bei allen bedanken, die zum Werden dieser Arbeit beigetragen haben.

Mein besonderer Dank gilt Frau Professor Dr.-Ing. F. Deuerler für die hervorragende Betreuung, die sich in ihrer kompetenten fachlichen Unterstützung bei der Durchführung der Arbeit und der dabei bewiesenen Geduld und Nervenstärke äußerte.

Prof. Dr.-Ing. Habil. P. Winzer, Prof. Dr.-Ing. Dr.-Ing. e.h. B.-J. Vorath und Prof. Dr.-Ing. A. Meyna danke ich für die Erstellung ihrer Gutachten und ihre Mitwirkung im Promotionsausschuss.

Prof. Dr.-Ing. H. Frenz, FH Gelsenkirchen, Standort Recklinghausen verdanke ich die Anregung zum Thema dieser Dissertation. Das Gelingen dieser Arbeit beruht zu einem wesentlichen Teil auf seiner Unterstützung und den vielen mit ihm geführten fachlichen Diskussionen.

Allen ungenannten Helfern des Instituts für Eignungsprüfung, Herten, der FH Gelsenkirchen, Abt. Recklinghausen und der Bergischen Universität Wuppertal danke ich für ihren Einsatz und ihre Hilfsbereitschaft.

Herrn Leit. Baudir. i. R. Dipl.-Ing. S. Langer danke ich für die Übersetzung der Zusammenfassung ins Französische.

Ohne die Unterstützung und Hilfe meiner Mutter und meines Bruders wäre diese Arbeit niemals fertig gestellt worden.

Inhaltsverzeichnis

1	Einführung	1
2	Problemstellung und Zielsetzung	5
3	Eignungsprüfungen: Stand der Technik	7
3.1	Ziele von Eignungsprüfungen	8
3.2	Bedeutung von Eignungsprüfungen für die Akkreditierung von Prüflaboratorien	8
3.3	Planung und Durchführung von Eignungsprüfungen	12
3.3.1	Definitionsphase	12
3.3.2	Entwicklungsphase	19
4	Statistische Verfahren zur Auswertung von Eignungsprüfungen	29
4.1	Homogenitätsprüfung	29
4.1.1	Einfaktorielle Varianzanalyse	31
4.1.2	Alternative Vorgehensweise zur Homogenitätsprüfung	37
4.1.3	Stabilitätsprüfung	39
4.2	Bestimmung des Stoffwertes und seiner Messunsicherheit	40
4.3	Statistische Auswertung von Ringversuchen	45
4.3.1	Statistisches Grundmodell nach DIN ISO 5725-1	45
4.3.2	Datenauswertung nach DIN ISO 5725-2	46
4.3.2.1	Grafische Vereinbarungsprüfung	47
4.3.2.2	Numerische Ausreißerprüfung	48
4.3.2.3	Berechnung des Gesamtmittelwertes und der Varianzen	50
4.3.3	Robuste Datenauswertung nach ISO/DIS 13528	50
4.4	Statistische Leistungsbewertung von Eignungsprüfungen	53
4.4.1	Berechnung der Standardabweichung $\hat{\sigma}$ zur Beurteilung der Leistung	53
4.4.2	Berechnung der Leistungsstatistik	56
4.4.3	Grafische Methoden zur Darstellung der Leistungsstatistik für mehrere Messungen	61
5	Auswertung und Diskussion der durchgeführten Eignungsprüfungen	67
5.1	Eignungsprüfung Zugversuch	69
5.2	Eignungsprüfungen mit Härteprüfverfahren	77
5.2.1	Eignungsprüfung Härteprüfung nach Vickers	78
5.2.1.1	Untersuchungsprogramm	80
5.2.1.2	Programm der Eignungsprüfung	81
5.2.1.3	Materialauswahl und Homogenitätsprüfung	82

5.2.1.4	Abweichung des Labormittelwertes vom Stoffwert	83
5.2.2	Eignungsprüfung Härteprüfung nach Brinell	88
5.2.2.1	Programm der Eignungsprüfung	90
5.2.2.2	Bestimmung der Messunsicherheit des Prüfverfahrens	91
5.3	Eignungsprüfung Metallographie	95
5.3.1	Untersuchungsprogramm	100
5.3.2	Programm der Eignungsprüfung	101
5.3.3	Materialauswahl und Homogenitätsprüfung	102
5.3.4	Abweichung des Laborwertes vom Stoffwert	105
5.4	Eignungsprüfung Kerbschlagbiegeversuch nach Charpy	116
5.4.1	Untersuchungsprogramm	119
5.4.2	Programm der Eignungsprüfung	120
5.4.3	Materialauswahl und Homogenitätsprüfung	120
5.4.4	Abweichung des Labormittelwertes vom Stoffwert	123
5.4.5	Statistische Beurteilung der Ergebnisse: Wiederholbarkeit	129
5.4.6	Genauigkeit der Probenfertigung	130
5.5	Eignungsprüfung Emissionsspektroskopie	132
5.5.1	Untersuchungsprogramm	133
5.5.2	Programm der Eignungsprüfung	133
5.5.3	Materialauswahl und Homogenitätsprüfung	134
5.5.4	Abweichung des Laborwertes vom Stoffwert	135
6	Zusammenfassung und Ausblick	142
7	Literaturverzeichnis	145
8	Definitionen und Erläuterungen verwendeter Begriffe	153
9	Aufstellung der verwendeten Abkürzungen	156
10	Verzeichnis der verwendeten Formelzeichen	157

1 Einführung

Heutzutage sind mehr Produkte und Dienstleistungen auf dem Markt erhältlich als jemals zuvor. Daher ist der Bedarf an Verbraucherschutz so groß wie nie. Dieser Verbraucherschutz wird durch die Zertifizierung, Abnahme und Prüfung von Produkten und Dienstleistungen gewährleistet. Eine zweite Möglichkeit ist das Erbringen dieser Leistungen unter Anwendung eines zertifizierten Qualitätsmanagementsystems [1]. Der Verbraucher kann jedoch nicht überprüfen, ob diese in seinem Interesse angewandten Maßnahmen richtig durchgeführt werden.

Akkreditierung ist ein Prozess, in dem eine Organisation, die eine oder mehrere der oben genannten Aufgaben wahrnimmt, ihre Kompetenz einem Dritten (der Akkreditierungsstelle) gegenüber nachweist. Der Nachweis wird durch die sachgerechte Anwendung nationaler und internationaler Normen oder Verfahrensregeln erbracht. Darüber hinaus führt die Akkreditierungsstelle eine Beurteilung dieser Organisation vor Ort durch [2], z. B. durch eine Laborbegehung und/oder ein Audit. Ziel der Akkreditierung ist die internationale Anerkennung und Übernahme von Prüfberichten und Zertifikaten, die von den akkreditierten Stellen erstellt wurden. Die Akkreditierung führt weltweit zum Abbau technischer Handelshemmnisse und ist somit eine wichtige Voraussetzung für den freien Waren- und Dienstleistungsverkehr geworden [3]. Akkreditierungen werden von nationalen Akkreditierungsstellen wie z. B. der Deutsches Akkreditierungssystem Prüfwesen GmbH (DAP GmbH) durchgeführt. Die DAP GmbH hat zur Zeit ca. 730 Prüflaboratorien aus den unterschiedlichsten Arbeitsgebieten (Textilprüfung, Abfallwirtschaft, Lebensmittelüberwachung, Baustoff- und Betonprüfung, mechanisch-technologische Werkstoffprüfung und andere mehr) akkreditiert [4]. Die Laborakkreditierung stellt für den Laborkunden einen generellen Nachweis dar: das Prüflaboratorium besitzt die technische Kompetenz und Zuverlässigkeit, ein Prüfverfahren, für das es akkreditiert ist, richtig anzuwenden. Der Kunde kann die Prüfergebnisse dieses Prüflaboratoriums bedenkenlos verwenden, z. B. um die Güte von Produkten oder Dienstleistungen gegenüber staatlichen Stellen in ganz Europa nachzuweisen [5]. Grundlage der Laborakkreditierungen war in Europa von 1990 bis 2000 die DIN EN 45001 [6], im April 2000 wurde sie von der DIN EN ISO/IEC 17025 [7] „Allgemeine Anforderungen an die Kompetenz von Prüf- und Kalibrierlaboratorien“ abgelöst.

Geht man aus von der fehlerfreien Umsetzung der Anforderungen der DIN EN ISO/IEC 17025 [7] und einer sorgfältigen Überprüfung aller Kriterien bei der Labor-

begutachtung, so sollten die Prüfergebnisse dieses Labors immer innerhalb bestimmter vorgegebener Grenzen liegen. Dies ist jedoch nicht immer der Fall, da das Akkreditierungsverfahren ein summarischer Vorgang ist, der nicht alle in der Laborpraxis auftretenden Probleme abdecken kann. Die Begehung eines zu akkreditierenden Laboratoriums ist eine Stichprobe, deren Schwerpunkt auf der Überprüfung vorhandener Dokumente wie Verfahrens-, Arbeits- und Prüfanweisungen, Qualifikationsnachweisen und Kalibriernachweisen liegt. Ein weiteres Problem, das vor allem die chemische Analytik betrifft, ist die Anwendung von Prüfverfahren, die bislang nicht ausreichend validiert sind [8].

Die DIN EN ISO/IEC 17025 fordert daher „... die Teilnahme an einem geeigneten Programm für Vergleichsmessungen zwischen Laboratorien, soweit wie möglich,...“ [7].

Vergleichsmessungen (engl. interlaboratory comparisons) und Ringversuche (engl. round robin tests) werden zur Methodvalidierung, zur Charakterisierung von Referenzmaterialien und zur Überprüfung der Maßnahmen zur Qualitätssicherung in Laboratorien durchgeführt. Sie sind zum Beispiel bei der Entstehung und Einführung einer neuen Prüfnorm gängige Praxis [8, 9].

Eignungsprüfungen (engl. proficiency tests) dienen heutzutage zum Nachweis der Beherrschung eines Prüfverfahrens und zur Beurteilung der Kompetenz eines Prüflaboratoriums für die Anwendung eines bestimmten Prüfverfahrens [10, 11]. Bislang wurden sie überwiegend zur Qualifikation von Prüflaboratorien im gesetzlich geregelten Bereich herangezogen. Im Jahr 1993 betrug die Zahl der regelmäßig und öffentlich angebotenen Eignungsprüfungen weniger als 100 – die Fachgebiete Medizin und Metrologie sind in dieser Zahl nicht enthalten [12]. Betroffen sind Arbeitsfelder wie physikalisches Messwesen, Arbeitssicherheit, Umwelttechnik und Prüfungen, welche die Gesundheit der Bürger betreffen, zum Beispiel die Lebensmittelüberwachung, der Strahlenschutz, die Wasser-, Luft- und Abwasseranalytik und die klinisch-chemische Analytik [13 bis 17]. In der Bundesrepublik Deutschland ist der Umfang dieser Eignungsprüfungen von Bundesland zu Bundesland stark unterschiedlich [18].

Mögliche Formen der Eignungsprüfung sind [19]:

- Teilnahme an einer Vergleichsmessung / Ringversuch;
- Prüfung von Proben, die von Dritten (Behörden, Eignungsprüfungsveranstaltern) speziell für eine Eignungsprüfung herausgegeben worden sind;
- Messungen an geeigneten Proben, die der Begutachter des Akkreditierers im Rahmen einer Laborbegehung zur Verfügung stellt.

Mechanisch-technologische Prüfverfahren dienen zur Bestimmung der Festigkeitseigenschaften, der Oberflächeneigenschaften (Härte), der für die Formgebung maßgebenden Eigen-

schaften (Formänderungsvermögen) und der für den Zusammenbau maßgebenden Eigenschaften (Schweißbarkeit) [20]. Auf diesem Prüfgebiet sind staatliche Prüfeinrichtungen wie die Landes-Materialprüfämter (LMPA), Forschungseinrichtungen wie die Bundesanstalt für Materialforschung und –prüfung (BAM) in Berlin, Abnahmeorganisationen wie die Technischen Überwachungsvereine (TÜV), Herstellerprüflaboratorien und freie, kommerziell ausgerichtete Laboratorien tätig. Im Rahmen dieser Arbeit werden alle diese Anbieter von Prüfleistungen mit „Prüflaboratorium“ bezeichnet.

In Deutschland werden seit 1990 in Prüflaboratorien für mechanisch-technologische Prüfverfahren (Werkstoffprüfung) Begehungen durchgeführt und Akkreditierungen ausgesprochen [21]. Die Schwerpunkte dieser Begutachtungen im Rahmen des Akkreditierungsverfahrens liegen bislang auf der Überprüfung der fachlichen Kompetenz und des Qualitätsmanagementsystems hinsichtlich Normkonformität und Funktionsweise. Wie erfolgreich ein Labor eine Prüfnorm, für die es akkreditiert wurde, jedoch in die Prüfpraxis umsetzt, lässt sich aufgrund dieser Begutachtung allein nicht beurteilen [22]. Daher ist eine objektive Bewertung der am Markt erhältlichen Prüfergebnisse – und damit letztlich die Kompetenzbestätigung für ein Prüflabor – mit den heutigen Mitteln außerordentlich schwierig. Das stichprobenweise Überprüfen der Anwendung von Prüfverfahren während der Vor-Ort-Begutachtung durch Fachexperten kann naturgemäß nur einen Teil der Prüfungen abdecken.

Das zuständige Fachgremium, das Sektorkomitee „Mechanisch-technologische Prüfungen“ (SK-MTP) der Deutschen Akkreditierungssystem Prüfwesen (DAP) GmbH, Berlin hat 1996 beschlossen, dass Eignungsprüfungen als wesentliche Ergänzung zur Bewertung der Laborleistung durchgeführt werden sollen [21]. Da bis 1996 in Deutschland keine öffentlich zugänglichen Eignungsprüfungen für mechanisch-technologische Prüflaboratorien angeboten wurden, führte das SK-MTP gemeinsam mit der Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin die als Vorversuch konzipierte Eignungsprüfung „Zugversuch nach DIN EN 10002-1“ durch [21, 23]. An dieser Eignungsprüfung beteiligten sich insgesamt 91 Laboratorien aus 24 europäischen Ländern, darunter 50 Prüflaboratorien aus Deutschland.

Aufgrund des großen Interesses der Prüflaboratorien an dieser Eignungsprüfung und dem häufig geäußerten Wunsch der Teilnehmer nach weiteren Eignungsprüfungen für mechanisch-technologische Prüflaboratorien, beschloss die Deutsche Akkreditierungssystem Prüfwesen GmbH (DAP) weitere Eignungsprüfungen durchzuführen [24]:

Im Jahre 1999 die Eignungsprüfung Härteprüfung nach Vickers [24, 25], im Jahr 2000 die Eignungsprüfung Metallographie [26] und im Jahr 2001 die Eignungsprüfungen Emissionsspektrometrie [27], Härteprüfung nach Brinell [28] und Kerbschlagbiegeversuch [29].

Entwicklung, Gestaltung, Durchführung und Auswertung dieser Eignungsprüfungen stellen den wissenschaftlichen Gegenstand dieser Untersuchung dar.

2 Problemstellung und Zielsetzung

Technische Prüfungen gleich welcher Art gehören zu den wichtigsten Bausteinen der Ingenieur-tätigkeit. Sie sollen die technischen Nachweise erbringen, dass die erwartete und geforderte Gebrauchstauglichkeit und die vereinbarten und vorausgesetzten Leistungen auch erreicht werden [30]. Prüfungen werden an Rohstoffen, Vormaterialien und Endprodukten durchgeführt und dienen z. B. der Prozessüberwachung, als Wareneingang- oder -eingangsprüfung, als Konformitätsprüfung zum Nachweis des Erfüllens grundlegender Sicherheitsanforderungen und zu vielem anderen mehr. Korrekte Prüfergebnisse sind ein Eckpfeiler eines reibungslosen Handels und einer funktionierenden Gesellschaft. Basierend auf diesen Prüfergebnissen werden folgenschwere Entscheidungen gefällt, die weitreichende wirtschaftliche Auswirkungen und / oder Einfluss auf die öffentliche Sicherheit haben. Prüfergebnisse werden ebenfalls genutzt, um sachverständige Beurteilungen zu erstellen [31]. Sicherheitsfachkräfte in Unternehmen, die z. B. bezüglich Emissionsschutz, Gewässerschutz usw. den Aufsichts- und Genehmigungsbehörden gegenüber nachweispflichtig sind, benötigen zwingend Prüflaboratorien, die eine ausreichende Kompetenz zur Durchführung notwendiger Prüfungen besitzen. Geeignete Prüfungen dienen der Haftungsprophylaxe; korrekte, in Prüfbescheinigungen dokumentierte Prüfergebnisse sind geeignet, Produkthaftungsrisiken zu mindern, da so die Rückverfolgbarkeit eines Produktes ermöglicht wird [32].

Mit Hilfe der mechanisch-technologischen Werkstoffprüfung werden Materialkennwerte gewonnen, die für die Festigkeitsprüfung und -forschung, den Nachweis von Güteeigenschaften, die Erstellung von Prüfvorschriften, die Untersuchung von technischen Schäden und in der Werkstoffforschung verwendet werden [20]. Heutzutage wird aus Gründen der Wirtschaftlichkeit eine bessere Ausnutzung des Materials durch Erhöhung der zulässigen Streckgrenze angestrebt [33]. Dies erfordert parallel von Anbietern von Prüfleistungen, wie z. B. den Werkstoffprüfstellen der Hersteller, freien Werkstoffprüflaboratorien oder staatlichen Stellen wie z. B. der Bundesanstalt für Materialforschung und -prüfung (BAM) in Berlin, eine hohe Kompetenz in der mechanisch-technologischen Werkstoffprüfung. Europaweit besteht daher von Seiten staatlicher Stellen (Aufsichtsbehörden), den Akkreditierungsstellen, den Prüflaborkunden und den Prüflaborbetreibern ein großer Bedarf an Eignungsprüfungen auf dem Gebiet der mechanisch-technologischen Werkstoffprüfung [21, 24]. Dabei sind jedoch die unterschiedlichen „Kundenanforderungen“ zu beachten:

Staatliche Stellen, Akkreditierungsstellen und Laborkunden benötigen einen objektiv belastbaren und aussagefähigen Kompetenznachweis, der mit einem Minimum an technischem und zeitlichen Aufwand erbracht werden kann.

Die Forderung der Prüflaborbetreiber besteht aus einer nachvollziehbaren, fairen und belastbaren Beurteilung ihrer Prüfleistung bei gleichzeitiger Kosten- und Aufwandsoptimierung.

Umfangreiche und regelmäßige Eignungsprüfungen wurden bislang nur in der Analytik (Chemie, Medizin, Pharmazie) angeboten und wissenschaftlich untersucht. Die daraus resultierenden Verfahren zur Auswertung von Eignungsprüfungen sind jedoch auf die Verfahren der mechanisch-technologischen Werkstoffprüfung aus folgenden Gründen nicht übertragbar:

- Prüfgegenstände sind hier immer feste Gegenstände;
- mechanisch-technologische Prüfverfahren sind immer zerstörende Prüfverfahren;
- man erhält nur wenige zertifizierte Referenzmaterialien, die zudem sehr teuer sind;
- die Präzision der Prüfverfahren ist nicht bekannt.

Im Rahmen dieser Arbeit wurden Eignungsprüfungen für verschiedene mechanisch-technologische Werkstoffprüfverfahren konzipiert, europaweit angeboten und durchgeführt.

Anhand dieser Eignungsprüfungen wird erstmalig wissenschaftlich untersucht, in wie weit die bekannten Methoden zur Durchführung von Eignungsprüfungen und ihrer statistischen Auswertung geeignet sind, wenn sie auf mechanisch-technologische Prüfverfahren angewendet werden. Ziel ist es, Vorgehensweisen und Auswertemethoden zu erarbeiten, die zu prüftechnisch angemessenen und wissenschaftlich begründeten Eignungsprüfungen für mechanisch-technologische Prüflaboratorien führen. Diese Eignungsprüfungen sollen eine wissenschaftlich abgesicherte Methode zur Beurteilung der Laborkompetenz sein. Darüber hinaus wird eingehend untersucht, welcher Zusatznutzen aus den in Eignungsprüfungen gewonnenen Daten für die Weiterentwicklung der Prüfverfahren gezogen werden kann.

3 Eignungsprüfungen: Stand der Technik

Eignungsprüfung ist die Bestimmung der Leistungsfähigkeit eines Prüflaboratoriums mittels Vergleichsprüfungen [10, 11]. Die gebräuchlichen Arten von Eignungsprüfungen sind [11]:

- *Messvergleichsprogramme* werden mit einem zu messenden Prüfgegenstand durchgeführt, welcher der Reihe nach von einem teilnehmenden Labor zum nächsten übergeben wird. Die einzelnen Ermittlungsergebnisse werden mit einem vorgegebenen Bezugswert (z. B. einem zertifizierten Referenzwert) verglichen und beurteilt.
- *Ringversuchsprogramme* werden mit zufällig ausgewählten Unterproben aus einer Stoffquelle durchgeführt, die gleichzeitig an die teilnehmenden Laboratorien ausgegeben werden. Auf eine ausreichende Homogenität des ausgegebenen Probenmaterials ist besonders zu achten.
- *Prüfprogramme mit aufgeteilten Proben* werden häufig im Kunden-Lieferanten-Verhältnis bei geschäftlichen Transaktionen durchgeführt. Das Probematerial wird in zwei oder mehr Teile geteilt, wobei jedes teilnehmende Prüflaboratorium ein Teil von jeder Probe prüft.
- *Qualitative Programme* sind Eignungsprüfungen, bei denen Laboratorien aufgefordert sind, einen Bestandteil eines Prüfgegenstandes zu identifizieren.
- *Programme mit bekannten Werten* dienen zur Beurteilung der Fähigkeit eines Laboratoriums, einen Prüfgegenstand zu prüfen und Zahlenwerte für den Stoffwert (siehe Kap. 3.3 Begriffe) für den Vergleich zur Verfügung zu stellen.
- *Programme mit Teilprozessen* sind eine Sonderform, bei der die Fähigkeit eines Laboratoriums überprüft wird, Teilschritte eines Messverfahrens, wie z. B. die Probenahme, sachgerecht auszuführen.

3.1 Ziele von Eignungsprüfungen

Das Hauptziel von Eignungsprüfungen besteht darin, die teilnehmenden Prüflaboratorien mit Informationen und Hinweisen zu versorgen, damit diese ihre Prüfleistungen überwachen, beurteilen und die Qualität gegebenenfalls verbessern können [9]. Die Ergebnisse aus Eignungsprüfungen und daraus abgeleitete Schlussfolgerungen bilden die Basis zum Erkennen systematischer, laborspezifischer Fehlerquellen in der praktischen Laborarbeit. Eine praxisorientierte Qualitätsverbesserung wird erzielt durch Korrekturmaßnahmen, die aus den Ergebnissen der Eignungsprüfung abgeleitet werden. Sie werden auf Basis des Labor-Qualitätsmanagementsystems [7] umgesetzt. Die Arbeitsweise des Laboratoriums wird dahingehend korrigiert, dass bei wiederholter Teilnahme an Eignungsprüfungen eine signifikante Verbesserung der Laborleistung erreicht wird. Dabei sind auch wirtschaftliche Gesichtspunkte angemessen zu berücksichtigen [34].

Es gibt noch eine ganze Reihe von Sekundärzielen bzw. Nutzenanwendungen für Eignungsprüfungen [9, 34]:

- Vergleich von Laborergebnissen, die mit unterschiedlichen analytischen Verfahren gewonnen werden;
- Beurteilung neuer Prüfverfahren im Vergleich zu etablierten Prüfverfahren;
- Hilfsmittel für Akkreditierungsstellen und Regulierungsbehörden zur Beurteilung eines Prüflaboratoriums;
- Nachweis der Kompetenz eines Laboratoriums gegenüber einem Kunden;
- Zuweisung von Werten zu zertifizierten Referenzmaterialien;
- Bestimmung der Genauigkeit einer Prüfmethode;
- Hilfestellung bei der Feststellung der Gründe für Abweichungen zwischen verschiedenen Laboratorien.

Grundsätzlich ist es auch möglich, Eignungsprüfungsprogramme innerhalb eines Laboratoriums durchzuführen. Dabei wird festgestellt, ob die Prüfergebnisse von zwei oder mehreren Mitarbeitern systematische Unterschiede in Wiederholgenauigkeit und Präzision zeigen [34]. Sollte dies der Fall sein, erfolgt eine Nachschulung der betroffenen Mitarbeiter.

3.2 Bedeutung von Eignungsprüfungen für die Akkreditierung von Prüflaboratorien

Bei der Akkreditierung nach DIN EN ISO/IEC 17025 [7] beurteilen die Akkreditierungsstelle und der Fachbegutachter das Qualitätsmanagementsystem, die technische Ausrüstung (bzw. deren Zustand) eines Laboratoriums und die Befähigung des dort beschäftigten Personals.

Alle drei Bestandteile müssen einen ausreichend hohen Stand haben, damit die Qualität der akkreditierten Laboratorien gewährleistet ist [17, 34, 36]. Grundlage einer Akkreditierung ist immer eine Norm, eine allgemein anerkannte Verfahrensweise oder eine hausinterne Verfahrensvorschrift. Die Akkreditierungsstelle und der Fachbegutachter müssen eine unabhängige Beurteilung der Laboratoriumskompetenz sicherstellen.

Eignungsprüfungen sind ein sehr effizientes Hilfsmittel im Rahmen der Akkreditierung von Prüf- bzw. Eichlaboratorien [12, 35, 36]. Bild 3.2.1 zeigt schematisch die Elemente des Akkreditierungsprozesses.

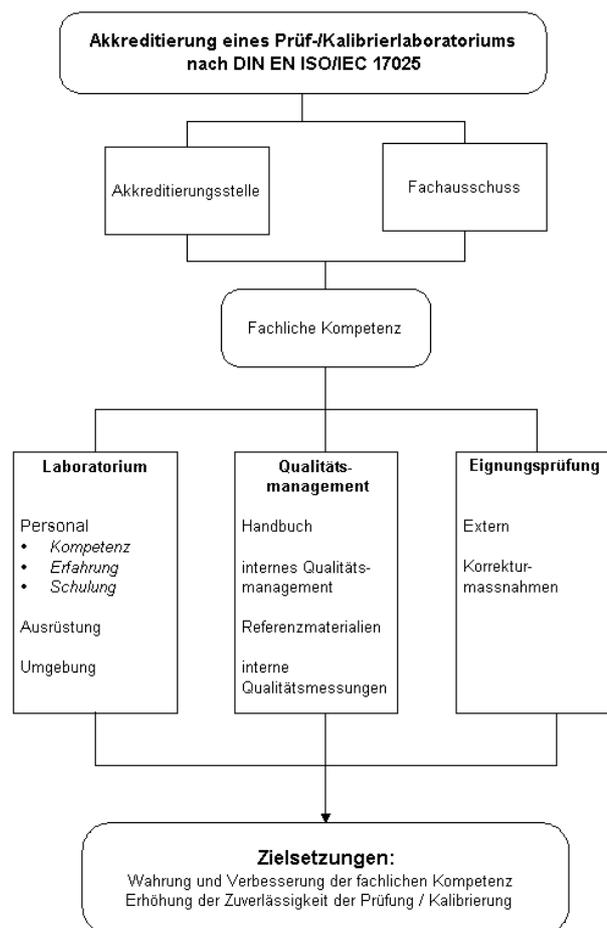


Bild 3.2.1: Schematische Darstellung der Elemente des Akkreditierungsprozesses nach [34]

In der DIN EN ISO/IEC 17025 [7] wird die Teilnahme akkreditierter Laboratorien und Bewerberlaboratorien an Eignungsprüfungen oder anderen Vergleichsversuchen zwischen Laboratorien empfohlen. Eine Akkreditierungsstelle kann, soweit dies zweckmäßig erscheint, von Prüflaboratorien die Teilnahme an Eignungsprüfungen verlangen [7]. Wenn die Ergebnisse aus der verlangten Eignungsprüfung nicht ausreichend sind, so ist die Akkreditierung bzw. der Fortbestand der Akkreditierung fraglich. Eine Akkreditierung kann jedoch nicht allein auf

Grund der Ergebnisse einer Eignungsprüfung gewährt oder verweigert werden [8]. Die Teilnahme an Eignungsprüfungen sollte als Teil der Überwachung betrachtet werden [36].

Eignungsprüfungsprogramme können entweder von der Akkreditierungsstelle selbst oder von einer beliebigen anderen kompetenten Stelle organisiert werden. Die Anforderungen an den Anbieter von Eignungsprüfungen sind im ISO/IEC Guide 43-1 [10] und in der Richtlinie ILAC-G13:2000 [37] zusammengestellt.

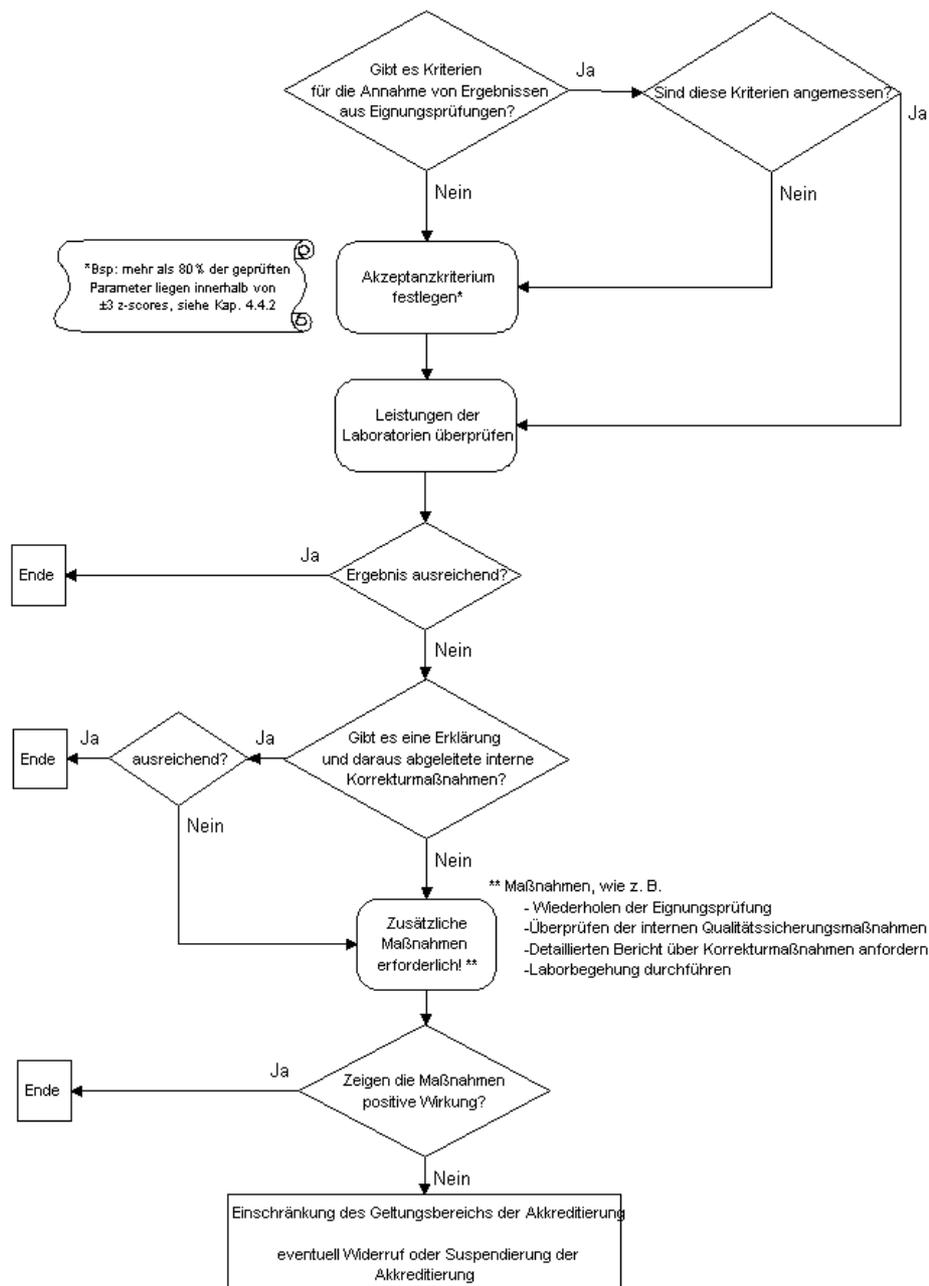


Bild 3.2.2: Vorgehensweise für die Beurteilung von Laboratorien durch eine Akkreditierungsstelle unter Verwendung von Eignungsprüfungen [35]

Bild 3.2.2 zeigt als Ablaufdiagramm die Vorgehensweise bei der Bewertung von Laboratorien durch eine Akkreditierungsstelle unter Verwendung von Eignungsprüfungen:

Die Akkreditierungsstelle überprüft, ob in der Eignungsprüfung ein belastbares Akzeptanzkriterium zur Bewertung der Laborleistung für das entsprechende Prüfverfahren festgelegt wurde. Dies kann z. B. ein zulässiger z-Score (vergleiche Kapitel 4.4.2) sein. Ist das Kriterium angemessen und belastbar, wird die Laborleistung damit bewertet. Ist die Leistung des Laboratoriums ausreichend, so akzeptiert die Akkreditierungsstelle diesen Nachweis und setzt das Akkreditierungsverfahren fort. Ist das Ergebnis unzureichend, wird vom Laboratorium eine Erklärung für die Ursache des schlechten Abschneidens verlangt und die Vorlage von eingeleiteten Korrekturmaßnahmen. Sind diese nicht ausreichend, stellt sich die Frage, ob zusätzliche Maßnahmen erforderlich sind. Dies kann z. B. die Wiederholung der Eignungsprüfung sein oder die Durchführung einer zusätzlichen Laborbegehung.

Wenn auch diese zusätzlichen Maßnahmen wirkungslos bleiben, kann die Akkreditierungsstelle eine Akkreditierung für das Prüfverfahren verweigern, zurückziehen oder suspendieren.

In Deutschland gibt es für Laboratorien, die im nicht geregelten Bereich tätig sind, keine gesetzlichen Anforderungen für die Teilnahme an Eignungsprüfungen. Im geregelten Bereich ergeben sich Teilnahmeanforderungen aus Verordnungen der Bundesländer:

Das Land Baden-Württemberg vergibt eine große Anzahl von Analyseaufträgen an staatliche und private Laboratorien auf dem Gebiet der Umweltanalytik [38]. Für Abwasseruntersuchungen nach Wassergesetz werden Laboratorien gemäß der „Verordnung des Ministeriums für Umwelt und Verkehr über sachverständige Stellen in der Wasserwirtschaft“ von der Landesanstalt für Umweltschutz zugelassen. So gelten in Baden-Württemberg die Anforderungen für „... Prüflaboratorien, die Proben entnehmen und physikalisch, physikalisch-chemisch oder chemisch untersuchen, unabhängig von ihrer Rechtsform...“ aus der Verordnung des Ministeriums für Umwelt und Verkehr über sachverständige Stellen in der Wasserwirtschaft [39]. Die Fachkompetenz ist durch eine Laborakkreditierung nach DIN EN ISO/IEC 17025 [7] nachzuweisen, außerdem muss das Prüflaboratorium „... sich nach den Vorgaben der Anerkennungsbehörde regelmäßig und auf eigene Kosten an Ringversuchen [zu] beteiligen und sich Laborbegutachtungen [zu] unterziehen.“ [39].

Unter der Bezeichnung AQS Baden-Württemberg sind alle Aktivitäten der Ministerien für Umwelt und Verkehr sowie für Ernährung und Ländlichen Raum in Baden-Württemberg zur Qualitätsüberwachung chemischer Analysenlaboratorien zusammengefasst [40]. Das Ministerium für Umwelt und Verkehr und das Ministerium für Ernährung und Ländlichen Raum Baden-Württemberg haben daher das Institut für Siedlungswasserbau, Wassergüte- und Abfallwirtschaft (ISWA) der Universität Stuttgart damit beauftragt, externe Qualitätssicherungsmaßnahmen für diese Bereiche durchzuführen. Die Hauptaufgabe besteht darin, Ring-

versuche auf dem Gebiet der Wasseranalytik [41, 42] anzubieten, die den speziellen Anforderungen der Analytischen Qualitätssicherung gerecht werden [38].

3.3 Planung und Durchführung von Eignungsprüfungen

3.3.1 Definitionsphase

Eine Eignungsprüfung wird entweder direkt von einer Akkreditierungsstelle oder im Auftrag durch einen besonders qualifizierten Eignungsprüfungsveranstalter durchgeführt [37].

Die Durchführung einer Eignungsprüfung nimmt sehr viel Zeit in Anspruch und bedeutet einen nicht unerheblichen Arbeitsaufwand. Daher ist der Planung der Eignungsprüfung besondere Sorgfalt zu widmen [34, 49, 50]. In Bild 3.3.1 sind alle wesentlichen Schritte dargestellt, die während der Planungsphase einer Eignungsprüfung durchlaufen werden.

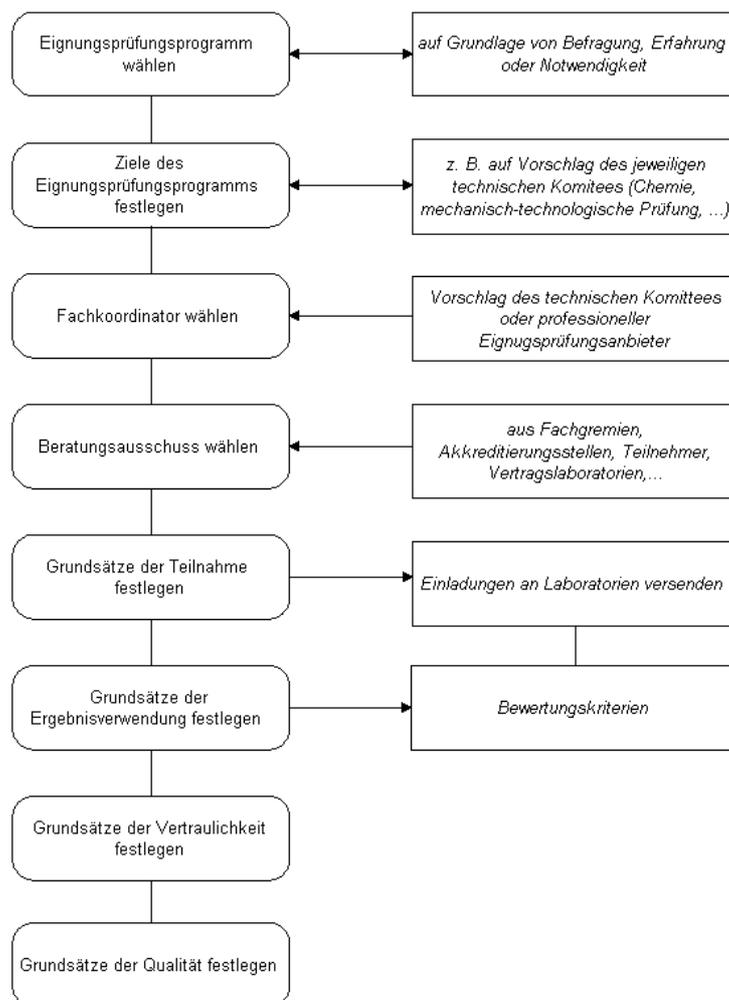


Bild 3.3.1: Planung einer Eignungsprüfung (schematisch) nach [34]

Die Prüfleistung eines Laboratoriums bei einer Eignungsprüfung geht im Rahmen der Akkreditierung in die Gesamtbeurteilung der Leistungsfähigkeit dieses Labors ein. Eine Eig-

nungsprüfung muss auf objektiven und belastbaren Methoden zur Bestimmung der Laborleistung beruhen, da eine Entscheidung über die Gewährung oder Beibehaltung einer Akkreditierung eventuell von der Leistung eines Laboratoriums in einer bestimmten Eignungsprüfung abhängt. Es wird empfohlen, nur an solchen Eignungsprüfungsprogrammen teilzunehmen, welche mindestens die Forderungen des ISO/IEC Guide 43-1 [10] erfüllen.

Bei der Auswahl einer geeigneten Eignungsprüfung sollten folgende Punkte berücksichtigt werden [34, 51, 52]:

- Prüfungen, Messungen oder Kalibrierungen sollten genau der Art von Prüfungen, Messungen oder Kalibrierungen entsprechen, die von den Teilnehmern angewendet werden;
- die Häufigkeit mit der die Eignungsprüfung stattfindet;
- bei Durchführung der Eignungsprüfung durch einen speziellen Anbieter soll die Akkreditierungsstelle Zugang zu den Ergebnissen der akkreditierten Teilnehmer haben sowie über die Einzelheiten der Programmgestaltung und Berichtserstellung informiert werden;
- Eignung der organisatorischen Logistik für die Eignungsprüfung wie Zeitplan, Betrachtungen zur Stabilität des Prüfgegenstandes, Verteilungspläne, usw.;
- Verfügbarkeit und Brauchbarkeit von Bewertungskriterien;
- Kosten der Eignungsprüfung: diese setzen sich zusammen aus Kosten für das Material, die Materialcharakterisierung, Homogenitätsprüfung, Transport, Verwaltung, Kosten der Akkreditierungsstelle, Kosten des Organisationsausschusses;
- Vorgehensweise zur Sicherstellung der Vertraulichkeit hinsichtlich der Teilnehmeridentität und der Ergebnisse;
- Pünktlichkeit bei der Berichterstattung über die Programmergebnisse;
- Vertrauen in die Eignung der verwendeten Prüfmaterialien, insbesondere deren Homogenität, Stabilität und ggf. Rückführbarkeit auf nationale oder internationale Normale.

Die Akkreditierungsstelle formuliert die Ziele und Kriterien für die Bewertung der Eignungsprüfungsergebnisse. Mögliche Ziele sind:

- Überprüfung der Prüfqualität eines akkreditierten Laboratoriums oder eines Kandidaten auf einem speziellen Prüfgebiet. Es werden diejenigen Laboratorien ermittelt, die gezielte Korrekturmaßnahmen ergreifen müssen, um die Qualität ihrer Arbeit zu verbessern.

- Entscheidung über die Gewährung einer Akkreditierung. Das Ergebnis der Eignungsprüfung darf jedoch nur ein Aspekt bei dieser Entscheidung sein [53].
- Im gesetzlich geregelten Bereich können die Registrierung und Anerkennung eines Laboratoriums direkt von der erfolgreichen Teilnahme an bestimmten Eignungsprüfungsprogrammen abhängen [39].

Der Akkreditierungsstelle wird empfohlen, wo es sinnvoll oder notwendig ist, eigene programmspezifische Ziele und Kriterien aufzustellen [35].

Normalerweise ist eine Akkreditierungsstelle nicht in der Lage, eine Eignungsprüfung mit eigenem Personal durchzuführen. Es werden technische Fachkräfte, Statistiker und ein Programmkoordinator benötigt, welche das Programm organisieren, lenken und durchführen. Daher ist die Akkreditierungsstelle auf die Zusammenarbeit mit einem professionellen Anbieter von Eignungsprüfungen angewiesen. Ein Anbieter von Eignungsprüfungen muss nach [37] eine Reihe von Forderungen erfüllen:

- langjährige Erfahrung auf dem jeweiligen Prüfgebiet;
- Kenntnisse von allen wesentlichen Aspekten der Organisation einer Eignungsprüfung;
- Erfahrung in der statistischen Versuchsplanung und der Anwendung statistischer Auswertemethoden; es wird empfohlen, gegebenenfalls einen Fachstatistiker hinzuzuziehen [54];
- Kenntnis aller wichtiger nationaler und internationaler Standards, die sich mit der Organisation, Durchführung und Auswertung von Eignungsprüfungen befassen.

Der Koordinator entwickelt einen Ablaufplan für die Eignungsprüfung und überwacht dessen Umsetzung. Dieser wird dabei durch einen Beratungsausschuss unterstützt. Der Beratungsausschuss besteht aus mindestens zwei Personen, die sich beispielsweise aus Fachgremien, Vertragslaboratorien, Teilnehmern an der Eignungsprüfung, Akkreditierungsstellen oder den Endnutzern der ermittelten Daten rekrutieren. Der Beratungsausschuss soll sich aktiv und dauerhaft an allen wichtigen Entscheidungen zur Durchführung und Organisation der Eignungsprüfung sowie zur Auswertung der Ergebnisse beteiligen. Die Aufgaben des Beratungsausschusses sind im einzelnen [34, 52]:

- Entwicklung und Überarbeitung von Verfahren für Planung, Durchführung, Analyse, Berichterstattung und Effektivität des Programms für die Eignungsprüfung;
- Identifizierung und Bewertung von ähnlichen Eignungsprüfungen, die von anderen Stellen organisiert wurden;
- Bewertung der Ergebnisse der Eignungsprüfung hinsichtlich der Leistungsfähigkeit der teilnehmenden Laboratorien;

- Beratung von Stellen, welche die technische Kompetenz von teilnehmenden Laboratorien beurteilen;
- Beratung von Teilnehmern, die offensichtlich auf Probleme stoßen;
- Schlichtung von Streitfällen zwischen Koordinator und Teilnehmern.

Eignungsprüfungsprogramme sollen möglichst allen Laboratorien offen stehen, die teilnehmen möchten [34]. Deshalb sollen Eignungsprüfungsanbieter die Grundsätze für die Teilnahme schriftlich offen legen. Diese Offenlegung erfolgt europaweit im European Proficiency Testing Information System (EPTIS), einer Datenbank, die zur Zeit (Stand 30.05.02) Informationen über 703 Eignungsprüfungen aus 27 verschiedenen Prüfgebieten und 52 Produkten aus 16 europäischen Staaten enthält [55].

In den Grundsätzen für die Teilnahme sollten folgende Punkte angesprochen werden [10, 34]:

- Ist die Teilnahme (je nach Umfang der Akkreditierung) freiwillig oder wird sie durch die Akkreditierungsstelle vorgeschrieben?
- Mit welcher Häufigkeit findet die Eignungsprüfung statt?
- Welche Akzeptanzkriterien wendet die Akkreditierungsstelle zur Bewertung der Laborleistung an? In vielen Fällen werden diese Akzeptanzkriterien aus den Ergebnissen bestimmt, die man während der Durchführung einer Eignungsprüfung ermittelt. In diesem Fall sollten die Laboratorien über die Prinzipien aufgeklärt werden, nach denen die Akzeptanzkriterien berechnet oder festgelegt werden.
- Ist es erforderlich, an einem Nachfolgeprogramm teilzunehmen, wenn ein Laboratorium in einer bestimmten Eignungsprüfung unzureichende Ergebnisse erzielt?
- Wie wird das Ergebnis bei der Entscheidung über eine Akkreditierung verwendet?
- Einzelheiten zur Wahrung der Vertraulichkeit;
- Teilnehmerkreis.

Die in einer Eignungsprüfung erzielten Ergebnisse können sowohl durch die Akkreditierungsstelle, einen Fachbegutachter im Akkreditierungsprozess und das Laboratorium selbst genutzt werden.

Es gibt jedoch Grenzen, was den Nutzen dieser Ergebnisse für einen Befähigungsnachweis angeht. Eine zufrieden stellende Leistung in einer speziellen Eignungsprüfung kann nur als Kompetenznachweis in diesem speziellen Fall gewertet werden. Ebenso kann ein unzureichendes Ergebnis in einem speziellen Programm als zufällige Abweichung vom normalen Kompetenzgrad eines Laboratoriums angesehen werden. Aus diesem Grund verwenden die Akkreditierungsstellen in ihren Akkreditierungsprozessen Eignungsprüfungen nur als ein Werkzeug zur Überprüfung der Laborkompetenz [34, 35].

Falls Laboratorien unzureichende Ergebnisse einreichen, stehen der Akkreditierungsstelle folgende Möglichkeiten offen:

- Das Laboratorium wird aufgefordert, innerhalb eines vereinbarten Zeitraumes seine Leistung selbst zu überprüfen und zu kommentieren.
- Das Laboratorium wird aufgefordert, sich einer weiteren Eignungsprüfung zu unterziehen. Werden solche Prüfungen durchgeführt, muss das Laboratorium bestätigen, dass es wirksame Korrekturmaßnahmen erarbeitet hat.
- Die Akkreditierungsstelle führt durch einen Fachbegutachter eine Bewertung des Laboratoriums vor Ort durch. Dabei soll die Wirksamkeit von Korrekturmaßnahmen überprüft werden.
- Einschränkung des Geltungsbereichs der Akkreditierung, Suspendierung der Akkreditierung oder, als ultima ratio, ein Widerruf der Akkreditierung.

Welche dieser Möglichkeiten letztlich Anwendung findet, hängt von der Vorgeschichte des Laboratoriums ab, d.h. vom Verlauf der letzten regulären Laborbegehung oder den Ergebnissen, die bislang bei Eignungsprüfungen erzielt wurden.

Die Akkreditierungsstelle ist für die Aufzeichnungen über das Abschneiden der Laboratorien bei Eignungsprüfungen in den Akkreditierungsunterlagen verantwortlich. Diese Unterlagen müssen einem Fachbegutachter bei einer Laborbegehung zur Verfügung gestellt werden. Der Fachbegutachter muss die Ergebnisse geeigneter Eignungsprüfungen bei der Beurteilung der technischen Kompetenz eines Laboratoriums berücksichtigen. Die Ergebnisse der Eignungsprüfungen werden bei der Laborbegehung besprochen: Der Fachbegutachter muss ein positives Ergebnis akzeptieren, wenn die Eignungsprüfung durch die Akkreditierungsstelle oder einen kompetenten Anbieter von Eignungsprüfungen aufgrund definierter Akzeptanzkriterien (vgl. Kapitel 4.3 und 4.4) durchgeführt wurde. Im Falle des Nichtbestehens der Eignungsprüfung für ein bestimmtes Prüfverfahren muss der Fachbegutachter die eingeleiteten Korrekturmaßnahmen sorgfältig überprüfen. Wenn die Wirksamkeit dieser Korrekturmaßnahmen nachgewiesen werden kann, wird die Kompetenz des Laboratoriums für diesen Fall positiv bewertet.

Hat das Laboratorium keine Korrekturmaßnahmen ergriffen, so muss der Fachbegutachter diese fordern und die Akkreditierung gegebenenfalls solange außer Kraft setzen, bis Korrekturmaßnahmen wirksam sind. Möglicherweise wird eine zusätzliche Laborbegehung notwendig. Wenn diese Maßnahmen nicht innerhalb einer akzeptablen Frist durchgeführt werden, sollte der Fachbegutachter für dieses Prüfverfahren keine Akkreditierung empfehlen. In jedem

Fall muss der Fachbegutachter seine Feststellungen und Empfehlungen an die Akkreditierungsstelle berichten.

Hat das Laboratorium an einer selbst organisierten Vergleichsprüfung teilgenommen, prüft der Fachgutachter die vorgelegte Dokumentation und beurteilt die Wirksamkeit dieser Vergleichsprüfung aufgrund seiner Fachkompetenz [12, 19, 34].

Die Ergebnisse, die ein Laboratorium bei einer Eignungsprüfung erzielt, sind für das Laboratorium ein Maß für seine Prüfkompetenz und die Qualität seiner Prüfergebnisse. Nach DIN EN ISO/IEC 17025 [7] muss ein Laboratorium an Eignungsprüfungen teilnehmen, sofern diese verfügbar und geeignet sind. Das Laboratorium muss die wesentlichen Phasen seiner Teilnahme dokumentieren. Diese Dokumentation muss vollständig sein und dem Fachbegutachter bei einer Laborbegehung vorgelegt werden können. Um die von den Laboratorien durchgeführten Eignungsprüfungen identifizieren zu können, empfiehlt es sich, eine Liste mit folgenden Punkten anzulegen:

- Datum der Durchführung der Eignungsprüfung,
- Organisator extern / intern,
- Probe oder Prüfgegenstand,
- Prüfmethode oder –verfahren,
- Kriterien für das Bestehen der Eignungsprüfung,
- Bekanntgabe des eigenen Ergebnisses,
- beabsichtigte Korrekturmaßnahmen, Durchführungsverantwortlicher, Datum, Unterschrift,
- bereits durchgeführte Maßnahmen und Feedback an den Anbieter der Eignungsprüfung.

Die Kommunikation zwischen dem teilnehmenden Laboratorium und dem Anbieter der Eignungsprüfung erfolgt per Schreiben, Newsletter, Bericht, mittels periodischer Treffen oder einer Kombination aus allem. Die Teilnehmer erhalten detaillierte Informationen über die Eignungsprüfung, in der Regel durch ein Schreiben, das die Aufgabenstellung und das Protokoll zur Erfassung der abgefragten bzw. ermittelten Daten enthält.

Der Anbieter einer Eignungsprüfung sollte die Teilnehmer zu positiven und negativen Rückmeldungen ermuntern, um so die Qualität der Eignungsprüfungen verbessern zu können. Teilnehmer, die der Meinung sind, ihre Leistungsbewertung sei fehlerhaft, sollten sich direkt an den Koordinator wenden.

Die Teilnehmer müssen die Unterlagen über ihre Leistung bei den Eignungsprüfungen aufbewahren. Besonders im Fall einer nicht ausreichenden Leistung sollten insbesondere Be-

richte über Korrekturmaßnahmen, die aus den Ergebnissen abgeleitet wurden, sowie Rückmeldungen aufbewahrt werden. Haben sich z. B. Änderungen im Prüfverfahren ergeben, so sind diese bei der nächsten Laborbegehung dem Fachbegutachter vorzuführen. Der Fachbegutachter beurteilt anhand der Aufzeichnungen des Laboratoriums die Leistungsfähigkeit des Laboratoriums und seine Fähigkeit aus den Ergebnissen einer Eignungsprüfung eine Verbesserung der Prüfqualität zu erzielen.

Die Laboratorien sind aufgefordert, ihre eigenen Schlussfolgerungen aus ihrem Abschneiden bei einer Eignungsprüfung zu ziehen und sich außerdem ein Urteil über die Organisation und die Gestaltung der Eignungsprüfung bilden. Ein Feedback an den Ausrichter der Eignungsprüfung sollte Angaben zu folgenden Dingen enthalten:

- Ursprung und Art der Proben / Prüfkörper,
- angewandte Prüfmethode und wo möglich eine Zuordnung der Ergebnisse zu den einzelnen Methoden,
- Organisation der Eignungsprüfung, z. B. statistisches Modell, Anzahl der Wiederholungen, zu prüfende Parameter, Art der Durchführung,
- Kriterien, die zur Bewertung der Leistung der Teilnehmer angewendet werden.

Der Anbieter der Eignungsprüfung und / oder die Akkreditierungsstelle können ein Treffen organisieren, bei dem die Einzelheiten zur Eignungsprüfung den Teilnehmern dargelegt und mit ihnen diskutiert wird.

Die zuständige Behörde oder Akkreditierungsstelle sollte von einem Eignungsprüfungsanbieter mindestens die Einhaltung der Anforderungen des ISO/IEC Guide 43-1 [10] fordern. Detaillierte Anforderungen an die Kompetenz von Anbietern von Eignungsprüfungen finden sich in ILAC-G13:2000 [37].

In allen Phasen einer Eignungsprüfung ist die Wahrung der Vertraulichkeit von höchster Bedeutung. Die Identität der teilnehmenden Laboratorien darf nur einem Minimum von Personen bekannt sein. Man erreicht dies zum Beispiel durch die Vergabe von Code-Nummern an die Laboratorien. Alle Verweise und Zuordnungen in den Abschlußberichten beziehen sich auf diese Codenummer, so dass die Anonymität des Teilnehmers gewahrt bleibt. Auf diese Weise kann eine Rufschädigung für ein Laboratorium bei schlechtem Abschneiden vermieden werden. Falls im Vorfeld mit den Teilnehmern vereinbart, kann der Koordinator nicht ausreichende Ergebnisse an die zuständige Akkreditierungsstelle weitergeben [10, 11].

Eignungsprüfungen dienen dem Nachweis der Laborkompetenz und damit letztlich den teilnehmenden Laboratorien. Trotzdem kann die Versuchung für einen Teilnehmer groß sein, einen zu optimistischen, falschen Eindruck über die eigene Leistungsfähigkeit zu erwecken.

Beispielsweise können Laboratorien diesen falschen Eindruck hervorrufen, wenn sie routinemäßig Einzelanalysen durchführen, aber den Mittelwert von wiederholten Bestimmungen angeben oder zusätzlich zu den in der Eignungsprüfung vorgesehenen Prüfungen Wiederholungen durchführen. Des Weiteren sollten Absprachen zwischen den Laboratorien vermieden werden.

Es ist nicht möglich, Fälschungen und Absprachen gänzlich auszuschließen, aber mit einigen Maßnahmen lässt sich das Risiko minimieren:

- Es werden keine Proben von einem teilnehmenden Laboratorium zum nächsten weitergegeben. Die Probe wird zum Organisator zurückgeschickt, der überprüft sie und sendet sie an den nächsten Teilnehmer.
- Es werden keine Proben oder Teile von Proben einer Eignungsprüfung zur Analyse an ein anderes Laboratorium weitergegeben.
- Ein Teilnehmer einer Eignungsprüfung nimmt wissentlich keine Probe oder Teile von Proben einer Eignungsprüfung an, die ihm von einem anderen Laboratorium zur Prüfung übergeben werden.
- Die Leitung oder das Personal eines Laboratoriums sprechen nicht mit anderen Laboratorien (oder Unternehmen) über die Probe einer Eignungsprüfung.
- Die Teilnehmer versuchen nicht, den Zielwert für eine Eignungsprüfung vom Anbieter zu erfragen.

Auch wenn der Koordinator der Eignungsprüfung alle Maßnahmen zur Verhinderung von Absprachen und Fälschungen von Ergebnissen getroffen hat, liegt es letztlich in der Eigenverantwortung der Teilnehmer, sie zu vermeiden [34].

3.3.2 Entwicklungsphase

Für den Erfolg einer Eignungsprüfung sind in der Entwicklungsphase Beiträge von Fachleuten verschiedener Fachrichtungen notwendig. Bevor man eine Eignungsprüfung durchführen kann, ist es notwendig, einen Plan zu entwerfen, zu dokumentieren und durch den Beratungsausschuss genehmigen zu lassen. Inhalt und Umfang dieses Plans werden im Folgenden diskutiert, Beispiele finden sich in der Literatur [23, 54 bis 59].

Das Personal zur Durchführung einer Eignungsprüfung muss Erfahrung in der Entwicklung, Realisierung und schriftlichen Dokumentation Laboratoriumsübergreifender Prüfungen haben sowie über eine ausreichende fachliche Qualifikation, ausreichende statistische Kenntnisse und entsprechende verwaltungstechnische Fähigkeiten verfügen. Je nach Art der ge-

planten Eignungsprüfung ist die Anleitung durch Fachleute mit detaillierten Kenntnissen und Erfahrungen bei der Anwendung bestimmter Prüfmethoden notwendig.

Die Mitglieder des Organisationsausschusses werden ernannt, und ihnen werden klar definierte Verantwortlichkeiten zugewiesen. Der Ausschuss sollte sich regelmäßig treffen, um das Programm der Eignungsprüfung zu aktualisieren und die Ergebnisse zu bewerten. Beispiele für Zuständigkeiten sind:

Programm-Koordinator:

- Leitung und offizielle Repräsentation des Organisationsausschusses;
- Zeichnungsbefugnis (Alleinvertretungsrecht) für Schreiben, Anweisungen, Einladungen usw.;
- Datenauswertung;
- Erstellung des Abschlussberichtes;
- Kommunikationsschnittstelle zu allen Beteiligten an einer Eignungsprüfung.

Statistiker:

- Auswahl der statistischen Methoden;
- statistische Auswertung der Ergebnisse;
- Darstellung der Ergebnisse;
- Beratung des Programm-Koordinators und des Organisationsausschusses.

Experten für:

- Dokumentation, Zeitpläne, Datenauswertung, Auswahl des Prüfgegenstandes, Homogenitätsprüfung, usw.

Ein Fachbegutachter kann verlangen, die fachliche Kompetenz des besuchten Labors mit Hilfe einer eigenen Eignungsprüfung nachzuweisen, z. B. indem er dem Laboratorium einen geeigneten Prüfgegenstand (Normal, zertifiziertes Referenzmaterial, ...) zur Prüfung übergibt.

Der Fachbegutachter ist verpflichtet, seine Kriterien „Bestanden“ / „Nicht-bestanden“ vorher schriftlich festzulegen, die Offenlegung erfolgt nach Durchführung der Prüfung. Die Forderungen müssen technisch belastbar und gerechtfertigt sein. Am besten beruhen diese Kriterien auf der Vorgehensweise bereits erprobter Eignungsprüfungen.

Diese von einer Einzelperson organisierte Eignungsprüfung muss mindestens die folgenden Forderungen erfüllen und ausreichend schriftlich dokumentiert sein:

- Art der Probe (Homogenität, vollständige Charakterisierung);

- Zeitpunkt der Prüfung und Bekanntgabe der Ergebnisse, möglichst im Rahmen der Laborbegehung, damit der Fachbegutachter sich von der Kompetenz des Fachpersonals überzeugen kann;
- Das Labor sollte ein Ergebnisblatt übergeben und ggf. einen Prüfbericht erstellen;
- Die Kosten für die Eignungsprüfung müssen dem Laboratorium vorher bekannt sein;
- Das Laboratorium erhält einen schriftlichen Bericht über seine Prüfergebnisse und die Kriterien, die der Fachbegutachter zur Beurteilung verwendet hat;
- Vorgehensweise, falls die definierten Akzeptanzkriterien nicht erfüllt wurden (Wiederholung der Eignungsprüfung; zusätzliche Laborbegehung; zeitweilige Suspendierung der Akkreditierung für dieses Prüfverfahren).

Die Geräte, die der Organisator der Eignungsprüfung verwendet, um den Prüfgegenstand vollständig zu charakterisieren, müssen mindestens dem hohen Qualitätsstandard entsprechen, der von der Eignungsprüfung gefordert wird. Es wird empfohlen, nur akkreditierte Laboratorien mit anerkannt hoher Kompetenz und entsprechender Ausrüstung für eine bestimmte Prüfung als Vertragslaboratorium zu benennen.

ISO/IEC Guide 43-1 [10] und ILAC-G13 [37] fordern die lückenlose Dokumentation der Entwicklung, Planung und Durchführung einer Eignungsprüfung.

Als erstes wird der oben genannte Plan der Eignungsprüfung schriftlich festgehalten. Er enthält:

- Name und Anschrift der Organisation, welche die Eignungsprüfung durchführt;
- Name und Anschrift des Koordinators und der anderen Personen, die an der Entwicklung und Durchführung der Eignungsprüfung beteiligt sind;
- Zweck der Eignungsprüfung;
- Teilnahmebedingungen;
- Name und Anschrift des Vertragslaboratoriums, das für die Durchführung einzelner wesentlicher Teile der Eignungsprüfung vor Versenden der Prüfgegenstände / Proben verantwortlich ist, z. B. Homogenitätsprüfung, Probenherstellung, Stabilitätsprüfung, usw.;
- Liste der potentiellen Teilnehmer (Zielgruppe der Eignungsprüfung). Diese Liste hat keinen Bezug zum Codesystem, mit dem die Anonymität der Teilnehmer garantiert wird;
- Beschreibung des gewählten Prüfgegenstandes / Probe und der Prüfungen sowie eine kurze Begründung, weshalb diese Wahl getroffen wurde;

- Beschreibung der Vorgehensweise bei Beschaffung, Verarbeitung, Überprüfung und Transport der Proben;
- Zeitplan für die verschiedenen Phasen der Eignungsprüfung, dieser Plan enthält die Anfangs- und Endtermine sowie den Zeitraum, in dem die Teilnehmer die Prüfung durchführen;
- Beschreibung der Informationen, die den Teilnehmern vorab überlassen werden;
- Einzelheiten über das anzuwendende Prüfverfahren;
- Grundlage für das gewählte statistische Modell und eventuell notwendige Ausreißerprüfungen;
- Das Bewertungsverfahren für die Laboratoriumsleistungen;
- Beschreibung des Umfangs der Prüfergebnisse und der Kriterien für ihre Bewertung.

Alle Schreiben, Papiere, Ergebnismeldungen, Ausgangsdaten und Rückmeldungen von den teilnehmenden Laboratorien müssen archiviert und mindestens 10 Jahre aufbewahrt werden. Diese Unterlagen dürfen der Öffentlichkeit nicht zugänglich sein.

Anbieter von Eignungsprüfungen sollten ein Qualitätsmanagementsystem (QM-System) auf Grundlage des Leitfadens ILAC-G13:2000 [37] und des ISO/IEC Guide 43-1 [10] aufbauen und unterhalten. Die Akkreditierungsstellen überprüfen beim Anbieter der Eignungsprüfungen regelmäßig die Anwendung und Einhaltung des QM-Systems. Es wird außerdem die Abgabe einer Qualitätserklärung durch den Anbieter von Eignungsprüfungen empfohlen, welche die folgenden im Anhang B des ISO/IEC Guide 43-1 [10] aufgeführten Punkte enthalten sollte:

Dokumentation der Qualitätssicherung nach ISO/IEC Guide 43-1 [10, 11, 34]:

- 1) Grundsätze der Qualitätssicherung;
- 2) Organisation der koordinierenden Stelle;
- 3) Qualifikation und Kompetenz des Personals;
- 4) Festlegung von Zuständigkeiten;
- 5) Lenkung der Dokumentation, Lenkung von Qualitätsaufzeichnungen;
- 6) Verfahren für Audit und Bewertung des QM-Systems;
- 7) Ziele, Umfang, Häufigkeit, Anwendungsbereich, statistische Gestaltung und Format der Eignungsprüfungen
- 8) Prozesslenkung:
 - Probenvorbereitung, Homogenitätsprüfung, Ausrüstung und Geräte, Verfahrensablauf zur Ermittlung von Stoffwerten, Lieferanten, Logistik und Handhabung der Proben, Datenanalyse, Klassifizieren von nicht auswertbaren Proben;

- 9) Erstellung und Herausgabe von Berichten;
- 10) Dokumentation und Archivierung von Berichten;
- 11) Umgang mit Kundenbeschwerden und –beanstandungen;
- 12) Grundsätze der Vertraulichkeit und Geheimhaltung;
- 13) Ethische Betrachtungen;
- 14) Angaben zum Computereinsatz, Hard- und Softwarepflege;
- 15) Sicherheit und andere Umweltfaktoren;
- 16) Vergabe von Unteraufträgen;
- 17) Teilnahmegebühren;
- 18) Teilnahmebedingungen;
- 19) Grundsätze über die Verwendung der Ergebnisse der Eignungsprüfung.

Für die Auswahl des Prüfverfahrens, das bei einer Eignungsprüfung angewandt wird, gibt es zwei grundlegend unterschiedliche Vorgehensweisen. So ist es vor allem in der chemischen Analytik üblich, die Teilnehmer das Prüfverfahren selbst auswählen zu lassen [27, 60 bis 62]. Das gewählte Verfahren entspricht dann in der Regel den Routineverfahren, welche die Laboratorien anwenden. Damit die Ergebnisse aus unterschiedlichen Verfahren miteinander verglichen werden können, fordert der Koordinator Einzelheiten zu dem jeweiligen Prüfverfahren an. Die zweite Möglichkeit ist die Festlegung eines national oder international genormten Prüfverfahrens durch den Organisationsausschuss, wie z. B. Durchführung der Prüfungen nach einem Standardverfahren zur Bestimmung der Genauigkeit von Prüfverfahren für Baustoffe nach ASTM C 802 [63] oder die Eignungsprüfung Härtemessung nach Brinell HBW 2001 [28].

Die Definition der Leistungskriterien ist Aufgabe der Akkreditierungsstelle, welche diese in der Regel an den Organisationsausschuss delegiert. Es gibt folgende Möglichkeiten zur Definition der Leistungskriterien [10, 11]:

- *Expertenkonsens*, d.h. der Beratungsausschuss entscheidet direkt ggf. unter Hinzuziehung weiterer Sachverständiger, ob die eingereichten Ergebnisse gebrauchsfähig sind. Dies ist der übliche Weg zur Beurteilung der Ergebnisse qualitativer Prüfungen.
- *Gebrauchsfähigkeit der Ergebnisse* (engl.: fitness for purpose), wie z. B. Vorgaben für die Leistung der Messmethode bezüglich systematischer Abweichung, Wiederhol- und Vergleichspräzision.
- *Statistische Bestimmung* von Werten, wie z. B. z-Scores (z-Noten) oder E_n -Zahlen, vgl. Kap. 4.4.2.

- *Konsens von Teilnehmern*, d. h. es wird ein Bereich definiert, der von einem bestimmtem Anteil der Teilnehmer erreicht wird (z. B. mittlerer prozentualer Anteil 95 % ausreichend).

Die Ergebnisse werden in der Regel in die Kategorien „Ausreichend“, „Fragwürdig“ und „Unzureichend“ eingeteilt.

Die Auswahl und Vorbereitung des Prüfgegenstandes für die Eignungsprüfung ist einer der schwierigsten Teile bei der Durchführung einer Eignungsprüfung [64]. Homogenität, Stabilität und die Weiterverarbeitbarkeit in Teilprobesätze sind wichtige Gesichtspunkte bei der Auswahl des Prüfgegenstandes für die Eignungsprüfung. Die in den ISO Guides 34 und 35 [65,66] festgelegte Verfahrensweise für die Zertifizierung von Referenzmaterialien sind ein guter Leitfaden, vgl. Kap. 4.1. Die Vorgehensweise bei der Auswahl eines Probematerials ist schematisch im Bild 3.3.2 dargestellt.

Das Material für die Proben einer Eignungsprüfung kann unter Umständen sehr teuer sein. Daher ist es sinnvoll, in der Vorauswahl mehrere mögliche Materialien in Betracht zu ziehen. Diese Materialien müssen sehr sorgfältig charakterisiert werden. Anschließend wird überprüft, ob sich aus dem vorgesehenen Material Proben präparieren lassen, die der Charakterisierung des Vormaterials weiterhin entsprechen. Das Material, welches die Anforderungen der Eignungsprüfung am besten erfüllt (unter anderem auch den Aspekt der Kostenminimierung), wird auf Homogenität und Stabilität geprüft (vgl. Kap. 4.1). Sind diese für die Zwecke der Eignungsprüfung ausreichend, kann mit der Eignungsprüfung begonnen werden. Andernfalls muss diese Schleife mit einem anderen Material erneut durchlaufen werden [34].

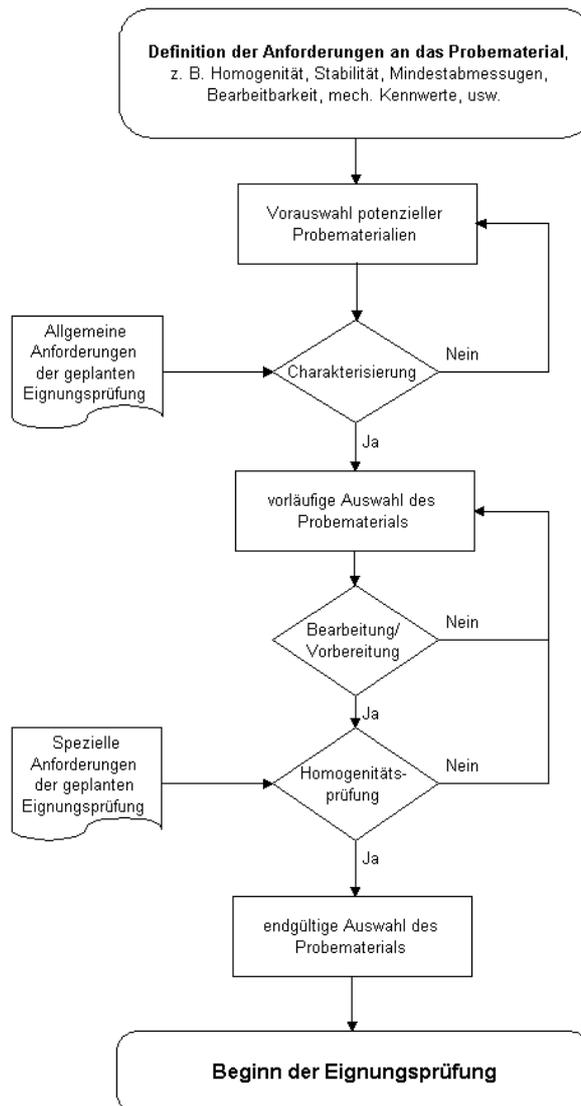


Bild 3.3.2 : Schematische Vorgehensweise bei der Auswahl eines Probematerials nach [26]

Jeder Eignungsprüfung liegt ein geeignetes statistisches Modell zugrunde. Das statistische Modell sollte, je nach Eignungsprüfung, Antworten auf die folgenden Fragen geben [34]:

- Wie viele Laboratorien müssen mindestens teilnehmen?
- Wie viele Proben soll jeder Teilnehmer prüfen?
- Wie viele Messungen muss jeder Teilnehmer an jeder Probe vornehmen?
- Welches Verfahren wird zur Ermittlung des Stoffwertes eingesetzt?
- Wie werden Ausreißer behandelt?
- Welche Wiederhol- und Vergleichspräzision ist bei der jeweiligen Prüfung zu erwarten?
- Wie groß sind die kleinsten Differenzen, die auf einem gewünschten Vertrauensniveau zwischen den Teilnehmern feststellbar sind?

In den Kapiteln 4.2 bis 4.4 werden verschiedene statistische Möglichkeiten zur Auswertung von Eignungsprüfungen vorgestellt.

Nach der endgültigen Festlegung eines Probematerials befasst man sich mit dem Probenmanagement. Darunter versteht man die Stichprobenauswahl, die Probennahme, die Herstellung von Prüfgegenständen, Sicherstellen der Rückverfolgbarkeit, Versand der Prüfgegenstände, die Handhabung der Proben, Zollformalitäten, usw. Dieser Abschnitt einer Eignungsprüfung ist schematisch im Bild 3.3.3 dargestellt.

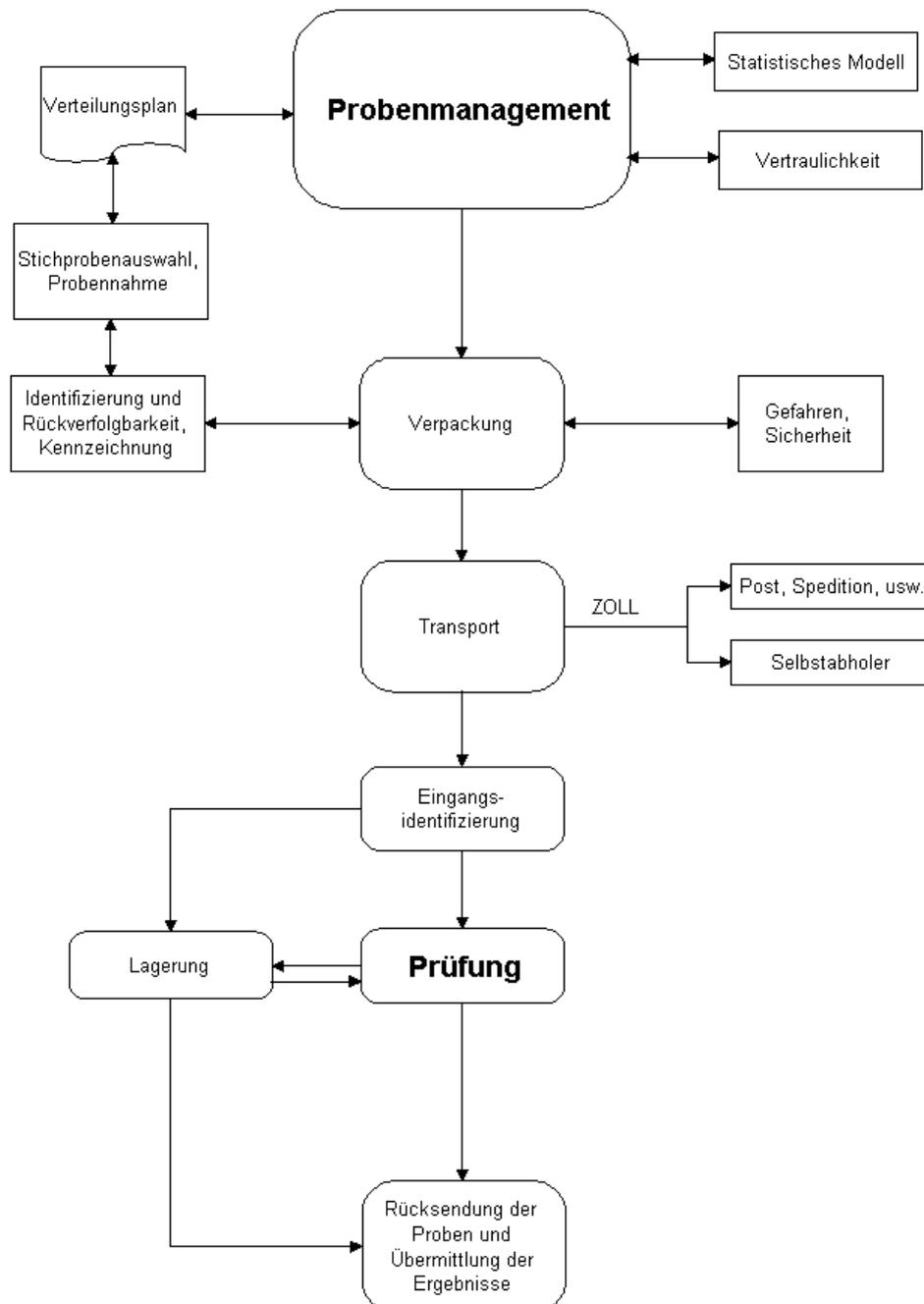


Bild 3.3.3: Probenmanagement, schematischer Ablauf nach [26]

Damit ein reibungsloser Ablauf sichergestellt ist, wird zu jeder Eignungsprüfung ein Zeitplan in Form eines Projektplans erstellt werden. Die Projektfortschritte werden protokolliert und die Planung an den Projektverlauf angepasst. Der Zeitplan sollte Angaben enthalten wie z. B. den Abschluss der Materialcharakterisierung und Homogenitätsprüfung, den Zeitpunkt des Versendens von Teilnahmeaufforderungen, Start der Eignungsprüfung, Ende der Eignungsprüfung, Erscheinen des Abschlussberichtes.

Um möglichst vielen Laboratorien die Teilnahme an einer Eignungsprüfung zu ermöglichen, unterrichtet der Ausrichter einer Eignungsprüfung im Rahmen der Einladung die potenziellen Teilnehmer über Art und Umfang der geplanten Eignungsprüfung, Kosten, Termine, benötigtes Equipment oder Personal. Zu Beginn der Eignungsprüfung erhalten die Teilnehmer gemeinsam mit dem Prüfgegenstand detaillierte und vollständige Anweisungen zur Durchführung der Eignungsprüfung. Zusätzlich zu diesen erhalten die Teilnehmer in der Regel standardisierte Datenerfassungsblätter, damit Art und Umfang der Laboratoriumsauskünfte vergleichbar sind. In manchen Fällen kann die Akkreditierungsstelle die Anfertigung eines Prüfberichtes fordern.

Die Ergebnisse der Eignungsprüfung werden in Form eines Abschlussberichtes veröffentlicht. Alle Angaben in diesem Bericht beziehen sich auf die Code-Nummern der Laboratorien, um so die Anonymität der Teilnehmer zu wahren. Der Bericht sollte folgende Angaben enthalten:

- Name und Anschrift der Organisation, welche die Eignungsprüfung anbietet;
- Namen von Personen und Organisationen, die an der Entwicklung und Durchführung des Programms beteiligt sind;
- Ausgabedatum;
- Berichtsnummer und eindeutige Kennzeichnung der Eignungsprüfung;
- eindeutige Beschreibung der verwendeten Prüfgegenstände sowie Einzelheiten zur Probenpräparation und Homogenitätsprüfung;
- Vorgehensweise bei Ermittlung des Stoffwertes;
- Prüfergebnisse, aufgeschlüsselt nach Laboratoriumscodes, wenn möglich graphisch aufbereitet;
- Schlusswort des Koordinators zur Gesamtleistung aller Teilnehmer.

Einige Anbieter von Eignungsprüfungen stellen den Teilnehmern zusätzlich Teilnahmezertifikate aus, in denen das Abschneiden des Laboratoriums zusammenfassend bewertet ist. Hat ein Laboratorium ausreichende Ergebnisse erzielt, so wird im Zertifikat „erfolgreich teilgenommen“ angegeben, andernfalls wird die Leistung mit „teilgenommen“ beschrieben.

Nachfolgetreffen, an denen der Koordinator, Mitglieder des Beratungsausschusses und die teilnehmenden Laboratorien teilnehmen, sind wünschenswert und sollten einmal jährlich stattfinden. Ein Vertreter der Akkreditierungsstelle sollte ebenfalls hinzugezogen werden. Die Präsentation der Ergebnisse erfolgt unter Wahrung der Anonymität der Teilnehmer. Es ist ausreichend Zeit für Diskussionsbeiträge und Rückmeldungen von Laboratorien vorzusehen.

Um Erfahrungen mit der Durchführung von Eignungsprüfungen zu sammeln, sollte der Beratungsausschuss nach Abschluss eine interne Diskussion über den Gesamtverlauf einer Eignungsprüfung führen. Dies können z. B. Überlegungen zur Gesamtleistung im Vergleich zur erwarteten Leistung oder zu bereits erbrachten Gesamtleistungen in früheren Eignungsprüfungen sein oder mögliche Fehlerquellen bei extremen Ergebnissen und Vorschläge zur Verbesserung. Alle Ergebnisse dieser Diskussion sollten sorgfältig dokumentiert werden, um ggf. bei der Planung der nächsten Eignungsprüfung berücksichtigt zu werden.

4 Statistische Verfahren zur Auswertung von Eignungsprüfungen

Die Ergebnisse von Eignungsprüfungen können in unterschiedlichster Weise dargestellt werden, wobei es große Variationsmöglichkeiten für die vorliegenden Daten und deren statistische Verteilung gibt. Für die Auswertung der gewonnenen Daten müssen geeignete statistische Methoden gewählt werden, denn es gibt kein genormtes Einheitsverfahren zur Auswertung von Eignungsprüfungen.

Um die Ergebnisse einer Eignungsprüfung auszuwerten und zu bewerten, werden bei jeder Eignungsprüfung vier allgemein gebräuchliche Schritte durchlaufen [10]:

- a) Überprüfung und Bewertung von Homogenität und Stabilität der Prüfgegenstände;
- b) Bestimmung des Stoffwertes und seiner Messunsicherheit;
- c) Berechnung der Leistungsstatistik;
- d) Bewertung der Leistung.

4.1 Homogenitätsprüfung

Homogenität ist ein Zustand, in dem eine einheitliche Struktur oder Zusammensetzung für eine oder mehrere festgelegte Eigenschaften vorliegt. Ein Referenzmaterial wird hinsichtlich einer bestimmten Eigenschaft als homogen angesehen, wenn der bei Prüfungen an Stichproben in einem festgelegten Umfang bestimmte Eigenschaftswert innerhalb der festgelegten Unsicherheitsgrenzen liegt.

Die Untersuchung des gesamten Materials verbietet sich naturgemäß, also bleibt nur die stichprobenartige Homogenitätsuntersuchung mit einem statistischen Signifikanztest (Hypothesentest). Da aber jedes statistische Prüfverfahren wiederum eine mathematische Unschärfe besitzt, die unter anderem von der Stichprobengröße abhängt, ist eine Aussage über Homogenität oder Nichthomogenität eines Materials immer nur eine Wahrscheinlichkeitsaussage.

Wird bei einer Homogenitätsprüfung kein signifikanter Unterschied zwischen den untersuchten Proben festgestellt, dann ist folgende Aussage erlaubt:

„Mit dem angewendeten Prüfverfahren kann für die untersuchte Stichprobengröße die Nullhypothese, es handele sich um ein homogenes Material, nicht widerlegt werden.“ [67]

Ein empfindlicheres Prüfverfahren und/oder eine höhere Stichprobengröße können jedoch durchaus zu einer Rückweisung der Nullhypothese führen.

Die Homogenität eines Materials kann deshalb nie bewiesen werden. Es kann lediglich auf einem definierten Signifikanzniveau eine Nicht-Homogenität festgestellt werden. Die Aussagen „homogen“ bzw. „nicht-homogen“ hängen daher auch bei mechanisch-technologischen Prüfungen ab vom Prüfverfahren und der Stichprobengröße, die zur Homogenitätsprüfung verwendet werden. Werden signifikante Unterschiede zwischen den Proben erkannt, dann ist aber in jedem Fall weiterhin zu prüfen, ob die Unterschiede auch relevant sind: Der Unterschied kann z. B. kleiner sein als die Wiederholbarkeit des Analyseverfahrens und ist dann nicht mehr relevant [68].

Die Homogenität des eingesetzten Probematerials muss vor Beginn des Versuches sichergestellt sein, es sei denn, es liegen ausreichende praktische Erfahrungen vor, nach denen die Homogenität des Probematerials keinen Einfluss auf die Prüfergebnisse der Teilnehmer hat. In so einem Fall sollten die Laboratorien Prüfergebnisse von mindestens zwei Proben oder Prüfgegenständen liefern, um so eine nachträgliche Verifizierung der Homogenität zu gestatten.

Grundsätzlich gibt es zwei Arten der Homogenitätsbestimmung, vergleiche Bild 4.1.1:

- Homogenität innerhalb einer Probe (engl. within-bottle homogeneity testing);
- Homogenität zwischen mehreren Proben (engl. between-bottle homogeneity testing).

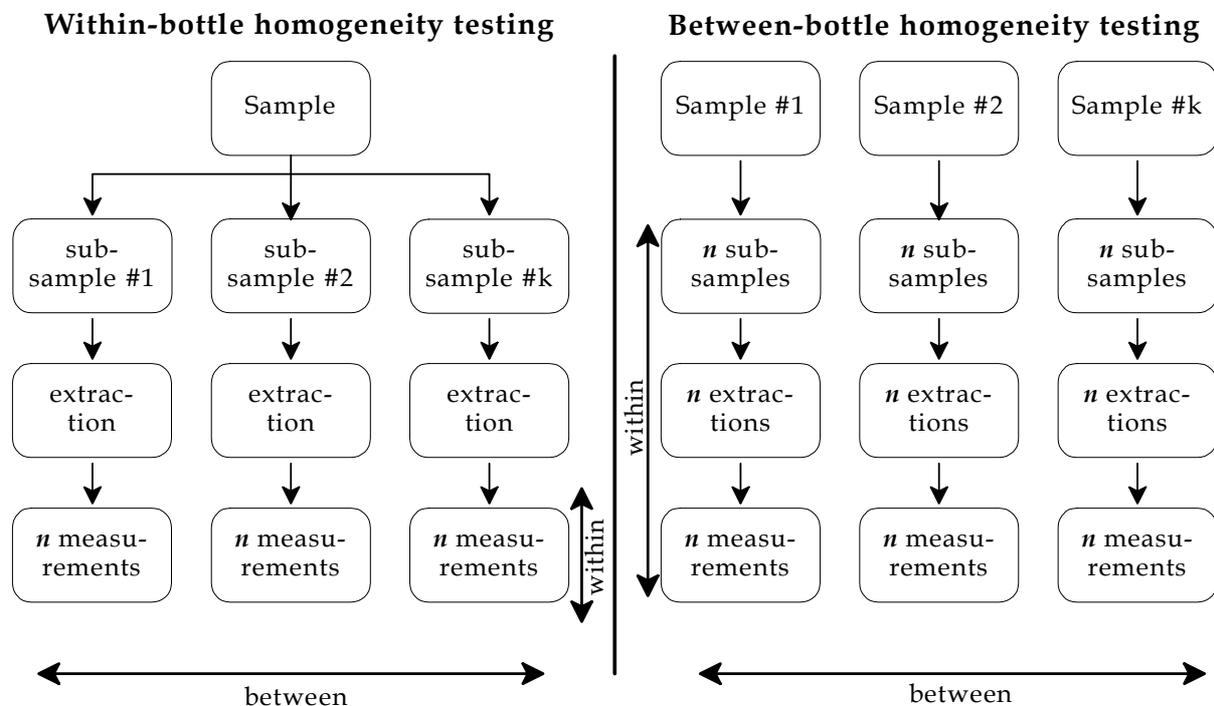


Bild 4.1.1: Versuchsaufbau für die Homogenitätsprüfung innerhalb einer Probe und zwischen mehreren Proben [34]

Die Zielsetzung der beiden Vorgehensweisen zur Homogenitätsprüfung ist unterschiedlich. Die Homogenitätsprüfung zwischen mehreren Proben liefert Informationen (in Form einer Unsicherheitsverteilung) über die Homogenität gemessen über mehrere Proben. Die Homogenitätsprüfung innerhalb einer Probe gibt an, wie groß die verwendete Probemenge sein muss, damit ausreichende Homogenität vorliegt [67]. Die Pfeile "between" und "within" beziehen sich auf jene Teile einer Messkette, die durch die Größen "Varianz zwischen den Gruppen" und "Varianz innerhalb einer Gruppe" bei einer einfaktoriellen Varianzanalyse beschrieben werden [34].

4.1.1 Einfaktorielle Varianzanalyse

Der klassische Weg der Homogenitätsprüfung ist die einfaktorielle Varianzanalyse. Sie dient zum Vergleich von mehr als zwei unabhängigen Stichproben hinsichtlich ihrer Mittelwerte, wobei die Werte der Stichproben normalverteilt sein müssen und über die Stichproben hinweg Varianzenhomogenität vorliegen muss. Die Überprüfung auf Varianzenhomogenität und Normalverteilung der Mittelwerte erfolgt vor Durchführung der Varianzanalyse mittels geeigneter statistischer Tests. Grundlage der Varianzanalyse ist die Zerlegung der Gesamtvarianz. Die Beschreibung des Rechenweges folgt [69].

Mit k = Anzahl der Stichproben (im folgenden Gruppen genannt), n = Gesamtzahl der Werte und x_{ij} als j -ter Wert der i -ten Stichprobe berechnet sich die Gesamtvarianz VAR aus

$$\text{VAR} = \frac{1}{n-1} \cdot \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 \quad (4.1.1).$$

Dabei ist \bar{x} der Mittelwert über alle Werte und n_i der Umfang der i -ten Stichprobe. Die Gesamtvarianz wird zerlegt in eine Varianz innerhalb der Gruppen und eine Varianz zwischen den Gruppen. Für die Summe der Abweichungsquadrate (SAQ) gilt die Beziehung

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 + \sum_{i=1}^k (n_i \cdot (\bar{x}_i - \bar{x})^2) \quad (4.1.2)$$

$\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$	=	SAQ_{gesamt}	<i>Aufsummierung der Abweichungen aller Werte vom Gesamtmittelwert \bar{x}.</i>
$\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$	=	$SAQ_{\text{innerhalb}}$	<i>Abweichung der Werte vom jeweiligen Gruppenmittel.</i>
$\sum_{i=1}^k (n_i \cdot (\bar{x}_i - \bar{x})^2)$	=	SAQ_{zwischen}	<i>Variabilität, die sich aus den Abweichungen der Gruppenmittel vom Gesamtmittel ergibt.</i>

Damit lässt sich die Gleichung (4.1.2) lesen als

$$SAQ_{gesamt} = SAQ_{innerhalb} + SAQ_{zwischen}.$$

Die mittleren Quadrate MQ ergeben sich, wenn man die Summen der Abweichungsquadrate durch ihre zugehörige Anzahl der Freiheitsgrade teilt. Liegen keine signifikanten Unterschiede vor, werden sich MQ (innerhalb) und MQ (zwischen) nur zufällig unterscheiden, was zu einer entsprechenden Prüfgröße führt, die einer F-Verteilung folgt.

Zur Berechnung der SAQ werden die Ausdrücke von Gleichung 4.1.2 umgeformt:

$$S_i = \sum_{j=1}^{n_i} x_{ij} \quad i=1, \dots, k \quad (4.1.3),$$

$$S = \sum_{i=1}^k S_i \quad (4.1.4),$$

$$SAQ (gesamt) = \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}^2 - \frac{S^2}{n} \quad (4.1.5),$$

$$SAQ (zwischen) = \sum_{i=1}^k \frac{S_i^2}{n_i} - \frac{S^2}{n} \quad (4.1.6),$$

$$SAQ (innerhalb) = SAQ (gesamt) - SAQ (zwischen) \quad (4.1.7),$$

$$df (zwischen) = k-1 \quad (4.1.8),$$

$$df (innerhalb) = n-k \quad (4.1.9),$$

$$MQ (zwischen) = \frac{SAQ(zwischen)}{df (zwischen)} \quad (4.1.10),$$

$$MQ (innerhalb) = \frac{SAQ(innerhalb)}{df (innerhalb)} \quad (4.1.11),$$

$$F = \frac{MQ(zwischen)}{MQ(innerhalb)} \quad (4.1.12).$$

Die errechneten Zwischenergebnisse trägt man in das in Tabelle 4.1.1 dargestellte Schema ein. Die Berechnung erfolgt durch leistungsstarke Software, die auch große Datenmengen problemlos bewältigt [70, 71].

Die zu prüfende Hypothese (Nullhypothese) lautet: “k Stichproben stammen aus nur einer Grundgesamtheit, eventuelle Mittelwertunterschiede zwischen den einzelnen Stichproben sind zufällig.“

Die mittleren Quadrate (MQ) sind Schätzer für die unbekannte Varianz der Grundgesamtheit. Wenn die Nullhypothese zutrifft, dann müssen beide berechneten Varianzen erwartungsgetreue Schätzfunktionen für die einzige Grundgesamtheit σ^2 sein oder anders ausgedrückt, ist $MQ_{innerhalb} = MQ_{zwischen}$ zu erwarten. Je größer der Varianzunterschied, desto eher muss die Nullhypothese verworfen werden.

Tabelle 4.1.1: Schema der einfaktoriellen Varianzanalyse [69]

Variabilität	Summe der Abweichungsquadrate (SAQ)	Freiheitsgrad (df)	Mittlere Quadrate (MQ)
gesamt	SAQ_{gesamt}	$n-1$	---
zwischen	$SAQ_{zwischen}$	$k-1$	$MQ_{zwischen} = SAQ_{zwischen} / (k-1)$
innerhalb	$SAQ_{innerhalb}$	$n-k$	$MS_{innerhalb} = SS_{innerhalb} / (n-k)$

Die Prüfgröße F ist F-verteilt mit (k-1, n-k) Freiheitsgraden. Für die drei klassischen Signifikanzniveaus mit der Irrtumswahrscheinlichkeit p (vgl. Tabelle 4.1.2) sind die kritischen F-Werte in Abhängigkeit von den Freiheitsgraden df1 und df2 tabelliert [69]. Signifikanz liegt vor, wenn die berechnete Prüfgröße F den zutreffenden Tabellenwert F_{krit} , in der Regel $F_{0,05}$, überschreitet.

Tabelle 4.1.2: Signifikanzgrenzen (Signifikanzniveaus), Irrtumswahrscheinlichkeit p [69]

$p \leq 0,05$	signifikant
$p \leq 0,01$	sehr signifikant
$p \leq 0,001$	höchst signifikant

Normalerweise werden Homogenitätsuntersuchungen mit dem Ziel durchgeführt, dass alle Einheiten eines Probematerials als „identisch“ anzusehen sind bzw. eine einzelne Probe in sich keine Unterschiede zeigt. Wie PAUWELS et al. in [67] am Beispiel der Vorgehensweise bei einem zertifizierten Referenzmaterial (CRM) zeigen, muss in der Praxis die zu klärende Frage lauten: Ist die Inhomogenität des Materials (entdeckt oder nicht entdeckt) signifikant gegenüber dem für das Material ermittelten Messunsicherheitsintervall?

Die verschiedenen Modelle der möglichen Ergebnisse bei einer Homogenitätsuntersuchung sind in den Bildern 4.1.2 bis 4.1.5 dargestellt:

- A. Signifikante Inhomogenität ($F > F_{0,05}$) entdeckt; diese ist aber vernachlässigbar gegenüber dem Messunsicherheitsintervall des CRM (Bild 4.1.2).
- B. Keine Inhomogenität ($F < F_{0,05}$) entdeckt, die mögliche Inhomogenität ist eventuell signifikant gegenüber dem Messunsicherheitsintervall des CRM (Bild 4.1.3).
- C. Signifikante Inhomogenität ($F > F_{0,05}$) entdeckt, diese ist ebenfalls signifikant gegenüber dem Messunsicherheitsintervall des CRM (Bild 4.1.4).
- D. Signifikante Inhomogenität ($F > F_{0,05}$) entdeckt, diese ist vergleichbar mit dem Messunsicherheitsintervall der Materialeigenschaften des CRM (Bild 4.1.5).

Erhält man bei Durchführung einer Homogenitätsuntersuchung (zwischen den Gruppen) ein Ergebnis Modell A (Bild 4.1.3), so kann die Inhomogenität vernachlässigt werden [67]. Ein solches Ergebnis erzielt man bei einer Eignungsprüfung dann, wenn die Präzision des Messverfahrens zur Homogenitätsprüfung signifikant höher ist als das in der Eignungsprüfung angewandte Messverfahren. Ergibt sich ein Ergebnis vom Modell C, sollte

diese Gruppe des Materials entweder zurückgewiesen oder jedes einzelne Element der Gruppe individuell zertifiziert werden.

Am häufigsten ist ein Ergebnis vom Modell D. In diesem Fall sollte Inhomogenität als Bestandteil (u_{zwischen}) in die Gesamt-Messunsicherheit U des CRM einfließen. U berechnet sich gemäß DIN V ENV 13005 [72] aus der kombinierten Standardunsicherheit u_c und einem Erweiterungsfaktor k nach der Gleichung

$$U = k \cdot u_c = k \cdot \sqrt{\sum u_i^2 + u_{\text{zwischen}}^2} \quad (4.1.13).$$

Es steht u_i als Symbol für alle Messunsicherheiten, die aus anderen Quellen stammen.

Bei der Homogenitätsprüfung innerhalb einer Gruppe gelten die Grundsätze der Prüfung zwischen Gruppen. Die experimentell ermittelte Messunsicherheit u_{exp} ergibt sich aus der Messunsicherheit für die Messung u_{Messung} und der Materialinhomogenität $u_{\text{innerhalb}}$ nach der Gleichung

$$u_{\text{exp}}^2 = u_{\text{Messung}}^2 + u_{\text{innerhalb}}^2 \quad (4.1.14).$$

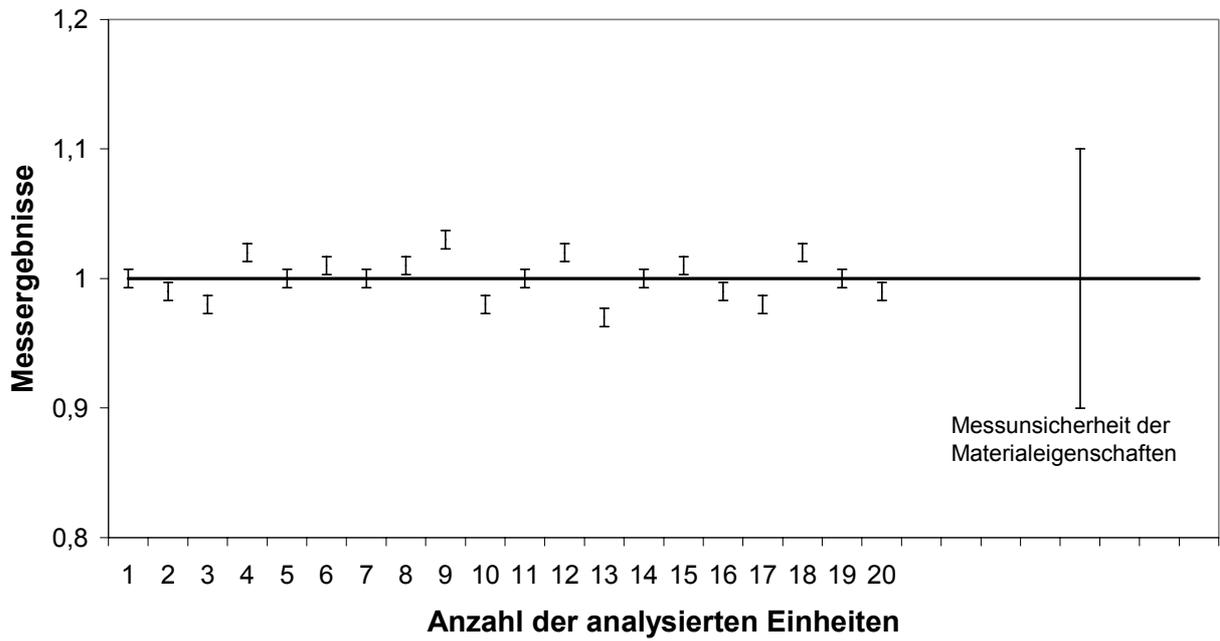


Bild 4.1.2: Homogenitätsuntersuchung an CRM, Ergebnis Modell A [67]

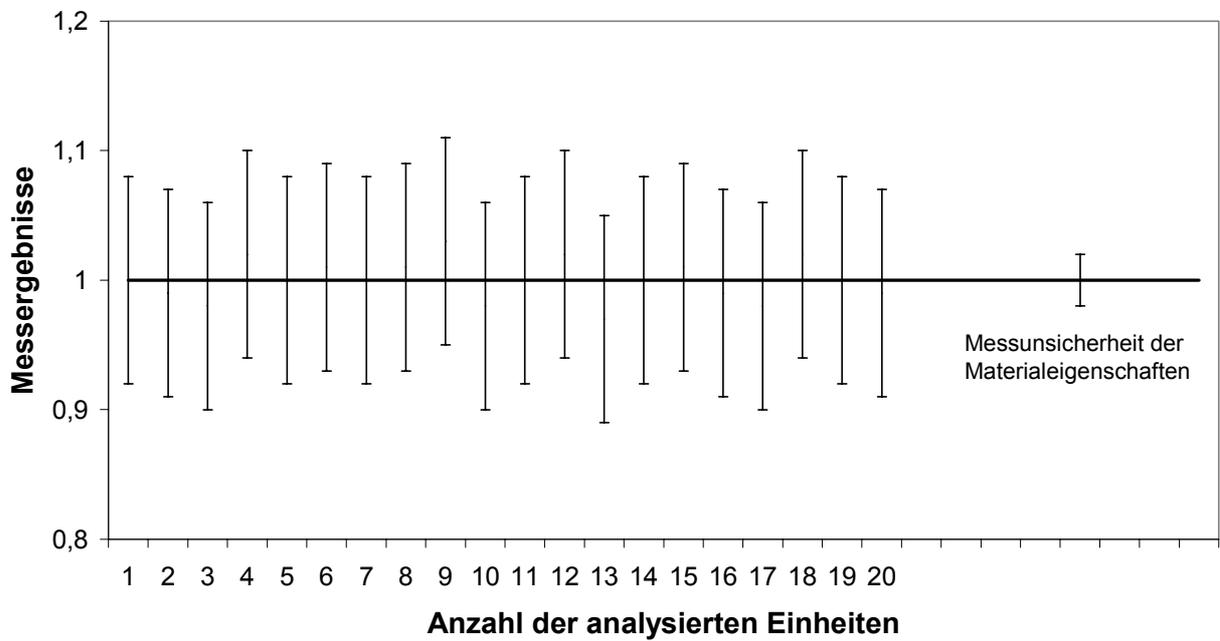


Bild 4.1.3: Homogenitätsuntersuchung an CRM, Ergebnis Modell B [67]

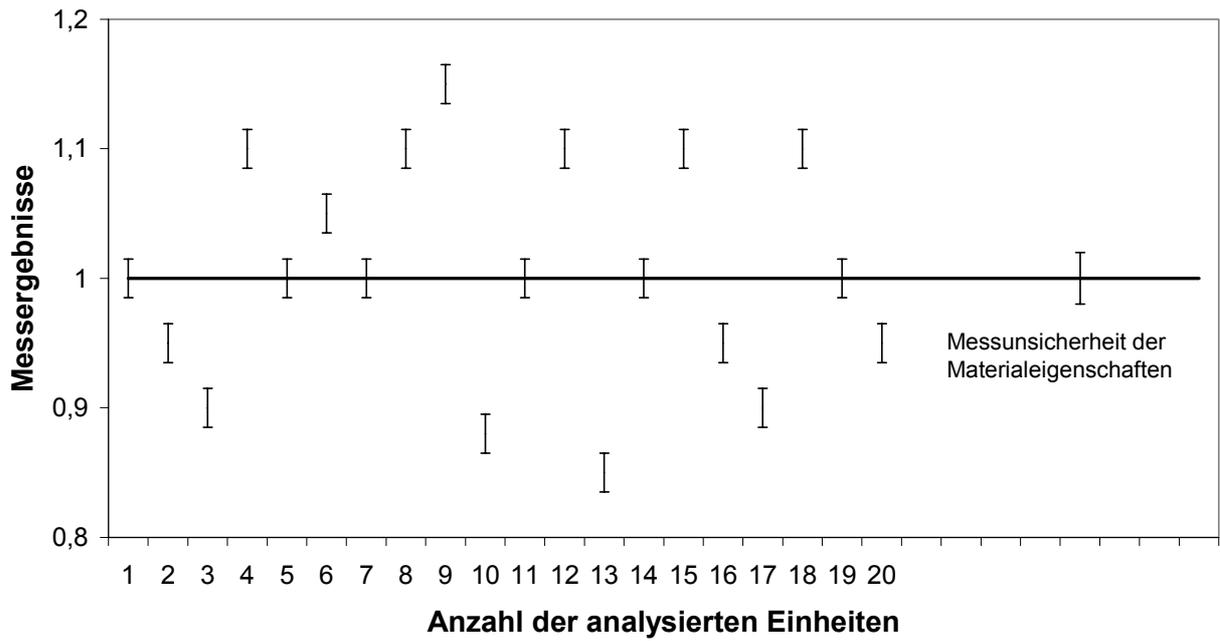


Bild 4.1.4: Homogenitätsuntersuchung an CRM, Ergebnis Modell C [67]

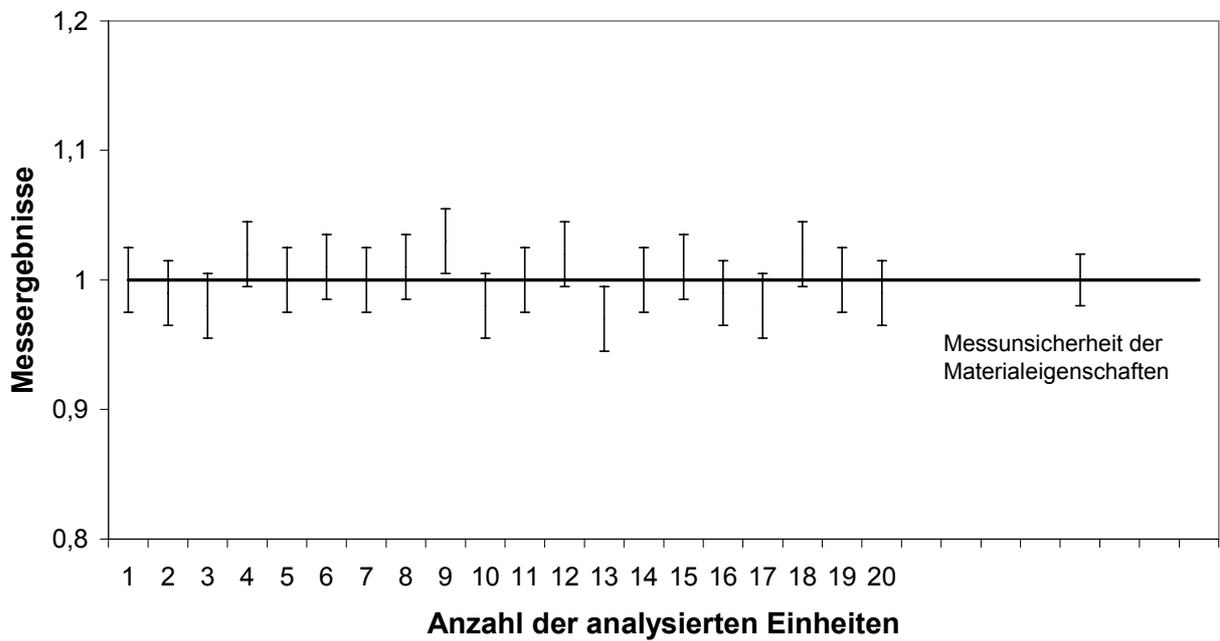


Bild 4.1.5: Homogenitätsuntersuchung an CRM, Ergebnis Modell D [67]

4.1.2 Alternative Vorgehensweise zur Homogenitätsprüfung

In ISO/DIS 13528:2002 [73] wird folgende vereinfachte Vorgehensweise zur Homogenitätsprüfung vorgeschlagen:

- a) Zur Durchführung der Homogenitätsprüfung wird ein geeignetes Laboratorium ausgewählt. Wenn es nicht notwendig ist die Homogenitätsprüfung an jeder Messgröße durchzuführen, wähle man ein Prüfverfahren für die Homogenitätsprüfung und diejenige(n) Materialkenngröße(n), die für die Homogenität zwischen den Proben wesentlich ist (sind).
- b) Die Proben für die Eignungsprüfung werden vorbereitet und versandfertig gemacht. Es müssen sowohl für die Teilnehmer als auch für die Homogenitätsprüfung ausreichend Proben vorhanden sein.
- c) Nach einem Zufallsverfahren wird aus dem versandfertigen Probensatz eine Stichprobe mit der Größe g entnommen, mit $g \geq 10$.
- d) Aus jeder der g Stichproben werden zwei Probemengen hergestellt. Die dabei verwendete Präparationstechnik muss geeignet sein, die Nichthomogenität zwischen den Probemengen zu minimieren.
- e) Die $2g$ Probemengen werden in zufälliger Reihenfolge einzeln unter Wiederholbedingungen geprüft.
- f) Berechne den gemeinsamen Mittelwert \bar{x} , die Standardabweichung innerhalb der Stichproben s_w und die Standardabweichung zwischen den Stichproben s_s , wie in den Gleichungen (4.1.18) bis (4.1.23) beschrieben.

Zur Bewertung der Homogenitätsprüfung schlägt [73] folgende Kriterien vor: Die Standardabweichung zwischen den Stichproben s_s wird mit der Standardabweichung zur Beurteilung der Eignungsprüfung $\hat{\sigma}$ verglichen. Die Proben werden als ausreichend homogen erachtet, wenn gilt

$$s_s \leq 0,3 \hat{\sigma} \quad (4.1.15).$$

Wenn diese Bedingung nicht eingehalten wird, gibt es folgende Möglichkeiten fortzufahren:

- a) Überprüfen der Vorgehensweise bei der Probenpräparation, mit dem Ziel, diese möglicherweise zu verbessern.
- b) Es wird eine Anzahl von Proben an jeden Teilnehmer der Eignungsprüfung verschickt und ein Messresultat an jeder Probe verlangt. Durch die Heterogenität der Proben wird die Standardabweichung innerhalb der Stichproben ansteigen auf:

$$\sigma_{r1} = \sqrt{\sigma_r^2 + s_s^2} \quad (4.1.16).$$

σ_{r1} wird anstelle von s_r bei der Bestimmung der Anzahl der Wiederholungsmessungen verwendet.

- c) Die Standardabweichung zwischen den Stichproben s_s wird bei der Berechnung der Standardabweichung für die Eignungsprüfung $\hat{\sigma}$ berücksichtigt:

$$\hat{\sigma} = \sqrt{\hat{\sigma}_1^2 + s_s^2} \quad (4.1.17).$$

Dabei ist $\hat{\sigma}_1$ die Standardabweichung für die Eignungsprüfung, bei der die Heterogenität des Materials nicht berücksichtigt wurde.

Die Berechnungsformeln für die Homogenitätsprüfung sind folgende:

Die Daten der Homogenitätsprüfung werden dargestellt durch $x_{t,k}$ mit

$t \dots$ Nummer der Stichprobe ($t = 1, 2, \dots, g$)

$k \dots$ Anzahl der Probemengen ($k = 1, 2$).

Berechne den Mittelwert einer Stichprobe aus:

$$\bar{x}_t = \frac{x_{t,1} + x_{t,2}}{2} \quad (4.1.18).$$

Berechne die Spannweiten zwischen den Probemengen aus:

$$w_t = |x_{t,1} - x_{t,2}| \quad (4.1.19).$$

Berechne den Gesamtmittelwert aus:

$$\bar{x} = \frac{\sum_{t=1}^g \bar{x}_t}{g} \quad (4.1.20).$$

Berechne die Standardabweichung der Stichprobenmittelwerte vom Gesamtmittelwert aus:

$$s_x = \sqrt{\frac{\sum_{t=1}^g (\bar{x}_t - \bar{x})^2}{g-1}} \quad (4.1.21).$$

Berechne die Standardabweichung innerhalb der Stichprobe aus

$$s_w = \sqrt{\frac{\sum_{t=1}^g w_t^2}{2g}} \quad (4.1.22).$$

Berechne die Standardabweichung zwischen den Stichproben aus

$$s_s = \sqrt{s_x^2 - (s_w^2/2)} \quad (4.1.23).$$

4.1.3 Stabilitätsprüfung

Damit eine Eignungsprüfung unter kontrollierten Bedingungen abläuft, ist es wichtig, dass das verwendete Probenmaterial für den Zeitraum der Eignungsprüfung gleich bleibende Eigenschaften hat. Man spricht hier von der Stabilität der Materialeigenschaften. Die Eigenschaften der Werkstoffe, die mittels mechanisch-technologischer Werkstoffprüfverfahren untersucht werden, sind stabil (Ausnahme: Kunststoffe), daher ist eine Stabilitätsprüfung in der Regel nicht notwendig. Bei Bedarf kann das folgende Verfahren angewandt werden [73]:

- a) Für die Stabilitätsprüfung wird dasselbe Laboratorium wie für die Homogenitätsprüfung eingesetzt. Es wird dieselbe Messmethode verwendet und dieselbe(n) Messgröße(n) gemessen.
- b) Zwischen der Homogenitätsprüfung und der Stabilitätsprüfung lässt man einen Zeitraum verstreichen, der erfahrungsgemäß etwa der Verzögerung entspricht, die bei der Prüfung der Proben im Rahmen der Eignungsprüfung entsteht.
- c) Es wird eine Zufallsstichprobe der Größe g entnommen, mit $g \geq 3$.
- d) Aus jeder Stichprobe werden mit derselben Methode wie bei der Homogenitätsprüfung zwei Probemengen präpariert.
- e) Die $2g$ Probemengen werden in zufälliger Reihenfolge unter Wiederholbedingungen geprüft. Für jede Probe erhält man ein Messergebnis $y_{t,k}$.
- f) Berechne den Gesamtmittelwert \bar{y} der Messungen der Stabilitätsprüfung.
- g) Vergleiche den Gesamtmittelwert der Homogenitätsprüfung \bar{x} mit dem Gesamtmittelwert der Stabilitätsprüfung \bar{y} . Es liegt ausreichende Stabilität des Probenmaterials vor, wenn die Beziehung

$$|\bar{x} - \bar{y}| \leq 0,3 \hat{\sigma} \quad (4.1.24)$$

erfüllt ist. Andernfalls sollten die Probenpräparationstechnik und die Aufbewahrungsbedingungen für das Probenmaterial verbessert werden.

Solange die Veränderung des Materials ein Prozess ist, der mehr oder weniger unabhängig von den Umgebungsbedingungen abläuft bzw. für alle Teilnehmer einer Eignungsprüfung in etwa gleich ist, kann man das Material für eine Eignungsprüfung verwenden. In diesem Fall muss ein Zeitpunkt oder eine Frist festgesetzt werden, an dem oder innerhalb der die Proben geprüft werden müssen.

Wird ein Stoffwert für eine Eignungsprüfung ermittelt, muss darauf geachtet werden, dass

- (1) der Stoffwert unter den selben Bedingungen ermittelt wird wie die Prüfergebnisse der teilnehmenden Laboratorien oder
- (2) die Messunsicherheit aufgrund der Instabilität in die Gesamtmessunsicherheit des Stoffwertes hineingerechnet wird.

4.2 Bestimmung des Stoffwertes und seiner Messunsicherheit

Der Stoffwert X (engl. assigned value) eines Prüfgegenstandes ist derjenige Wert, mit dem die Prüfergebnisse der teilnehmenden Laboratorien verglichen werden. Dieser wird den Teilnehmern erst bekannt gegeben, wenn die Auswertung der Eignungsprüfung abgeschlossen ist. In manchen Fällen kann es zweckmäßig sein, den teilnehmenden Laboratorien vor der Prüfung Zielbereiche zu empfehlen [11].

Der Stoffwert X hat eine Standardmessunsicherheit u_X , die vom Verfahren abhängt, mit dem der Stoffwert ermittelt wurde. Die Standardabweichung der Eignungsprüfung $\hat{\sigma}$ wird verwendet, um die systematische Abweichungskomponente (engl. bias) der Laboratorien zu bewerten, die in der Eignungsprüfung gefunden wurden. Wenn die Standardmessunsicherheit des Stoffwertes u_X im Vergleich zur Standardabweichung der Eignungsprüfung $\hat{\sigma}$ zu groß ist, besteht für einige Laboratorien die Gefahr unzulässiger Ungenauigkeit bei der Bestimmung des Stoffwertes, ohne dass die Ursache hierfür den Laboratorien angelastet werden kann. Solange die Beziehung

$$u_X \leq 0,3 \hat{\sigma} \quad (4.2.1)$$

erfüllt ist, kann die Messunsicherheit des Stoffwertes vernachlässigt werden und muss bei der Bewertung der Eignungsprüfung nicht berücksichtigt werden.

Falls Gleichung (4.2.1) nicht erfüllt wird, muss entweder eine Methode zur Bestimmung des Stoffwertes gewählt werden, die zur Erfüllung von Gleichung (4.2.1) führt, oder die Messunsicherheit des Stoffwertes muss bei der Leistungsbewertung (siehe Kapitel 4.3.2) berücksichtigt werden. Außerdem müssen die Teilnehmer der Eignungsprüfung informiert werden, dass die Messunsicherheit des Stoffwertes nicht vernachlässigt werden kann [73].

Üblicherweise werden fünf Verfahren zur Bestimmung des Stoffwertes angewendet, die bezüglich der Genauigkeit hierarchisch geordnet sind [34, 74]. Die Messunsicherheit der Stoffwerte nimmt von Methode A zu Methode E hin zu. Der Koordinator legt gemeinsam mit dem Beratungsausschuss fest, welche Bestimmungsmethode verwendet wird.

Methode A: bekannter Wert durch

- (1) Expertenübereinkunft (Ergebnisse, die in Kategorien eingeteilt werden können);
- (2) besondere Herstellung des Prüfgegenstandes, z. B. Verdünnen oder Mischen.

Die Methode A (2) wird vor allem dann angewendet, wenn auf diese Art einzelne Proben hergestellt werden können, deren Zusammensetzung ermittelt werden soll. Wird der Stoffwert X aus der Zusammenstellung der Inhaltsstoffe bestimmt, schätzt man die Standardmessunsicherheit u_X durch die Kombination der einzelnen Messunsicherheiten wie in [73] beschrieben ab.

Für eine Eignungsprüfung für mechanisch-technologische Prüfverfahren ist diese Methode in der Regel nicht anwendbar. Eine Ausnahme stellt die Korngrößenmessung in der Metallographie dar, bei der es möglich ist, auf grafischem Wege ein Gefügebild mit definierter Korngröße zu simulieren. Ein denkbares anderes Gebiet ist eine Eignungsprüfung für Prüfmaschinen-Software, bei der bekannte Datensätze ausgewertet werden müssen. Hier liegen jedoch noch keine praktischen Erfahrungen vor.

Methode B: Zertifizierter Referenzwert

Wenn in der Eignungsprüfung ein zertifiziertes Referenzmaterial (engl. certified reference material CRM) verwendet wird, so wird sein zertifizierter Referenzwert als Stoffwert X verwendet. Als Standardmessunsicherheit u_X wird die Angabe zur Messunsicherheit übernommen, die im Zertifikat angegeben wird [11]

Methode C: Referenzwert

Der Stoffwert wird durch Analyse, Messung oder Vergleich des Prüfgegenstandes mit einem zertifizierten Referenzmaterial oder Normal (zurückführbar auf ein nationales oder internationales Normal) ermittelt [73].

Das als Probenmaterial vorgesehene Referenzmaterial (RM) wird versandfertig vorbereitet. Eine Anzahl Proben wird zufällig entnommen und wird von einem Referenzlaboratorium unter Wiederholbedingungen zusammen mit einem zertifizierten Referenzmaterial (CRM) geprüft. Der Stoffwert X_{RM} wird aus der Kalibrierung gegen den zertifizierten Referenzwert (oder das Normal) gewonnen.

Die Standardmessunsicherheit $u_{X, RM}$ wird aus dem Vergleich der Ergebnisse der Messreihe mit Referenzmaterial und zertifiziertem Referenzmaterial u_D berechnet, wobei die Messunsicherheit des CRM $u_{X, CRM}$ berücksichtigt wird,

$$u_{X, RM} = \sqrt{u_{X, CRM}^2 + u_D^2} \quad (4.2.2).$$

Methode D: Konsenswert aus den Ergebnissen besonders qualifizierter Laboratorien

Wie bei Methode C wird das Probenmaterial versandfertig vorbereitet. Dann werden einige Proben zufällig entnommen und von einer Gruppe besonders qualifizierter Laboratorien^A (Expertengruppe, engl. expert laboratories) geprüft. Alternativ kann diese Gruppe von besonders qualifizierten Laboratorien auch Teilnehmer an einer Runde einer Eignungsprüfung sein. Der Stoffwert und seine Messunsicherheit wird dann nach Abschluss dieser Runde bestimmt. Der Stoffwert X wird als robuster Mittelwert (siehe Kapitel 4.3.2) aus den Ergebnissen der Gruppe der besonders qualifizierten Laboratorien berechnet.

Jedes der p besonders qualifizierten Laboratorien führt eine Messung x_i auf dem Probenmaterial durch und gibt eine Schätzung der Standardmessunsicherheit u_i an. Wird der Stoffwert X als robuster Mittelwert ermittelt, siehe Gleichungen (4.3.26) bis (4.3.33), so lässt sich die Standardmessunsicherheit des Stoffwertes u_X dann wie folgt berechnen:

$$u_x = \frac{1,23}{p} \cdot \sqrt{\sum_{i=1}^p u_i^2} \quad (4.2.3).$$

Falls die besonders qualifizierten Laboratorien keine Messunsicherheit angeben oder die Messunsicherheit nicht durch eine unabhängige Stelle (d. h. eine Akkreditierungsstelle) validiert wurden, kann man die Messunsicherheit auch wie in Methode E beschrieben ermitteln.

Das Problem bei der Anwendung von Methode D besteht in möglicherweise unerkannten systematischen Abweichungen in den Resultaten der Expertengruppe. Außerdem können die gemeldeten Messunsicherheiten unzuverlässig sein.

Methode E: Konsenswert aus den Ergebnissen aller Teilnehmer

Bei dieser Vorgehensweise wird der robuste Mittelwert berechnet aus den Ergebnissen aller Teilnehmer als Stoffwert X für das Probenmaterial einer Eignungsprüfung verwendet. Der robuste Mittelwert wird entweder wie in Kapitel 4.3.3, Algorithmus A beschrieben, berechnet oder mit einer anderen statistisch abgesicherten Methode bestimmt. Liegt eine Normalverteilung der Daten vor, ist es möglich, den Stoffwert und die Standardmessunsicherheit des Stoffwertes mit einer klassischen statistischen Methode, siehe Kapitel 4.3.2, zu berechnen [49, 54].

Verwendet man den Algorithmus A zur Bestimmung des Stoffwertes, berechnet sich die Standardmessunsicherheit des Stoffwertes u_X aus Gleichung (4.2.4)

^A [11] bezeichnet diese Laboratorien als Sachverständigenlaboratorien, „...die eine nachweisbare Kompetenz bei der Bestimmung der zu prüfenden Messgröße haben [sollten und] validierte Verfahren anwenden, die als äußerst präzise und genau bekannt und mit anderen allgemein angewendeten Verfahren vergleichbar sind.“

$$u_x = 1,23 \cdot \frac{s^*}{\sqrt{p}} \quad (4.2.4).$$

Die robuste Standardabweichung s^* wird ebenfalls nach Algorithmus A iterativ berechnet. Die Anzahl der teilnehmenden Laboratorien ist p .

Die Methode E weist zwei Nachteile auf:

- a) Es gibt Fälle, in denen sich aus den Werten der Teilnehmer kein wirklicher Konsenswert ermitteln lässt, z. B. bei Vorliegen einer zweigipfeligen Verteilung. Ein Beispiel, bei dem die Bildung eines Konsenswertes nicht möglich war, findet sich in Kap. 5.3.
- b) Der Konsenswert wird durch eine stark ausgeprägte systematische Abweichungskomponente verfälscht, die z. B. auf starke Unterschiede im Kompetenzniveau der teilnehmenden Laboratorien zurückzuführen ist, die sich umso mehr bemerkbar macht, je weniger Laboratorien teilnehmen. Diese systematische Abweichungskomponente wird durch die nach Gleichung (4.2.4) bestimmte Messunsicherheit nicht ausreichend erfasst. Außerdem kann das Ergebnis eines Laboratoriums, das die geprüfte Messmethode signifikant besser beherrscht als die anderen Teilnehmer, aufgrund des verfälschten Konsenswertes irrtümlich mit „nicht ausreichend“ bewertet werden, ohne dass dieses Problem erkannt wird.

Wird der Stoffwert X nach den Methoden A bis D bestimmt, sollte nach jeder Runde der Eignungsprüfung der robuste Mittelwert x^* (vgl. Kap. 4.3, Gleichungen (4.3.26) bis (4.3.31)), der aus den Resultaten dieser Runde berechnet wurde, mit dem Stoffwert verglichen werden [73]. Dabei muss gelten

$$x^* - X \leq 2 u_{(x^*-X)} \quad (4.2.5).$$

Die Standardunsicherheit von $u_{(x^*-X)}$ wird abgeschätzt aus der Gleichung

$$u_{(x^*-X)} = \sqrt{\frac{(1,23 \times s^*)^2}{p} + u_x^2} \quad (4.2.6),$$

wobei s^* gleich der robusten Standardabweichung und p gleich der Anzahl Laboratorien ist. Wenn Gleichung (4.2.5) nicht erfüllt ist, sollte die Ursache dafür untersucht werden.

Mögliche Ursachen können sein:

- systematische Abweichung in der Messmethode;
- eine gemeinsame systematische Abweichung bei allen Laboratorien;
- Fehler bei der Abschätzung der Grenzen einer Messmethode (nur bei Methode A);
- Systematische Abweichung in den Resultaten der besonders qualifizierten Laboratorien (nur bei Methode D).

Wenn die Anzahl der Wiederholungsmessungen n bei einer Eignungsprüfung mindestens gleich 2 ist, ist der erste Schritt zur Auswertung der Ergebnisse die Berechnung von Mittelwert und Standardabweichung eines jeden Laboratoriums. Diese wiederum werden dann zur Berechnung der Leistungsstatistik herangezogen. Obwohl alle Teilnehmer aufgefordert sind, die gleiche Anzahl von Wiederholungsmessungen durchzuführen, gibt es immer wieder Laboratorien, die weniger Messergebnisse rückmelden, z. B. wenn die Messungen nicht fachgerecht ausgeführt wurden und nicht wiederholt werden können.

Gibt ein Laboratorium mindestens $n^*=0,59 n$ Messwerte^B an, können Mittelwert und Standardabweichung aus diesen n^* Messungen zur Berechnung der Leistungsstatistik herangezogen werden. Wird diese Bedingung nicht erfüllt, darf dieses Laboratorium z. B. nicht zur Berechnung eines Konsenswertes nach Methode E herangezogen werden, da sich die Standardabweichung des Mittelwertes unzulässig vergrößert. [10, 11, 34, 48, 49, 54, 72 bis 74]

^B Begründung: Vermindert man die Anzahl der Wiederholungsmessungen von n auf $n^*=0,59$, folgt daraus eine Vergrößerung der Standardabweichung des Mittelwertes um den Faktor 1,3. Dieser Faktor wird als akzeptabel betrachtet. Da es sich bei dieser Regel um eine willkürliche Festlegung handelt, kann in begründeten Fällen von ihr abgewichen werden [73].

4.3 Statistische Auswertung von Ringversuchen

4.3.1 Statistisches Grundmodell nach DIN ISO 5725-1

Die Auswertung nach DIN ISO 5725-2 [54] wird von MÄDER [49] als „klassisches Verfahren der Ringversuchsauswertung“ bezeichnet. Das dieser Auswertung zu Grunde liegende statistische Modell wird in DIN ISO 5725-1 [48] beschrieben:

Grundlage zur Schätzung der Genauigkeit (Richtigkeit und Präzision) eines Messverfahrens ist die Annahme, jedes einzelne Messergebnis x sei die Summe von drei Komponenten:

$$x = m + B + e \quad (4.3.1).$$

Dabei ist für das jeweils untersuchte Material

- m der Gesamtmittelwert (Erwartungswert),
- B die systematische Abweichungskomponente des Laboratoriums unter Wiederholbedingungen,
- e die zufällige Abweichung, die bei jeder Messung unter Wiederholbedingungen vorkommt.

Die Größe B wird als Summe aus zufälligen und systematischen Anteilen betrachtet. Die Varianz von B wird als „Varianz zwischen den Laboratorien“ bezeichnet und ausgedrückt als

$$\text{var}(B) = \sigma_L^2 \quad (4.3.2)$$

und schließt Unterschiede zwischen den Bearbeitern und zwischen den Geräteausstattungen ein.

Die zufällige Abweichung e steht für eine bei jeder Messwertermittlung auftretende zufällige Abweichungskomponente. Das Verfahren nach DIN ISO 5725-2 [54] geht von einer näherungsweise Normalverteilung von e aus. Innerhalb eines Einzellaboratoriums wird die Varianz unter Wiederholbedingungen „Varianz innerhalb des Laboratoriums“ genannt und ausgedrückt als

$$\text{var}(e) = \sigma_w^2 \quad (4.3.3).$$

In verschiedenen Laboratorien hat σ_w^2 unterschiedliche Werte. Für ein ordnungsgemäß durchgeführtes vereinheitlichtes Messverfahren geht DIN ISO 5725-2 [54] jedoch von vernachlässigbar kleinen Unterschieden aus. Auf Basis dieser Annahme ist es gerechtfertigt, einen gemeinsamen Wert für die „Varianz innerhalb eines Laboratoriums“ für alle das Messverfahren anwendenden Laboratorien festzulegen. Dieser Wert wird als „Wiederholvarianz“ bezeichnet und wird durch den arithmetischen Mittelwert aus allen „Varianzen innerhalb eines Laboratoriums“ abgeschätzt:

$$\sigma_r^2 = \overline{\text{var}(e)} = \overline{\sigma_w^2} \quad (4.3.4).$$

In diesem Grundmodell wird die Wiederholvarianz unmittelbar als die Varianz der zufälligen Abweichung e gemessen. Die Vergleichsvarianz hängt von der Summe aus der Wiederholvarianz und der „Varianz zwischen den Laboratorien“ ab. Zwei Größen sind als Präzisionsmaße erforderlich:

Die Wiederholstandardabweichung

$$\sigma_r = \sqrt{\text{var}(e)} \quad (4.3.5)$$

und die Vergleichsstandardabweichung

$$\sigma_R = \sqrt{\sigma_L^2 + \sigma_r^2} \quad (4.3.6).$$

Gegebenenfalls notwendige Erweiterungen zum statistischen Grundmodell werden in den weiteren Teilen 3 bis 6 der DIN ISO 5725 beschrieben.

In der statistischen Praxis ist der wahre Wert einer Standardabweichung σ in der Regel unbekannt und wird durch einen Stichprobenschätzwert ersetzt. Zur Verdeutlichung wird das Symbol σ durch das Symbol s ersetzt. Daraus folgt:

s_L^2 ist der Schätzwert für die „Varianz zwischen den Laboratorien“;

s_w^2 ist der Schätzwert für die „Varianz innerhalb eines Laboratoriums“;

s_r^2 ist der arithmetische Mittelwert von s_w^2 und zugleich Schätzwert für die Wiederholvarianz. Dieser arithmetische Mittelwert wird aus den Ergebnissen aller teilnehmenden Laboratorien berechnet, die nach dem Aussondern von Ausreißern übrig bleiben.

s_R^2 ist der Schätzwert für die Vergleichsvarianz gemäß der Gleichung

$$s_R^2 = s_L^2 + s_r^2 \quad (4.3.7).$$

4.3.2 Datenauswertung nach DIN ISO 5725-2

Zur Auswertung der Daten einer Eignungsprüfung werden diese zu den Gesamtkennwerten der Eignungsprüfung zusammengefasst. Dies ist folgerichtig, da alle Messwerte der selben Grundgesamtheit entstammen, da ja die Prüfungen an identischen Proben erfolgten und die Laboratorien nach einem vereinheitlichten Prüfverfahren arbeiten. Man kann verschiedene Stichproben jedoch nur zu einer vereinigten Stichprobe zusammenfassen, wenn Varianzenhomogenität und die Homogenität der Mittelwerte gegeben ist. Das Hauptproblem aller Eignungsprüfungen besteht jedoch in der Inhomogenität der ermittelten Daten [49]. Daher werden vor der eigentlichen Auswertung die Ergebnisse auf Vereinbarkeit und Ausreißer geprüft. DIN ISO 5725-2 [54] gibt zwei mögliche Ansätze an, die grafische Vereinbarkeitsprüfung oder die Anwendung numerischer Ausreißertests.

4.3.2.1 Grafische Vereinbarkeitsprüfung

Die grafische Vereinbarkeitsprüfung verwendet zwei Prüfgrößen: „Mandels h-Statistik“ und „Mandels k-Statistik“.

Wenn von $i = 1, 2, \dots, p$ Laboratorien Prüfungen auf $j = 1, 2, \dots, q$ Merkmalsniveaus durchgeführt wurden, berechnet man die Vereinbarkeitsprüfgröße zwischen den Laboratorien h wie folgt:

$$h_{i,j} = \frac{\bar{x}_{i,j} - \bar{x}_j}{\sqrt{\frac{1}{p_j - 1} \cdot \sum_{i=1}^{p_j} (\bar{x}_{i,j} - \bar{x}_j)^2}} \quad (4.3.8).$$

$\bar{x}_{i,j}$ ist der Mittelwert des Laboratoriums i für das Merkmalsniveau j . \bar{x}_j ist der Gesamtmittelwert für das Merkmalsniveau j .

Die Werte für $h_{i,j}$ werden getrennt nach Laboratorien, gruppiert nach den unterschiedlichen Merkmalsniveaus gemeinsam in einem Diagramm dargestellt. Ein Beispiel findet sich in Kapitel 5.5, Bild 5.5.2.

Die Vereinbarkeitsprüfgröße innerhalb des Laboratoriums k berechnet man nach:

$$k_{i,j} = \frac{s_{i,j} \cdot \sqrt{p_j}}{\sqrt{\sum s_{i,j}^2}} \quad (4.3.9)$$

$s_{i,j}$ ist die Standardabweichung des Laboratoriums i für das Merkmalsniveau j . Ansonsten gilt die Nomenklatur von Gleichung (4.3.8).

Die Werte für $k_{i,j}$ werden getrennt nach Laboratorien, gruppiert nach den unterschiedlichen Merkmalsniveaus, gemeinsam in einem Diagramm dargestellt. Ein Beispiel findet sich in Kapitel 5.5, Bild 5.5.1.

Die Prüfung der grafischen Darstellungen kann für bestimmte Laboratorien spezifische Ergebnismuster anzeigen, die sich markant von denen der anderen Laboratorien unterscheiden, z. B. eine durchgängig hohe Streuung innerhalb der Messwerte eines Labors.

Die Werte dieses Laboratoriums können dann entweder vorläufig zurückgehalten oder aus dem Versuch entfernt werden oder die Prüfung wird (wenn möglich) wiederholt.

In Tabelle 6 und 7 der DIN ISO 5725-2 [54] sind für 1%- und 5%-Vertrauensniveaus Indikatoren für die Prüfgrößen h und k tabelliert. Diese Indikatoren sollten als Gerade in die grafischen Darstellungen von h und k eingetragen werden und dienen als Anhaltspunkte bei der Prüfung der Daten auf Muster.

4.3.2.2 Numerische Ausreißerprüfung

Ein Ausreißer wird von DIN ISO 5725-1 [48] definiert als ein „Wert unter einer Serie von Werten, der mit den anderen Werten dieser Serie nicht verträglich ist“.

Für die Behandlung von Ausreißern wird empfohlen, die in der DIN ISO 5725-2 [54] vorgestellten Tests durchzuführen, um Fastausreißer (engl. Straggler) oder Ausreißer zu erkennen. Die beiden Tests verfolgen unterschiedliche Ziele:

Der Cochran-Test ist ein Test auf Streuung innerhalb eines Labors und sollte als erster angewendet werden. Er prüft die Gültigkeit der Annahme, zwischen den Laboratorien bestünden nur kleine Unterschiede bezüglich der Streuungen innerhalb eines Labors [75]. Der Test beruht auf der Auswertung einer Reihe von p Standardabweichungen s_i , die alle aus derselben Anzahl n von Wiederholergebnissen berechnet wurden. Die Prüfgröße C nach Cochran ist dann

$$C = \frac{s_{\max}^2}{\sum_{i=1}^p s_i^2} \quad (4.3.10),$$

s_{\max} ist die größte Standardabweichung dieser Reihe.

Diese Prüfgröße C wird jeweils mit den tabellierten Signifikanzgrenzen $V(\alpha)$ für eine Irrtumswahrscheinlichkeit von $\alpha=5\%$ und $\alpha=1\%$ verglichen:

- wenn $V(5\%) < C \leq V(1\%)$ liegt ein Fastausreißer vor.
- wenn $C > V(1\%)$ liegt ein statistisch hoch signifikanter Ausreißer vor.

Wurde die größte Standardabweichung als Ausreißer eingestuft, sollte der Wert zurückgewiesen werden, der Test wird für die verbleibenden Werte wiederholt. Es kann jedoch zu einer übermäßigen Rückweisungsquote kommen, wenn die Annahme der Normalverteilung der Daten nicht hinreichend angenähert ist.

Der Grubbs-Test [76, 77] untersucht die Streuung zwischen den Laboratorien. Um zu ermitteln, ob der größte Wert x_p ein Ausreißer ist, wird aus der Datenreihe $x_i = 1, 2, \dots, p$ angeordnet in aufsteigender Reihenfolge die Prüfgröße G_p errechnet:

$$G_p = \frac{x_p - \bar{x}}{s} \quad (4.3.11).$$

Dabei ist \bar{x} das arithmetische Mittel dieser Datenreihe und s ihre Standardabweichung.

Diese Prüfgröße G_p wird jeweils mit den tabellierten Signifikanzgrenzen $V(\alpha)$ für eine Irrtumswahrscheinlichkeit von $\alpha=5\%$ und $\alpha=1\%$ verglichen.

- wenn $V(5\%) < G_p \leq V(1\%)$ liegt ein Fastausreißer (Straggler) vor.
- wenn $G_p > V(1\%)$ liegt ein statistisch hoch signifikanter Ausreißer vor.

Prüft man den kleinsten Wert der Datenreihe auf Signifikanz, so ist die Prüfgröße

$$G_1 = \frac{\bar{x} - x_l}{s} \quad (4.3.12).$$

Soll geprüft werden, ob die zwei größten Beobachtungen Ausreißer sind, errechnet sich die Prüfgröße G aus

$$G = \frac{s_{p-1,p}^2}{s_0^2} \quad (4.3.13),$$

mit

$$s_0^2 = \sum_{i=1}^p (x_i - \bar{x})^2 \quad (4.3.14)$$

und

$$s_{p-1,p}^2 = \sum_{i=1}^{p-2} (x_i - \bar{x}_{p-1,p})^2 \quad (4.3.15)$$

und

$$\bar{x}_{p-1,p} = \frac{1}{p-2} \cdot \sum_{i=1}^{p-2} x_i \quad (4.3.16).$$

Zur Prüfung der zwei kleinsten Werte der Datenreihe auf Signifikanz rechnet man

$$G = \frac{s_{1,2}^2}{s_0^2} \quad (4.3.17),$$

mit

$$s_{1,2}^2 = \sum_{i=3}^p (x_i - \bar{x}_{1,2})^2 \quad (4.3.18)$$

und

$$\bar{x}_{1,2} = \frac{1}{p-2} \cdot \sum_{i=3}^p x_i \quad (4.3.19).$$

Die Signifikanzgrenzen $V(\alpha)$ sind tabelliert und finden sich z. B. in der DIN ISO 5725-2 [54].

Wurde der größte arithmetische Mittelwert als Ausreißer eingestuft, sollte der Wert zurückgewiesen werden, der Test wird für die verbleibenden Werte wiederholt und zwar mit dem kleinsten Mittelwert. Erst wenn der Grubbs-Test für einen Ausreißer keine Ausreißer mehr anzeigt, wird der beschriebene Test für zwei Ausreißer angewendet.

Grundsätzlich gilt: die Aussagen von Ausreißer-Tests sollten sorgfältig geprüft werden, da es vorkommen kann, dass ein Test einen signifikanten Ausreißer nicht entdeckt [78, 79].

4.3.2.3 Berechnung des Gesamtmittelwertes und der Varianzen

Für jedes Merkmalsniveau j wird aus p Laboratoriumsmittelwerten der Gesamtmittelwert berechnet [54]:

$$\hat{m}_j = \bar{x}_j = \frac{\sum_{i=1}^p n_{ij} \bar{x}_{ij}}{\sum_{i=1}^p n_{ij}} \quad (4.3.20).$$

Für jedes Merkmalsniveau j werden drei Varianzen berechnet:

die Wiederholvarianz:

$$s_{rj}^2 = \frac{\sum_{i=1}^p (n_{ij} - 1) s_{ij}^2}{\sum_{i=1}^p (n_{ij} - 1)} \quad (4.3.21),$$

die Varianz zwischen den Laboratorien:

$$s_{Lj}^2 = \frac{s_{dj}^2 - s_{rj}^2}{n_j} \quad (4.3.22),$$

mit

$$s_{dj}^2 = \frac{1}{p-1} \cdot \sum_{i=1}^p n_{ij} (\bar{x}_{ij} - \bar{x}_j)^2 \quad (4.3.23)$$

und

$$s_{Lj}^2 = \frac{1}{n_j} \cdot \left[\sum_{i=1}^p n_{ij} - \frac{\sum_{i=1}^p n_{ij}^2}{\sum_{i=1}^p n_{ij}} \right] \quad (4.3.24).$$

Ergibt sich bei diesen Berechnungen zufallsbedingt für s_{Lj}^2 ein negativer Wert, sollte dieser mit dem Wert Null angenommen werden. Die Vergleichsvarianz berechnet sich aus

$$s_{Rj}^2 = s_{rj}^2 + s_{Lj}^2 \quad (4.3.25).$$

4.3.3 Robuste Datenauswertung nach ISO/DIS 13528

Wie ANSCOMBE [80] ca. 1960 feststellte, ist die wahre Verteilung der untersuchten Daten in der Regel unbekannt. Bei Abweichung von der angenommenen Verteilung (meist Normalverteilung) ist die Wirksamkeit der meisten klassischen statistischen Tests oder Schätzer jedoch mehr oder weniger instabil [49]. Die so genannten robusten Verfahren lassen Abweichungen von der idealen Verteilung zu um den Preis einer verringerten Effektivität. Wendet man den U-Test von Mann und Whitney, der bei nicht gegebener Normalverteilung oder beim Vorliegen ordinalskaliertter Daten den t-Test nach Student ersetzt, auf

normalverteilte Werte an, so besitzt er eine Effizienz von 95% des t-Tests, bei großen Fallzahlen von 95,5% [69]. Betrachtet man ein robustes Verfahren als eine Art Police, um sich gegen die Unwägbarkeiten durch die Abweichung von der Normalverteilung zu versichern, muss man sich im Vorfeld entscheiden, wie viel Effizienz zu opfern man bereit ist und wie groß die Abweichungen sind, gegen die man sich versichern will [49, 75].

Werden robuste statistische Methoden zur Auswertung einer Eignungsprüfung herangezogen, besteht das Ziel darin, Werte für die Standardabweichung der Wiederhol- und Vergleichspräzision zu berechnen, die nicht von Ausreißern beeinflusst werden. Wird die Eignungsprüfung nach einer robusten Methode ausgewertet, sollten die Ausreißertests und Übereinstimmungsprüfungen durchgeführt und ausgewertet werden, die zu untersuchenden Daten bleiben jedoch unverändert. ISO/DIS 13528 [73] greift auf die in DIN ISO 5725-5 [81] beschriebenen Algorithmen A und S zurück.

Algorithmus A

Eine Menge von p Daten wird in steigender Reihenfolge sortiert: $x_1, x_2, \dots, x_b, \dots, x_p$.

Der robuste Mittelwert dieser Daten wird als x^* bezeichnet, die robuste Standardabweichung mit s^* . Die Startwerte von x^* und s^* ($i=1, 2, \dots, p$) werden wie folgt berechnet:

$$x^* = \text{Median } x_i \quad (4.3.26),$$

$$s^* = 1,483 \text{ Median } |x_i - x^*| \quad (4.3.27).$$

Zur Verbesserung der Werte von x^* und s^* berechnet man einen Faktor ϕ

$$\phi = 1,5 s^* \quad (4.3.28).$$

Für jedes x_i ($i=1, 2, \dots, p$) berechnet man

$$x_i^* = \begin{cases} x^* - \phi & \text{wenn } x_i < x^* - \phi \\ x^* + \phi & \text{wenn } x_i > x^* + \phi \\ x_i & \text{übrige Werte} \end{cases} \quad (4.3.29)$$

Anschließend werden die neuen Werte für x^* und s^* berechnet:

$$x^* = \frac{\sum x_i^*}{p} \quad (4.3.30),$$

$$s^* = 1,134 \cdot \sqrt{\frac{\sum (x_i^* - x^*)^2}{p-1}} \quad (4.3.31).$$

Die Summation erfolgt über i .

Die Werte für x^* und s^* können iterativ ermittelt werden, d.h. die Verbesserung für x^* und s^* wird solange wiederholt, bis der Prozess konvergiert. Konvergenz ist dann erreicht, wenn die dritte signifikante Stelle von einem Iterationsschritt zum nächsten unverändert bleibt.

Algorithmus S

Dieser Algorithmus wird auf die Standardabweichungen oder Variationsbreiten (engl. ranges) innerhalb der Laboratorien angewandt. Eine Menge von p Daten wird in steigender Reihenfolge sortiert: $w_1, w_2, \dots, w_i, \dots, w_p$. Diese Daten können Standardabweichungen oder Variationsbreiten sein. Der robuste gemeinsame Wert dieser Daten wird als w^* bezeichnet, die zu jedem w_i gehörigen Freiheitsgrade werden mit v bezeichnet. Ist w_i eine Variationsbreite, so ist $v=1$. Wenn w_i eine Standardabweichung ist, dann ergibt sich $v=n-1$. Die Werte für ξ und η sind in ISO/DIS 13528, Tabelle 19 [73] aufgeführt und werden in DIN ISO 5725-5 [81] hergeleitet.

Der Startwert von w^* wird wie folgt ermittelt:

$$w^* = \text{Median } w_i \quad (i = 1, 2, \dots, p) \quad (4.3.32).$$

Zur Verbesserung von w^* berechnet man:

$$\psi = \eta w^* \quad (4.3.33).$$

Für jedes w_i ($i=1, 2, \dots, p$) berechnet man

$$w_i^* = \begin{cases} \psi & \text{wenn } w_i < \psi \\ w_i & \text{sonstiges} \end{cases} \quad (4.3.34).$$

Abschließend wird der neue Wert für w^* berechnet aus:

$$w^* = \xi \cdot \sqrt{\frac{\sum (w_i^*)^2}{p}} \quad (4.3.35).$$

Der Wert für w^* kann iterativ ermittelt werden, d.h. die Verbesserung für w^* wird solange wiederholt, bis der Prozess konvergiert. Konvergenz ist dann erreicht, wenn die dritte signifikante Stelle von einem Iterationsschritt zum nächsten unverändert bleibt.

4.4 Statistische Leistungsbewertung von Eignungsprüfungen

Ergebnisse aus Eignungsprüfungen müssen oft in eine Leistungsstatistik transformiert werden, um den Vergleich mit festgelegten Zielen zu ermöglichen. Die Abweichung vom Stoffwert soll auf eine Weise ermittelt und dargestellt werden, die einen Vergleich mit den Leistungskriterien zulässt. Die notwendigen Maßnahmen können von „keine Aufarbeitung erforderlich“ bis zu „komplexe statistische Transformation erforderlich“ reichen. Die Leistungsgrößen müssen für alle Teilnehmer der Eignungsprüfung aussagekräftig sein. Daher sollten sich diese Leistungsgrößen auf die angewandte Prüfung beziehen, üblich in der Anwendung und gut verständlich sein.

Zur Berechnung von Leistungsstatistiken werden häufig Variabilitätsgrößen verwendet, wie z. B. die Standardabweichung, der Variationskoeffizient oder relative Standardabweichung, Perzentile, der Medianwert der absoluten Abweichung oder andere bestätigte Maße [10].

4.4.1 Berechnung der Standardabweichung $\hat{\sigma}$ zur Beurteilung der Leistung

In ISO/DIS 13528:2002 (E) [73] werden sechs Methoden zur Bestimmung der Standardabweichung der Eignungsprüfung $\hat{\sigma}$ vorgeschlagen.

Die einfachste Methode zur Bestimmung der Standardabweichung $\hat{\sigma}$ ist die Verwendung eines vorgeschriebenen Wertes. Dieser Wert kann z. B. eine gesetzlich vorgeschriebene Größe sein oder ein Wert, der für die Bearbeitung einer bestimmten Aufgabe unbedingt eingehalten werden muss. Diese Methode hat den Vorteil, direkt zu einer Aussage über die Gebrauchstauglichkeit der Messmethode zu führen.

Eine weitere Methode ist die Festlegung von $\hat{\sigma}$ auf einen Wert, der ein Leistungsniveau widerspiegelt, das vom Beratungsausschuss, der Akkreditierungsstelle oder von Laborkunden als notwendig für die teilnehmenden Laboratorien erachtet wird. Auch hier wird eine direkte Aussage zur Gebrauchstauglichkeit der Messmethode erreicht.

Bei diesen beiden Vorgehensweisen zur Bestimmung von $\hat{\sigma}$ besteht die Gefahr, einen Wert für $\hat{\sigma}$ festzulegen, der im Verhältnis zur Wiederholbarkeit der Messmethode unrealistisch ist. Vorausgesetzt es sind Informationen zur Wiederholbarkeit und Genauigkeit eines Messverfahrens bekannt, kann man mittels der folgenden Rechenschritte abschätzen, ob der Wert für $\hat{\sigma}$ realistisch ist.

Bekannt sind

σ_R = die Vergleichsstandardabweichung

σ_r = die Wiederholstandardabweichung.

Damit berechnet man die Standardabweichung zwischen den Laboratorien σ_L aus

$$\sigma_L = \sqrt{\sigma_R^2 - \sigma_r^2} \quad (4.4.1)$$

und ermittelt anschließend den Wert ϕ nach Gleichung 4.4.2, indem man den gewählten Wert für $\hat{\sigma}$ und die Werte für σ_L und σ_r einsetzt:

$$\hat{\sigma} = \sqrt{(\phi \times \sigma_L)^2 + (\sigma_r^2/n)} \quad (4.4.2)$$

mit n = Anzahl der Wiederholungsmessungen.

Wenn der für ϕ gefundene Wert klein ist (d. h. $\phi < 0,5$), so liegt der gewählte Wert für $\hat{\sigma}$ außerhalb der Vergleichspräzision, welche die Laboratorien normalerweise in der Praxis erzielen können.

Der Wert für die Standardabweichung der Eignungsprüfung $\hat{\sigma}$ kann auch aus einem generellen Modell für die Vergleichspräzision abgeleitet werden. Als Beispiel sei hier die Horwitz-Kurve [41, 82] genannt. Horwitz gibt ein Modell für die Vergleichspräzision an, aus dem sich die Standardabweichung der Vergleichspräzision aus dem Ausdruck

$$\sigma_R = 0,02 c^{0,8495} \quad (4.4.3)$$

ergibt, mit c als Konzentration des Analyten in %.

In der mechanisch-technologischen Werkstoffprüfung gibt es diese Art von Modellen bislang nicht.

Der Wert für die Standardabweichung der Eignungsprüfung $\hat{\sigma}$ kann auf Grundlage der Resultate eines (früheren) Ringversuches zur Bestimmung der Genauigkeit eines Messverfahrens (engl. precision experiment) gewonnen werden. Ein standardisiertes Messverfahren und bekannte Werte für die Wiederhol- und Vergleichspräzision vorausgesetzt, lässt sich $\hat{\sigma}$ wie folgt berechnen.

Bekannt sind: σ_R = die Vergleichsstandardabweichung

σ_r = die Wiederholstandardabweichung.

Die Standardabweichung zwischen den Laboratorien σ_L lässt sich aus Gleichung (4.4.1) berechnen. Anschließend kann man $\hat{\sigma}$ berechnen mit

$$\hat{\sigma} = \sqrt{\sigma_L^2 + (\sigma_r^2/n)} \quad (4.4.4)$$

mit n = Anzahl der Wiederholungsmessungen.

Die Bestimmung der Standardabweichung der Eignungsprüfung $\hat{\sigma}$ auf Grundlage der Daten der Teilnehmer aus der gleichen Runde einer laufenden Eignungsprüfung setzt eine ausreichende Anzahl von Teilnehmern in jeder Runde voraus. Verwendet wird die robuste

Standardabweichung, die z. B. nach Gleichung (4.3.26) bis (4.3.31)^C aus den Ergebnissen der Teilnehmer berechnet wird.

Zur Überprüfung der Leistung der Laboratorien und zur Messung der positiven Effekte eines Eignungsprüfungsschemas^D kann man folgende Routine durchführen:

Aus den Ergebnissen jeder Runde eines Eignungsprüfungsschemas werden die Werte für die Wiederhol- und Vergleichspräzision mittels einer robusten Methode, z. B. nach DIN ISO 5725-5 [81], berechnet. Diese Werte werden grafisch als Zeitreihe dargestellt, wenn möglich gemeinsam mit den in einem Ringversuch zur Bestimmung der Genauigkeit eines Messverfahrens gewonnenen Werte. Diese Darstellungen werden ausgewertet. Weichen die Werte aus dem Eignungsprüfungsschema um den Faktor 2 oder mehr von den Werten aus dem Ringversuch zur Bestimmung der Genauigkeit eines Messverfahrens ab, ist es Aufgabe des Koordinators, die Gründe hierfür zu finden. Wenn die Darstellung keine Verbesserung der Genauigkeit mit der Zeit zeigt, so ist dies ein deutliches Zeichen für die Wirksamkeit des Eignungsprüfungsschemas. Gründe, weshalb ein Eignungsprüfungsschema keine Verbesserung der Präzision mehr zeigt, sind:

- Die teilnehmenden Laboratorien untersuchen die Ursachen für das Warnsignal „fragwürdiges Ergebnis“ oder das Aktionssignal „unzureichendes Ergebnis“ nicht oder die eingeleiteten Korrekturmaßnahmen sind ausreichend;
- Die teilnehmenden Laboratorien sind nicht in der Lage, die Ursachen für die Warn- bzw. Aktionssignale zu identifizieren;
- Die Messmethode befindet sich in einem statistisch stabilen Zustand und liefert zuverlässige Messergebnisse.

^C Zur Berechnung der robusten Standardabweichungen können auch andere Verfahren angewandt werden, solange sie auf einer gesunden statistischen Basis beruhen und im Abschlussbericht erläutert werden.

^D Unter einem Eignungsprüfungsschema versteht man eine Abfolge von Eignungsprüfungen für ein Prüfverfahren.

4.4.2 Berechnung der Leistungsstatistik

Eine Berechnung der Leistungsstatistik ist bei qualitativen Ergebnissen in der Regel nicht erforderlich [10]. Im Folgenden werden die bei quantitativen Ergebnissen häufig verwendeten Bewertungsmethoden in einer Reihenfolge zunehmenden Transformationsgrades vorgestellt [73].

1) Schätzung der systematischen Abweichungskomponente des Laboratoriums

Der von einem teilnehmenden Laboratorium gelieferte Messwert (oder der Mittelwert der Messwerte) sei mit x bezeichnet. Es errechnet sich die Schätzung der systematischen Abweichungskomponente D eines Laboratoriums nach der Gleichung

$$D = x - X \quad (4.4.5),$$

wobei X der Stoffwert ist.

Um das Vorzeichen der systematischen Abweichungskomponente D nicht zu verschleiern, sollten auf $|D|$ oder D^2 beruhende Leistungsstatistiken vermieden werden.

Wenn das Resultat eines Laboratoriums eine Schätzung der systematischen Abweichungskomponente D ergibt, für die $D \geq 3,0\hat{\sigma}$ oder $D \leq -3,0\hat{\sigma}$ ist, wird das Ergebnis mit „unzureichend“ bewertet und das Laboratorium erhält ein Aktionssignal. Ist D größer als $2,0\hat{\sigma}$ oder kleiner als $-2,0\hat{\sigma}$ wird das Ergebnis als „fragwürdig“ bezeichnet und das Laboratorium erhält ein Warnsignal. Ein unzureichendes Ergebnis in einer Eignungsprüfung oder zwei fragwürdige Ergebnisse in zwei Eignungsprüfungen nacheinander (zum selben Prüfverfahren) gelten als Beweis für das Auftreten einer Anomalie, die eine nähere Untersuchung erfordert. Im Sinne der Warn- bzw. Aktionssignale wirkt dieses Kriterium analog zum z-Score, vgl. Gleichung (4.4.7).

Die Begründung für die Multiplikatoren 2,0 und 3,0 ist folgende: Wenn X und $\hat{\sigma}$ gute Schätzer des Mittelwertes bzw. der Standardabweichung der Verteilung sind, aus der x entnommen wurde, dann sind die Werte von D näherungsweise normalverteilt mit einem Mittelwert von Null und einer Standardabweichung von $\hat{\sigma}$. Dann liegen nur ca. 0,3% der Laborabweichungen außerhalb des Intervalls $-3,0\hat{\sigma} \geq D \geq 3,0\hat{\sigma}$ und nur ca. 5% der Laborabweichungen sind außerhalb des Intervalls $-2,0\hat{\sigma} < D < 2,0\hat{\sigma}$ zu erwarten. Aufgrund dieser niedrigen Wahrscheinlichkeiten geht man von einer guten Chance aus, die Ursache für Anomalie aufgrund eines Aktionssignals auffindig zu machen.

Wird die Standardabweichung für die Eignungsprüfung $\hat{\sigma}$ nach den Methoden „vorgeschriebener Wert“ oder „Festlegung“ (vgl. Kap. 4.3.1) ermittelt, kann $\hat{\sigma}$ signifikant

von der Standardabweichung der Vergleichspräzision abweichen. In diesem Falle treffen die Wahrscheinlichkeiten von 0,3% und 5% nicht mehr zu.

II) Prozentuale Differenz

Mit der Legende zu Gleichung (4.4.5) wird die prozentuale Differenz $D_{\%}$ nach folgender Gleichung berechnet:

$$D_{\%} = \frac{x-X}{X} \cdot 100\% \quad (4.4.6).$$

Die Interpretation der Ergebnisse ist dieselbe wie bei Methode I):

Wenn das Resultat eines Laboratoriums eine systematische Abweichungskomponente $D_{\%}$ ergibt, $D_{\%} \geq 3,0\% \hat{\sigma} / X$ oder $D_{\%} \leq -3,0\% \hat{\sigma} / X$, wird das Ergebnis mit „unzureichend“ bewertet und das Laboratorium erhält ein Aktionssignal. Ist $D_{\%}$ größer als $2,0\% \hat{\sigma} / X$ oder kleiner als $-2,0\% \hat{\sigma} / X$, wird das Ergebnis als „fragwürdig“ bezeichnet und das Laboratorium erhält ein Warnsignal. Ein unzureichendes Ergebnis in einer Eignungsprüfung oder zwei fragwürdige Ergebnisse in zwei Eignungsprüfungen nacheinander (zum selben Prüfverfahren) gelten als Beweis für das Auftreten einer Anomalie, die eine nähere Untersuchung erfordert.

III) Rangzahl oder prozentuale Rangzahlen

Nehmen an einer Eignungsprüfung p Laboratorien teil, erhält das Laboratorium mit dem niedrigsten Resultat die Rangzahl 1, das Laboratorium mit dem nächst höheren Resultat die Rangzahl 2 bis hin zum Laboratorium mit dem höchsten Resultat, das die Rangzahl p erhält. Berichten zwei oder mehr Laboratorium das gleiche Ergebnis, so erhalten sie alle den Mittelwert der Rangzahl. Wenn in der Eignungsprüfung verschiedene Parameter gemessen werden, wird für jeden eine eigene Rangliste erstellt. Bezeichnet man die Rangzahlen mit $i=1,2,..,p$, berechnen sich die prozentualen Rangzahlen aus $100x(i-1/2)/p$.

Der Interpretation von Rangzahlen oder prozentualen Rangzahlen liegt keine Annahme zu Grunde, die Daten folgten einer bestimmten Verteilung. Ihre Entstehung lässt den Stoffwert oder die Standardabweichung der Eignungsprüfung für die Beurteilung der Eignungsprüfung außer Betracht. Diese Methoden verschaffen einen schnellen Überblick, welche Laboratorien die Extremresultate liefern. Ihr größter Nutzen liegt in den ersten Runden eines Eignungsprüfungsschemas, um die Laboratorien mit dem größten Verbesserungspotenzial zu identifizieren.

ISO/IEC Guide 43-1 [10] warnt davor, die Berichterstattung über die Laborleistungen bei einer Eignungsprüfung in Form einer Rangtabelle vorzunehmen, da diese irreführend sein und zu Fehldeutungen führen kann.

Im Rahmen dieser Arbeit findet dieses Verfahren keine Anwendung.

IV) Z-Scores

Unter Verwendung der Legende zu Gleichung (4.4.4) wird der z-Score z („z-Note“) nach folgender Gleichung berechnet:

$$z = \frac{x - X}{\hat{\sigma}} \quad (4.4.7).$$

$\hat{\sigma}$ ist die Standardabweichung der Eignungsprüfung.

Wenn das Resultat eines Laboratoriums einen z-Score z ergibt $z \geq 3,0$ oder $z \leq -3,0$, wird das Ergebnis mit „unzureichend“ bewertet, und das Laboratorium erhält ein Aktionssignal. Ist z größer als 2,0 oder kleiner als -2,0 wird das Ergebnis als „fragwürdig“ bezeichnet und das Laboratorium erhält ein Warnsignal. Ein unzureichendes Ergebnis in einer Eignungsprüfung oder zwei fragwürdige Ergebnisse in zwei Eignungsprüfungen nacheinander (zum selben Prüfverfahren) gelten als Beweis für das Auftreten einer Anomalie, die eine nähere Untersuchung erfordert.

Ist die Anzahl der teilnehmenden Laboratorien sehr groß (> 100 Laboratorien), besteht bei Verwendung der Grenzwerte 2,0 bzw. 3,0 die Gefahr, unnötigen Aufwands bei der Untersuchung aller Warn- und Aktionssignale.

Die Gefahr gar keine Signale zu bekommen besteht dann, wenn die Anzahl der teilnehmenden Laboratorien sehr klein (< 10 Laboratorien) ist. In diesem Fall empfiehlt [73] die Auswertung grafischer Aufzeichnungen, die einen Überblick über die z-Scores aus mehreren Runden von Eignungsprüfungen über ein Prüfverfahren geben.

V) E_n -Zahlen

E_n -Zahlen werden üblicherweise bei Messvergleichsprogrammen verwendet. Sie werden berechnet nach

$$E_n = \frac{x - X}{\sqrt{U_{Lab}^2 + U_{ref}^2}} \quad (4.4.8).$$

Der Stoffwert X wird von einem Referenzlaboratorium ermittelt. U_{ref} ist die erweiterte Messunsicherheit des Stoffwertes X , U_{lab} bezeichnet die erweiterte Messunsicherheit des Messergebnis x , das ein Laboratorium gemeldet hat. Da bei der Berechnung von E_n -Zahlen die erweiterten Messunsicherheiten statt der Standardabweichungen im Nenner stehen, wird üblicherweise ein Wert von 1,0 als kritischer Wert verwendet. ISO/IEC Guide 43-1 [10] gibt an:

$|E_n| \leq 1$... Ergebnis ausreichend

$|E_n| \geq 1$... Ergebnis unzureichend.

Wenn die erweiterte Messunsicherheit mit einem Erweiterungsfaktor von 2,0, vgl. [73], berechnet wird, entspricht eine kritische E_n -Zahl von 1,0 einem kritischen Wert von 2,0 bei Verwendung von z-Scores.

Eine Auswertung mittels E_n -Zahlen sollte mit großer Vorsicht durchgeführt werden, wenn die Kompetenz der Teilnehmer bei Ermittlung der Messunsicherheit sehr gering ist oder es keine einheitliche Vorgehensweise gibt. Allerdings ist die Einbeziehung der Messunsicherheiten in die Auswertung einer Eignungsprüfung ein sehr wichtiges Mittel, um das Verständnis der Laboratorien für die schwierige Thematik „Messunsicherheit“ zu verbessern. Da für die Bestimmung der Messunsicherheit von mechanisch-technologischen Prüfverfahren zur Zeit in der Regel noch keine ausreichenden Daten vorliegen, ist eine Anwendung dieses oder der folgenden Verfahren VI) und VII) im Rahmen dieser Arbeit nicht möglich.

Es ist jedoch offensichtlich, dass eine Auswertung mit E_n -Zahlen nur dann sinnvoll ist, wenn im Rahmen der Prüfnorm oder im Rahmen der „Spielregeln“ der Eignungsprüfung maximal zulässige Werte für die Messunsicherheit U_{Lab} eines Verfahrens festgelegt werden müssen. Gleiches gilt für U_{ref} , da das Bestehen einer Eignungsprüfung ansonsten durch Angabe einer möglichst hohen Messunsicherheit, die ja im Nenner der Gleichung (4.4.8) steht, sichergestellt werden kann.

VI) z' -Scores

Unter Verwendung der Legende zu Gleichung (4.3.4) wird der z' -Score z' nach folgender Gleichung berechnet:

$$z' = \frac{x - X}{\sqrt{\hat{\sigma}^2 + u_x^2}} \quad (4.4.9).$$

u_x ist die Standardmessunsicherheit des Stoffwertes X .

Die z' -Scores werden genauso interpretiert wie die z-Scores (siehe Abschnitt IV) und verwenden die selben kritischen Werte von 2,0 und 3,0.

Anwendung des z' -Scores:

Wie ein Vergleich zwischen den Gleichungen (4.4.7) und (4.4.9) zeigt, sind die z' -Scores für eine Eignungsprüfung um den konstanten Faktor $\hat{\sigma} / \sqrt{\hat{\sigma}^2 + u_x^2}$ kleiner als die korrespondierenden z-Scores. Solange der Richtwert zur Eingrenzung der Messunsicherheit des Stoffwertes, Gleichung (4.2.1) eingehalten wird, nimmt dieser konstante Faktor Werte von

$$0,96 \leq \hat{\sigma} / \sqrt{\hat{\sigma}^2 + u_x^2} \leq 1,00 \quad (4.4.10)$$

an. Der Unterschied zwischen z-Scores und z' -Scores kann dann vernachlässigt werden.

Ist die Bedingung nach Gleichung (4.2.1) jedoch nicht erfüllt, wird die Differenz zwischen dem z-Score und z'-Score so groß, dass einige z-Scores bereits die kritischen Grenzen 2,0 und 3,0 überschreiten und so zu Warn- bzw. Aktionssignalen führen, während die entsprechenden z'-Scores noch unterhalb der kritischen Werte liegen. Bei der Entscheidung, ob z-Scores durch z'-Scores ersetzt werden sollen, gibt es folgende Aspekte zu bedenken:

- a) Erfüllt die Messunsicherheit des Stoffwertes die Gleichung (4.2.1)? Wenn ja, ist von den z'-Scores kein Zusatznutzen zu erwarten.
- b) Wird Gleichung (4.2.1) nicht erfüllt, stellt sich die Frage, ob der Vorteil der z'-Scores überwiegt gegenüber dem zusätzlichen Rechenaufwand und den Schwierigkeiten, die möglicherweise entstehen, wenn die z'-Scores den Teilnehmern erklärt werden.
- c) Wie scharf sind die Konsequenzen, die ein Laboratorium zu erwarten hat, wenn es ein Warn- oder Aktionssignal erhält? Werden die Ergebnisse der Eignungsprüfung dazu verwendet, Laboratorien von der Durchführung bestimmter Prüfaufgaben auszuschließen?

VII) ζ (zeta)-Scores

Zur Zeit ist es normalerweise nicht üblich, die Angaben der teilnehmenden Laboratorien über die Unsicherheit ihrer Messungen bei der Auswertung von Eignungsprüfungen zu berücksichtigen. Mittels ζ -Score gibt ISO/DIS 13528:2002 (E) [73] die Möglichkeit, diese Angaben bei einer Eignungsprüfung zu berücksichtigen, wenn sie vorhanden sind. Dies wird in Zukunft der Regelfall sein, da DIN EN ISO/IEC 17025 [7] mittlerweile von akkreditierten Laboratorien Angaben zur Messunsicherheit der angewandten Messverfahren verlangt.

Unter Verwendung der Legende zu Gleichung (4.3.4) wird der ζ -Score ζ nach folgender Gleichung berechnet:

$$\zeta = \frac{x - X}{\sqrt{u_x^2 + u_X^2}} \quad (4.4.11)$$

Dabei ist u_x die Schätzung des Laboratoriums für die Unsicherheit seines Messergebnisses x und u_X die Standardmessunsicherheit des Stoffwertes X .

Wenn ein effektives System zur Validierung der laboreigenen Schätzungen von Messunsicherheiten betrieben wird, kann man die ζ -Scores anstelle der z-Scores verwenden. Sie sollten dann auf dieselbe Art ausgelegt werden wie die z-Scores (vgl. Abschnitt IV) unter Verwendung derselben kritischen Werte 2,0 und 3,0. Gibt es ein solches System nicht, sollten ζ -Scores nur in Verbindung mit z-Scores verwendet werden und zwar als Hilfsmittel zur Verbesserung der Laborleistung.

VIII) Kombinierte Leistungsbewertung

Es ist üblich, in einer Eignungsprüfung Ergebnisse für mehrere zu prüfende Messgrößen oder Prüfgegenstände zu erhalten [23, 25, 58]. In diesem Fall wird für jede Messgröße oder jeden Prüfgegenstand eine eigene Bewertung wie in den Abschnitten I) bis VII) beschrieben durchgeführt. Eine Ausnahme sind Eignungsprüfungen, in denen zwei ähnliche Materialien verwendet wurden, mit der Absicht, diese als Youden-Paar zu verwenden. In diesem Fall werden die Ergebnisse in Form eines Youden-Plot (s. Kap. 4.4.3) dargestellt.

Darüber hinaus empfiehlt [73] die Verwendung grafischer Methoden (s. Kap. 4.4.3) zusätzlich zur rechnerischen Auswertung.

Es wird dringend davon abgeraten zur abschließenden Beurteilung einer Eignungsprüfung, die mehrere Messgrößen oder Prüfgegenstände umfasst, einen mittleren z-Score oder ähnliches zu bilden, da dieser das Gesamtergebnis unter Umständen vollständig verfälscht.

Die Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin empfiehlt, die jeweiligen Messwerte einzeln zu bewerten („bestanden“ – „nicht bestanden“) und den Ringversuch insgesamt als bestanden zu werten, wenn der Anteil der bestandenen Parameter mindestens 80% beträgt [83, 84]. Diese Vorgehensweise wird im Kapitel 5.5 am Beispiel der Eignungsprüfung Emissionsspektroskopie gezeigt.

4.4.3 Grafische Methoden zur Darstellung der Leistungsstatistik für mehrere Messungen

Die Leistungsbewertungen, die in jeder Runde einer Eignungsprüfung ermittelt werden, sollten durch grafische Darstellungen visualisiert werden. Diese grafischen Darstellungen sollten den Teilnehmern zugänglich gemacht werden, damit diese ihre eigene Leistungsfähigkeit im Verhältnis zu den anderen Laboratorien beurteilen können. Die Teilnehmer werden durch die Code-Nummern repräsentiert, damit jedes Laboratorium sein eigenes Ergebnis problemlos findet und die Vertraulichkeit gewahrt bleibt. Die grafischen Darstellungen sollten auch dem Organisationsausschuss zur Verfügung stehen, damit dieser die Wirksamkeit der Eignungsprüfung verfolgen und ggf. die Veränderung eines Leistungskriteriums veranlassen kann [73].

Histogramme

Bild 4.4.1 zeigt schematisch ein Beispiel für ein Histogramm, das die z-Scores einer Messung aus einer Eignungsprüfung enthält. In das Histogramm werden Geraden bei $\pm 2,0$ und $\pm 3,0$ eingetragen, um die Bewertungskriterien der Eignungsprüfung darzustellen. Die Klassenbreite sollte 0,3 bis 0,5 betragen und die Spannweite des Diagramms von $-6,0$ bis $6,0$. Aus Gründen der Übersichtlichkeit, kann es manchmal notwendig sein, die Spannweite zu verändern.

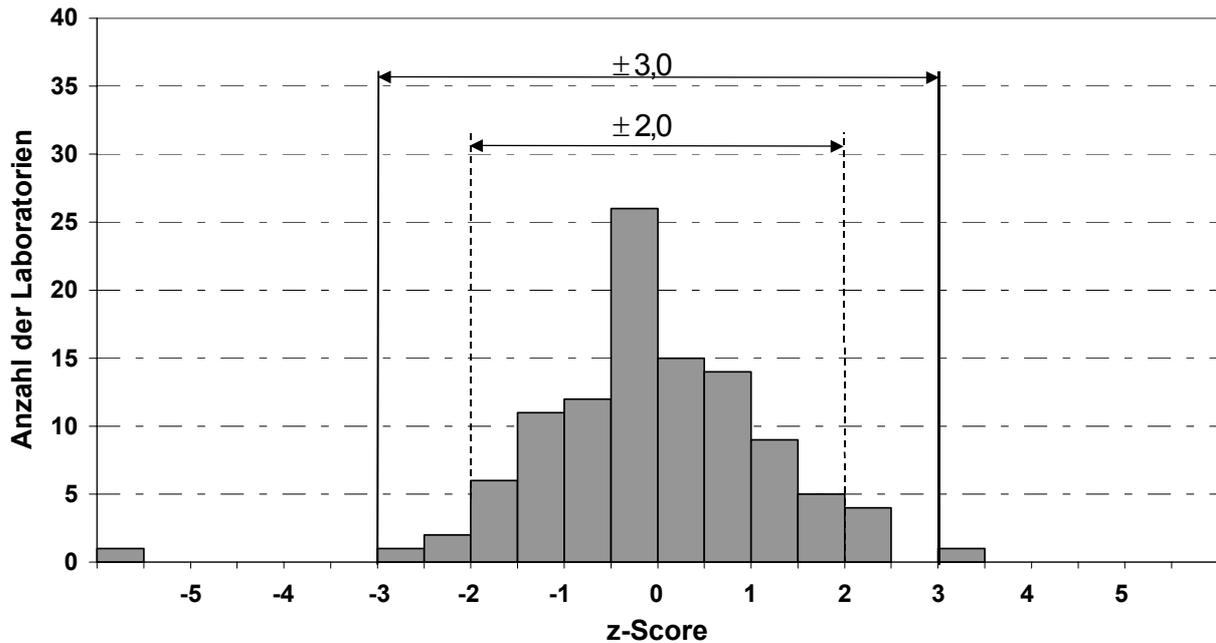


Bild 4.4.1: Histogramm; z-Scores

Liegt das Histogramm mit einigen wenigen Ausnahmen innerhalb der $\pm 2,0$ -Grenzen, so sollte die Bewertung der Eignungsprüfung verschärft werden, indem $\hat{\sigma}$ verkleinert wird. Wenn zu viele Laboratorien die $\pm 3,0$ -Grenzen überschreiten muss untersucht werden, ob dies nicht auf eine zu scharfe Bewertung der Eignungsprüfung zurückzuführen ist. In Tabelle 4.4.1 sind die Diagrammparameter für unterschiedliche Arten der Leistungsstatistik zusammengestellt:

Tabelle 4.4.1: Diagrammparameter für Histogramme von Leistungsbewertungen

Art der Leistungsstatistik	Klassenbreite	Warngrenze	Aktionsgrenze	Spannweite
Schätzung der systematischen Abweichungskomponente D	$0,3 \hat{\sigma}$ bis $0,5 \hat{\sigma}$	$\pm 2,0 \hat{\sigma}$	$\pm 3,0 \hat{\sigma}$	$\pm 6,0 \hat{\sigma}$
Prozentuale Differenz D%	$3\% \hat{\sigma} / X$ bis $5\% \hat{\sigma} / X$	$\pm 2\% \hat{\sigma} / X$	$\pm 3\% \hat{\sigma} / X$	$\pm 6\% \hat{\sigma} / X$
z-Score	0,3 bis 0,5	$\pm 2,0$	$\pm 3,0$	$\pm 6,0$

Säulendiagramm (engl. bar-plot)

Um ein solches Säulendiagramm zu erstellen, werden z. B. die z-Scores jedes Teilnehmers gemeinsam in einer Rubrik aufgetragen, siehe Bild 4.4.2. Säulendiagramme sind eine einfache Möglichkeit, die statistischen Kenngrößen für ähnliche Charakteristika in einem Diagramm gleichzeitig darzustellen. Bei dieser Darstellungsform kann man auf einen Blick erkennen, ob die Kenngrößen eines Laboratoriums gemeinsame Tendenzen zeigen, z. B. ob ein Teilnehmer außergewöhnlich viele hohe z-Scores hat, was generell auf ein schwaches Gesamtabschneiden des Laboratoriums schließen lässt [49, 73].

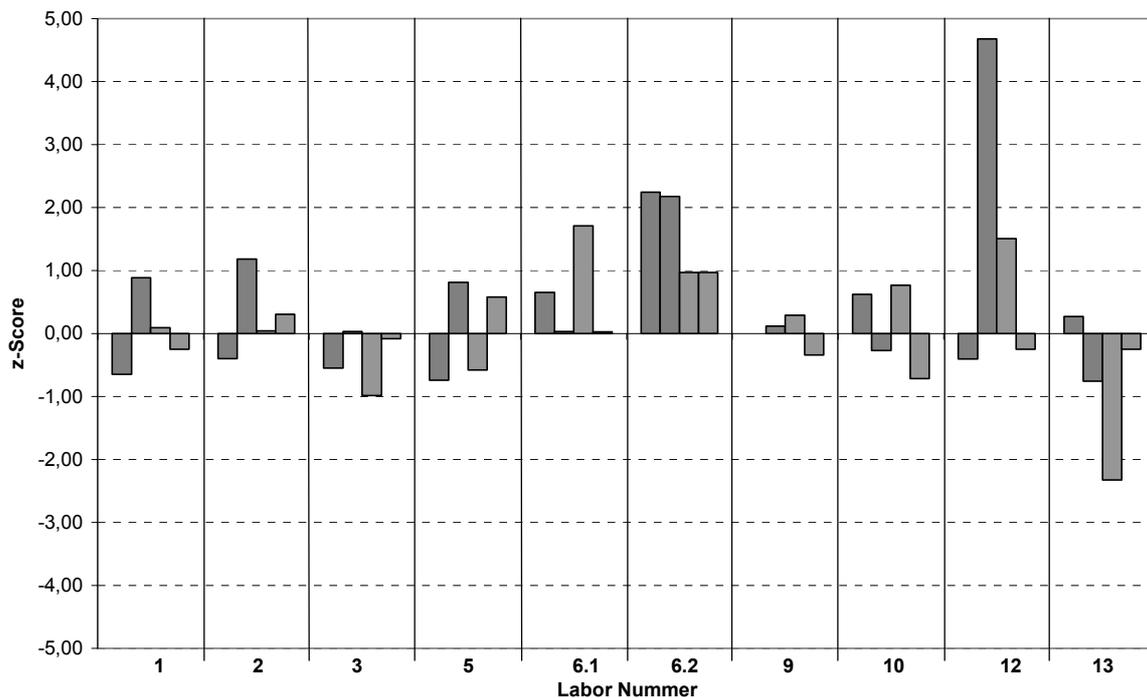


Bild 4.4.2: Säulendiagramm; Emissionsspektrometrie, z-Scores für vier analysierte Elemente, Daten aus [27]

Youden Diagramme

Werden in einer Runde einer Eignungsprüfung zwei ähnliche Materialien *A* und *B* geprüft, kann ein Youden Diagramm zur Beurteilung der Ergebnisse herangezogen werden. Die Konstruktion des Youden Diagramms erfolgt, indem man die z-Scores, die man am Material *A* erzielt hat, gegen die z-Scores aufträgt, die man am Material *B* erzielt hat. Eine Vertrauensellipse, deren Berechnung im folgenden beschrieben wird, erleichtert die Interpretation des Diagramms. Die Vorteile des Youden Diagramms sind die einfache Auswertung, schnelles Identifizieren systematischer Abweichungen und geringe Manipulations-

möglichkeiten. Es können jedoch keine Verfahrenskenndaten ermittelt werden, und man erhält erst sinnvolle Ergebnisse bei mehr als 30 Teilnehmern [49, 85].

Berechnung der Vertrauensellipse nach [73]

Die am Material A ermittelten Prüfergebnisse werden mit $x_{A1}, x_{A2}, \dots, x_{Ap}$ bezeichnet.

Die am Material B ermittelten Prüfergebnisse werden mit $x_{B1}, x_{B2}, \dots, x_{Bp}$ bezeichnet.

Die Anzahl der teilnehmenden Laboratorien ist p .

Zuerst werden die Mittelwerte \bar{x}_A, \bar{x}_B sowie die Standardabweichungen s_A, s_B der beiden Datensätze berechnet und anschließend der Korrelationskoeffizient $\hat{\rho}$ [86].

$$\hat{\rho} = \frac{\text{cov}(x_{A,i}, x_{B,i})}{\sqrt{\text{var}(x_{A,i}) \cdot \text{var}(x_{B,i})}} \quad (4.4.11).$$

Die z-Scores für die beiden Materialien berechnet man aus:

$$z_{A,i} = \frac{x_{A,i} - \bar{x}_A}{s_A} \quad i=1,2,\dots,p \quad (4.4.12),$$

$$z_{B,i} = \frac{x_{B,i} - \bar{x}_B}{s_B} \quad i=1,2,\dots,p \quad (4.4.13).$$

Der kombinierte z-Score für beide Materialien ergibt sich aus

$$z_{A,B,i} = \sqrt{z_{A,i}^2 - 2\hat{\rho}z_{A,i}z_{B,i} + z_{B,i}^2} \quad (4.4.14).$$

Es werden als standardisierte Variablen definiert:

$$z_A = \frac{x_A - \bar{x}_A}{s_A} \quad (4.4.15),$$

$$z_B = \frac{x_B - \bar{x}_B}{s_B} \quad (4.4.16).$$

Mit den standardisierten Variablen kann man die Vertrauensellipse darstellen als

$$z_A^2 - 2\hat{\rho}z_A z_B + z_B^2 = (1 - \hat{\rho}^2)T^2 \quad (4.4.17)$$

mit

$$T^2 = 2\{(p-1)/(p-2)\}F_{(1-\alpha);2;(p-1)} \quad (4.4.18).$$

$F_{(1-\alpha);2;(p-1)}$ ist das tabellierte $(1-\alpha)$ -Quantil der F-Verteilung [69] mit 2 und $(p-1)$ Freiheitsgraden. Die Vertrauensellipse kann in einem Diagramm, dessen Achsen die z-Scores z_A und z_B sind, konstruiert werden, indem eine Reihe von Punkten im Intervall $-T \leq z_A \leq T$ gezeichnet wird.

$$z_B = \hat{\rho} \cdot z_A \pm \sqrt{(1 - \hat{\rho}^2) \cdot (T^2 - z_A^2)} \quad (4.4.19)$$

Um die Vertrauensellipse in ein Diagramm einzutragen, dessen Achsen die Originaleinheiten der Messung tragen, werden die o. g. Punkte mit Hilfe der Gleichungen (4.4.20 bzw. 21) transformiert.

$$x_A = \bar{x}_A + s_A \cdot z_A \quad (4.4.20)$$

$$x_B = \bar{x}_B + s_B \cdot z_B \quad (4.4.21)$$

In Bild 4.4.3 zeigt ein Youden Diagramm, das unter Verwendung von Daten einer Eignungsprüfung „Zugversuch“ [87] entstanden ist.

Das Laboratorium außerhalb des 1%-Vertrauensbereiches hat ein Aktionssignal erhalten, das Laboratorium zwischen dem 1%- und dem 5%-Vertrauensbereich ein Warnsignal.

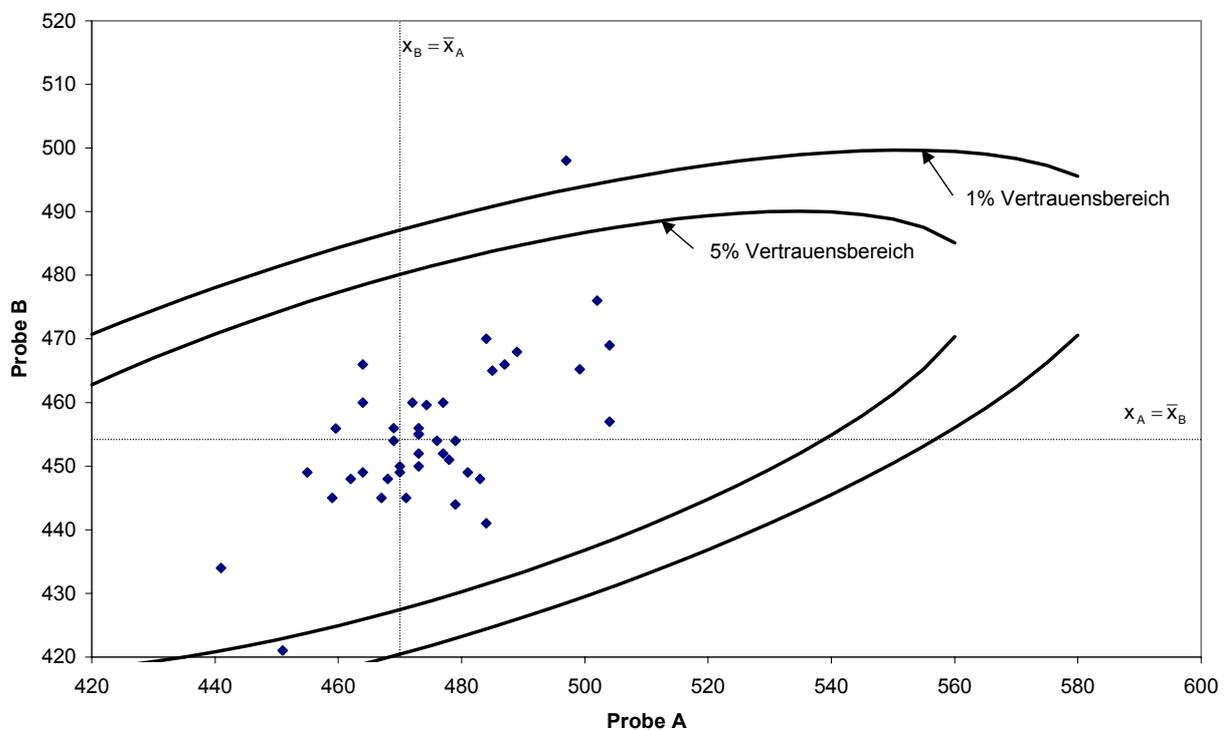


Bild 4.8: Beispiel für ein Youden Diagramm mit 1% und 5% Vertrauensellipse

Der kombinierte z-Score nach Gleichung (4.4.15) ist ein Hilfsmittel zur Interpretation des Youden Diagramms. Die höchsten kombinierten z-Scores bestimmen die Signifikanzstufen bei der Berechnung der Vertrauensellipse. Manchmal kann es notwendig sein, einen oder mehrere fern ab liegende Punkte zu eliminieren und die Ellipse erneut zu berechnen. Die kombinierten z-Scores helfen dann bei der Entdeckung der zu eliminierenden Punkte.

Bei der Interpretation eines Youden Diagramms sind folgende Punkte zu beachten [73, 49]:
 Direkte Auswertung: Sind die Punkte homogen um den Schnittpunkt der Geraden $y = \bar{x}_A$ und $x = \bar{x}_B$ (siehe Bild 4.8) verteilt, so liegen nur zufällige Fehler vor. Man untersucht das

Diagramm auf Punkte, die weit entfernt von den übrigen Daten liegen. Wendet ein Laboratorium das Prüfverfahren nicht richtig an, d. h. es liegt eine systematische Abweichung vor, so liegt der Punkt entlang der Hauptachse der Vertrauensellipse. Punkte, die weit entfernt von der Hauptachse liegen, werden durch eine schlechte Wiederholpräzision hervorgerufen.

Nach [57] eignet sich dieses Verfahren besser bei Präzisionsuntersuchungen. Im Rahmen dieser Arbeit findet der Youden-Plot keine Anwendung. Um eine fachgerechte Auswertung vornehmen zu können, ist es notwendig zwei geeignete Materialien auf unterschiedlichen Niveaus zu qualifizieren. Es entstehen höhere Kosten, denen jedoch kein technisch relevanter Zusatznutzen gegenübersteht.

5 Auswertung und Diskussion der durchgeführten Eignungsprüfungen

Die Grundlage für die Verwendbarkeit eines Stoffes als Werk- oder Baustoff bildet sein Festigkeitsverhalten. Darüber hinaus spielen jedoch auch andere Eigenschaften, wie z. B. die technologischen Eigenschaften (z. B. Formbarkeit, Schweißbarkeit), chemische Eigenschaften (z. B. Korrosionsbeständigkeit) oder physikalische Eigenschaften (z. B. elektrische Leitfähigkeit), eine wichtige Rolle bei der Werkstoffauswahl. Ein sicherer und wirtschaftlicher Einsatz der Werkstoffe ist nur dann gewährleistet, wenn die Eigenschaften genau bekannt sind. Daraus ergibt sich die Notwendigkeit, diese Eigenschaften durch Werkstoffprüfungen zu erproben und zu überwachen. Ziel der Werkstoffprüfung wird es stets bleiben, die Eigenschaften der Werkstoffe durch Ermittlung der maßgebenden Stoffwerte vollständig zu beschreiben [20]. Um diese Aufgabe zu erfüllen, wurden seit etwa 1860 eine Reihe von Prüfverfahren entwickelt [88].

Im Rahmen dieser Arbeit wurde untersucht, für welche mechanisch-technologischen Prüfverfahren die meisten Laboratorien akkreditiert sind. Durch Befragung von Laboratorien und Akkreditierungsstellen wurden die folgenden Prüfverfahren der mechanisch-technologischen Werkstoffprüfung ermittelt:

- der Zugversuch;
- die Härteprüfung nach Vickers, Brinell oder Rockwell;
- der Kerbschlagbiegeversuch nach Charpy;
- Metallographische Untersuchungen^E.

In vielen Werkstoffprüflaboratorien wird außerdem die Prüfung der chemischen Zusammensetzung von metallischen Werkstoffen mittels Emissionsspektrometrie durchgeführt. In allen im Rahmen dieser Arbeit durchgeführten Eignungsprüfungen wurden Prüfverfahren für metallische Werkstoffe angewandt.

Bis 1996 gab es nur wenige Eignungsprüfungen für mechanisch-technologische Werkstoffprüfungen, die nur einem ausgewählten Teilnehmerkreis zugänglich waren [21], oder wissenschaftlich wenig belastbare Voruntersuchungen.

^E Metallographische Untersuchungen zählen streng genommen nicht zur mechanisch-technologischen Werkstoffprüfung, sondern stellen ein eigenständiges Prüfgebiet dar. Sie sind jedoch aus Anwendersicht untrennbar mit der MTP verbunden.

Die Auswertung der als Vorversuch konzipierten Eignungsprüfung „Zugversuch nach DIN EN 1002-1, RZ DAP-1“ [23] zeigte die Notwendigkeit, die aus der Analytik bekannten Methoden zur Bewertung von Eignungsprüfungen systematisch auf ihre Anwendbarkeit hin zu untersuchen.

Im Rahmen dieser wissenschaftlichen Untersuchung werden systematisch Methoden untersucht und bewertet, die eine belastbare und abgesicherte Beurteilung der Kompetenz von Prüflaboratorien zur Durchführung von mechanisch-technologischen Prüfverfahren ermöglichen. Ebenfalls Bestandteil dieser Arbeit sind Untersuchungen, in wie weit Parameter, die von den Vorgaben der Prüfnormen nicht erfasst werden, bei der Bewertung von Eignungsprüfungen berücksichtigt werden müssen. Auf dieser Grundlage wurden für die wichtigsten mechanisch-technologischen Prüfverfahren Eignungsprüfungen konzipiert und durchgeführt. Die zeitliche Abfolge des Prüfprogramms wurde folgendermaßen festgelegt:

- 1999 Eignungsprüfung Härteprüfung nach Vickers [25]
- 2000 Eignungsprüfung Metallographie [26]
- 2001 Eignungsprüfung Emissionsspektrometrie [27]
Eignungsprüfung Härteprüfung nach Brinell [28]
Eignungsprüfung Kerbschlagbiegeversuch [29]

An diesen Eignungsprüfungen nahmen jeweils zwischen 70 und 115 Laboratorien aus bis zu 24 europäischen und außereuropäischen Ländern teil. Zwischen 70% und 90% der Teilnehmer sind entweder nach DIN EN ISO/IEC 17025 [7] oder nach DIN EN 45001 [6] akkreditiert. Etwa 50 % der Teilnehmer unterhält ein Qualitätsmanagementsystem nach DIN EN ISO 9001 [91], während Qualitätsmanagementsysteme z. B. nach VDA 6.1 [92] nur von einem sehr geringen Anteil (<10%) der Teilnehmer installiert worden sind. Zwei Drittel der Teilnehmer sind entweder freie Laboratorien oder Industrielaboratorien, das restliche Drittel verteilt sich auf staatliche Institutionen und Abnahmeorganisationen.

Die Ergebnisse der Eignungsprüfungen werden von einer Reihe europäischer Akkreditierungsgesellschaften anerkannt und im Akkreditierungsverfahren, wie im Kapitel 3.2 beschrieben, verwendet.

5.1 Eignungsprüfung Zugversuch

Der Zugversuch ist in der DIN EN 10002-1 [93] europaweit genormt. Der Versuch besteht darin, eine Probe durch eine Zugbeanspruchung zu dehnen, im Allgemeinen bis zum Bruch. Dabei werden die an der Probe angreifende Kraft und die Verlängerung der Probe gemessen. In der industriellen Prüfpraxis^F werden üblicherweise einer oder mehrere der folgenden Verformungs- bzw. Festigkeitskennwerte bestimmt:

Die *Bruchdehnung* A ist die bleibende Verlängerung der Messlänge nach dem Bruch, d. h. die Messlänge nach dem Bruch wird vermindert um die Ausgangsmesslänge und auf die Ausgangsmesslänge bezogen. Sie wird in Prozent angegeben.

Die *Brucheinschnürung* Z ist die größte Änderung des Probenquerschnitts nach dem Bruch, d. h. der Anfangsquerschnitt wird vermindert um den Querschnitt nach dem Bruch und auf den Anfangsquerschnitt bezogen. Sie wird in Prozent angegeben.

Die *0,2 %-Dehngrenze* $R_{p0,2}$ in MPa und/oder die *1,0 %-Dehngrenze* $R_{p1,0}$ in MPa: dies ist die Spannung σ , bei der $\epsilon_p=0,2\%$ bzw. $\epsilon_p=1,0\%$ beträgt, siehe Bild 5.1.1.

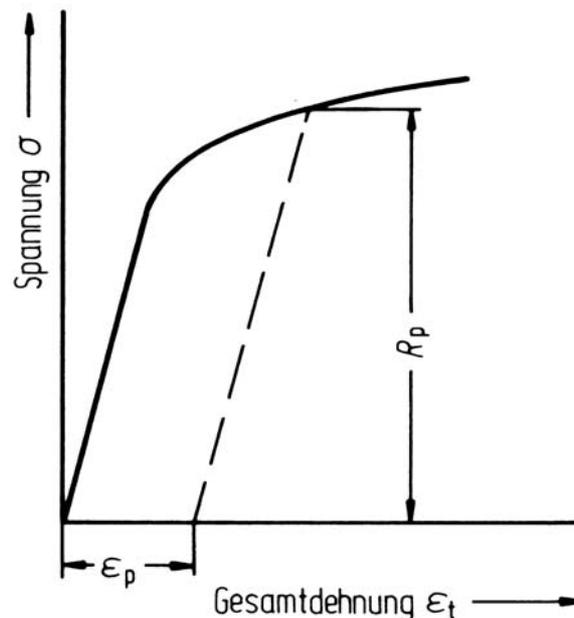


Bild 5.1.1: Definition von Dehngrenzen R_p nach [100]

Die *obere Streckgrenze* R_{eH} in MPa: das ist die Spannung σ in dem Moment, bei dem der erste deutliche Kraftabfall auftritt. Die *untere Streckgrenze* R_{eL} in MPa ist die kleinste Spannung im Fließbereich, wobei Einschwingvorgänge nicht berücksichtigt werden. Die

^F In der industriellen Prüfpraxis dient der Zugversuch zur Ermittlung bzw. Überprüfung der in den entsprechenden Werkstoffnormen vorgegebenen mechanischen Eigenschaften von Erzeugnissen, z. B. bei Abnahmeprüfungen oder Schadenuntersuchungen.

Zugfestigkeit R_m in MPa ist die Spannung, die sich aus der Höchstkraft bezogen auf den Anfangsquerschnitt ergibt [93], siehe Bild 5.1.2.

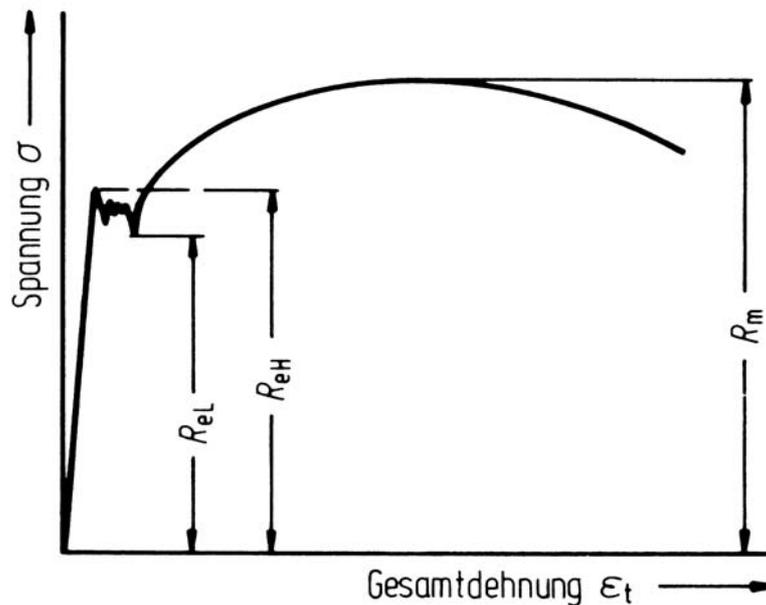


Bild 5.1.2: Spannungs-Dehnungs-Kurve eines Werkstoffes mit ausgeprägter Streckgrenze nach [100]

Eignungsprüfung „Zugversuch nach DIN EN 10002-1“, 1996

Die Eignungsprüfung „Zugversuch nach DIN EN 10002-1, RZ DAP-1“ [23] war als Vorversuch konzipiert, um festzustellen, in wie weit die aus der chemischen Analytik bekannten Auswerteverfahren auf einen Versuch der mechanisch-technologischen Werkstoffprüfung angewandt werden können.

Der Versuch wurde mit Stangenmaterial aus dem Werkstoff 115 CrV3 (Werkstoff-Nr. 1.2210) durchgeführt, an dem im Vorfeld umfangreiche Untersuchungen zur Werkstoffhomogenität vorgenommen wurden. Die 91 teilnehmenden Laboratorien ermittelten die in Tabelle 5.1.1 zusammengestellten Kennwerte. Die Berechnung des Stoffwertes X und der Standardabweichung der Eignungsprüfung $\hat{\sigma}$ erfolgte als Konsenswert aus den Teilnehmerergebnissen nach DIN ISO 5725 [54], die Bewertung der Laborleistung wurde mit z-Scores (s. Kap. 4.4) vorgenommen.

Die Auswertung der Eignungsprüfung bringt folgende grundsätzlichen Probleme zu Tage: Die Auswertung der Ergebnisse nach DIN ISO 5725 [54] führt zu „ausreichenden“ Ergebnissen, die bis zu $\pm 7,5\%$ vom Stoffwert abweichen dürfen. Ein Ergebnis muss mehr als $\pm 10\%$ vom Stoffwert abweichen, damit es mit „unzureichend“ bewertet wird. Diese in der Eignungsprüfung akzeptierten Abweichungen sind auf einem Niveau von 400 MPa in der Prüfpraxis jedoch bedenklich.

Tabelle 5.1.1: Zusammenstellung der Ergebnisse der Eignungsprüfung RZ DAP-1, 1996 [21], die Werte sind gerundet

Kennwert	Anzahl der Werte	Stoffwert X	Standardabweichung $\hat{\sigma}$	Grenzen für				Gemeldete Werte	
				$z=-3$	$z=-2$	$z=2$	$z=3$	MIN	MAX
$R_{p0,2}$ MPa	513	400,4	14,2	357,9	372,1	428,7	442,9	342	530
R_m MPa	508	673,6	9,1	646,2	655,4	691,8	700,9	643	759
A %	531	24,67	1,57	20,0	21,5	27,8	29,4	20,5	33,8
Z %	489	45,4	2,2	38,7	40,9	49,9	52,1	32	59
E GPa		191	14,5	147	162	220	234	102	260

Die Laborinterne Qualitätssicherung funktioniert nicht immer zuverlässig. So wurden offensichtlich falsche Rechenergebnisse (Bestimmung der Bruchdehnung Z) nicht entdeckt.

Die anzuwendende Prüfnorm ist in einigen Laboratorien nicht ausreichend bekannt. Der Werkstoff der Eignungsprüfung zeigte in keinem Bereich einen linearen Spannungs-Dehnungs-Verlauf. Nach der Prüfnorm DIN EN 10002-1 [93] musste daher die 0,2%-Dehngrenze mit Hilfe einer Teilentlastung und einer daraus resultierenden Hystereseschleife bestimmt werden. Einige Laboratorien führten jedoch eine manuelle Bestimmung durch. Andere verließen sich auf die vollautomatische Bestimmung durch ein zur Prüfmaschine gehörendes Rechnersystem; die Kennwerte lassen sich jedoch anhand der Rohdaten nicht verifizieren [23].

Die Bestimmung des Elastizitäts-Moduls (E-Modul) bereitet größte Schwierigkeiten, da das Verfahren nicht genormt ist. In der Literatur wird für den E-Modul von Stahl 206,0 GPa angegeben [94]. Die eingereichten Ergebnisse reichen von 102 GPa bis 260 GPa. Nahezu die Hälfte der Teilnehmer hat die von den Prüfmaschinenrechnern ausgegebenen Kennwerte kritiklos als Kennwert weitergegeben, obwohl der ermittelte Wert für einen Stahl offensichtlich falsch war [23]. Diese unsachgemäße Vorgehensweise ist übrigens in der Prüfpraxis sehr häufig zu beobachten.

Die große Spannweite zulässiger Ergebnisse zeigt die Notwendigkeit, Methoden zur statistischen Bewertung von Eignungsprüfungen besonders im Hinblick auf ihre technisch relevante Aussagefähigkeit zu untersuchen.

EA Interlaboratory Comparison MT 1, Tensile Test of Metals [95]

Diese Eignungsprüfung zum Zugversuch nach DIN EN 10002-1 [93] wurde durch das Czech Accreditation Institute (CAI), Prag organisiert und durchgeführt. Als Versuchswerkstoff wurde ein nicht näher spezifizierter unlegierter Stahl eingesetzt, mit überwiegend ferritischem Gefüge und einem Perlitanteil von 5 bis 7 %. Hauptbestandteile der chemischen Analyse sind Kohlenstoff (0,088 %), Mangan (0,46 %) und Silizium (0,20 %). Dieser Werkstoff besitzt eine ausgeprägte Streckgrenze, daher müssen die Teilnehmer anstelle einer Dehngrenze die obere Streckgrenze R_{eH} bestimmen. Als Konsenswert aus den Teilnehmerergebnissen wurden der Stoffwert X und die Standardabweichung der Eignungsprüfung $\hat{\sigma}$ sowohl mit der klassischen statistischen Methode nach DIN ISO 5725 [54] als auch mit einer robusten statistischen Methode berechnet. Die robuste statistische Methode verwendet den Median aller eingereichten Ergebnisse zur Berechnung des Stoffwertes X , die Standardabweichung der Eignungsprüfung $\hat{\sigma}$ wird aus dem normalisierten Quartilsabstand aller eingereichten Ergebnisse berechnet. Der normalisierte Quartilsabstand (nIQR) errechnet sich aus der Formel $nIQR=(Q3-Q1)0,7413$. Dieser Wert ist mit der Standardabweichung vergleichbar [87].

Die Kennwerte obere Streckgrenze R_{eH} und die 0,2%-Dehngrenze $R_{p0,2}$ fließen als Bemessungskennwerte direkt in die Auslegung realer Bauteile ein. Da die Ausnutzung von Werkstoffeigenschaften z. B. mit Hilfe von Leichtbaukonzepten zur Senkung des Energiebedarfs oder zur Steigerung von Nutzlasten von Fahrzeugen, ständig an Bedeutung gewinnt, ist die möglichst genaue und zuverlässige Bestimmung dieser Festigkeitskennwerte im Zugversuch besonders wichtig [96].

Die Ergebnisse der Eignungsprüfungen RZ DAP-1 und EA MT 1 sind für den Parameter R_{eH} bzw. $R_{p0,2}$ in Tabelle 5.1.2 gegenübergestellt. In dieser Tabelle sind die Stoffwerte, die Standardabweichung der Eignungsprüfung $\hat{\sigma}$, die oberen Grenzwerte (OG) für $z=2$, die unteren Grenzwerte (UG) für $z=-2$ und die Spannweite R (=OG-UG) der ausreichenden Ergebnisse eingetragen. In die Spalte REL R_{eH} bzw. REL $R_{p0,2}$ sind die Standardabweichung für die Eignungsprüfung $\hat{\sigma}$, die Grenzwerte für $z=2$ bzw. $z=-2$ und die Spannweite R relativ zum jeweiligen Stoffwert aufgetragen, um die Ergebnisse miteinander vergleichen zu können.

Tabelle 5.1.2: Gegenüberstellung der Ergebnisse der Eignungsprüfung EA MT 1 und RZ DAP-1

	<i>EA MT 1</i>		<i>RZ DAP-1</i>	
	R_{eH} in MPa	REL R_{eH} in %	R_{p0,2} in MPa	REL R_{p0,2} in %
<i>Auswertung nach DIN ISO 5725-2 [54]</i>				
Stoffwert X_{ISO}	312,9	---	399,2	---
$\hat{\sigma}_{ISO}$	33	10,5	14,4	3,6
z_{ISO} = 2	379	121,1	428	107,2
z_{ISO} = -2	247	78,9	370	92,7
R_{ISO}	132	42,2	58	14,5
<i>Auswertung mit robuster statistischer Methode</i>				
Stoffwert X_{rob}	319	---	399	---
$\hat{\sigma}_{rob}$	32,6	10,2	14,1	3,5
z_{rob} = 2	384	120,4	427	107,0
z_{rob} = -2	254	79,6	372	93,0
R_{rob}	130	40,8	50	12,5

Aus Tabelle 5.1.2 kann man folgendes entnehmen:

- Verwendet man die Auswertung nach einer robusten statistischen Methode anstelle der Auswertung nach DIN ISO 5725-2 [54], wird die Spanne der zulässigen Werte R kleiner.
- In der Eignungsprüfung EA MT 1 ist das Ergebnis eines Laboratoriums auch dann noch ausreichend, wenn es mehr als 20% nach oben oder unten vom Stoffwert abweicht. In der Eignungsprüfung RZ-DAP 1 ist eine maximale Abweichung von $\pm 7\%$ zulässig – diese zulässige Abweichung ist vom Standpunkt der technischen Relevanz eines Prüfergebnisses bereits bedenklich.
- Bei der Eignungsprüfung RZ-DAP 1 beträgt der Stoffwert X etwa 399 MPa und zwar sowohl nach der Ermittlung mittels der robusten statistischen Methode als auch nach der Ermittlung mittels DIN ISO 5725-2 [54]. Bei der Eignungsprüfung EA MT 1 weichen diese Werte leicht voneinander ab.

Eine mögliche Ursache für die große Streuung der zulässigen Werte in der Eignungsprüfung EA MT 1 liegt in der für diese Prüfaufgabe nicht ausreichenden Homogenität des Materials.

In Tabelle 5.1.3 sind die Ergebnisse der Homogenitätsprüfung für beide Eignungsprüfungen einander gegenübergestellt. Die Homogenitätsprüfung der Eignungsprüfung EA MT 1 für R_{eH} wurde an 15 Proben durchgeführt, die insgesamt 15 Einzelwerte lieferten. Die Homogenitätsprüfung der Eignungsprüfung RZ-DAP 1 für $R_{p0,2}$ wurde an 47 Stangen durchgeführt und basiert auf insgesamt 92 Einzelwerten. Für den eigentlichen Versuch wurden aufgrund der Ergebnisse der Homogenitätsprüfung die besten 36 Stangen ausgewählt.

Tabelle 5.1.3: Gegenüberstellung der Ergebnisse der Homogenitätsprüfung
Eignungsprüfung EA MT 1 und RZ DAP-1

		Min in MPa	Max in MPa	R in MPa	MW in MPa	Median in MPa	s in MPa	S _{REL} in %	R _{rel} in %
<i>RZ DAP-1</i>	R _{p0,2}	378	433	55	404,9	405,5	9,2	2,3	13,6
<i>EA MT 1</i>	R _{eH}	253	325	72	292,3	297	21,4	7,3	24,6

Die Bilder 5.1.3 und 5.1.4 zeigen für die jeweiligen Eignungsprüfungen die Verteilung der Einzelergebnisse der Homogenitätsprüfung. Zur besseren Vergleichbarkeit wurden die ermittelten Werte klassifiziert und die Klassenhäufigkeiten wurden über der Klassenmitte aufgetragen. Die Klasseneinteilung ist äquidistant und beträgt in beiden Darstellungen 10 MPa.

Die Homogenitätsprüfung der Eignungsprüfung RZ DAP-1 (Bild 5.1.3) zeigt gleichmäßig um den Mittelwert der Verteilung liegende Einzelwerte. Die Spannweite R der Einzelwerte beträgt 55 MPa, das sind bezogen auf den Mittelwert 13,6%. Die Standardabweichung von 9,2 MPa beträgt 2,3% des Mittelwertes. Der arithmetische Mittelwert und Median sind praktisch identisch und liegen in der Klasse mit der größten Häufigkeit.

Die Homogenitätsprüfung der Eignungsprüfung EA MT 1 zeigt, dass die Einzelwerte in einer ungleichmäßigen, schiefen Verteilung vorliegen. Arithmetisches Mittel und Median unterscheiden sich geringfügig. Sie liegen nicht in der Klasse der größten Häufigkeit. Die Spannweite der Einzelwerte beträgt 72 MPa, das sind 24,6% bezogen auf den Mittelwert. Die Standardabweichung von 21,4 MPa beträgt 7,3% des Mittelwertes. Die Standardabweichung als Maß für die Streuung setzt sich prinzipiell aus einer verfahrensbedingten Streuung (Gerät, Probenfertigung, Bediener) und einer werkstoffbedingten Streuung (Inhomogenität) zusammen.

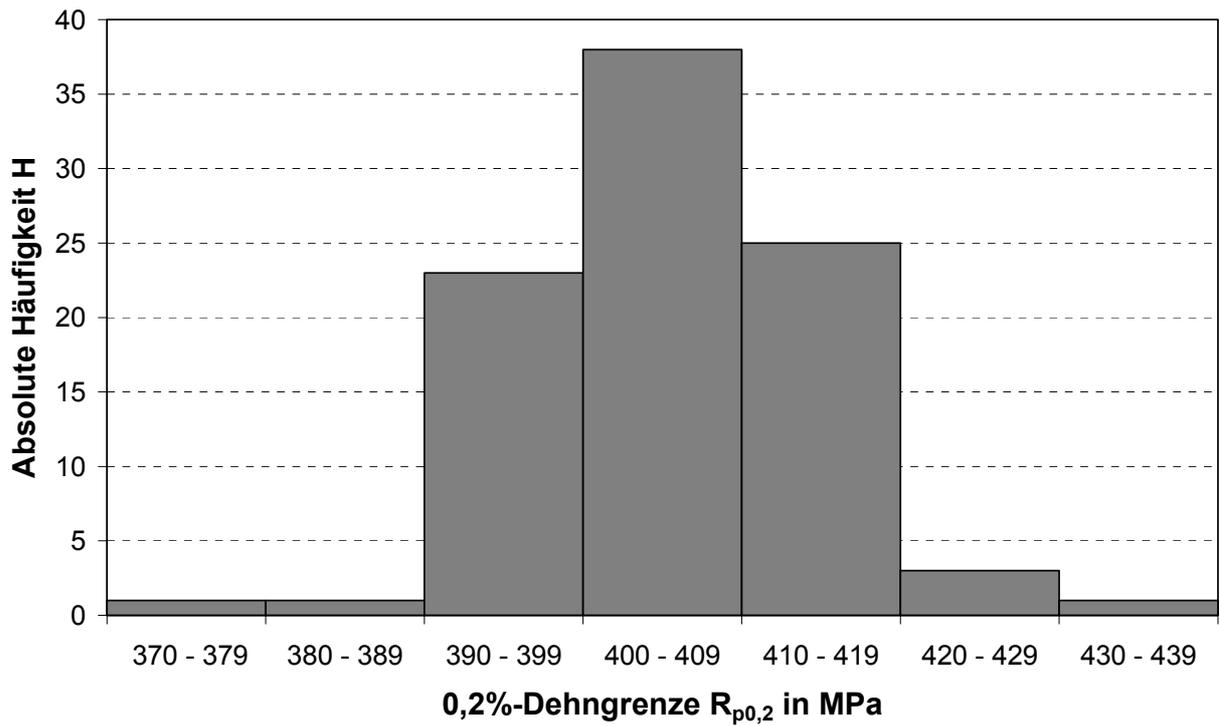


Bild 5.1.3: Homogenitätsprüfung RZ DAP-1, Verteilung der 92 Einzelwerte

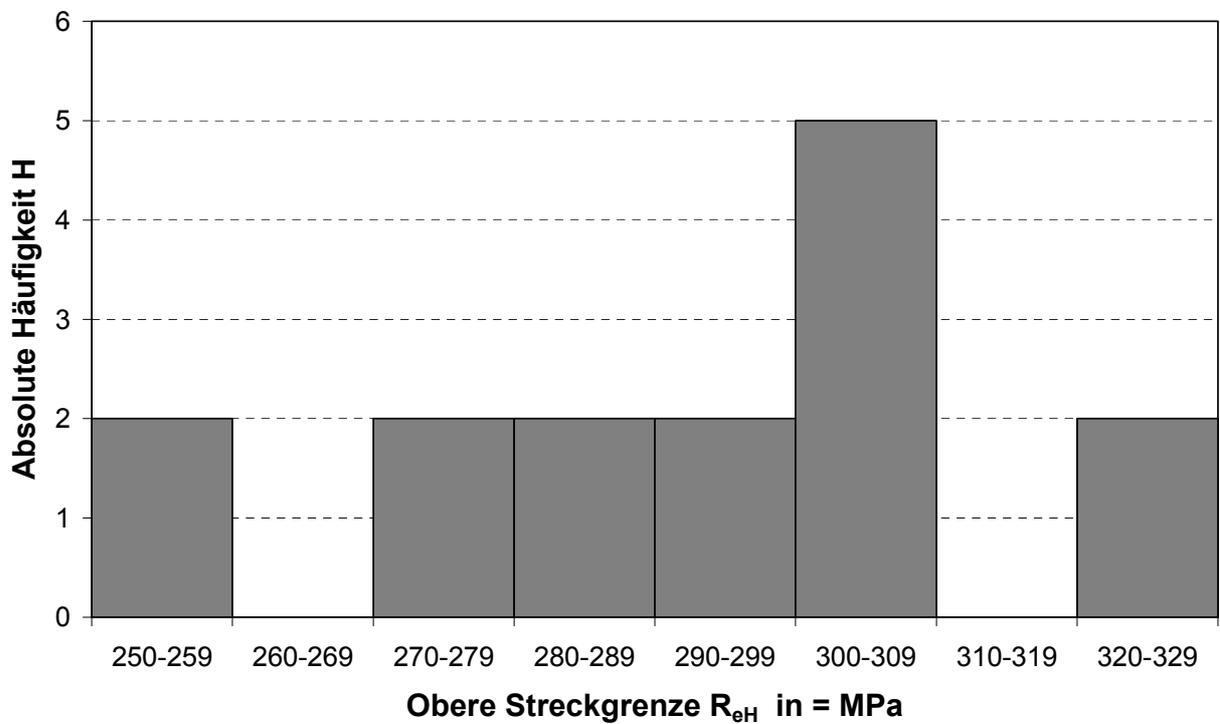


Bild 5.1.4: Homogenitätsprüfung EA MT 1, Verteilung der 15 Einzelwerte

Geht man bei der Homogenitätsprüfung für eine Eignungsprüfung in einem Expertenlaboratorium von einer geringen verfahrensbedingten Streuung aus (vereinheitlichtes Prüfverfahren, besonders qualifiziertes Personal, kalibrierte Prüfmaschinen), so muss bei einer großen Standardabweichung der untersuchte Kennwert für dieses Material stark streuen. Das Aussehen der Verteilung (Bild 5.1.4) lässt ebenfalls auf eine starke Streuung der oberen Streckgrenze schließen. Außerdem ist die Anzahl der geprüften Proben im vorliegenden Fall sicherlich nicht ausreichend.

Der vorliegende Werkstoff ist für eine Eignungsprüfung „Zugversuch“ für den Parameter „obere Streckgrenze“ nicht ausreichend homogen.

Eine weitere mögliche Ursache für die große Streuung der zulässigen Werte in der Eignungsprüfung EA MT 1 liegt darin, dass die obere Streckgrenze R_{eH} stark von der Prüfgeschwindigkeit abhängt. Nach DIN EN 10002-1 [93] muss die Querschnittsgeschwindigkeit innerhalb des elastischen Bereichs und bis zur oberen Streckgrenze so konstant wie möglich gehalten werden. Für Stahl muss die Spannungszunahmegeschwindigkeit zwischen 6 und 60 MPa s^{-1} liegen. Der Vorschlag der Arbeitsgruppe ECISS/TC 1/WG 1, basierend auf einer Arbeit von AEGERTER et al. [97], zur Überarbeitung der EN 10002-1 sieht vor, im elastischen Bereich die Steuerung der Prüfmaschine mittels der Dehngeschwindigkeit vorzunehmen.

Die vorgeschlagene Dehngeschwindigkeit von $0,00025 s^{-1} \pm 20\%$ stellt eine drastische Verminderung der zulässigen Geschwindigkeit und starke Einschränkung des zulässigen Geschwindigkeitsbereichs dar, um eine möglichst große Vergleichbarkeit der Prüfergebnisse für R_{eH} bzw. $R_{p0,2}$ sicherzustellen.

Für die Ausrichtung einer Eignungsprüfung empfiehlt es sich, (bis dieser Vorschlag in die Normung einfließt) die Spannungs- bzw. Dehnungszunahmegeschwindigkeit in den Randbedingungen für die Eignungsprüfung entsprechend einzuschränken. Die Parameter müssen in Vorversuchen ermittelt werden.

5.2 Eignungsprüfungen mit Härteprüfverfahren

Härte wird üblicherweise definiert als der Widerstand eines Stoffes gegen das Eindringen eines anderen Körpers. Die Härteprüfung liefert schnell und praktisch zerstörungsfrei Anhaltswerte für die statische Festigkeit eines Werkstoffes, über sein Verschleißverhalten und die Regelmäßigkeit einer Wärmebehandlung. Man kann die Härte nach folgenden grundsätzlichen Methoden ermitteln:

- Ritzen der Oberfläche;
- Eindringen eines Prüfkörpers unter statischer Belastung;
- Eindringen eines Prüfkörpers unter dynamischer Belastung;
- Rückprall infolge elastischen Verhaltens des Prüfstücks.

Die weiteste Verbreitung und größte Bedeutung haben die statischen Prüfverfahren, die ein Standardprüfverfahren aller Laboratorien für mechanisch-technologische Werkstoffprüfung sind. Bei diesen Prüfverfahren wird eine Prüfkraft stoßfrei aufgebracht und wirkt ruhend eine vorgegebene Zeit auf die Probe ein. Nach der Entlastung wird ein bleibender Eindruck vermessen und die Härte als Quotient aus Prüfkraft und Eindruckfläche berechnet, dies sind die Verfahren Härteprüfung nach Vickers und Härteprüfung nach Brinell.

Eine andere Möglichkeit ist das Messen der Eindringtiefe bei vorgegebener Prüfkraft und die direkte Umwandlung in einen Härtewert, die Härteprüfung nach Rockwell.

Das modernste Härteprüfverfahren ist die Martenshärte HM (früher Universalhärte HU). Sie wird durch Indentereindruckversuche ermittelt, bei denen die Kraft und die Eindringtiefe kontinuierlich während der Be- und Entlastung registriert werden [98].

5.2.1 Eignungsprüfung Härteprüfung nach Vickers

Das Prüfverfahren Härteprüfung nach Vickers ist in der DIN EN ISO 6507-1 [99] genormt. Ein Diamant-Eindringkörper in Form einer geraden Pyramide mit quadratischer Grundfläche und einem festgelegten Winkel zwischen den gegenüberliegenden Flächen wird in die Oberfläche einer Probe gedrückt. Zur Ermittlung der Vickershärte werden die Diagonalen d_1 und d_2 des Eindrucks vermessen, der sich nach Wegnahme der Prüfkraft F auf der Prüffläche befindet, vgl. Bild 5.2.1.1. Die Vickershärte ist proportional dem Quotienten aus der Prüfkraft und der Oberfläche des Eindrucks, der als gerade Pyramide mit quadratischer Grundfläche und gleichem Winkel wie der Eindringkörper angenommen wird [99].

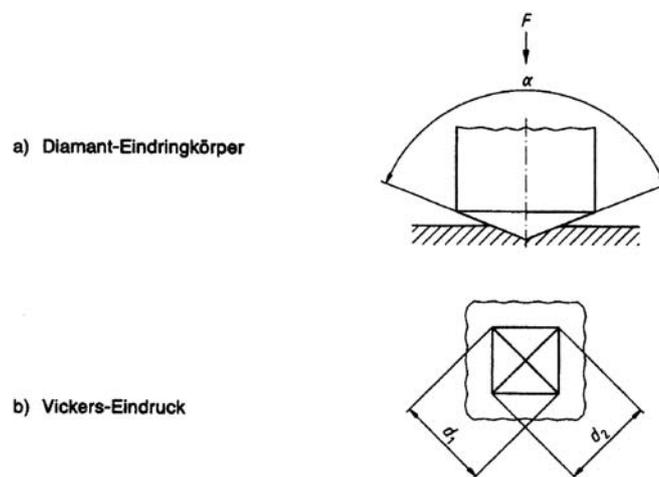


Bild 5.2.1.1: Härteprüfung nach Vickers, Verfahrensprinzip (schematisch) aus [99]

Die wesentlichen Größen zur Bestimmung der Vickershärte sind in Tabelle 5.2.1.1 zusammengestellt.

Tabelle 5.2.1.1: Größen zur Bestimmung der Vickershärte aus [99]

Symbol	Benennung
α	Winkel zwischen gegenüberliegenden Flächen des pyramidenförmigen Eindringkörpers (136°)
F	Prüfkraft in N
d	arithmetischer Mittelwert der beiden Diagonalenlängen d_1 und d_2 (nach Bild 1) in mm
HV	<p>Vickershärte</p> $= \text{Konstante} \times \frac{\text{Prüfkraft}}{\text{Oberfläche des Eindrucks}}$ $= 0,102 \frac{2 F \sin \frac{136^\circ}{2}}{d^2} \approx 0,1891 \frac{F}{d^2}$
ANMERKUNG: Konstante = $\frac{1}{g_n} = \frac{1}{9,806\,65} \approx 0,102$	

Die Härteprüfung nach Vickers ist in Abhängigkeit von der Prüfkraft F in drei Bereiche eingeteilt: Mikrohärtebereich (0,099 N bis 0,98 N), Kleinkraftbereich (1,96 N bis 29,4 N) und konventioneller Härtebereich (49 N bis 981 N). Da die Eindrücke geometrisch immer ähnlich bleiben, sind die ermittelten Härtewerte im Normalfall unabhängig von der Größe der Prüfkraft. Im Kleinkraft- und im Mikrohärtebereich ist eine Zunahme der Härtewerte zu beobachten, da der Anteil elastischer Verformungen größer wird [100]. Die zulässigen Prüfkraften sind in DIN EN ISO 6507-1 [99] tabelliert.

Die Prüfung erfolgt an einer glatten, ebenen Oberfläche, die frei von Verunreinigungen sein muss. Die Oberflächengüte der Prüffläche muss die exakte Messung der Eindruckdiagonalen ermöglichen. Die Probe muss in einer starren Aufnahme liegen und darf sich während der Prüfung nicht verschieben. Der Eindringkörper wird auf die Oberfläche aufgesetzt und die Prüfkraft ist stoß- und erschütterungsfrei aufzubringen. In Tabelle 5.2.1.2 sind weitere Prüfbedingungen zusammengefasst.

Es werden die Längen der beiden Diagonalen gemessen. Für die Berechnung der Vickershärte ist der arithmetische Mittelwert der beiden zu verwenden. An ebenen Oberflächen darf die Differenz der beiden Diagonalen nicht mehr als 5% betragen. Die Ermittlung des Prüfergebnisses erfolgt entweder rechnerisch nach der Formel aus Tabelle 5.2.1.1 oder über die in DIN EN ISO 6507-1, Anhang C [99] enthaltenen Tabellen.

Tabelle 5.2.1.2: Härteprüfung nach Vickers, Prüfbedingungen [99]

Prüftemperatur	$23 \pm 5 \text{ °C}$
Zeit zum Aufbringen der Prüfkraft*	2 bis 8 sec
Einwirkdauer der Prüfkraft*	10 bis 15 sec
Mindestdicke der Probe	min. 1,5 d
Abstand eines Prüfeindrucks (Mittelpunkt) vom Rand (Stahl, Kupfer, Kupferlegierungen)	min. 2,5 d
Abstand eines Prüfeindrucks (Mittelpunkt) vom Rand (Blei, Zinn, Leichtmetalle und deren Legierungen)	min. 3 d
Abstand zweier nebeneinander liegender Prüfeindrücke (Mittelpunkte) (Stahl, Kupfer, Kupferlegierungen)	min. 3 d
Abstand zweier nebeneinander liegender Prüfeindrücke (Mittelpunkte) (Blei, Zinn, Leichtmetalle und deren Legierungen)	min. 6 d

* gilt für den konventionellen Härtebereich

Die Härteprüfung nach Vickers ist fast universell anwendbar und eignet sich sowohl für sehr weiche als auch für sehr harte Werkstoffe, ebenso für dünne Teile und Schichten. Die Härtewerte betragen zwischen 3 HV (z. B. Blei) und 1500 HV (Hartmetall). Der wesentliche Vorteil des Härteprüfverfahrens nach Vickers liegt darin, auch bei kleinen Eindrücken die Diagonalen des Eindruckquadrates genau ausmessen zu können.

5.2.1.1 Untersuchungsprogramm

Für die Eignungsprüfung EPHV-1 HV 10/30 [25] wurden folgende Aufgaben festgelegt, zu denen die teilnehmenden Laboratorien eine Leistung erbringen sollten:

- (1) Härtemessung nach Vickers: Abweichung des Mittelwertes vom Stoffwert.
- (2) Härtemessung nach Vickers: Wiederholbarkeit.
- (3) Vermessen von Referenzhärteeindrücken.

Die Homogenitätsuntersuchung der Proben erfolgte nach der in Tabelle 5.2.1.3 mit „X“ gekennzeichneten Methode. Auf die im Rahmen der Eignungsprüfung gewonnenen Daten wurden die in den Tabellen 5.2.1.4 bis 5.2.1.5 mit „X“ gekennzeichneten Methoden auf die Leistungskriterien (1) und (2) angewendet und anschließend kritisch untersucht und bewertet:

Tabelle 5.2.1.3: Homogenitätsuntersuchung

Einfaktorielle Varianzanalyse (Kap. 4.1.1)	-
Nach ISO/DIS 13528:2002 (Kap. 4.1.2)	-
Andere Vorgehensweisen (Kap. 4.1)	X

Tabelle 5.2.1.4: Bestimmung des Stoffwertes X

Methode (nach Kapitel 4.2)	
Bekannter Wert (Methode A)	-
Zertifiziertes Referenzmaterial (Methode B)	X
Referenzmaterial (Methode C)	-
Konsenswert Experten (Methode D)	-
Konsenswert Teilnehmer (Methode E)	-

Tabelle 5.2.1.5: Berechnung der Standardabweichung der Eignungsprüfung $\hat{\sigma}$

Methode (nach Kapitel 4.4.1)	
Vorgeschriebener Wert	X
Festlegung eines erwünschten Leistungsniveaus	-
Präzisionsexperiment nach DIN ISO 5725-2	-
Konsenswert, robuste Datenauswertung	-

In Tabelle 5.2.1.6 sind die untersuchten Verfahren zur Berechnung der Leistungsstatistik zusammengestellt.

Tabelle 5.2.1.6: Berechnung der Leistungsstatistik

Methode (nach Kapitel 4.4.2)	
Schätzung der systematischen Abweichungskomponente D	-
Prozentuale Differenz D%	X
z-Scores	-
Histogramm	-
Säulendiagramm	-

5.2.1.2 Programm der Eignungsprüfung

Die Teilnehmer erhielten als Probematerial eine zertifizierte Härtevergleichsplatte der Materialprüfungsanstalt Nordrhein-Westfalen (MPA NRW), Dortmund, siehe Bild 5.2.1.2. Auf dieser Härtevergleichsplatte sollten an vorgegebenen Prüforten je sechs Härtemessungen HV 10 und sechs Härtemessungen HV 30 durchgeführt werden. Zusätzlich waren in die Härtevergleichsplatte von der MPA NRW Dortmund Referenzhärteeindrücke eingebracht worden, die von den Teilnehmern vermessen werden sollten. Das Vermessen der Referenzeindrücke erlaubt den Prüflaboratorien ihre Geräteoptik zu überprüfen. Auf jeder Härtevergleichsplatte prüfen bis zu sieben Laboratorien.

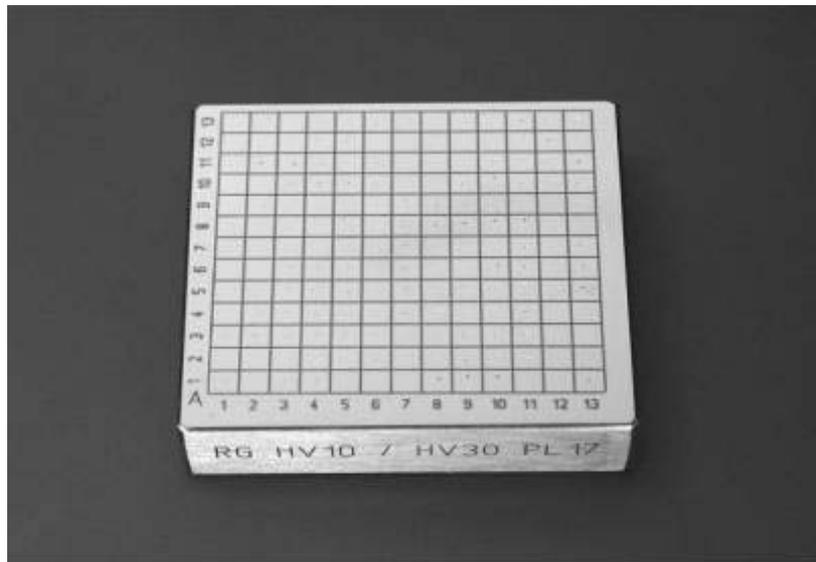


Bild 5.2.1.2: Härtevergleichsplatte mit Raster

5.2.1.3 Materialauswahl und Homogenitätsprüfung

Bei der Eignungsprüfung wurden insgesamt achtzehn von der MPA NRW Dortmund zertifizierte Härtevergleichsplatten als Referenzmaterial verwendet. Die Härte der Platten betrug 390 bis 405 HV 10 bzw. 390 bis 405 HV 30. Sie wurden mittels eines Gitternetzes jeweils in 13x13 gleich große Prüffelder aufgeteilt, um den Laboratorien anhand einer Prüfschablone eindeutige Prüforte zuzuweisen.

Obwohl zertifizierte Härtevergleichsplatten verwendet wurden, ist eine zufällig ausgewählte Platte auf Homogenität geprüft worden. Drei Prüfer, denen der Referenzwert nicht bekannt war, haben mit derselben Prüfmaschine die Härte HV 10 in sämtlichen 169 Feldern gemessen. Die Ergebnisse dieser Prüfung sind in Tabelle 5.2.1.7 zusammengefasst und dem Referenzwert der MPA NRW Dortmund für die Härtevergleichsplatte gegenübergestellt.

Vergleicht man für die einzelnen Prüfer die Kennzahlen der Verteilungen, so ergibt sich folgendes Bild: in allen drei Fällen liegt eine symmetrische Verteilung vor (Mittelwert \approx Median) bei gleichzeitig geringer Streuung (die Standardabweichung s bzw. die Spannweite der Ergebnisse R sind klein). Außerdem sind die Anforderungen der DIN EN ISO 6507-2, Tabelle 4 und Tabelle 5 bezüglich der zulässigen Abweichung vom Referenzwert und Wiederholbarkeit (vgl. Kap. 5.1.2.4) für alle drei Versuchsreihen erfüllt. Aufgrund dieses Ergebnisses werden die zertifizierten Härtevergleichsplatten für die Verwendung als Referenzmaterial in diesem Ringversuch als ausreichend homogen angesehen [24].

Tabelle 5.2.1.7: Ergebnisse der Homogenitätsprüfung

			<i>Referenzwert</i>	Prüfer 1	Prüfer 2	Prüfer 3
Diagonalenmessung	Mittelwert \bar{D}	μm	<i>214,359</i>	213,747	212,659	214,147
	Standardabweichung s_D	μm	<i>0,1267</i>	0,3636	0,2782	0,3278
	Median \tilde{D}	μm	<i>214,3875</i>	213,831	212,756	214,268
	$R_D = D_{\max} - D_{\min}$	μm	<i>0,345</i>	1,915	1,574	1,212
	Wiederholpräzision	%	<i>0,16</i>	0,90	0,74	0,57
Härtemessung	Mittelwert \bar{H}	HV 10	<i>403,6</i>	405,9	410,2	404,3
	Standardabweichung s_H	HV 10	<i>0,48</i>	1,42	1,00	1,28
	Median \tilde{H}	HV 10	<i>403,5</i>	406	410	404
	$R_H = H_{\max} - H_{\min}$	HV 10	<i>1,3</i>	7	6	5
	Abweichung vom Referenzwert	%	---	0,29	0,79	0,10

5.2.1.4 Abweichung des Labormittelwertes vom Stoffwert

Bei Verwendung eines zertifizierten Referenzmaterials wird der zertifizierte Referenzwert jeder Härtevergleichsplatte als Stoffwert X verwendet. Gleiches gilt sinngemäß für die vorgegebenen Härteeindrücke. Die DIN EN 6507-2 [101] gibt die zulässigen Abweichungen bei der Prüfung der Prüfmaschine vor, so dass keine Berechnung der Standardabweichung der Eignungsprüfung $\hat{\sigma}$ notwendig ist. Da die Laboratorien nicht alle auf derselben Härtevergleichsplatte messen, werden die Prüfergebnisse auf den Referenzwert bezogen. Die Auswertung erfolgt über die Bestimmung der prozentualen Differenz $D\%$. Am Beispiel der Härtemessung HV 10 werden im Folgenden die Ergebnisse der Eignungsprüfung vorgestellt:

Vermessen der Referenzeindrücke

Mit dem Ausmessen vorgegebener Diagonalen sollte die von den Laboratorien verwendete Geräteoptik überprüft werden. Die zulässige Abweichung beim Vermessen der vorgegebenen sechs Härteeindrücke wurde auf $\pm 1\%$ festgelegt. Diese Abweichung entspricht rechnerisch einer Abweichung von rd. $\pm 2\%$ für einen aus den gemessenen Diagonalen ermittelten Härtewert. In Bild 5.2.1.3 sind die Ergebnisse der Prüfungen zusammengestellt.

Die Ergebnisse von 28 % der Teilnehmer weichen um mehr als 1% vom Stoffwert ab. Beim Analysieren der Darstellung (Bild 5.2.1.3) fällt auf, dass die von den Laboratorien ermittelten Werte nicht symmetrisch um den Stoffwert verteilt sind. 80 % der ausgewerteten Ergebnisse ist kleiner als der Stoffwert, nur 20 % der Werte sind größer oder gleich dem Stoffwert. Eine offensichtliche Ursache für diese ungleiche Verteilung war nicht zu erkennen.

Die Wiederholpräzision der Messungen wird in Anlehnung an DIN EN ISO 6507-2, Tabelle 4 [101] bestimmt. Als Kriterium wird herangezogen $\frac{d_{\max} - d_{\min}}{\bar{d}} * 100\% \leq 2\%$, mit $d_{\max/\min}$ größte/kleinste Eindruckdiagonale, \bar{d} Mittelwert aus allen Eindruckdiagonalen.

In Bild 5.2.1.3 sind die Ergebnisse aller Laboratorien graphisch dargestellt. Neun Laboratorien weisen eine Wiederholpräzision von 100 % auf. Dies Ergebnis ist in Kenntnis der Einflussgrößen des Prüfverfahrens nicht nachvollziehbar.

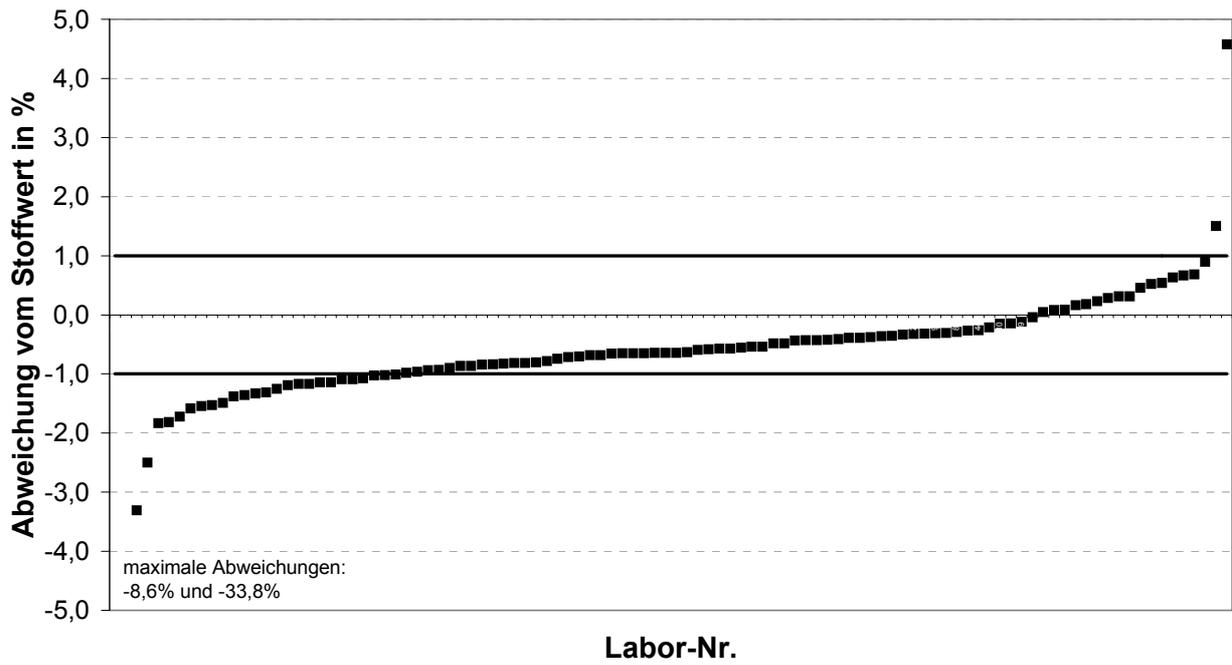
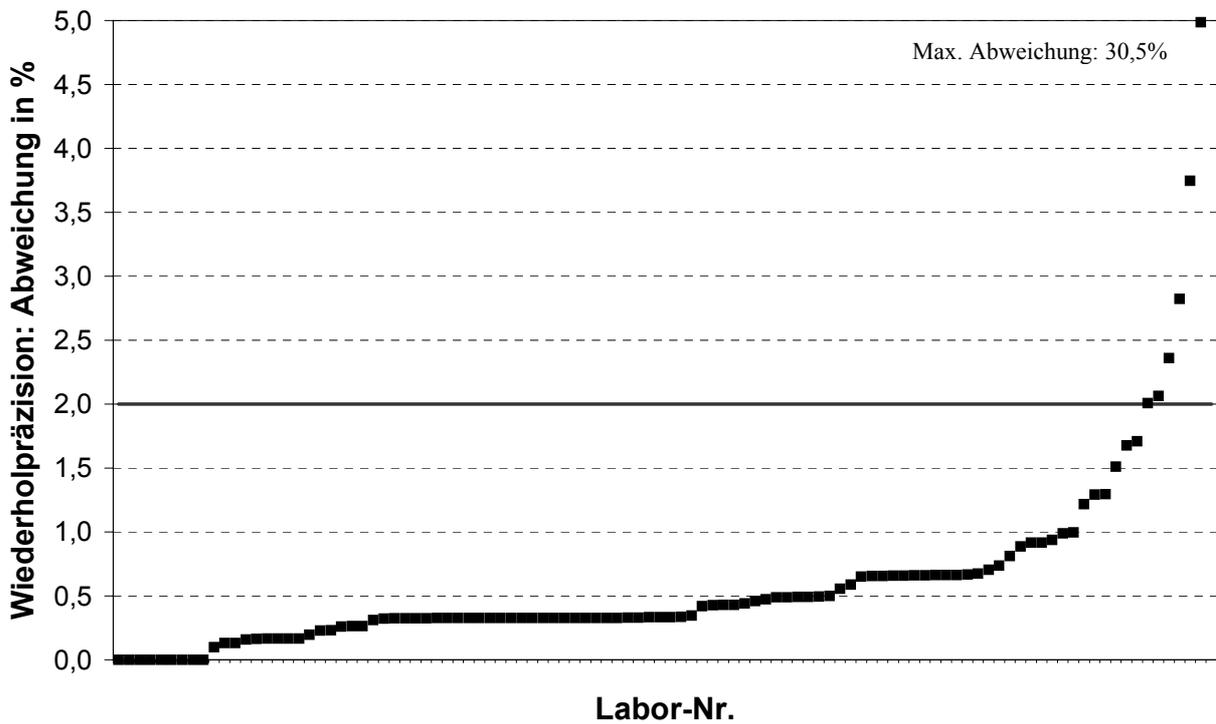


Bild 5.2.1.3: Vermessen der Referenzeindrücke, Abweichung vom Stoffwert X



Härtemessung HV 10

Als Abweichung der Härteprüfmaschine nach DIN EN ISO 6507-2, Tabelle 5 [101] bezeichnet man die Differenz zwischen dem Mittelwert der gemessenen Härtewerte und der Härte der verwendeten Härtevergleichsplatte. Diese Abweichung wird in Prozent angegeben. Für das Verfahren HV 10 sind maximal $\pm 3\%$ zulässig. Die Ergebnisse sind in Bild 5.2.1.6 zusammengestellt. Die eingereichten Ergebnisse sind gleichmäßig um den Stoffwert verteilt, nur 7 von 106 Laboratorien (6,6 %) erfüllen nicht die Anforderung der Norm. Wie nachträgliche Untersuchungen zeigten, sind die Fehler entweder auf fehlerhaftes Ausmessen der Eindrücke oder auf eine falsche Wahl der Prüfkraft zurückzuführen.

Die Wiederholpräzision der Prüfmaschine (DIN EN ISO 6507-2, Tabelle 4, [101]) wird durch die Differenz zwischen der größten und kleinsten Eindruckdiagonalen^G dargestellt. Bezieht man diese Differenz auf den Mittelwert aller Eindruckdiagonalen (eines Labors), ist eine Abweichung von 2% zulässig (Bedingung ①). Gleichzeitig dürfen die ermittelten Härtewerte (Bedingung ②) höchstens 11 HV 10 voneinander abweichen.

In den Bildern 5.2.1.6 (Bedingung ①) und 5.2.1.7 (Bedingung ②) sind die Ergebnisse zusammengestellt. Die Ergebnisse wurden mit beiden Bedingungen auf eine ausreichende Wiederholpräzision verglichen. Dabei zeigte sich, dass 3,8 % der Teilnehmer die Bedingung ① nicht erfüllen; die schärfere Bedingung ② wird von 7,5 % der Teilnehmer nicht eingehalten.

Vier Laboratorien, die gemäß Bedingung ① noch eine ausreichende Wiederholpräzision erzielen, erfüllen Bedingung ② nicht mehr.

Bei vier dieser Laboratorien ergab eine Nachmessung, dass die Eindrücke bei ordnungsgemäßer Ausmessung eine geringere Streuung aufweisen als die Streuung, die sich mit den von den Prüflaboratorien ermittelten Messwerten ergibt.

Neun Laboratorien geben eine Wiederholpräzision von 100 % an. Unter Berücksichtigung der Einflussgrößen des Messverfahrens sind diese Ergebnisse nicht nachvollziehbar. Einige Laboratorien gaben zwölf bis auf $5\mu\text{m}$ identische Ergebnisse an. Ein Nachmessen dieser Eindrücke konnte diese Ergebnisse nicht reproduzieren, zumal einige der Eindrücke geometrisch offensichtlich nicht vollkommen waren.

^G Eindruckdiagonale ist hier das arithmetische Mittel aus den jeweils zwei an einem Prüfeindruck gemessenen Eindruckdiagonalen.

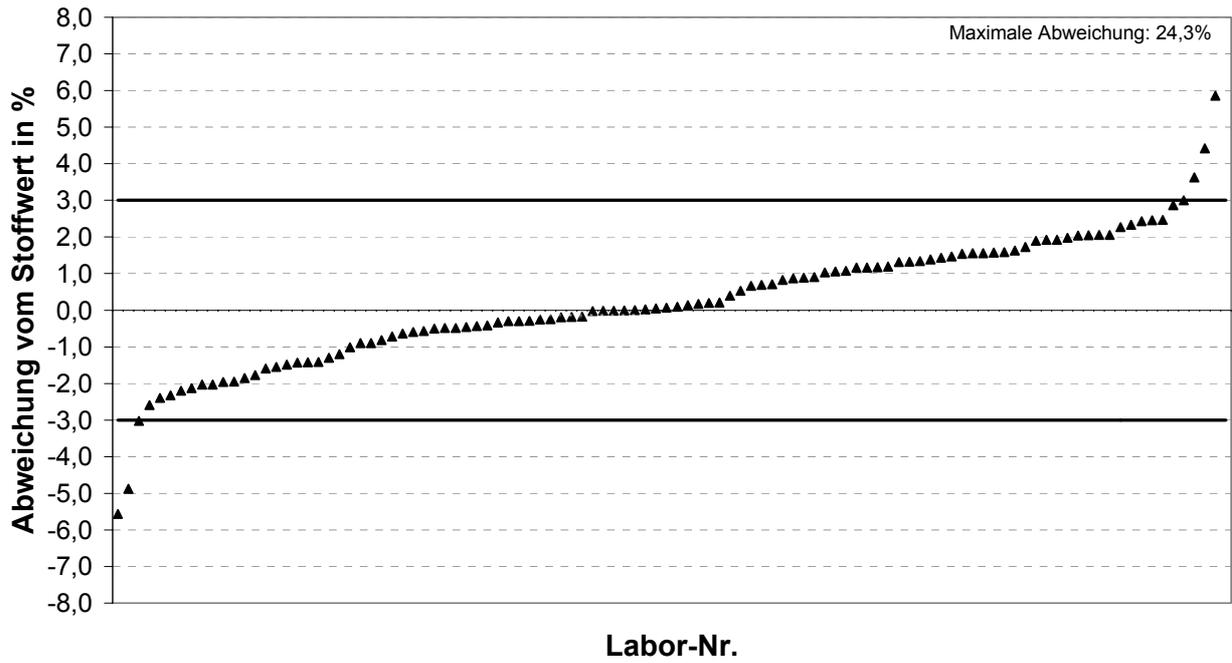


Bild 5.2.1.5: Härteprüfung HV 10, Abweichung vom Stoffwert

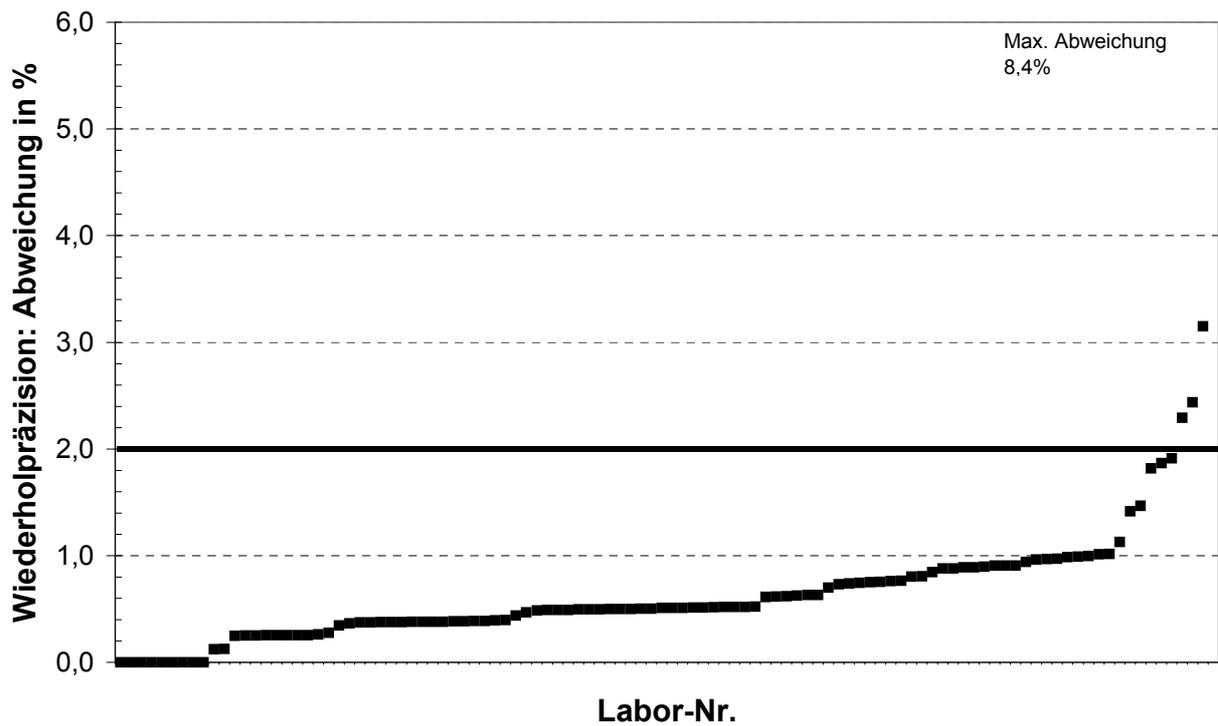


Bild 5.2.1.6: Härteprüfung HV 10, Wiederholpräzision, Bedingung ①

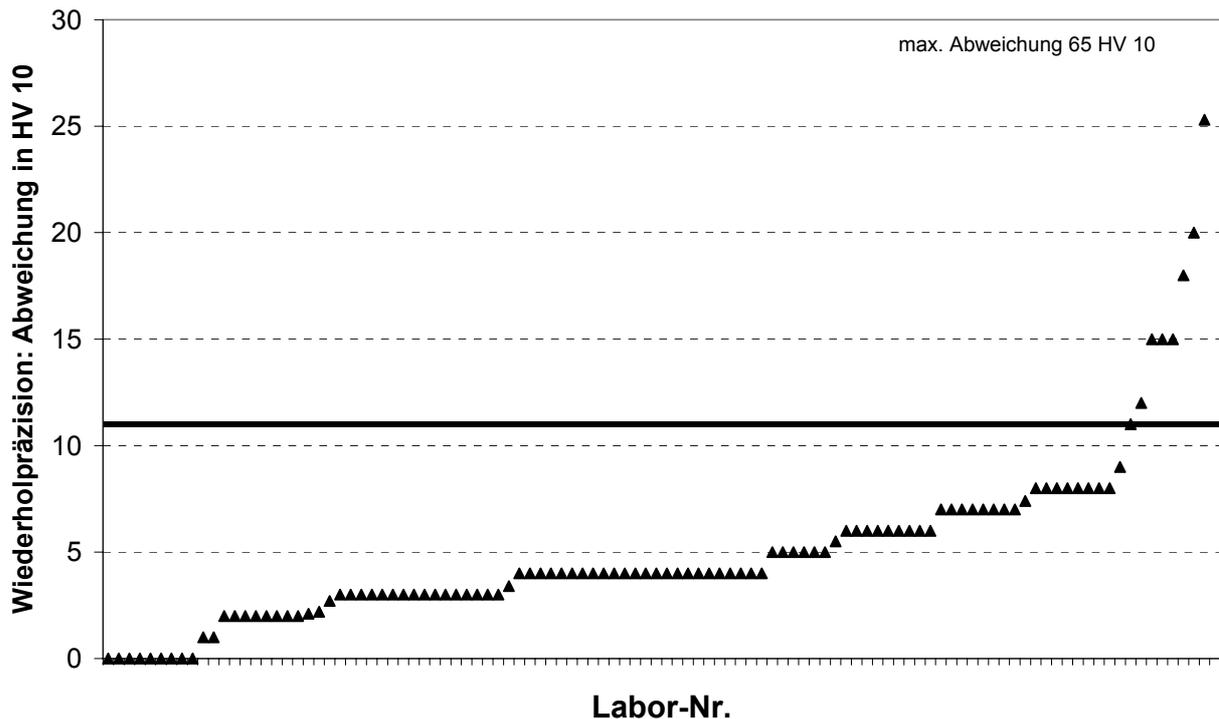


Bild 5.2.1.7: Härteprüfung HV 10, Wiederholpräzision, Bedingung ②

Die Verwendung von zertifiziertem Referenzmaterial und die Beurteilung einer Eignungsprüfung nach Normvorgaben, führt zu einer relativ einfach durchzuführenden Auswertung der Eignungsprüfung, da der rechnerische Aufwand gering gehalten wird. Die Vorgaben sind technisch relevant, den Teilnehmern bekannt und daher leicht nachzuvollziehen.

Das Ergebnis der Eignungsprüfung weist erstmalig auf ausreichender Datenbasis nach, dass die Vorgaben der DIN EN 6507-2 [101] bezüglich der zulässigen Abweichung der Prüfmaschine und der Wiederholgenauigkeit zutreffend sind. Die Vorgaben sind konservativ, denn die Ergebnisse der meisten Teilnehmerergebnisse liegen noch relativ weit von den vorgegebenen Grenzen entfernt.

5.2.2 Eignungsprüfung Härteprüfung nach Brinell

Das Prüfverfahren Härteprüfung nach Brinell ist in der DIN EN ISO 6506-1 [102] genormt. Es eignet sich für metallische Werkstoffe bis zu einer Grenzhärte von 650 HBW.

Eine Hartmetallkugel mit dem Durchmesser D wird mit der Prüfkraft F senkrecht auf die Oberfläche einer Probe gedrückt, siehe Bild 5.2.2.1.

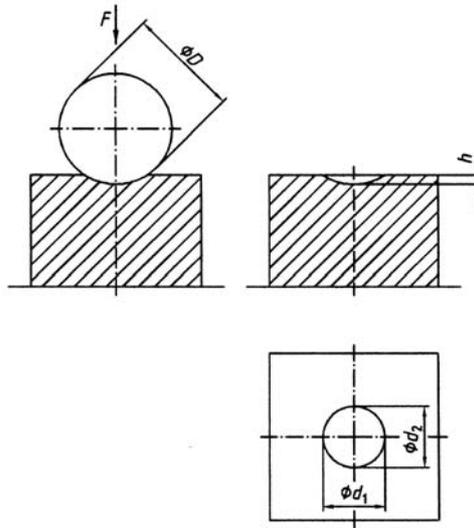


Bild 5.2.2.1: Härteprüfung nach Brinell, Verfahrensprinzip (schematisch) aus [102]

Der Eindruckdurchmesser d , der sich nach Wegnahme der Prüfkraft F auf der Prüffläche befindet, wird gemessen. Die Brinellhärte ist proportional dem Quotienten aus der Prüfkraft und der gekrümmten Oberfläche des Kugeleindrucks. Die Oberfläche wird als kalottenförmig angenommen mit einem Krümmungsradius, der dem halben Durchmesser D der verwendeten Kugel entspricht. Die wesentlichen Größen zur Bestimmung der Brinellhärte sind in Tabelle 5.2.2.1 zusammengestellt.

Die Größe der Prüfkraft F hängt ab vom gewählten Kugeldurchmesser D und dem zu prüfenden Werkstoff. Damit die Eindrücke ähnlich werden, gibt die DIN EN ISO 6506-1 [102] für verschiedene Werkstoffe bestimmte Beanspruchungsgrade $0,102 F/D^2$ (in N/mm^2) vor. So beträgt der Belastungsgrad für Stähle $30 N/mm^2$, für Leichtmetalle und ihre Legierungen (Härte $<35 HBW$) $2,5 N/mm^2$. Die Härte ist unabhängig von der Prüfkraft F und dem Kugeldurchmesser D , solange der Eindruckdurchmesser d zwischen $0,24 D$ und $0,6 D$ liegt [102]. Größere und kleinere Eindrücke oder andere Belastungsgrade führen zu falschen und nicht mehr direkt miteinander vergleichbaren Prüfergebnissen [100]. Die Belastungsgrade, die zulässigen Kugelgrößen und zulässigen Prüfkraften sind in der DIN EN ISO 6506-1 [102] tabelliert.

Die Prüfung erfolgt an einer glatten, ebenen Oberfläche, die frei von Verunreinigungen sein muss. Die Oberflächengüte der Prüffläche muss die exakte Messung des Eindruckdurchmessers ermöglichen.

Tabelle 5.2.2.1: Größen zur Bestimmung der Brinellhärte aus [102]

Formelzeichen	Benennung	Einheit
D	Durchmesser der Kugel	mm
F	Prüfkraft	N
d	Mittlerer Durchmesser des Eindrucks $\left(d = \frac{d_1 + d_2}{2}\right)$	mm
d_1, d_2	Durchmesser des Eindrucks, gemessen um 90°	mm
h	Eindrucktiefe $= \frac{D - \sqrt{D^2 - d^2}}{2}$	mm
HBW	Brinellhärte $= \text{Konstante} \cdot \frac{\text{Prüfkraft}}{\text{Oberfläche des Eindrucks}}$ $= 0,102 \cdot \frac{2F}{\pi D (D - \sqrt{D^2 - d^2})}$	
$0,102 F/D^2$	Beanspruchungsgrad	N/mm ²
ANMERKUNG: Konstante = $\frac{1}{g_n} = \frac{1}{9,80665} = 0,102$		
g_n Fallbeschleunigung		

Die Probe muss in einer starren Aufnahme liegen und darf sich während der Prüfung nicht verschieben. Der Eindringkörper wird auf die Oberfläche aufgesetzt und die Prüfkraft ist stoß- und erschütterungsfrei aufzubringen. In Tabelle 5.2.2.2 sind weitere Prüfbedingungen zusammengefasst.

Tabelle 5.2.2.2: Härteprüfung nach Brinell, Prüfbedingungen [102]

Prüftemperatur	23 ± 5 °C
Zeit zum Aufbringen der Prüfkraft	2 bis 8 sec
Einwirkdauer der Prüfkraft	10 bis 15 sec
Abstand eines Prüfeindrucks (Mittelpunkt) vom Rand	min. 2,5 d
Abstand zweier nebeneinander liegender Prüfeindrücke (Mittelpunkte)	min. 3 d

Die Mindestdicke der Probe ist in Abhängigkeit vom Eindruckdurchmesser in der Prüfnorm tabelliert (DIN EN ISO 6506-1, Anhang B [102]). Die Ermittlung des Prüfergebnisses erfolgt entweder rechnerisch nach der in Tabelle 5.2.2.1 angegebenen Formel oder über die in DIN EN ISO 6506-1, Anhang C [102] enthaltenen Tabellen.

5.2.2.1 Programm der Eignungsprüfung

Analog zur Eignungsprüfung „Härteprüfung nach Vickers“, vgl. Kapitel 5.2.1, wurde den Teilnehmer als Probematerial eine zertifizierte Härtevergleichsplatte der MPA NRW Dortmund zur Verfügung gestellt. Auf dieser Härtevergleichsplatte sollten an vorgegebenen Prüferten je fünf Härtemessungen HBW 2,5/187,5 durchgeführt werden. Um den Laboratorien Gelegenheit zu geben, die Optik der eingesetzten Prüfmaschinen zu überprüfen, wurden auf der Härtevergleichsplatte durch die MPA NRW Dortmund Referenzeindrücke eingebracht. Die Eignungsprüfung wurde in aufeinander folgenden Runden durchgeführt, so dass auf jeder Härtevergleichsplatte bis zu sieben Laboratorien geprüft haben.

Die Eignungsprüfung ergab für die einzelnen Kriterien die in Tabelle 5.2.2.3 zusammengestellten Ergebnisse:

Tabelle 5.2.2.3: Härteprüfung nach Brinell, Ergebnisse der Eignungsprüfung [28]

	Anzahl TN	Positive Ergebnisse
Grenzabweichung der Prüfmaschine	90	82%
Wiederholpräzision der Prüfmaschine	85	97%

Darüber hinaus wurde die Durchführung der Prüfung hinsichtlich der Normvorgaben (kleinster Messschritt, Lastaufbringung und –einwirkung) betrachtet. Dabei stellte sich heraus, dass 33 Laboratorien zum kleinsten Messschritt entweder keine Angaben machen konnten oder nach eigenen Angaben die Vorgaben der Norm nicht einhalten. Zehn Laboratorien halten nach eigenen Angaben die Zeitvorgaben für die Lastaufbringung bzw. –einwirkung nicht ein.

Insgesamt zeigt das Ergebnis der Eignungsprüfung die überwiegend gute Qualität der Durchführung der Härteprüfung nach Brinell gemäß DIN EN ISO 6506-1 [102] in Europa. Die auf die Prüfnorm bezogenen Laborergebnisse lassen den Schluss zu, dass die Ausbildung der Prüferinnen und Prüfer, der Zustand und die Kalibrierung der eingesetzten Prüfgeräte ein hohes Niveau erreicht haben. Als Grund für die Prüfqualität kann die regelmäßige Kalibrierung der Geräte und die von den Gutachtern geforderten Härtevergleichsplatten zur regelmäßigen indirekten Prüfung der Geräte angeführt werden.

Die überprüften Ergebnisse, die nach den Angaben der Laboratorien außerhalb der zulässigen Grenzen liegen, ergab, dass häufig Messfehler vorliegen. Die von der MPA NRW Dortmund nachgemessenen Eindrücke hätten in einigen Fällen die Normkriterien erfüllt. Dort liegt ein Verbesserungspotential für die Teilnehmer.

Die Durchführung und Auswertung der Eignungsprüfung HBW 2001 [28] führt im Rahmen dieser Arbeit zu keiner neuen Erkenntnis, die über die im Kap. 5.2.1 bereits diskutierten Punkte hinausgeht.

Als zusätzlich Nutzenanwendung dieser Eignungsprüfung wurde die Messunsicherheit der Prüflaboratorien für dieses Prüfverfahren nach UNCERT Code of Practice No. 14 (COP 14) [103] bestimmt. Zu diesem Zweck erhielten die Teilnehmer ein handelsübliches Material mit ähnlicher Härte wie die Härtevergleichsplatte, an dem sie die Härte bestimmen sollten.

5.2.2.2 Bestimmung der Messunsicherheit des Prüfverfahrens

Bei allen Messungen muss man bei Versuchswiederholungen mit unterschiedlichen Ergebnissen rechnen, man spricht von einer Streuung der Messergebnisse. Oft werden mit Messergebnissen, aufgrund einer hohen Ablesegenauigkeit, Messgenauigkeiten vorgetäuscht, die gar nicht vorhanden sind. Daher ist eine der wichtigsten Bedingungen für einen erfolgreichen Messvorgang, über die Mess- und Aussagegenauigkeit der ermittelten Daten Bescheid zu wissen [104, 105].

Der besondere Wert der Eignungsprüfung HBW 2001 [28] liegt in der Verwendung der Ergebnisse zur Überprüfung und Beurteilung des Verfahrens zur Bestimmung der Messunsicherheit auf Basis des COP 14 [103]. Die Abschätzung der Messunsicherheit nach COP 14 [103, 106, 107] erfordert eine Messung auf einem Probematerial, welches möglichst keine Härtevergleichsplatte sein sollte.

$$U = 2 * \sqrt{u_E^2 + u_{\bar{X}_{CRM}}^2 + u_{CRM}^2 + u_H^2 + u_{MS}^2 + u_x^2} \quad (5.2.1)$$

$$\tilde{U} = \frac{U}{\bar{X}_{CRM}} * 100\% \quad (5.2.2)$$

mit

U	Erweiterte Messunsicherheit
\tilde{U}	Relative erweiterte Messunsicherheit
u_E	Standardmessunsicherheit aus der Grenzabweichung (1σ) der Härteprüfmaschine
$u_{\bar{X}_{CRM}}$	Messunsicherheit (1σ) der Härteprüfmaschine zum Kalibrieren der Härtevergleichsplatte
u_{CRM}	Standardmessunsicherheit der Härtevergleichsplatte
u_H	Standardmessunsicherheit der Härteprüfmaschine bei Messung auf der Härtevergleichsplatte
u_{MS}	Standardmessunsicherheit durch den kleinsten Ableseschritt
u_x	Standardmessunsicherheit bei der Prüfung eines Probematerials

Es lagen Messergebnisse von 88 Laboratorien vor. Für alle Laboratorien, welche die Anforderungen der DIN EN ISO 6506-2 [108] erfüllen, wurden die in Bild 5.2.2.2 in einer Übersicht zusammengestellten erweiterten Messunsicherheiten U für die Messung auf dem Probematerial nach den Gleichungen (5.2.1) und (5.2.2) berechnet. Der Rechenweg ist mit einem Beispiel in Tabelle 5.2.2.3 tabellarisch zusammengefasst.

Tabelle 5.2.2.3: Berechnung von $U = 2 * \sqrt{u_E^2 + u_{\bar{X}_{CRM}}^2 + u_{CRM}^2 + u_{\bar{H}}^2 + u_{ms}^2 + u_x^2}$

Beispiel: Härtevergleichsplatte mit einem Mittelwert von **246,8 ± 1,5 HBW 2,5/187,5**. Die Einzelwerte der Kalibrierung betragen 246,9 - 245,8 - 246,3 - 247,9 - 247,0 HBW 2,5/187,5; die resultierende Standardabweichung ist $s_{X_{CRM}} = 0,79$ HBW 2,5/187,5.

Rechenschritt	Beschreibung	Formelzeichen	Formel	Fundort / Nachweis	Beispiel [..] = HBW 2,5/187,5
1	Standardmessunsicherheit aus der Grenzabweichung (1σ) der Härteprüfmaschine	u_E	$u_E = \frac{u_{E,2\sigma}}{2,8}$	$u_{E,2\sigma}$ nach DIN EN ISO 6506-2, Tabelle 2	$u_E = \frac{0,02 * 246,8}{2,8} = 1,763$
2	Messunsicherheit (1σ) der Härteprüfmaschine zum Kalibrieren der Härtevergleichsplatte	$u_{\bar{X}_{CRM}}$	$u_{\bar{X}_{CRM}} = \frac{u_{\bar{X}_{CRM,2\sigma}}}{2}$	$u_{\bar{X}_{CRM,2\sigma}}$ gemäß Kalibrierschein der Härtevergleichsplatte	$u_{\bar{X}_{CRM}} = \frac{1,5}{2} = 0,75$
3	Mittelwert und Standardabweichung der Messungen beim Kalibrieren der Härtevergleichsplatte	\bar{X}_{CRM} , $s_{X_{CRM}}$	$\bar{X}_{CRM} = \frac{1}{n} * \sum_{i=1}^n X_{i,CRM}$ $s_{X_{CRM}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_{i,CRM} - \bar{X}_{CRM})^2}$	X_i gemäß Kalibrierschein der Härtevergleichsplatte	$\bar{X}_{CRM} = 246,8$ $s_{X_{CRM}} = 0,79$
4	Standardmessunsicherheit der Härtevergleichsplatte	u_{CRM}	$u_{CRM} = \frac{t * s_{X_{CRM}}}{\sqrt{n}}$	t=1,15 (Student-Faktor) für n=5 und α=68,3%	$u_{CRM} = \frac{1,15 * 0,79}{\sqrt{5}} = 0,406$
5	Mittelwert und Standardabweichung der Messungen auf der Härtevergleichsplatte	\bar{H} , s_H	$\bar{H} = \frac{1}{n} * \sum_{i=1}^n H_i$ $s_H = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (H_i - \bar{H})^2}$	H_i gemäß DIN EN ISO 6506, Teil 2, Abs. 5.6	Einzelmessungen: 246,0–245,0–246,0–246,0–246,0 $\bar{H} = 245,8 / s_H = 0,45$
6	Standardmessunsicherheit der Härteprüfmaschine bei Messung auf der Härtevergleichsplatte	$u_{\bar{H}}$	$u_{\bar{H}} = \frac{t * s_H}{\sqrt{n}}$	t=1,15 (Student-Faktor) für n=5 und α=68,3%	$u_{\bar{H}} = \frac{1,15 * 0,45}{\sqrt{5}} = 0,23$

Rechenschritt	Beschreibung	Formelzeichen	Formel	Fundort / Nachweis	Beispiel [..] = HBW 2,5/187,5
7	Mittelwert und Standardabweichung bei der Prüfung eines Probematerials	\bar{x} , s_x	$\bar{x} = \frac{1}{n} * \sum_{i=1}^n x_i$ $s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$	x_i : 5 Messungen auf Probematerial	Einzelmessungen: 288,0–290,0–285,0–285,0–282,0 $\bar{x} = 286,0$ $s_x = 3,08$
8	Standardmessunsicherheit bei der Prüfung eines Probematerials	u_x	$u_x = \frac{t * s_x}{\sqrt{n}}$	t=1,15 (Student-Faktor) für n=5 und $\alpha=68,3\%$	$u_x = \frac{1,15 * 3,08}{\sqrt{5}} = 1,584$
9	Standardmessunsicherheit durch den kleinsten Ableseschritt	u_{ms}	$u_{ms} = \frac{ms}{\sqrt{3}}$	ms=1 HBW	$u_{ms} = \frac{1}{\sqrt{3}} = 0,577$
10	Berechnung der erweiterten Messunsicherheit	U	$U = 2 * \sqrt{u_E^2 + u_{X_{CRM}}^2 + u_{CRM}^2 + u_H^2 + u_x^2}$	Schritte 1 bis 8, $\alpha=95\%$	$U = 2 * \sqrt{1,76^2 + 0,75^2 + 0,39^2 + 0,23^2 + 0,577^2 + 1,59^2}$ $U = 5,13$
11	Ergebnisprotokollierung		$\bar{X} = \bar{x} \pm U$	Schritte 7 und 9 $\alpha=95\%$	$\bar{X} = 286,0 \pm 5,13$
ALTERNATIV					
12 A	Berechnung der erweiterten relativen Messunsicherheit	\tilde{U}	$\tilde{U} = \frac{U}{X_{CRM}} * 100\%$	Schritte 1 bis 9, $\alpha=95\%$	$\tilde{U} = \frac{5,13}{246,8} * 100\% = 2,1\%$
13	Ergebnisprotokollierung		$\bar{X} = \bar{x} \pm U$	Schritte 7 und 11A $\alpha=95\%$	$\bar{X} = 286,0 \text{ HBW } 2,5/187,5 \pm 2,1\%$

5.3 Eignungsprüfung Metallographie

Die Metallographie befasst sich mit der Herstellung von Schliffbildern, um aus diesen Aufschluss über den Gefügebau eines metallischen Werkstoffes zu erhalten und den Zusammenhängen zwischen Aufbau und Eigenschaften in physikalischer, chemischer, kristallographischer und technologischer Hinsicht [100, 109, 110]. Man unterscheidet makroskopische und mikroskopische^H Verfahren.

Die makroskopischen Verfahren dienen zur Identifizierung grundsätzlicher Gefügeunterschiede, wie z. B. Schweißnähten, Wärmebehandlungszonen oder zum Sichtbarmachen von Faserverläufen, der Verteilung von Seigerungen und anderem mehr [109]. Sie beruhen auf der Wirkung chemischer Reaktionen an der Oberfläche, die dazu geschliffen sein muss (Makroschliff).

Schliffe mit polierter Oberfläche werden im Lichtmikroskop untersucht. Die Verwendung des Lichtmikroskops bedingt eine sehr geringe Tiefenschärfe von etwa 0,1 µm bei 1000facher Vergrößerung. Daher muss eine vollkommen ebene, polierte Schliiffläche hergestellt werden, damit das Gefüge im Schliffbild einheitlich scharf erscheint (Mikroschliff). Im Mikroschliff enthüllt sich häufig die ganze Geschichte der mechanischen und thermischen Behandlungen, denen die Probe im Laufe der Fertigung und anschließend unterworfen war [110]. Die Proben werden in aufeinander abgestimmten Arbeitsgängen geschliffen und abschließend poliert. In den wenigsten Fällen sind die Gefügebestandteile bereits nach dem Polieren durch unterschiedliche Färbung zu erkennen (Grau- und Temperguß, Kupferlegierungen). Meistens muss das Gefüge durch Ätzen sichtbar gemacht werden. Die für unterschiedliche Metalle erforderlichen Ätzmittel findet man in der Literatur [89]. Das Ätzmittel kann nach DOMKE [109] auf mehrere Arten wirken:

- Unterschiedliche Färbung der Gefügebestandteile, z. B. durch Bildung kohlenstoffreicher, dunkler Niederschläge (Kornfärbung);
- Kornflächenätzung, hervorgerufen von verschiedenartigen Rauigkeiten durch unterschiedliche Orientierung der Kristallgitter;
- Korngrenzenätzung durch unterschiedliche Löslichkeit der Gefügebestandteile im Ätzmittel und anisotropem Verhalten der Kristallite gegenüber dem Ätzangriff (Reliefbildung);

^H Die zu den metallographischen Untersuchungsverfahren zählenden Verfahren der Elektronenmikroskopie werden hier nicht besprochen, da sie nicht zur Bestimmung von Korngrößen und Gefügebestandteilen eingesetzt werden.

- Ausbildung elektrochemischer Lokalelemente zwischen Kristallen und den unedleren Verunreinigungen auf den Korngrenzen, wodurch diese aufgelöst werden bevor der Angriff auf den Kornflächen sichtbar wird.

Der Mikroschliff ermöglicht die Identifizierung von Gefügebestandteilen, die Abschätzung des Volumenanteils dieser Gefügebestandteile und die Bestimmung der Korngröße [100].

Die Bestimmung der Korngröße nach DIN 50601, „Ermittlung der Ferrit- oder Austenitkorngröße von Stahl und Eisenwerkstoffen“ [111] ist ein in der Prüfpraxis häufig angewendetes metallographisches Prüfverfahren. Nach DIN 50601 [111] ist ein Korn ein einzelner räumlich ausgedehnter Kristallit innerhalb eines metallischen Werkstoffs mit einheitlich ausgerichteter Gitter. Ein Korn besteht aus den Phasen Austenit oder Ferrit. Beim Perlit wird ein Bereich gleicher Lamellenausrichtung als Kolonie bezeichnet. Eine solche Kolonie kann als Korn betrachtet werden, wenn nichts anderes vereinbart wird. In der Schnittebene eines Mikroschliffes erscheinen die einzelnen Körner als Flächen, die sich durch ausgeprägte Korngrenzen gegen die anders gerichteten Nachbarkörner abheben.

Die Korngrößenkennzahl G ist ein Kennwert für die Schnittfläche der in der metallographischen Schnittebene geschnittenen Körner. Zur Kennzeichnung ihrer Größe werden sie einer Serie von Vergleichsbildern zugeordnet oder es wird ihr mittlerer Durchmesser innerhalb einer Messstrecke bestimmt.

Im Rahmen der Eignungsprüfung Met 1/00 [26] konnten Prüfleistungen zu den drei folgenden Verfahren zur Bestimmung der Korngröße erbracht werden¹:

- (F1) Ferritisches Gefüge: Vergleich mit einer Bildreihentafel (Prüfung DIN 50601-Verfahren 1);
- (A2) Austenitisches Gefüge: Linienschnittverfahren (Prüfung DIN 50601-V2);
- (F4) Ferritisches Gefüge: Flächenausählverfahren (Prüfung DIN 50601-V4).

gestellt. Im Abbildungsmaßstab 100:1 sind die Gefügebilder I bis VIII so nummeriert, dass ihre Nummern der Korngrößen-Kennzahl G entsprechen.

Beim Vergleich mit einer Bildreihentafel (Verfahren 1) wird die Korngrößen-Kennzahl G durch Vergleich mit einer achtstufigen Bildreihe von Gefügebildern nach Abschnitt 7.1 der DIN 50601 [111] ermittelt. In Bild 5.3.1 sind die Gefügebilder II bis IV zur Veranschaulichung nebeneinander gestellt. Die Korngrößen-Kennzahl G leitet sich von der gezähl-

¹ Das Verfahren nach Snyder und Graff (Prüfung DIN 50601-V3) ist eine spezielle Anwendung des Linienschnittverfahrens zur Bestimmung der Austenitkorngröße von Schnellarbeitsstählen und Stählen mit besonders feinem Korn in gehärtetem und angelassenem Zustand, bei der eine Vergrößerung von 100:1 nicht ausreicht [111].

ten Anzahl m der auf einer Fläche von 1 mm^2 der metallographischen Schliiffebene vorhandenen Körner ab. Nach Definition ist $G=1$ für $m=16$; die anderen Indizes ergeben sich nach der Gleichung (5.3.1).

$$m = 8 \times 2^G \quad (5.3.1).$$

Grundlage einer Auswertung sind mindestens 20 Blickfelder in regelloser Anordnung oder in einem Rasternetz liegender Anordnung.

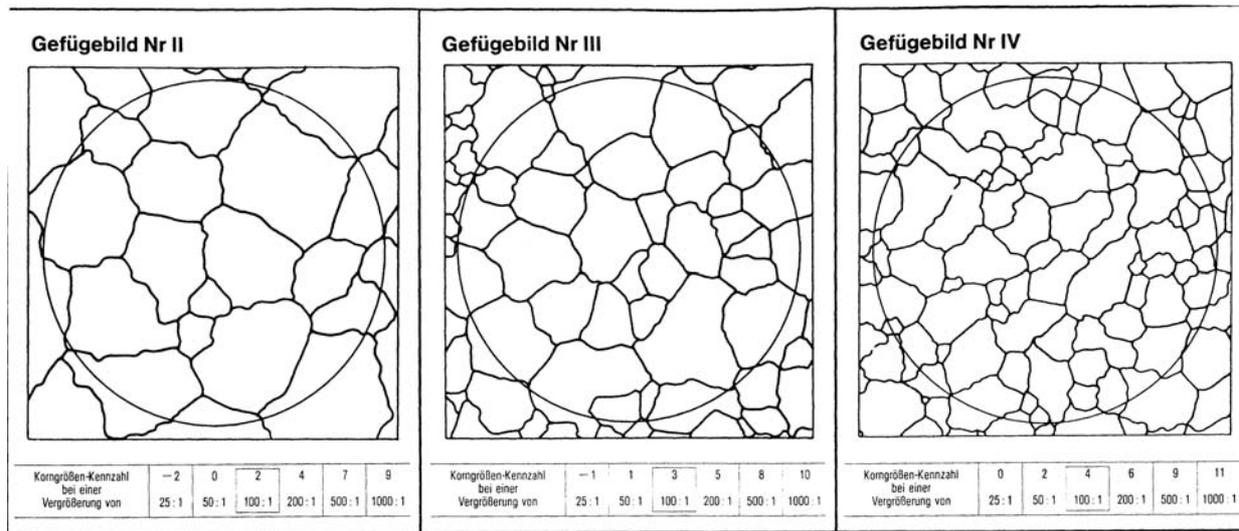


Bild 5.3.1: Gefügebilder II bis IV nach DIN 50601 [111]

Unter den Gefügebildern sind jeweils die Korngrößen-Kennzahlen angegeben, die sich bei einer von 100:1 abweichenden Vergrößerung ergeben. Die Korngrößen-Kennzahl G wird als ganze Zahl angegeben, da man bei diesem Verfahren selten eine Genauigkeit erreicht, die besser als ein ganzer Zahlenschritt der Kennzahlreihe ist. Das Verfahren ist für die meisten betrieblichen Einsatzzwecke ausreichend und wird daher in der Praxis den anderen Verfahren gegenüber bevorzugt, da es den geringsten Aufwand in der Anwendung fordert.

Beim Linienchnittverfahren (Verfahren 2) wird die von einer Messlinie durchquerte Anzahl Körner bestimmt und zwar durch Zählung im Okular, auf der Mattscheibe des Mikroskops oder auf einer photographischen Aufnahme eines für das Gefüge typischen Blickfeldes. Die Messstrecke ist normalerweise gerade, kann aber auch kreisförmig sein, üblicherweise wird eines der in Bild 5.3.2 dargestellten Messgitter eingesetzt.

Das Messgitter wird einmal je Blickfeld angewandt. In regelloser Verteilung wird eine ausreichende Anzahl von Blickfeldern ausgemessen. Mit diesem mittelnden Verfahren ist die Kennzeichnung von Mischkorn nicht möglich.

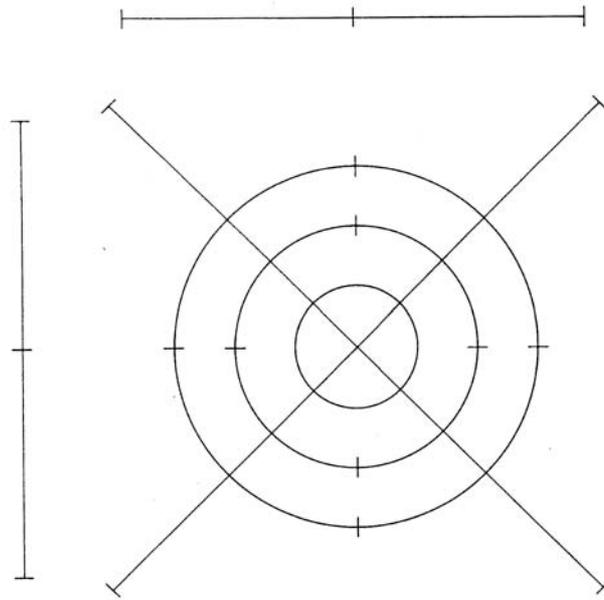


Bild 5.3.2: Messgitter für das Linienschnittverfahren (unmaßstäbliche Verkleinerung) [111]

Das Linienschnittsegment L_s ist der Abschnitt einer das Korn durchlaufenden Messlinie. Die Schnittpunkte dieser Messlinie mit den Korngrenzen werden als Linienschnittpunkte bezeichnet. Ist N_0 die Anzahl der Körner oder Korngrenzen, die von einer Linie der Länge L_0 in mm, gemessen auf der Probe, geschnitten werden, ergibt sich der Mittelwert des Linienschnitt-Segments \bar{L}_s wie folgt:

$$\bar{L}_s = \frac{L_0}{N_0} \text{ in mm} \quad (5.3.2).$$

L_0 und N_0 können sich als Summe mehrerer Einzelmessungen von Anzahlen N mit Messlinien L ergeben: $N_0 = \sum N$; $L_0 = \sum L$.

Beim *Linienschnittverfahren unter Verwendung von Geraden* setzt sich die gesamte Messstrecke aus den vier geraden Teilstücken nach Bild 5.3.2 zusammen, die eine Gesamtlänge von 500 mm haben. Die Längen werden unter der Berücksichtigung der Vergrößerung im Mikroskop in wahre Längen auf der Probe umgerechnet. Pro Blickfeld sollten mindestens 50 Linienschnittpunkte ausgewertet werden.

Verwendet man das *Kreisschnittverfahren*, besteht die Messstrecke aus allen drei konzentrischen Kreisen nach Bild 5.3.2 mit einer Gesamtlänge von 500 mm. Es müssen mindestens 50 Linienschnittpunkte pro Blickfeld ausgewertet werden. Wird nur der größte Kreis, der einen Umfang von 250 mm hat, angewandt, sind mindestens 25 Linienschnittpunkte pro Blickfeld erforderlich.

Das Flächenausählverfahren (Verfahren 4) stellt die theoretische Grundlage des Vergleichsverfahrens mit Bildreihen dar. Beim Flächenausählverfahren wird die Anzahl der Körner innerhalb des Kreises im Okular, auf der Mattscheibe des Mikroskops oder auf einer photographischen Aufnahme gezählt. Üblicherweise beträgt die Vergrößerung 100:1.

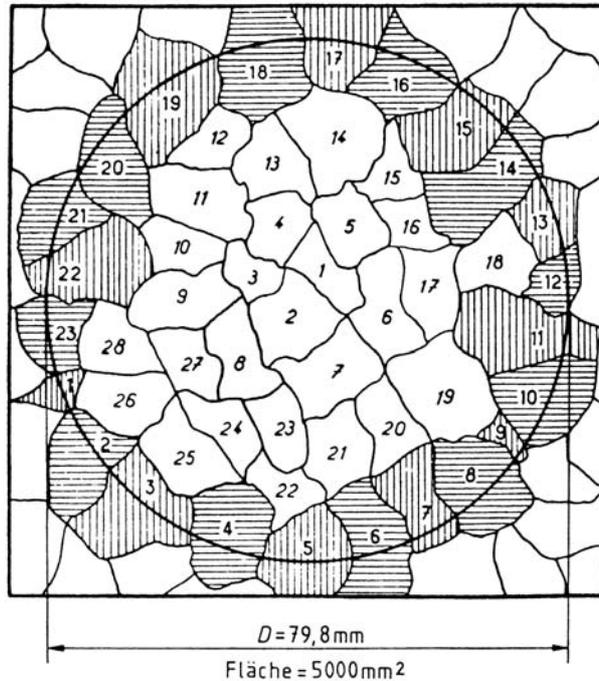


Bild 5.3.3: Beispiel für das Flächenausählverfahren [111]

In einem Kreis von 79,8 mm Durchmesser (Flächeninhalt = 5000 mm²) sollten mehr als 50 Körner gezählt werden können. Bei einer Vergrößerung von 100:1 entspricht dies einer wahren Größe der beobachteten Probenfläche von 0,5 mm². Man bestimmt die Anzahl der Körner n_1 , die vollständig innerhalb des Kreises liegen, sowie die Anzahl der Körner n_2 , die von der Kreislinie geschnitten werden. Für die Gesamtzahl der Körner gilt für die Zählung bei einer Vergrößerung g

$$n_g = n_{1(g)} + \frac{1}{2} n_{2(g)} \quad (5.3.3).$$

Ist A die ausgewertete Fläche, so ist die Anzahl der Körner je mm² Schnittfläche

$$m = \frac{n_g \cdot g^2}{A} \text{ in mm}^{-2} \quad (5.3.4).$$

Bei der Vergrößerung 100:1 und der Wahl des oben angegebenen Kreises ist $A=5000 \text{ mm}^2$ und die Gesamtzahl der Körner n_{100} . Damit ergibt sich

$$m = 2 n_{100} \text{ in mm}^{-2} \quad (5.3.5).$$

Jedem Wert der Korngrößen-Kennzahl G entspricht ein Nennwert von m . Der nach Gleichung 5.3.4 berechnete Wert von m wird einem ganzzahligen Wert von G innerhalb der in DIN 50601, Tabelle A.2 [111] angegebenen Grenzwerte zugeordnet.

Für das Beispiel aus Abbildung 5.3.3 ergibt sich mit

$$g = 100 \text{ (Vergrößerung 100:1)}$$

$$n_1=28 \text{ Körner, } n_2=23 \text{ Körner, } A=5000 \text{ mm}^2$$

$$n_g = 39,5 \text{ mm}^{-2} \quad (\text{s. 5.3.3})$$

$$m = 79 \text{ mm}^{-2} \quad (\text{s. 5.3.5})$$

$$G = 3 \quad (\text{s. DIN 50601, Tab. A.2}).$$

5.3.1 Untersuchungsprogramm

Als Leistungskriterium, nach dem die Laboratorien in der Eignungsprüfung Metallographie Met 1/00 [26] beurteilt werden, eignet sich nur die Abweichung des Laborwertes vom Stoffwert. Auf die im Rahmen der Eignungsprüfung gewonnenen Daten wurden die in den Tabellen 5.3.1 bis 5.3.3 gekennzeichneten Methoden angewendet, kritisch untersucht und bewertet.

Tabelle 5.3.1: Bestimmung des Stoffwertes X

Methode (nach Kapitel 4.2)	Prüfaufgabe		
	F1	F4	A2
Bekannter Wert (Methode A)	-	-	-
Zertifiziertes Referenzmaterial (Methode B)	-	-	-
Referenzmaterial (Methode C)	-	-	-
Konsenswert Experten (Methode D)	X	X	-
Konsenswert Teilnehmer (Methode E)	X	X	X

Tabelle 5.3.2: Berechnung der Standardabweichung der Eignungsprüfung $\hat{\sigma}$

Methode (nach Kapitel 4.4.1)	Prüfaufgabe		
	F1	F4	A2
Vorgeschriebener Wert	X	X	-
Festlegung eines erwünschten Leistungsniveaus	-	-	-
Präzisionsexperiment nach DIN ISO 5725-2	-	-	X
Konsenswert, robuste Datenauswertung	-	-	X

Tabelle 5.3.3: Berechnung der Leistungsstatistik

Methode (nach Kapitel 4.4.2)	Prüfaufgabe		
	F1	F4	A2
Schätzung der systematischen Abweichungskomponente D	X	X	-
Prozentuale Differenz D%	-	-	-
z-Scores	-	-	X
Histogramm	X	X	-
Säulendiagramm	-	-	-

5.3.2 Programm der Eignungsprüfung

Die Teilnehmer der Eignungsprüfung erhielten je einen Ausdruck einer Photographie eines ferritischen sowie eines austenitischen Gefüges in der Vergrößerung 100:1. An diesen Gefügebildern war die Korngröße nach DIN 50601 [111] zu bestimmen. Die DIN 50601 [111] stimmt, wie eine Reihe anderer in Europa gebräuchlicher Normen, z. B. UNI 3245 (1987), EURONORM 103 (1971), BS 4490 (1989) mit der ISO 643 (09.83) [114] überein. Die Durchführung der Prüfung nach diesen Normen war freigestellt, musste aber im Prüfbericht angegeben werden.

Als zusätzliche Aufgabe erhielten die Laboratorien einen zu präparierenden Schliff, an dem der Volumenanteil des Perlits am Gefüge ermittelt werden sollte. Das anzuwendende Prüfverfahren war den Laboratorien freigestellt, z. B. nach ASTM E 562-99 [112] oder ISO 9042-88 [113] oder einem Hausverfahren.

5.3.3 Materialauswahl und Homogenitätsprüfung

Bei der Auswahl des Probenmaterials wurden folgende Punkte besonders berücksichtigt:

- die Gefügebilder sollten typische Gefüge zeigen, die in der Prüfpraxis häufig auftauchen;
- das Gefügebild Ferrit sollte eine einheitliche Korngröße aufweisen, um die Voraussetzungen zur Anwendung der Verfahren 1 und 4 nach DIN 50601 [111] einzuhalten;
- das Gefügebild Austenit sollte die Zählung von mehr als 50 Körner ermöglichen;
- das Material zur Bestimmung des Volumenanteils des Perlits sollte häufig in der Prüfpraxis vorkommen, damit die Präparation des Schliffes mit einem Standardverfahren durchgeführt werden kann. Außerdem sollte anhand des Schliffbildes eine Plausibilitätsbetrachtung für das errechnete Ergebnis durchgeführt werden können.

Die Verwendung photographischer Aufnahmen zur Anwendung der metallographischen Prüfverfahren ermöglicht, allen Laboratorien identische Proben zur Untersuchung zur Verfügung zu stellen. Eine Homogenitätsuntersuchung ist daher nicht erforderlich.

Als Werkstoff für die Prüfaufgaben F1 und F4 wurde Reinst Eisen ausgewählt. Das Schliffbild zeigt das rein ferritische Gefüge (Bild 5.3.4).

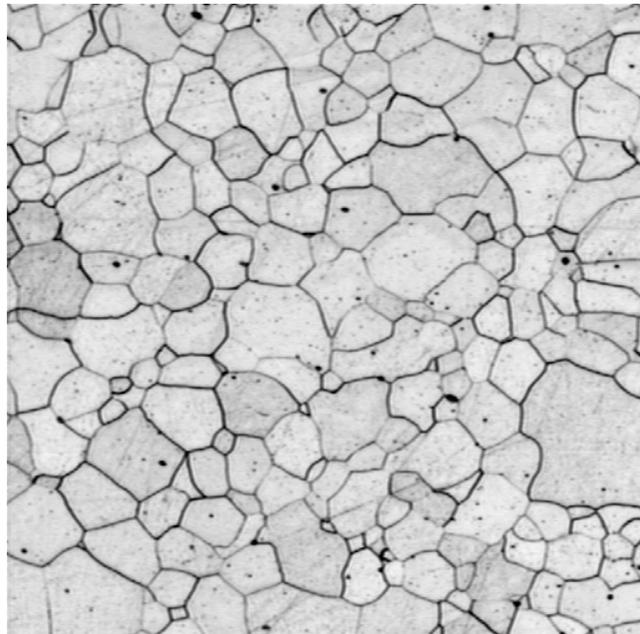


Bild 5.3.4: Ferritisches Gefüge; Reinst Eisen; unmaßstäbliche Vergrößerung aus [26]

Für die Prüfaufgabe A2 wurde der Werkstoff X5CrNi18-10 (1.4301) nach DIN 17440 [115] ausgewählt, dessen chemische Zusammensetzung Tabelle 5.3.4 wiedergibt. Das

austenitische Gefüge zeigt Bild 5.3.5. Da das Gefüge Zwillinge enthält, setzt die Auswertung nach dem Linienschnittverfahren beim Bearbeiter metallographische Fachkenntnisse voraus. Nach DIN 50601 [111] werden Zwillingsgrenzen nicht mitgezählt, und an einigen Stellen des Bildes ist die Entscheidung über Zwillings- oder Korngrenze nicht leicht zu treffen.

Tabelle 5.3.4: Schmelzenanalyse (Soll) des Werkstoffs X5CrNi18-10 (1.4301) [115]

Element	C	Si max.	Mn max.	P max.	S max.	N max.	Cr	Ni
Massenanteil in %	≤ 0,07	1,0	2,0	0,045	≤ 0,015	≤ 0,11	17,0-19,5	8,0-10,5

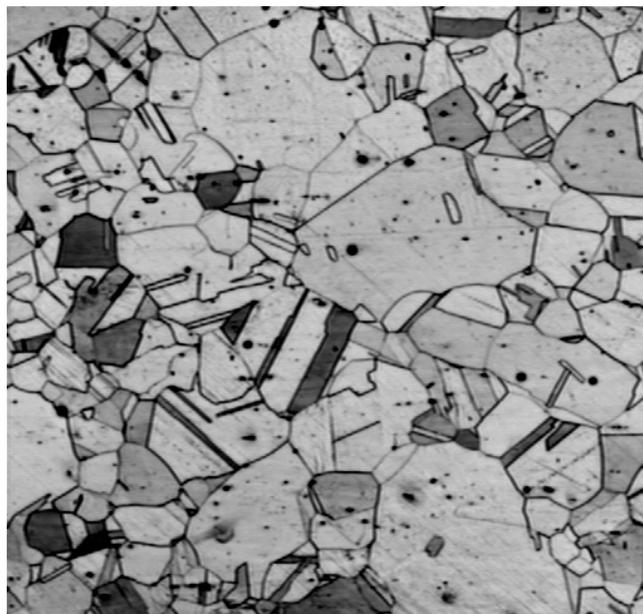


Bild 5.3.5: Austenitisches Gefüge; X5CrNi18-10 (1.4301) nach DIN 17440 [115]; unmaßstäbliche Vergrößerung aus [26]

Die für die Eignungsprüfung angefertigten Schlitze sind aus Stangenmaterial des Baustahls S235JO gemäß DIN EN 10025 [116] an der Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin angefertigt und dort einer Homogenitätsprüfung unterzogen worden. In Tabelle 5.3.5 ist die chemische Zusammensetzung des Materials wiedergegeben. Dieser Stahl ist vergleichbar mit dem Werkstoff St 37-3 U gemäß DIN 17100 [90].

Tabelle 5.3.5: Schmelzenanalyse (Soll) des Werkstoffs S235JO (1.0114) [116]

Element	C max.	Mn max.	P max.	S max.	N max.
Massenanteil in %	0,17	1,40	0,04	0,04	0,009

Die Homogenität des Materials wurde mittels Linienschnittverfahren nach DIN 50601 [111] an 5% des Materials bestimmt. Die durchgeführte Schlißpräparation entspricht der einer konventionellen vollautomatischen Probenpräparation, und als Ätzmittel wurde 2%ige

alkoholische Salpetersäure (Nital) verwendet, Bild 5.3.6. Das Ergebnis der Untersuchung ist in Tabelle 5.3.6 zusammengefasst.



Bild 5.3.6: Ferritisch-perlitisches Gefüge; S235JO (1.0114) nach DIN EN 10025 [116]; unmaßstäbliche Vergrößerung

Tabelle 5.3.6: Homogenitätsuntersuchung des Probenmaterials (Schliff)

Prüfnorm	Spanne der ermittelten Ergebnisse	Mittelwert	Standardabweichung
DIN 50601-V2	$\overline{L}_s = 0,0162 \text{ bis } 0,0172 \text{ mm}$	0,01678 mm	0,0004 mm

Die mittleren Linienschnittsegmente \overline{L}_s befinden sich alle in dem der Korngröße G=8 zugeordneten Intervall $0,0200 \text{ mm} \leq \overline{L}_s < 0,0141 \text{ mm}$ (DIN 50601, Tabelle A.2), bei einer Standardabweichung von $0,4 \mu\text{m}$. Die Standardabweichung als Maß für die Streuung setzt sich prinzipiell aus einer verfahrensbedingten Streuung (Gerät, Bediener, Probenfertigung) und einer werkstoffbedingten Streuung (Inhomogenität) zusammen. Die Inhomogenität der Korngröße bezüglich des Linienschnittsegments \overline{L}_s beträgt maximal $0,4 \mu\text{m}$. Daher kann für die Eignungsprüfung von einer ausreichenden Homogenität des Materials ausgegangen werden.

Gestalt und Volumen der Körner in einem Stahl- und Eisenwerkstoff sind niemals gleich. Selbst bei Gleichheit erscheint in der Schnittebene ein Gemenge verschieden großer Schnittflächen, weil die Körner an unterschiedlichen Stellen angeschnitten werden [111].

Die Verfahren nach ASTM E 562-99 [112] und nach ISO 9042-88 [113] sind Verfahren zur statistischen Abschätzung des Volumenanteils eines Gefügebestandteiles in einem

metallischen Werkstoff. Über einen Bildausschnitt im Mikroskop oder eine photographische Aufnahme eines Gefüges wird ein in den Normen definiertes Gitter gelegt. Die Anzahl der Gitterpunkte, die über dem zu untersuchenden Gefügebestandteil liegen wird gezählt und anschließend durch die Gesamtzahl der Gitterpunkte dividiert. Das Verfahren beruht auf der Erkenntnis, dass ein definiertes Gitter, welches systematisch über ein zweidimensionales Gefügebild gelegt wird, eine statistische Schätzung des Volumenanteils eines bestimmten Gefügeanteils ermöglicht, wenn eine ausreichende Anzahl von Auswertungen an repräsentativen Stellen des Gefüges vorgenommen wird [112].

5.3.4 Abweichung des Laborwertes vom Stoffwert

Im Folgenden werden die in Kapitel 4 vorgestellten Modelle zur Ermittlung des Stoffwertes X und zur Berechnung der Standardabweichung der Eignungsprüfung $\hat{\sigma}$ auf die Daten der Eignungsprüfung Met 1/00 [26] systematisch angewendet.

Prüfaufgabe F1 (Schliff Ferrit, Vergleichsverfahren)

Die Standardabweichung der Eignungsprüfung $\hat{\sigma}$ ergibt sich aus DIN 50601 [111], wonach die Auswertung nach dem Vergleichsverfahren selten eine Genauigkeit erreicht, die besser als ein ganzer Zahlenschritt der Kennzahlenreihe ist. Betrachtet man Bild 5.3.1, so ist der Unterschied zwischen Gefügebildern, die zwei oder mehr Korngrößen-Kennzahlen auseinander liegen so groß, dass er deutlich zu erkennen ist. Daher ist diese Festlegung technisch sinnvoll und wird für die Eignungsprüfung herangezogen, d.h. $\hat{\sigma} = 1G$.

Zur Bestimmung des Stoffwertes X wurden zwei Varianten untersucht und beurteilt, die Ergebnisse sind in Tabelle 5.3.7 zusammengestellt:

- Variante F1-1 (Methode D), Expertenlaboratorium: BAM, Berlin
- Variante F1-2 (Methode E), Konsenswert aus den Teilnehmerergebnissen

Die Beurteilung der Leistung erfolgt direkt über die Abweichung vom Stoffwert.

Tabelle 5.3.7: Prüfaufgabe F1, Abweichung vom Stoffwert: Vergleich der Varianten

	Stoffwert X	STABW $\hat{\sigma}$
Variante	Korngrößenkennzahl G	
F1-1	3	1
F1-2.1	3	1
F1-2.2	3	1

Für die Variante F1-1 ermittelte das Expertenlaboratorium der BAM, Berlin mittels Tafel B aus EURONORM 103-71 (identisch mit der Bildreihentafel nach DIN 50601 [111]) die Korngrößenkennzahl $X_{F1-1}=3$.

Zur Ermittlung des Stoffwertes nach Variante F1-2.1 werden nur Ergebnisse berücksichtigt, die normkonform angegeben worden waren. Ergebnisangaben wie $G=2$ bis 3 oder $G=2$ und 4 oder $G=3,5$ sind nicht normkonform. Die Korngröße wird als ganze Zahl angegeben. Außerdem bleiben in einem einzelnen Blickfeld bis zu 30 % Flächenanteil einer vom Mittelwert des Vergleichsbildes abweichenden Korngröße unberücksichtigt, so dass es bei dem vorliegenden Bild nicht notwendig ist, zwei oder mehr Korngrößenzahlen G anzugeben. In Bild 5.3.7 ist die Verteilung der Ergebnisse als Histogramm dargestellt.

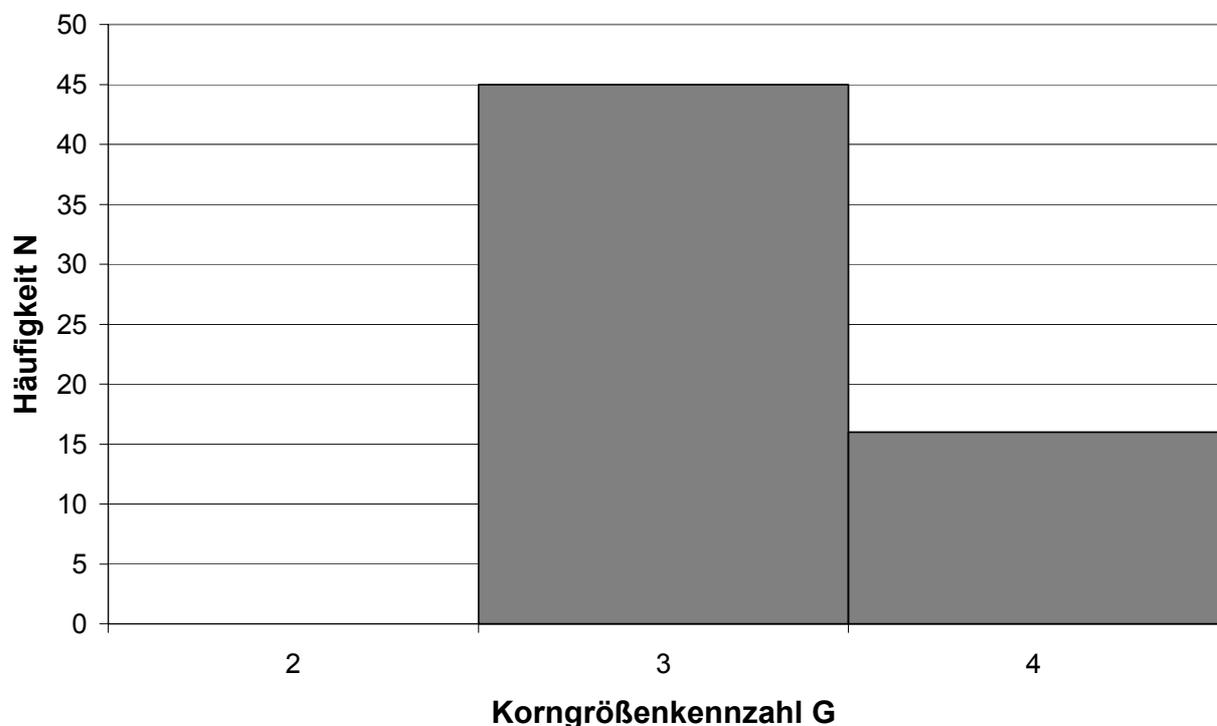


Bild 5.3.7: Variante F1-2.1, Photographie Ferrit: Verteilung der Ergebnisse

Es wurden die Ergebnisse von 61 Laboratorien berücksichtigt.

Da die Korngrößenkennzahl G immer eine ganze Zahl ist, ergibt die Berechnung des arithmetischen Mittelwertes oder des Medianes im vorliegenden Fall kein technisch sinnvolles Ergebnis. Ein weiteres Lagemaß ist der Modalwert (Modus, häufigster Wert), der die Ausprägung angibt, welche die größte Häufigkeit in einer Beobachtungsreihe besitzt [117]. Der mittels Modalwert ermittelte Stoffwertes $X_{F1-2.1}$ ergibt eine technisch sinnvolle

Aussage, und die Ermittlung des Wertes ist von den Teilnehmern problemlos nachzuvollziehen. Es ergibt sich ein Stoffwert $X_{F1-2.1} = 3$.

In Variante F1-2.2 werden alle eindeutigen Ergebnisse berücksichtigt, um bei der Ermittlung des Stoffwertes auf möglichst viele Teilnehmerergebnisse zurückgreifen zu können. Dazu werden die Ergebnisse, die entweder einen Zwischenwert (z. B. $G=3,5$) oder eine Spanne (z. B. $G= 3$ bis 4) angeben zu gleichen Teilen auf die direkt darüber bzw. darunter liegende Klasse aufgeteilt, d. h. das Ergebnis $G=3,5$ wird zur Hälfte in der Klasse $G=3$ und zur Hälfte in der Klasse $G=4$ berücksichtigt.

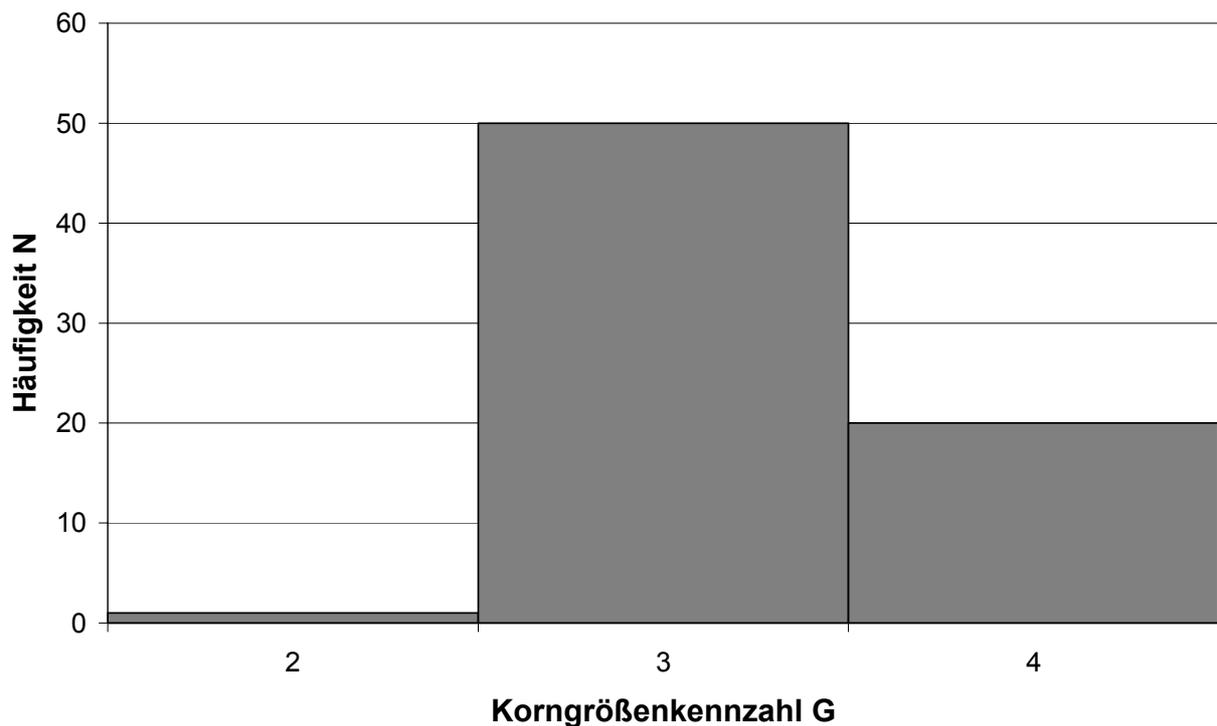


Bild 5.3.8: Variante F1-2.2, Photographie Ferrit: Verteilung der Ergebnisse

Bei dieser Vorgehensweise werden die Ergebnisse von insgesamt 71 Teilnehmern berücksichtigt. Bild 5.3.8 zeigt das Histogramm. Der Stoffwert $X_{F1-2.2}$ wird mittels Modalwert ermittelt, er ergibt sich zu $X_{F1-2.2} = 3$. Wie Tabelle 5.3.7 zu entnehmen ist, ergeben sich bei allen Varianten die selben Werte für X und $\hat{\sigma}$. Die Beurteilung der Ergebnisse erfolgt anhand der Festlegung eines oberen Akzeptanzlimits $G=4$ und eines unteren Akzeptanzlimits $G=2$. Diese Festlegung ist nachvollziehbar und technisch aussagekräftig.

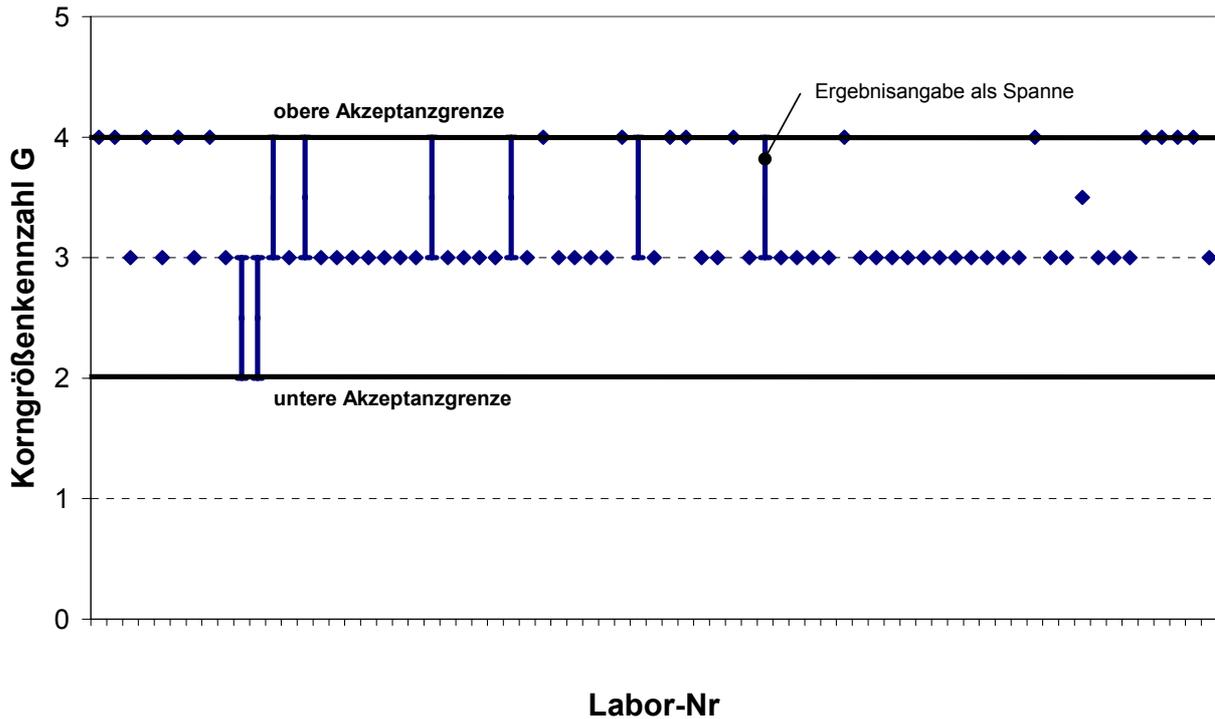


Bild 5.3.9: Prüfaufgabe F1, Photographie Ferrit: Bewertung der Ergebnisse

In Bild 5.3.9 sind die Ergebnisse der Prüfaufgabe F1 dargestellt. Alle Ergebnisse liegen innerhalb der vorgegebenen Akzeptanzgrenzen. An dieser Stelle zeigt sich deutlich, dass nicht alle Laboratorien mit den Vorgaben der Prüfnorm vertraut sind: Ein Laboratorium gibt zwei Werte ($G=2$ und 4 , in Bild 5.3.9 nicht berücksichtigt) an, acht Laboratorien geben eine Ergebnisspanne an und ein Laboratorium meldet den Wert $G=3,5$.

Prüfaufgabe F4 (Schliff Ferrit, Flächenauszählverfahren)

Für das Auszählverfahren gilt nach DIN 50601 [111] ebenfalls, dass die Auswertung selten eine Genauigkeit erreicht, die besser als ein ganzer Zahlenschritt der Kennzahlenreihe ist. Jeder Korngrößenkennzahl G ist in der Norm eindeutig eine Anzahl der Körner je $0,5 \text{ mm}^2$ zugeordnet, die Ergebnisangabe erfolgt als ganze Zahl. Diese Festlegung wird für die Standardabweichung der Eignungsprüfung $\hat{\sigma}$ übernommen, da sie technisch sinnvoll und in der Prüfpraxis erprobt ist, d. h. $\hat{\sigma}=1G$.

Zur Bestimmung des Stoffwertes X wurden zwei Varianten untersucht und beurteilt, die Ergebnisse sind in Tabelle 5.3.8 zusammengestellt:

- Variante 1 (Methode D), Expertenlaboratorium: BAM, Berlin
- Variante 2 (Methode E), Konsenswert aus den Teilnehmerergebnissen

Die Beurteilung der Leistung erfolgt direkt über die Abweichung vom Stoffwert.

Tabelle 5.3.8: F4, Abweichung vom Stoffwert: Vergleich der Varianten

	Stoffwert X	STABW $\hat{\sigma}$
Variante	Korngrößenkennzahl G	
F4-1	4	1
F4-2.1	3	1
F4-2.2	3	1

Das Expertenlaboratorium der BAM, Berlin ermittelte die Korngrößenkennzahl $X_{F4-1}=4$. Zur Ermittlung des Stoffwertes nach Variante F4-2.1 werden nur Ergebnisse berücksichtigt, die angegeben worden waren wie in DIN 50601 [111] vorgegeben. Danach ist die Korngröße immer eine ganze Zahl. In Bild 5.3.10 ist die Verteilung der Ergebnisse in Form eines Histogrammes dargestellt. Der Stoffwertes $X_{F4-2.1}=3$ ist der Modalwert aus den Ergebnissen von 24 Laboratorien.

Variante F4-2.2 berücksichtigt alle eindeutigen Ergebnisse, um bei der Ermittlung des Stoffwertes auf möglichst viele Teilnehmerergebnisse zurückgreifen zu können. Dazu werden die Ergebnisse, die entweder einen Zwischenwert (z. B. $G=2,75$) oder eine Spanne (z. B. $G=3$ bis 4) angeben, den richtigen Korngrößenkennzahlen gemäß DIN 50601, Tabelle A.2. [111] zugeordnet, d. h. das Ergebnis $G=3,5$ wird mit $G=3$ bewertet und bei der Bestimmung des Stoffwertes berücksichtigt. Bei dieser Vorgehensweise werden die Ergebnisse von insgesamt 32 Teilnehmern berücksichtigt. Diese Vorgehensweise widerspricht der Forderung, die Ergebnisse der Teilnehmer an Eignungsprüfungen unverändert zu bewerten. Sie lässt sich wie folgt rechtfertigen: nach Tabelle A.2. [111] entsprechen der Korngrößenkennzahl $G=3$ eine Menge von 24 bis 48 Körnern je $0,5 \text{ mm}^2$. Gibt ein Labor als Ergebnis z. B. $G=2,75$ an, so bedeutet dies, dass das Labor weniger Körner gezählt hat, als dem unteren Grenzwert von 24 Körnern je $0,5 \text{ mm}^2$ entsprechen, die normgerechte Korngrößenkennzahl ist also $G=2$.

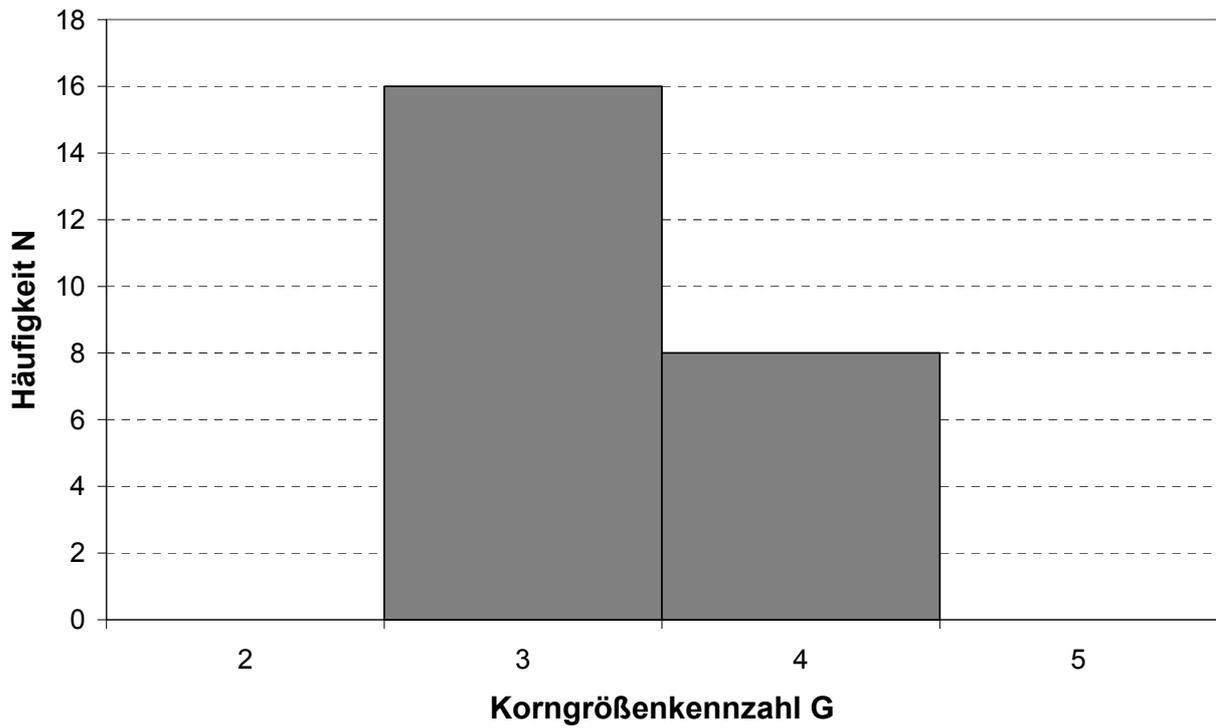


Bild 5.3.10: Variante F4-2.1, Photographie Ferrit: Verteilung der Ergebnisse

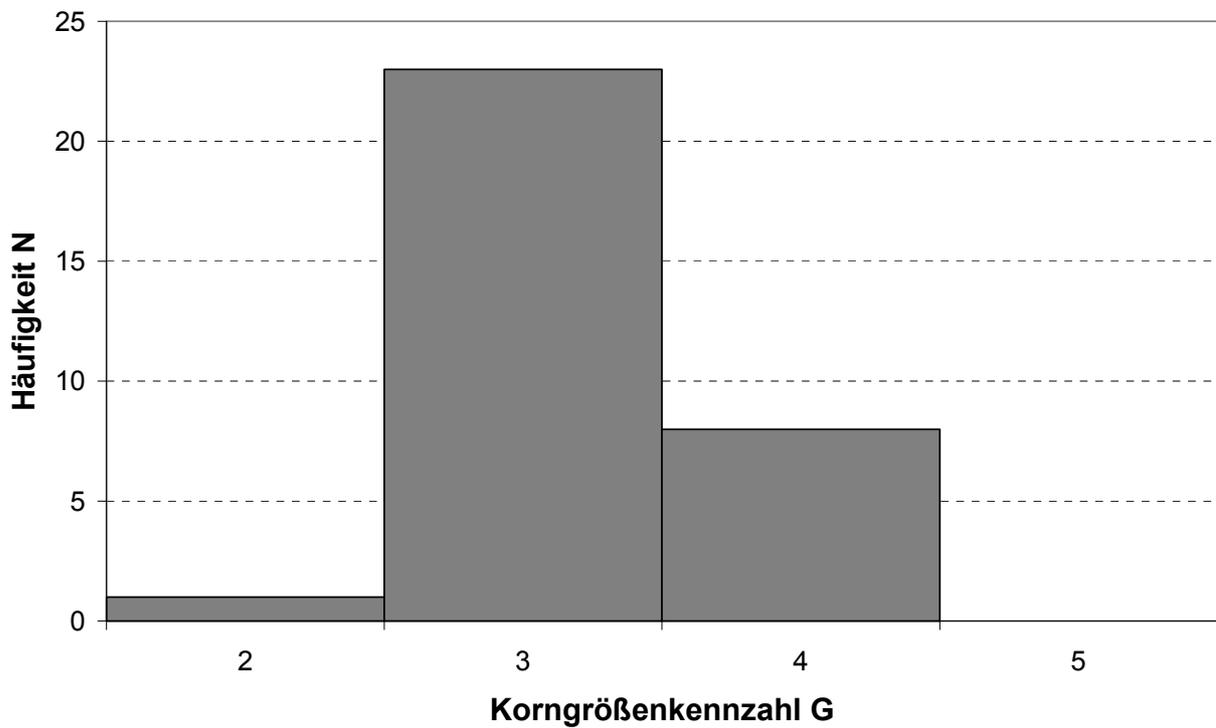


Bild 5.3.11: Variante F4-2.2, Photographie Ferrit: Verteilung der Ergebnisse

Die nicht normkonforme Angabe des Ergebnisses ist vermutlich auf das Bemühen des Laboratoriums zurückzuführen, es besonders gut zu machen, gepaart mit unzureichender Normenkenntnis. Bild 5.3.11 zeigt das Histogramm.

Der mittels Modalwert ermittelte Stoffwertes $X_{F4-2,2}=3$ ergibt eine technisch sinnvolle Aussage.

Wie Tabelle 5.3.9 zu entnehmen ist, ergibt sich für Variante F4-1 ein anderer Wert für X als bei den anderen Varianten. $\hat{\sigma}$ ist in allen Varianten gleich. Die Beurteilung der Ergebnisse erfolgt anhand der Festlegung eines oberen Akzeptanzlimits $X+1G$ und eines unteren Akzeptanzlimits $X-1G$ und basiert auf dem Konsenswert aus den Teilnehmerergebnissen. Diese Festlegung ist nachvollziehbar und technisch aussagekräftig. In Bild 5.3.12 sind die Ergebnisse der Prüfaufgabe F4 dargestellt.

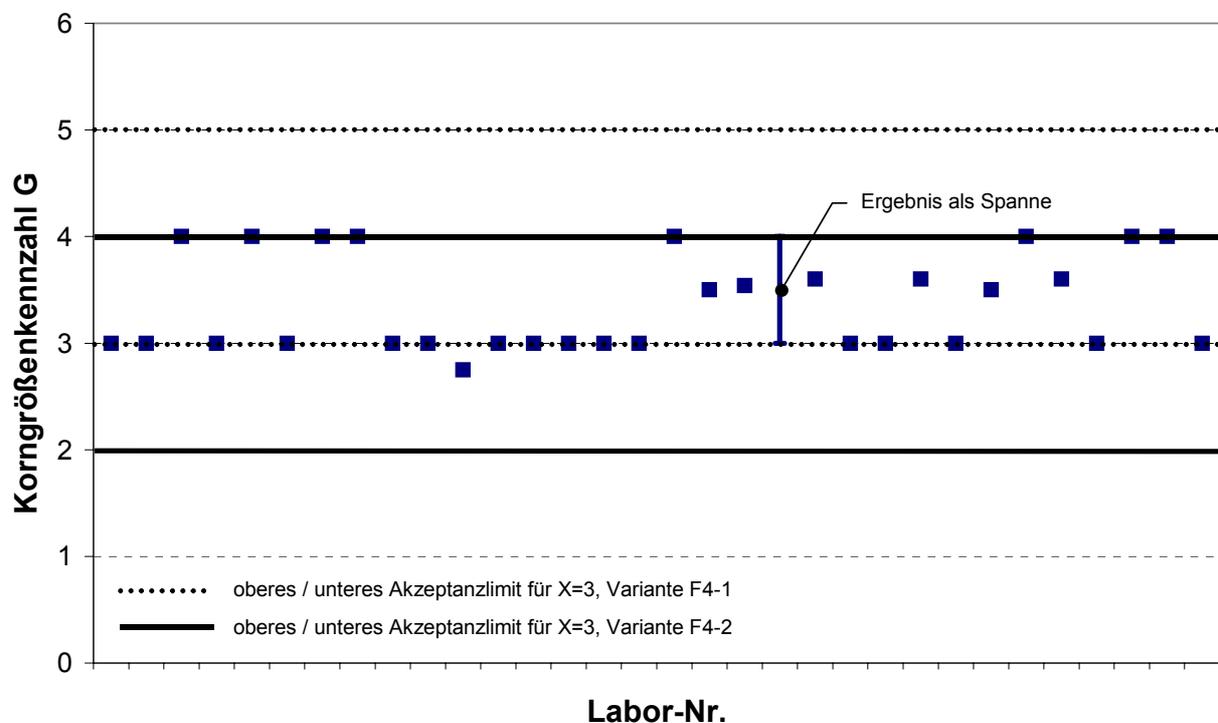


Bild 5.3.12: Prüfaufgabe F4, Photographie Ferrit: Bewertung der Ergebnisse

Verwendet man Variante F4-1, liegt ein Laboratorium außerhalb der Akzeptanzgrenzen. Gegen die Bestimmung des Stoffwertes mittels Expertenlaboratoriums spricht vor allem die Tatsache, dass es bislang keine Festlegungen gibt, welche konkreten Anforderungen ein Labor erfüllen muss, um Expertenstatus für ein bestimmtes Prüfverfahren zu erlangen. In diesem besonderen Fall gibt das zum Expertenlabor bestimmte Laboratorium für dieselbe Probe unterschiedliche Korngrößenkennzahlen an, obwohl das Flächenauszählverfahren die

Grundlage für das Vergleichsverfahren bildet und sich die Ergebnisse direkt in einander überführen lassen.

Daher wird die Bewertung der Eignungsprüfung mit den Werten der Varianten F4-2.1 bzw. F4-2.2 vorgenommen. Alle Laboratorien liegen innerhalb der Akzeptanzgrenzen. Wie in Bild 5.3.12 zu sehen, ist die Angabe des Prüfergebnisses bei 8 von 32 Teilnehmern nicht normkonform.

Prüfaufgabe A2 (Schliff Austenit, Linienschnittverfahren)

In Prüfaufgabe A2 sollte die mittlere Länge des Linienschnittsegments \bar{L}_s bestimmt werden. Die Beurteilung der Leistung erfolgt mittels z-Score (siehe Kapitel 4.4.2). Der Stoffwert in Prüfaufgabe A2 wird auf Grundlage eines Konsenswertes aus den Ergebnissen aller 56 Teilnehmer (Methode E, siehe Kap. 4.2) bestimmt. Es werden folgende Varianten gegenübergestellt:

- Variante A2-1: Berechnung von \bar{X} und $\hat{\sigma}$ mittels klassischer Statistik
- Variante A2-2: Berechnung von \bar{X} und $\hat{\sigma}$ nach Kap. 4.3.2

Die Ausreißerprüfung nach Cochran entfällt, da das Ergebnis aus einem Einzelwert besteht, d. h. Streuungen innerhalb eines Laboratoriums nicht untersucht werden können. Die Ausreißerprüfung nach Grubbs (Streuungen zwischen den Laboratorien) ergab einen Ausreißer. Außerdem wurden die Daten mittels Kolmogoroff-Smirnow-Test [69, 117] auf Vorliegen einer Normalverteilung überprüft. Mit einer statistischen Irrtumswahrscheinlichkeit $\alpha=10\%$ liegt eine Normalverteilung vor.

Variante A2-1:

Der als Ausreißer detektierte Messwert wurde zur Berechnung des Stoffwertes \bar{X} und der Standardabweichung der Eignungsprüfung $\hat{\sigma}$ aus dem Datensatz entfernt. Damit wird in dieser Variante der Stoffwert aus dem arithmetische Mittel aus den verbleibenden Daten bestimmt, $\bar{X}_{A2-1} = 0,127$ mm.

Die Standardabweichung der Eignungsprüfung $\hat{\sigma}$ ist gleich der Standardabweichung dieses Datensatzes, $\hat{\sigma}_{A2-1} = 0,024$ mm.

Variante A2-2:

Bei der Berechnung des Stoffwertes \bar{X} und der Standardabweichung $\hat{\sigma}$ wurden alle Daten beibehalten. Nach Gleichungen (4.3.26) bis (4.3.31) ergeben sich $\bar{X}_{A2-2} = 0,126$ mm bzw. $\hat{\sigma}_{A2-2} = 0,023$ mm.

Für beide Varianten wird unter Verwendung von Gleichung (4.4.6) für jedes Laboratorium der z-Score z berechnet:

$$z = \frac{x - X}{\hat{\sigma}}$$

In Tabelle 5.3.9 sind die beiden Varianten einander gegenübergestellt.

Tabelle 5.3.9: Abweichung vom Stoffwert, Warn- und Aktionssignale: Vergleich der Varianten

Variante	Stoffwert X	STABW $\hat{\sigma}$	Anzahl der Teilnehmer mit	
	in mm		Warnsignal $2 < z < 3$	Aktionssignal $ z \geq 3$
A2-1	0,127	0,024	3	1
A2-2	0,126	0,023	4	1

Bild 5.3.13 zeigt die Verteilung der Ergebnisse für die Variante A2-1. Die Verwendung von Variante A2-2 führt zu einer minimalen Verschiebung der Laborergebnisse gegenüber $z=0$ bei gleichzeitiger Aufspreizung der Kurve in x-Richtung, so dass der im Diagramm mit einem Pfeil gekennzeichnete Wert über die Grenze $z=3$ verschoben wird. Das Laboratorium erhält statt eines Warnsignals ein Aktionssignal.

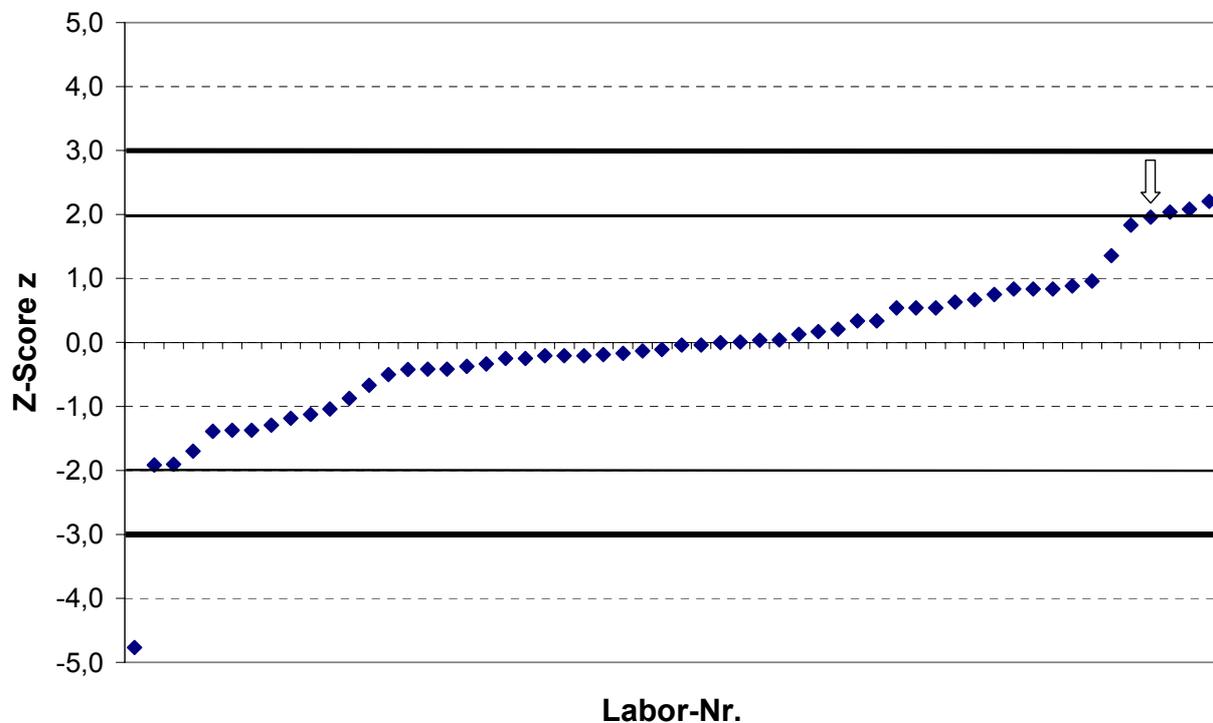


Bild 5.3.13: Prüfaufgabe A2, Variante A2-1 Photographie Austenit: Bewertung der Ergebnisse

Unter den vorliegenden Bedingungen (wenig Ausreißer, Daten sind normalverteilt) können beide Varianten als gleichwertig betrachtet werden. Variante A2-2 ist geringfügig schärfer. Laboratorien, deren Ergebnis sehr dicht an einer der Grenzwerte liegt, sollten im Grenzfall den für sie ungünstigeren Fall annehmen und auch dann nach den Ursachen für Abweichungen suchen und diese abstellen, wenn die Eignungsprüfung formal erfüllt ist.

Bestimmen des Volumenanteils Perlit am ferritischen Schliff

Die Bestimmung des Volumenanteils Perlit am ferritischen Schliff wurde von 77 Laboratorien durchgeführt. Es kamen mindestens 6 unterschiedliche Verfahren zum Einsatz. Wie Bild 5.3.14 zu entnehmen ist, kann man aufgrund der Daten keinen sinnvollen Konsenswert aus den Teilnehmerdaten ermitteln. Daher wird auf eine Bewertung der Ergebnisse verzichtet.

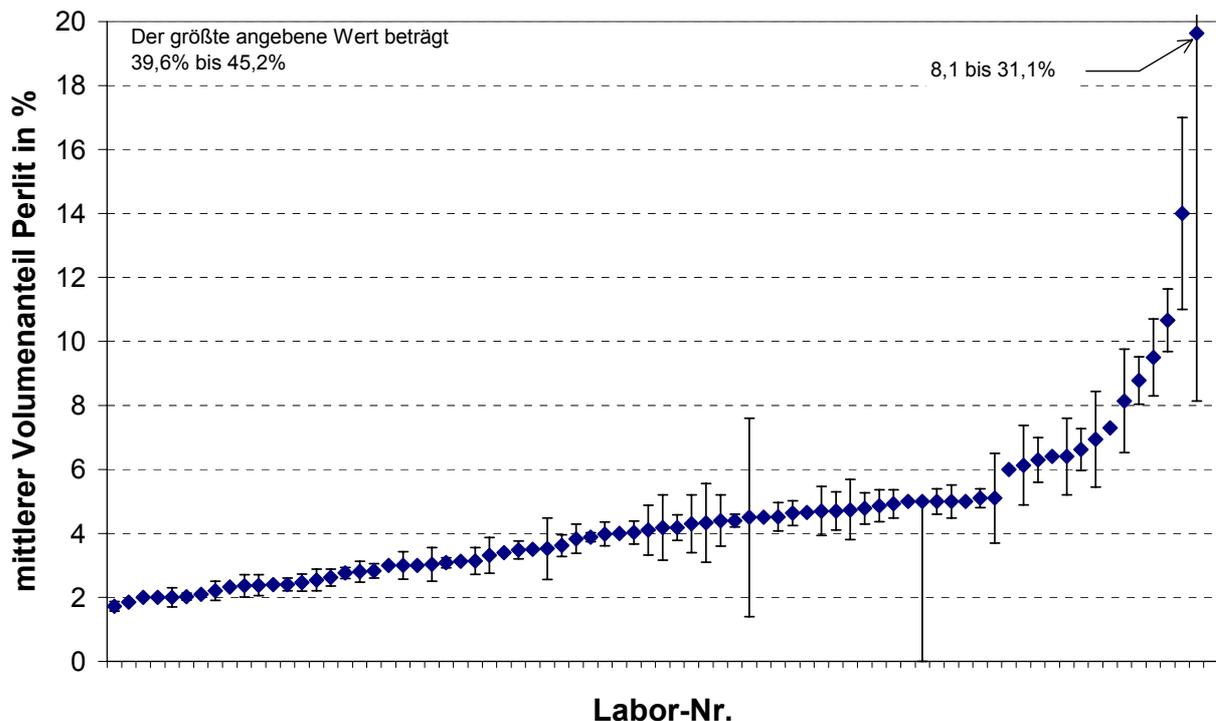


Bild 5.3.14: Bestimmung des Volumenanteils Perlit am Schliff, Ergebnisse

Die im Rahmen der Werkstoffauswahl durchgeführte Volumenbestimmung des Perlits nach ISO 9042-88 [113] durch die BAM, Berlin ergab einen Mittelwert von $2,7 \pm 0,2\%$.

Aus diesem Versuch kann man folgende Schlüsse ziehen:

- Die Fachkenntnisse in einzelnen Laboratorien sind nicht ausreichend. Als Versuchswerkstoff war ein im Laboralltag häufig vorkommender Werkstoff ausgewählt worden, um eine Plausibilitätsprüfung der gewonnenen Ergebnisse zu

ermöglichen. Diese Plausibilitätsbetrachtung ist zumindest von einigen Bearbeitern nicht durchgeführt worden, denn sonst wäre aufgefallen, dass ein Volumenanteil des Perlits von mehr als 10% nicht mit dem Gefügebild (Bild 5.3.15) in Einklang zu bringen ist. Dies ist außerdem ein deutlicher Hinweis auf ein nicht wirksames Qualitätssicherungssystem, das eigentlich verhindern soll, dass solche Ergebnisse an einen Kunden weitergereicht werden. Ein Laborkunde erhält in diesem Fall ein unbrauchbares und offensichtlich falsches Prüfergebnis.

- Die Prüfverfahren werden nicht ausreichend beherrscht; bei Anwendung von ASTM E 562-99 [112] werden Ergebnisse von 2,2% bis 42,5% ermittelt, bei Anwendung von ISO 9042-88 (E) [113] werden Ergebnisse von 2,0% bis 14,0% ermittelt, die übrigen Prüfverfahren erzielen Ergebnisse zwischen 1,9% und 8,8%.

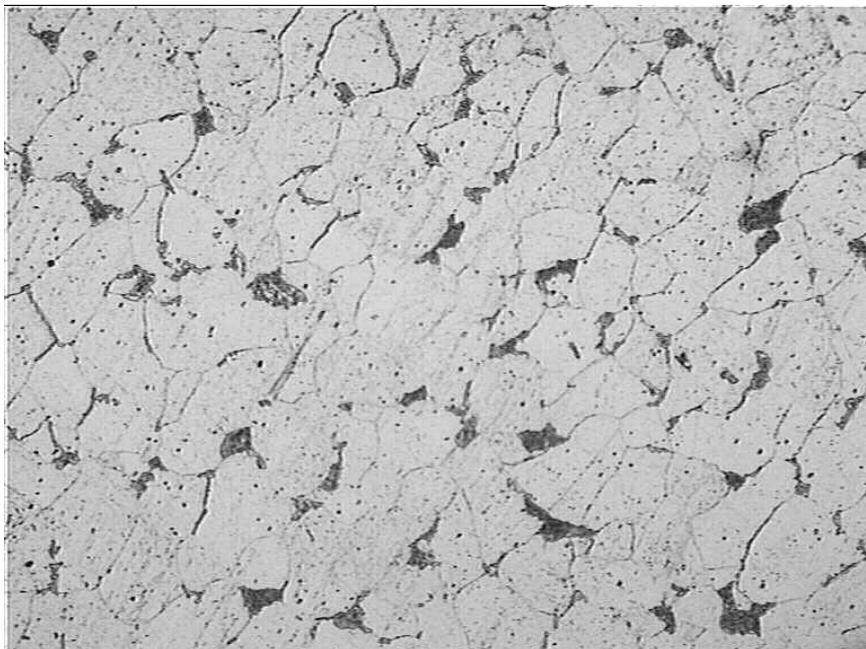


Bild 5.3.15: Schliff zur Bestimmung des Volumenanteils Perlit, $V = \text{ca. } 250:1$
Werkstoff: S235JO (\approx St 37-3)

5.4 Eignungsprüfung Kerbschlagbiegeversuch nach Charpy

Die Kerbschlagprüfung (engl. charpy impact test, CIT) eignet sich vorwiegend zur Feststellung der Trennbruchneigung eines Werkstoffes und für die Überwachung der Güte und Gleichmäßigkeit von Wärmebehandlungen [100]. Bei der Prüfung wird eine in der Mitte gekerbte Probe, die auf zwei Auflagern und gegen zwei Widerlager liegt, unter den in DIN EN 10045-1 [118] beschriebenen Bedingungen durch einen Pendelhammer mit einem einzigen Schlag durchschlagen. Die dabei verbrauchte Schlagarbeit A_V wird in Joule gemessen, Bild 5.4.1.

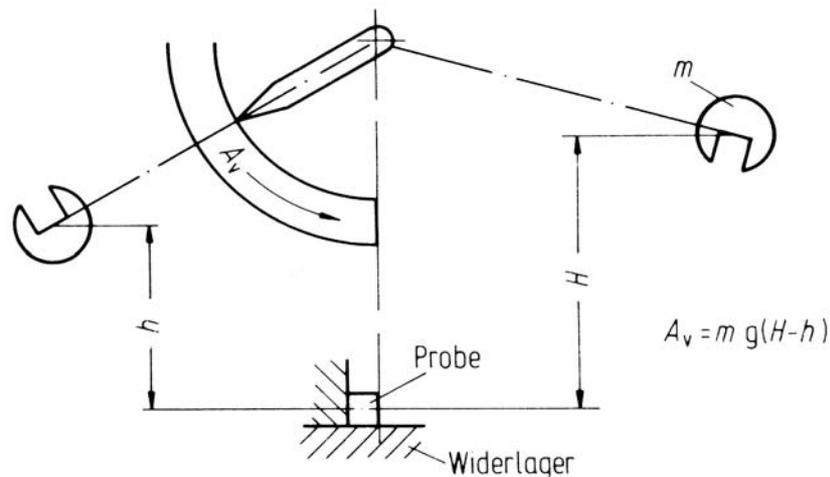


Bild 5.4.1: Schematische Kerbschlagarbeit-Temperatur-Kurve nicht-austenitischer Stähle nach [100]

Diese verbrauchte Schlagarbeit, die so genannte Kerbschlagarbeit, ist ein Maß für die Widerstandsfähigkeit der Werkstoffe gegen schlagartige Beanspruchung [118]. Die Kerbschlagarbeit kann nicht als Größe für die Festigkeitsberechnung verwendet werden. Sie ist stark von der Probenform abhängig, so dass Werte, die an Proben unterschiedlicher Geometrie ermittelt wurden, nicht miteinander verglichen werden können [100]. In der industriellen Anwendung wird die Kerbschlagarbeit in der Regel als Mittelwert von drei Einzelprüfungen angegeben.

Die besondere Bedeutung des Kerbschlagbiegeversuchs beruht auf einer Eigenschaft nicht-austenitischer Stähle: der im Bild 5.4.2 wiedergegebenen Abhängigkeit der Kerbschlagarbeit von der Prüftemperatur.

In der Hochlage treten duktile Brüche (Verformungsbrüche) auf, in der Tieflage spröde Brüche (Trennbrüche) und im Bereich des Steilabfalls vorwiegend Mischbrüche. Im Bereich des Steilabfalls streuen die gemessenen Werte der Kerbschlagarbeit stark, im Vergleich dazu treten in Hoch- und Tieflage geringere Streuungen der Messwerte auf [100].

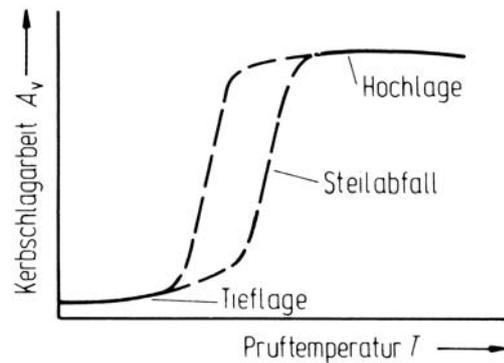


Bild 5.4.2: Schematische Kerbschlagarbeit-Temperatur-Kurve nicht-austenitischer Stähle nach [100]

LASKA und FELSCH geben in [119] schematisch die verschiedenen Einflussgrößen auf die Kerbschlagzähigkeit an, siehe Bild 5.4.3. Die Kerbschlagzähigkeit ist die auf die Probenfläche (im Kerbgrund) bezogene Kerbschlagarbeit.

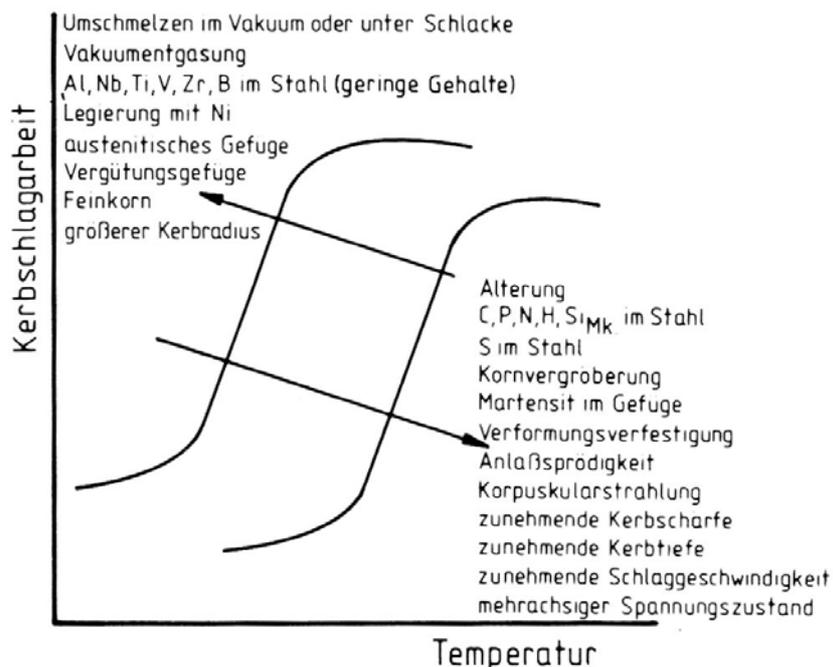


Bild 5.4.3: Schema der Einflussgrößen beim Kerbschlagbiegeversuch nach [119]

Die nach DIN EN 10045-1 [118] vorgegebene Normalprobe hat einen quadratischen Querschnitt von 10 mm Seitenlänge und ist auf allen Seiten bearbeitet, die Angaben für die zulässige Oberflächenrauigkeit sind in Bild 5.4.4 eingetragen. Die Probe besitzt in der Mitte einen Kerb, der einer der beiden in Bild 5.4.4 gezeigten Formen (U-Kerb, V-Kerb) entspricht. Die beiden Probenformen unterscheiden sich in der unterschiedlichen Kerbschärfe und Kerbtiefe. Die Proben unterscheiden sich zum einen in der Kerbgeometrie und zum anderen in der Höhe im Kerbgrund (Maß ④ in Bild 5.4.4), der für die Charpy-V-Probe 8 mm beträgt und für die Charpy-U-Probe nur 5 mm.

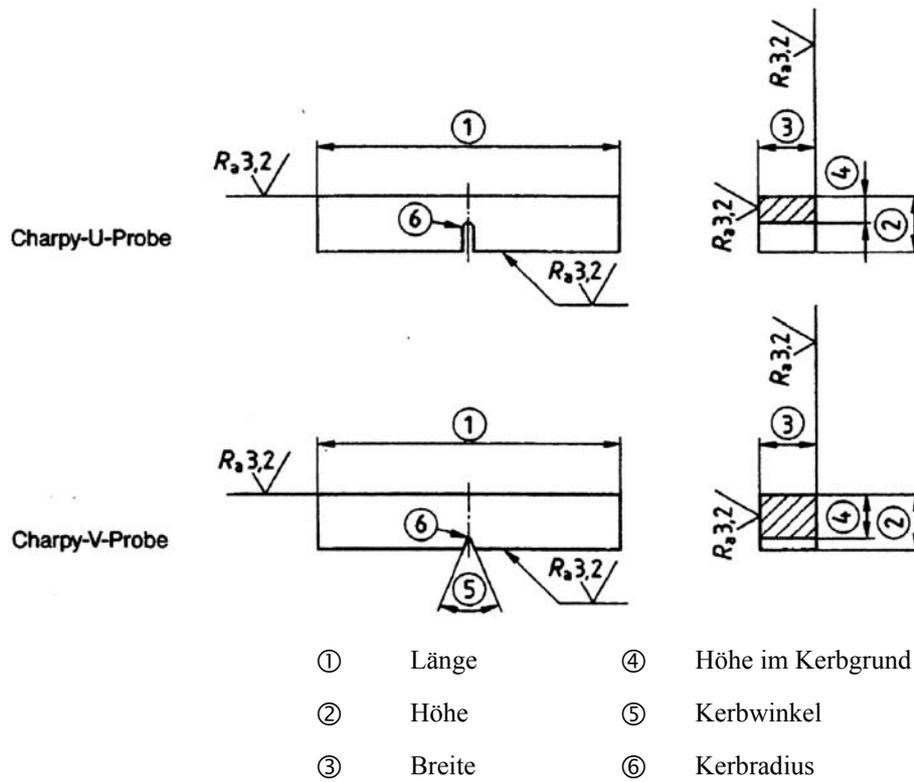


Bild 5.4.4: Probenformen für den Kerbschlagbiegeversuch nach Charpy aus [118]

5.4.1 Untersuchungsprogramm

Für die Eignungsprüfung CIT 2001 [29] wurden folgende Kriterien festgelegt, zu denen die teilnehmenden Laboratorien eine Leistung erbringen sollten:

- Abweichung des Mittelwertes vom Stoffwert (Kap. 5.4.4),
- Wiederholbarkeit (Kap. 5.4.5),
- Genauigkeit der Probenfertigung (Kap. 5.4.6).

Die Homogenitätsuntersuchung des Probenmaterials wurde nach den in Tabelle 5.4.1 gekennzeichneten Methoden vorgenommen. Auf die im Rahmen der Eignungsprüfung gewonnenen Daten wurden die in den Tabellen 5.4.2 bis 5.4.3 gekennzeichneten Methoden auf die Leistungskriterien (1) und (2) angewandt, anschließend kritisch untersucht und bewertet:

Tabelle 5.4.1: Homogenitätsuntersuchung

Methode	
Einfaktorielle Varianzanalyse (Kap. 4.1.1)	X
Nach ISO/DIS 13528:2002 (Kap. 4.1.2)	-
Andere Vorgehensweisen (Kap. 4.1)	X

Tabelle 5.4.2: Bestimmung des Stoffwertes X

Methode (nach Kapitel 4.2)	
Bekannter Wert (Methode A)	-
Zertifiziertes Referenzmaterial (Methode B)	-
Referenzmaterial (Methode C)	-
Konsenswert Experten (Methode D)	X
Konsenswert Teilnehmer (Methode E)	X

Tabelle 5.4.3: Berechnung der Standardabweichung der Eignungsprüfung $\hat{\sigma}$

Methode (nach Kapitel 4.4.1)	Kriterium	
	(1)	(2)
Vorgeschriebener Wert	-	-
Festlegung eines erwünschten Leistungsniveaus	X	-
Präzisionsexperiment nach DIN ISO 5725-2	X	-
Konsenswert, robuste Datenauswertung	X	-

In Tabelle 5.4.4 sind die untersuchten Verfahren zur Berechnung der Leistungsstatistik zusammengestellt.

Tabelle 5.4.4: Berechnung der Leistungsstatistik

Methode (nach Kapitel 4.4.2)	Kriterium	
	(1)	(2)
Schätzung der systematischen Abweichungskomponente D	-	X
Prozentuale Differenz D%	-	-
z-Scores	X	-
Histogramm	X	-
Säulendiagramm	-	-

5.4.2 Programm der Eignungsprüfung

Die Laboratorien erhielten Material zur eigenständigen Fertigung von maximal sieben Proben mit V-Kerb nach DIN EN 10045-1, Tabelle 2. Von diesen sieben Proben sollten sechs bei Raumtemperatur gemäß DIN EN 10045-1 [118] geprüft werden.

An der siebten Probe sollten die in DIN EN 10045-1, Tabelle 2 [118] vorgegebenen Probenmaße überprüft werden. Die Laboratorien sollten auf der Basis dieser Ergebnisse die Normkonformität ihrer Proben mit einer Ja- / Nein- Aussage bestätigen. Diese Probe wurde von einem akkreditierten Prüflabor des Materialprüfungsamt des Landes Brandenburg nachgemessen. Dieser Teil der Eignungsprüfung erlaubte den Teilnehmern, ihre Probenfertigung zu verifizieren.

5.4.3 Materialauswahl und Homogenitätsprüfung

Bei der Auswahl des Probenmaterials wurde auf folgende Punkte besonders geachtet:

- die von den Teilnehmern zu ermittelnde Kerbschlagarbeit muss sicher in der Hochlage liegen;
- die Wärmebehandlung des Probematerials muss sehr gleichmäßig sein (Homogenität);
- das Material soll gut zerspanbar sein;
- die Abmessungen des Vormaterials sollen möglichst nahe an den Hauptdimensionen der Kerbschlagprobe liegen.

Als Probematerial wurde ein Grobblech mit 15 mm Dicke aus dem legierten Edelstahl 16Mo3 (Werkstoff-Nr. 1.5415) entsprechend DIN EN 10028-2 [120] im Lieferzustand N (normalgeglüht) ausgewählt. Das Blech ist mit einem Abnahmeprüfzeugnis 3.1.B nach DIN EN 10204 [121] belegt.

Tabelle 5.4.5.1: Schmelzenanalyse (Soll) des Werkstoffs 16Mo3

Element	C	Si max.	Mn	P max.	S max.	Cr max.	Cu max.	Mo	Ni max.
Massenanteil in %	0,12- 0,20	0,35	0,40- 0,90	0,030	0,025	0,30	0,30	0,25- 0,35	0,30

Tabelle 5.4.5.2: Mechanische Eigenschaften (Soll) des Werkstoffs 16Mo3 (Querprobe)

Streckgrenze R_{eH} in N/mm ²	Zugfestigkeit R_m in N/mm ²	Bruchdehnung A in %	Kerbschlagarbeit bei 20 °C in J (Mittelwert aus 3 Proben)
275	440 bis 590	24	31

Das Zeugnis gibt Ergebnisse für zwei Probensätze mit je drei Einzelwerten zwischen 103 und 115 Joule an, die jeweiligen Mittelwerte betragen 105 bzw. 107 Joule, der Gesamtmittelwert beträgt 106,3 Joule und die Gesamtstandardabweichung 4,5 Joule.

Das Material wurde in Walzrichtung mittels Wasserstrahlschneidens in drei Probeabschnitte A, B, C aufgeteilt. Während der Vor-Erprobung stellte sich ein signifikanter Unterschied zwischen den Eigenschaften von Abschnitt A gegenüber denen der Abschnitte B und C heraus. Daher wurde für die Homogenitätsprüfung nur Material aus den Abschnitten B und C untersucht, das in der Folge bei der Eignungsprüfung verwendet wurde.

Die erste Runde der Homogenitätsprüfung wurde vom Landesmaterialprüfamt (LMPA) Sachsen-Anhalt in Magdeburg, der MPA / TU Darmstadt und der Bundesanstalt für Materialforschung und -prüfung (BAM) in Berlin durchgeführt. Die Homogenitätsprüfung an neun Probensätzen aus je sechs Einzelproben ergab Mittelwerte von 103 bis 118 Joule bei Standardabweichungen von 1,9 bis 4,9 Joule. Die abschließende Homogenitätsprüfung erfolgte bei der LMPA Sachsen-Anhalt, Magdeburg, die als Referenzlaboratorium diente. Ein Referenzlaboratorium verfügt über ein geeignetes Pendelschlagwerk, qualifiziertes Personal, ausreichend Erfahrung und ist in der Lage, die Kerbschlagproben mit hoher Genauigkeit herzustellen.

Die Prüfung erfolgte an einem zufällig ausgewählten Probesatz aus sechs Einzelproben. Die Einzelwerte reichen von 102 bis 107 Joule, der Mittelwert beträgt 104,8 Joule bei einer Standardabweichung von 1,9 Joule.

Wie das Ergebnis des Referenzlaboratoriums zeigt, ist die Inhomogenität des verwendeten Probenmaterials bezüglich der Kerbschlagarbeit kleiner oder gleich 1,9 Joule. Die Standardabweichung als Maß für die Streuung setzt sich prinzipiell aus einer verfahrensbedingten Streuung (Gerät, Bediener, Probenfertigung) und einer werkstoffbedingten Streuung (Inhomogenität) zusammen. Daher wurde die Homogenität des Materials bezüglich der

Kerbschlagarbeit aufgrund der Größe der Standardabweichungen beurteilt. Das Probenmaterial ist daher für die Eignungsprüfung CIT 2001 ausreichend homogen.

Homogenitätsprüfung mittels einfaktorieller Varianzanalyse

Zu Vergleichszwecken wurden aus Material des Abschnittes B zwölf Probesätze ausgearbeitet und geprüft. Die Ergebnisse der Prüfung sind in Tabelle 5.4.6 zusammengestellt. Diese Prüfergebnisse wurden der Homogenitätsprüfung mittels einfaktorieller Varianzanalyse (ANOVA) mit einer statistischen Irrtumswahrscheinlichkeit $\alpha = 5\%$ unterzogen. In Tabelle 5.4.7 ist das Ergebnis zusammengestellt. Da die Prüfgröße F den kritischen F-Wert nicht überschreitet, ist das Material als homogen zu betrachten.

Tabelle 5.4.6: Probenwerkstoff 16Mo3, Abschnitt B, Prüfergebnisse

Probe-Nr.	Schlagarbeit in Joule			
	Probe 1	Probe 2	Probe 3	Mittelwert
B2	105	115	110	110
B4	105	111	118	111
B6	110	118	111	113
B8	114	113	110	112
B10	112	117	120	116
B12	113	112	117	114
B20	122	118	119	120
B22	110	113	120	114
B24	118	110	109	112
B26	109	119	113	114
B28	108	110	109	109
B30	110	112	109	110

Tabelle 5.4.7: Homogenitätsprüfung mittels einfaktorieller Varianzanalyse

ZUSAMMENFASSUNG

Probe-Nr.	Anzahl	Summe	Mittelwert	Varianz
B2	3	330	110,0	25,0
B4	3	334	111,3	42,3
B6	3	339	113,0	19,0
B8	3	337	112,3	4,3
B10	3	349	116,3	16,3
B12	3	342	114,0	7,0
B20	3	359	119,7	4,3
B22	3	343	114,3	26,3
B24	3	337	112,3	24,3
B26	3	341	113,7	25,3
B28	3	327	109,0	1,0
B30	3	331	110,3	2,3

ANOVA

Streuungsursache	Quadratsummen (SS)	Freiheitsgrade (df)	Mittlere Quadratsumme (MS)	Prüfgröße (F)	P-Wert	kritischer F-Wert
Unterschiede zwischen den Gruppen	283,6	11	25,785	1,565	0,173	2,216
Innerhalb der Gruppen	395,3	24	16,472			
Gesamt	679,0	35				

5.4.4 Abweichung des Labormittelwertes vom Stoffwert

Im Folgenden werden die in Kapitel 4 vorgestellten Modelle zur Ermittlung des Stoffwertes X und zur Berechnung der Standardabweichung der Eignungsprüfung $\hat{\sigma}$ auf die Daten der Eignungsprüfung CIT 2001 systematisch angewendet. Insgesamt haben 107 Laboratorien ein Ergebnis gemeldet.

Auf die eingereichten Ergebnisse der Eignungsprüfung CIT 2001 wurden die Ausreißerprüfungen nach Grubbs und nach Cochran angewendet wie in DIN ISO 5725-2 [54] empfohlen. Es wurden ein Cochran-Ausreißer, ein Cochran-Straggler sowie ein Grubbs-Ausreißer identifiziert. Der Grubbs-Ausreißer beruht auf Nichteinhalten der Prüfspezifikation, das Laboratorium fertigte die Proben mit U-Kerb statt des vorgeschriebenen V-Kerbes. In Variante 3 wurden die Ausreißer aus dem Datensatz entfernt, in den Varianten 1, 2 und 4 wurden sie beibehalten.

Variante 1:

In Variante 1 wird der Stoffwert X_I als Konsenswert aus den Teilnehmerergebnissen (Kap. 4.2, Methode E) berechnet. Die Standardabweichung der Eignungsprüfung $\hat{\sigma}$ wird als erwünschtes Leistungsniveau festgelegt (Kap. 4.4.1).

Als Stoffwert X_I wird der *Median der Labormittelwerte* (MED_{LAB}) verwendet, d.h. ein robuster Konsenswert aus den Ergebnissen aller teilnehmenden Laboratorien. Der Stoffwert beträgt $X_I = 109,8$ Joule.

Die (robuste) Standardabweichung der Eignungsprüfung $\hat{\sigma}_1$ wurde aus dem *Median der Standardabweichungen* MED_{Stabw} (= 4,2 Joule) und der Forderung der DIN EN 10045-2, Tabelle 3 [122] hergeleitet. Es wurde der Wert $\hat{\sigma}_1 = 4$ Joule (der nächstgelegene, ganzzahlige Wert) festgelegt.

Bei dieser Vorgehensweise ergibt sich die größte zulässige Abweichung zu $3 \times \hat{\sigma}_1 = 12$ J und ist damit größer als die nach DIN EN 10045-2, Tabelle 3 [122] vorgesehene zulässige Abweichung von 10 % des Stoffwertes $X = 11$ J.

Variante 2:

Berechnung von X_2 (Methode E) und $\hat{\sigma}_2$ (Konsenswert) unter Verwendung der robusten Datenauswertung nach ISO/DIS 13528 [421] bzw. DIN ISO 5725-5 [431].

Die Rechnung nach Gleichungen (4.3.26) bis (4.3.31) ergibt die robusten Schätzer $x^* = X_2 = 110,0$ Joule und $s^* = \hat{\sigma}_2 = 6,3$ Joule. Die Werte konvergieren bereits nach dem ersten Iterationsschritt.

Variante 3:

Berechnung von X_3 und $\hat{\sigma}_3$ unter Verwendung der klassischen Datenauswertung nach DIN ISO 5725-2 [54].

Die Daten der Ausreißer nach Grubbs und Cochran wurden zur Berechnung der Präzisionskenngrößen aus dem Datensatz entfernt.

- $\hat{m} = \bar{x} = 110,0$ Joule nach Gl. (4.3.20),
- $s_r^2 = 25,16$ Joule nach Gl. (4.3.21),
- $s_L^2 = 42,99$ Joule nach Gl. (4.3.22 bis 4.3.24),
- $s_R^2 = 47,60$ Joule nach Gl. (4.3.25).

Setzt man diese Werte in Gleichung (4.4.4) ein, so erhält man

$$\hat{\sigma}_3 = \sqrt{42,99 + (25,16 / 6)}$$

$$\hat{\sigma}_3 = 8,30$$

Der Stoffwert $X_3=110,0$ J ist gleich dem arithmetischen Gesamtmittelwert \hat{m} .

Variante 4:

In Variante 4 wird der Mittelwert der LMPA Sachsen-Anhalt, Magdeburg, die als Referenzlaboratorium für die Homogenitätsprüfung diente, als Stoffwert verwendet, also $X_4=104,8$ Joule. $\hat{\sigma}_4$ berechnet sich aus der größten zulässigen Abweichung der Prüfmaschine nach DIN EN 10045-2, Tabelle 3 [122] (10 % vom Stoffwert) : $\hat{\sigma}_4=3,5$ J.

Die Leistung der Laboratorien wird mittels z-Score bewertet, der nach Gleichung (4.4.6) für jedes Laboratorium und jede Variante berechnet wird. Die Interpretation der z-Scores erfolgt wie in Kapitel 4.4.2, Abschnitt IV beschrieben. In Tabelle 5.4.8 sind die Auswirkung der Varianten 1 bis 4 auf das Ergebnis der Eignungsprüfung gegenübergestellt.

Tabelle 5.4.8: Abweichung vom Stoffwert, Warn- und Aktionssignale: Vergleich der Varianten

Variante	Stoffwert X	STABW $\hat{\sigma}$	Anzahl der Teilnehmer mit	
	in Joule		Warnsignal $2 < z < 3$	Aktionssignal $ z \geq 3$
1	109,8	4	16	11
2	110,0	6,3	7	2
3	110,0	8,3	2	1
4	104,8	3,5	25	22

In den Bildern 5.4.5 bis 5.4.8 ist die Verteilung der z-Scores für die verschiedenen Varianten in Form von Histogrammen gegenübergestellt.

Die Variante 4 gibt die schärfste Beurteilung der Eignungsprüfungsdaten wieder. Hier zeigt sich die Problematik, die auftritt, wenn die Bestimmung des Stoffwertes einem oder mehreren Expertenlaboratorien überlassen wird. Im vorliegenden Beispiel ist der Stoffwert X_4 um eine Größenordnung von $\hat{\sigma}$ kleiner als der Stoffwert X_1 , daher verschieben sich die z-Scores nach rechts, siehe Bilder 5.4.5 und 5.4.8.

Verschärft wird die Situation durch die Festlegung von $\hat{\sigma}_4$ auf Basis einer Normforderung, die streng genommen nur für zertifizierte Referenzproben gültig ist. Daher erhalten 23% der Teilnehmer ein Aktionssignal, weitere 20% erhalten ein Warnsignal (siehe Tabelle 5.4.8). Die Variante 4 ist als zu scharfe Beurteilung der Ergebnisse zu verwerfen.

Die Variante 3 gibt die „ideale“ Auswertung der Eignungsprüfung auf Grundlage der klassischen Statistik wieder. 1 Teilnehmer erhält ein Aktionssignal (<1%), 2 Teilnehmer erhalten ein Warnsignal (<2%). Die Verteilung der z-Scores nähert sich sehr gut der Normalverteilung an. Diese Variante ist jedoch unter dem Aspekt „Fitness-for-Purpose“ zu verwerfen. Wäre die Eignungsprüfung nach Variante 3 ($110 \text{ J} \pm 25 \text{ J}$ bzw. $110 \text{ J} \pm 22\%$) ausgewertet und bewertet worden, wären alle Ergebnisse zwischen 85 J und 135 J zumindest nicht „unzureichend“. Solche großen zulässigen Abweichungen wären in der praktischen Anwendung des Prüfverfahrens vor allem dann problematisch, wenn der Stoffwert nahe am Bereich des Steilabfalls liegt. Der Kerbschlagbiegeversuch wird standardmäßig bei Abnahmeprüfungen eingesetzt, z. B. nach Produktnormen, gesetzlichen Anforderungen oder Anwenderforderungen. Hier besteht von Seiten der Kunden und Lieferanten ein großes Interesse an einem möglichst präzisen Prüfverfahren.

Variante 2 steht zwischen Variante 3 und Variante 1. Für Variante 2 gilt in schwächerem Maße dasselbe wie für Variante 3. Die Verteilung der z-Scores nähert sich gut der Normalverteilung an, 2 Teilnehmer erhalten ein Aktionssignal.

Auch hier spricht vor allem der Aspekt „Fitness-for-Purpose“ gegen eine solche Form der Auswertung, da die zulässigen Abweichungen vom Stoffwert aus Kundensicht immer noch zu groß sind.

Die Variante 1 mit $\hat{\sigma}_1=4$ Joule berücksichtigt einerseits die Forderung der DIN EN ISO 10045-2 [122] ($\hat{\sigma}_{\text{NORM}}=3,7 \text{ J}$) und andererseits den Einfluss des Probenmaterials und die momentane Leistungsfähigkeit der Laboratorien, indem den Median der Standardabweichung der Laboratorien ($\hat{\sigma}_{\text{Med LAB}}=4,2 \text{ J}$) zur Beurteilung herangezogen wird.

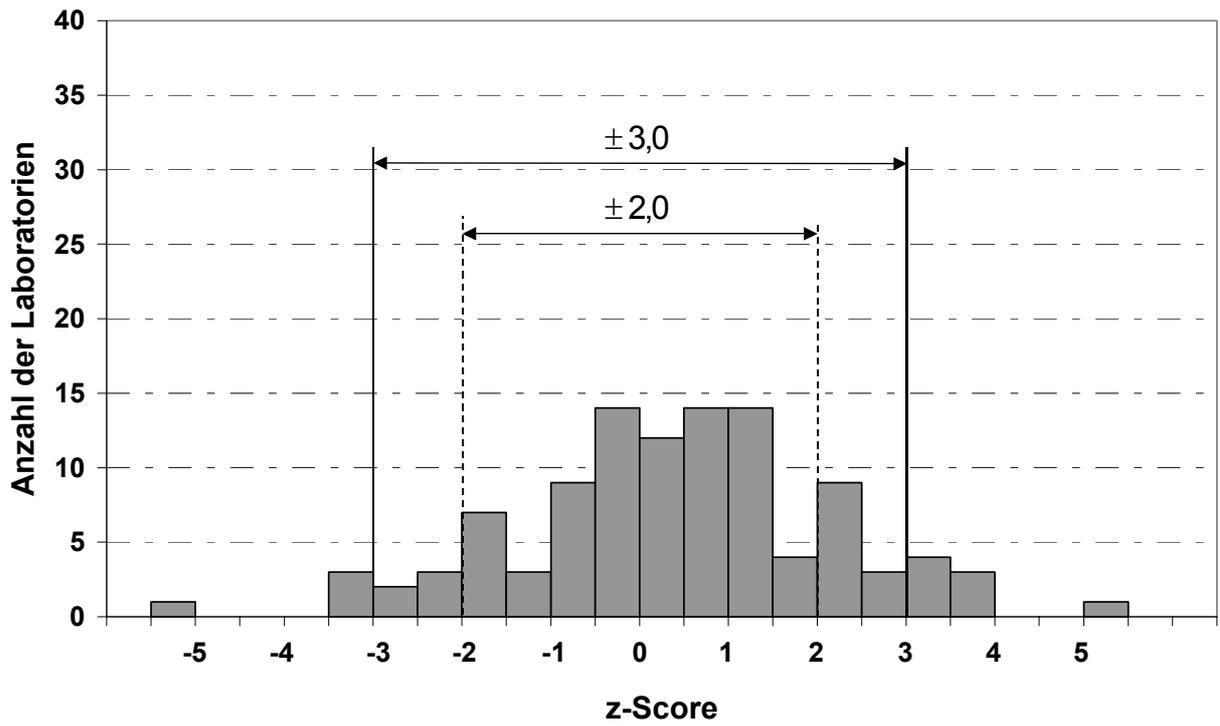


Bild 5.4.5: Variante 1: Abweichung vom Stoffwert, Verteilung der z-Scores

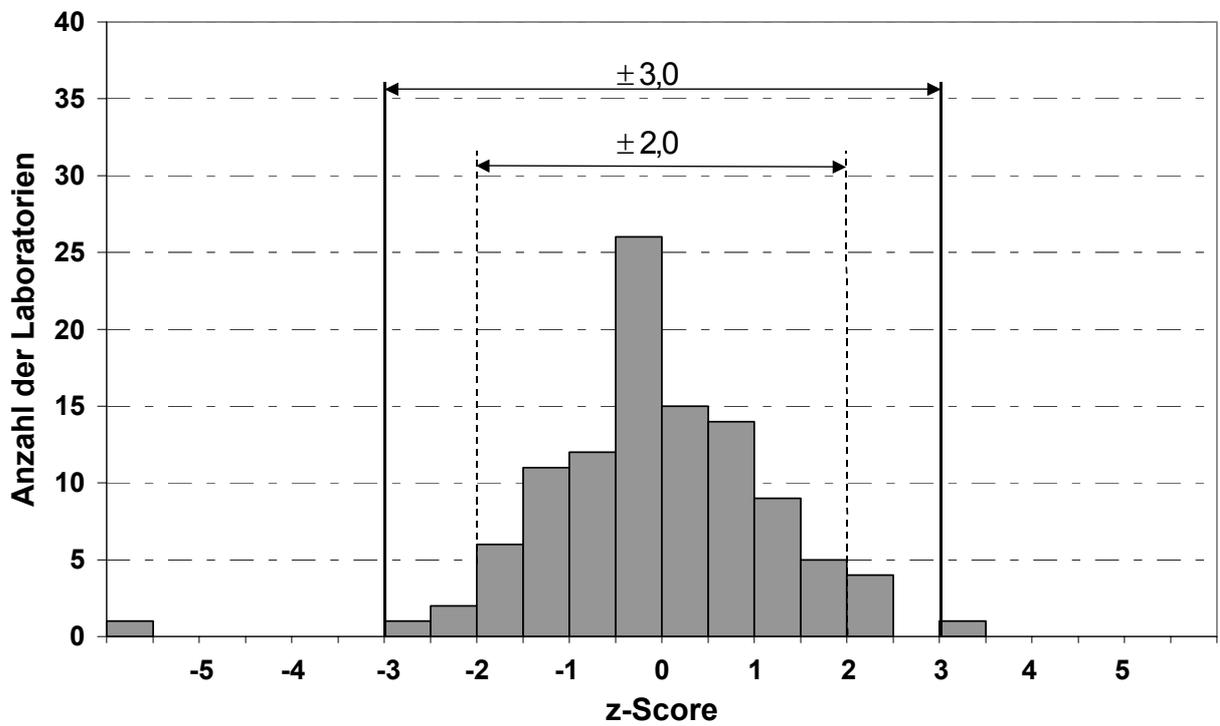


Bild 5.4.6: Variante 2: Abweichung vom Stoffwert, Verteilung der z-Scores

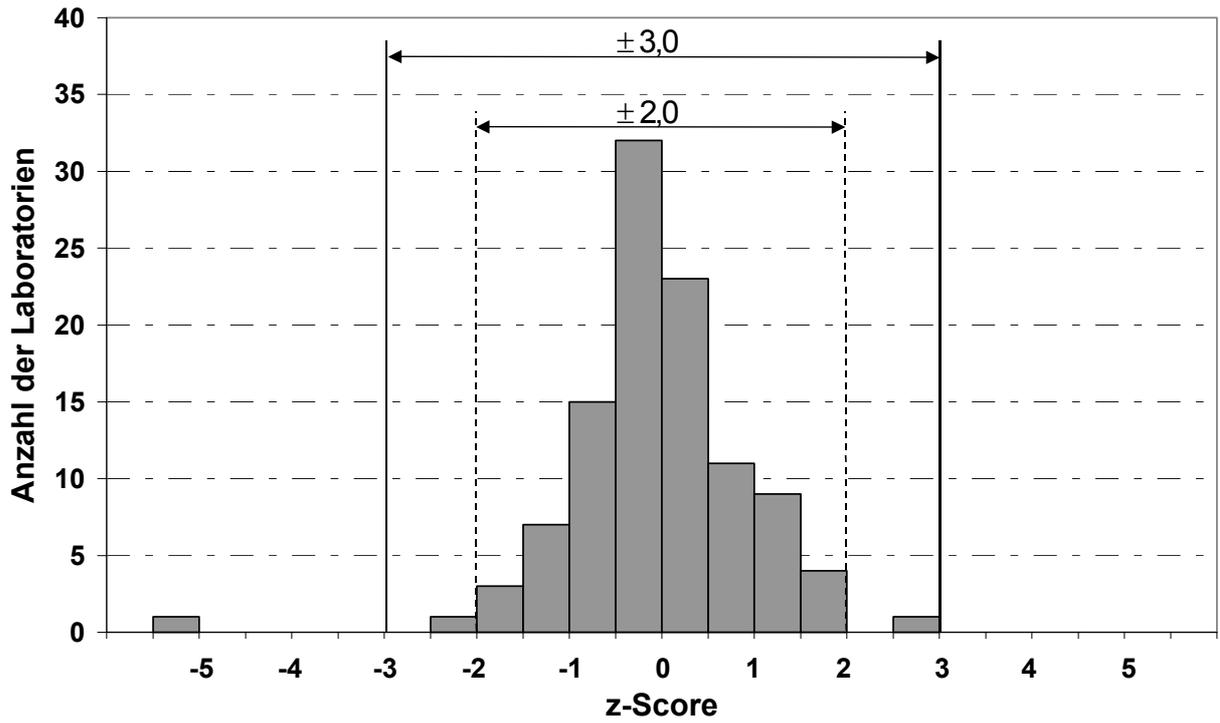


Bild 5.4.7: Variante 3: Abweichung vom Stoffwert, Verteilung der z-Scores

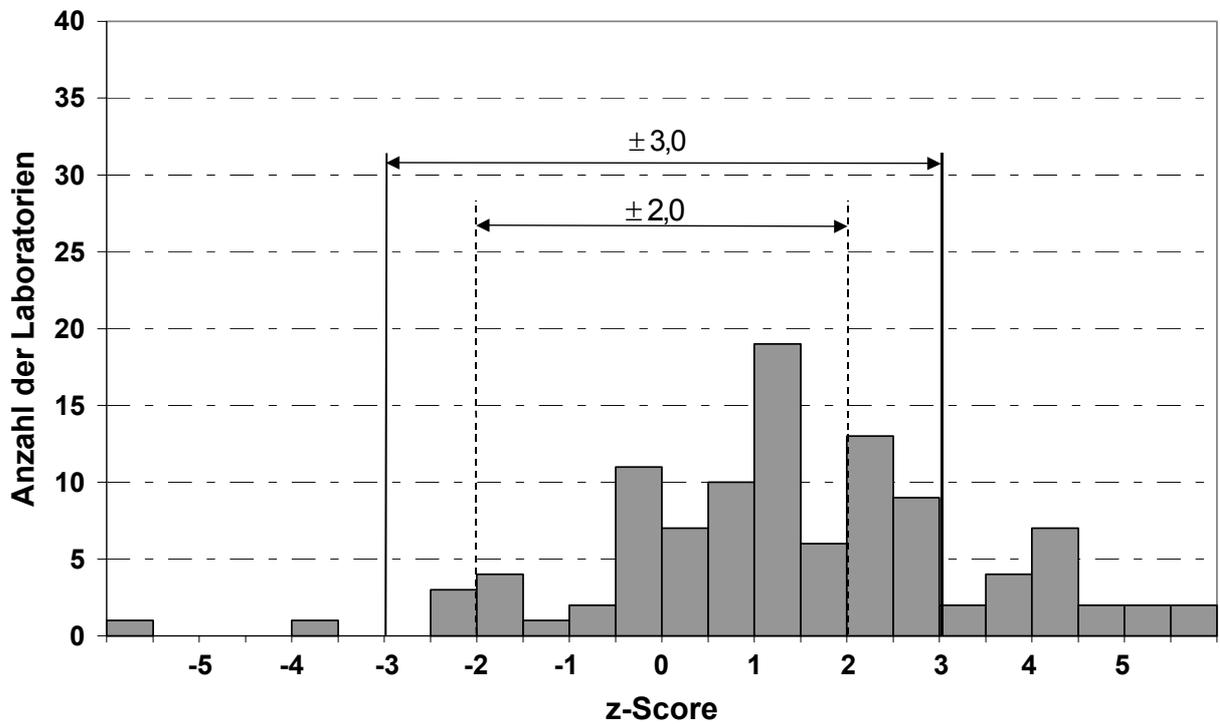


Bild 5.4.8: Variante 4: Abweichung vom Stoffwert, Verteilung der z-Scores

Die Abschätzung des Stoffwertes durch die in den Varianten 2 und 3 verwendeten Verfahren ist mathematisch gesehen zwar genauer als die nach Variante 1 durchgeführte Abschätzung von X_I durch den Median aller Labormittelwerte. Für praktische Zwecke ist X_I jedoch ausreichend genau und mit erheblich geringerem Rechenaufwand zu ermitteln.

Aus diesem Grund erfolgte die Auswertung der Eignungsprüfung CIT 2001 [29] nach Variante 1. Bild 5.4.9 zeigt die z-Scores aller Teilnehmer in der Übersicht. Die Grenzen für $|z|=2$ bzw. $|z|=3$ sind als waagerechte Linien in das Diagramm eingefügt.

Es sei an dieser Stelle jedoch ausdrücklich auf ein Problem hingewiesen, das sich bei der Bestimmung eines Konsenswertes aus den Ergebnissen aller Teilnehmer ergibt. Wenn an einer Eignungsprüfung außergewöhnlich viele „schlechte“ Laboratorien teilnehmen, erzielt ein besonders gutes Laboratorium möglicherweise ein unzureichendes Ergebnis, wenn sein Ergebnis entsprechend weit vom Konsenswert entfernt ist. Dieser Fall ist vor allem dann zu erwarten, wenn Prüfverfahren neu oder nicht ausreichend vereinheitlicht sind.

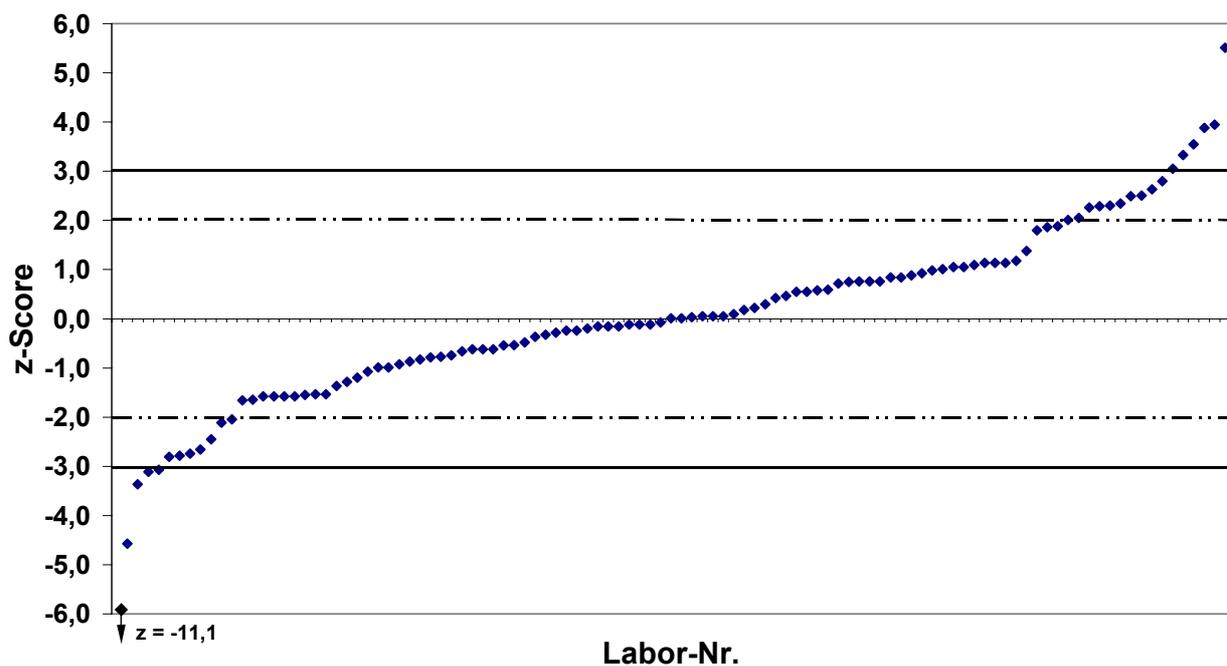


Bild 5.4.9: Auswertung der Eignungsprüfung CIT 2001: Abweichung vom Stoffwert

Die Beurteilung der Leistungsfähigkeit für das Kriterium „Abweichung vom Stoffwert“ mittels z-Scores ist gegenüber den Methoden „Schätzung der systematischen Abweichungskomponente D “ bzw. „Prozentuale Differenz $D\%$ “ übersichtlicher und für die teilnehmenden Laboratorien leichter verständlich.

Eine Auswertung nach E_n -Zahlen, z' -Scores oder ζ -Scores (vgl. Kap. 4.4.2) ist nicht möglich, da keine gesicherten Erkenntnisse über die individuelle Messunsicherheit für den Kerbschlagbiegeversuch vorliegen.

5.4.5 Statistische Beurteilung der Ergebnisse: Wiederholbarkeit

Die Wiederholbarkeit unter den vorgegebenen Prüfbedingungen wird durch die Differenz zwischen dem größten und kleinsten Einzelwert $x_{\max} - x_{\min}$ der verbrauchten Energie beschrieben. Gemäß DIN EN 10045-2, Tabelle 3 [122] ist die Wiederholbarkeit dann ausreichend, wenn die Beziehung

$$x_{\max} - x_{\min} \leq 0,15 X \quad (6.1.1)$$

erfüllt ist. Diese Vorgabe gilt streng genommen nur bei der Verwendung von Charpy-V-Referenzproben wie in der Norm beschrieben.

Da diese Bedingung nicht erfüllt ist und der Stoffwert $X = X_I = 119,8 J$ ein Konsenswert aus den Werten aller Teilnehmer ist, wird folgende Festlegung getroffen: Für die Eignungsprüfung CIT 2001 ist dann eine ausreichende Wiederholbarkeit erzielt, wenn die Gleichung

$$x_{\max} - x_{\min} \leq 0,2 X \quad (6.1.2)$$

erfüllt ist.

Damit hat ein Laboratorium in dieser Eignungsprüfung dann eine ausreichende Wiederholbarkeit erzielt, wenn $x_{\max} - x_{\min} \leq 22$ Joule ist. Sieben Laboratorien erzielten ein unzureichendes Ergebnis.

Eine Verwendung der in Tabelle 5.4.8 zusammengestellten Varianten ergibt für die Varianten 2 und 3 keine Änderung. Kommt Variante 4 zum Einsatz, ergibt sich ein Grenzwert von 21 Joule, d. h. drei weitere Laboratorien würden ein unzureichendes Ergebnis erzielen.

In Bild 5.4.10 sind die Ergebnisse aller Laboratorien und der Grenzwert für ausreichende Wiederholbarkeit aufgetragen.

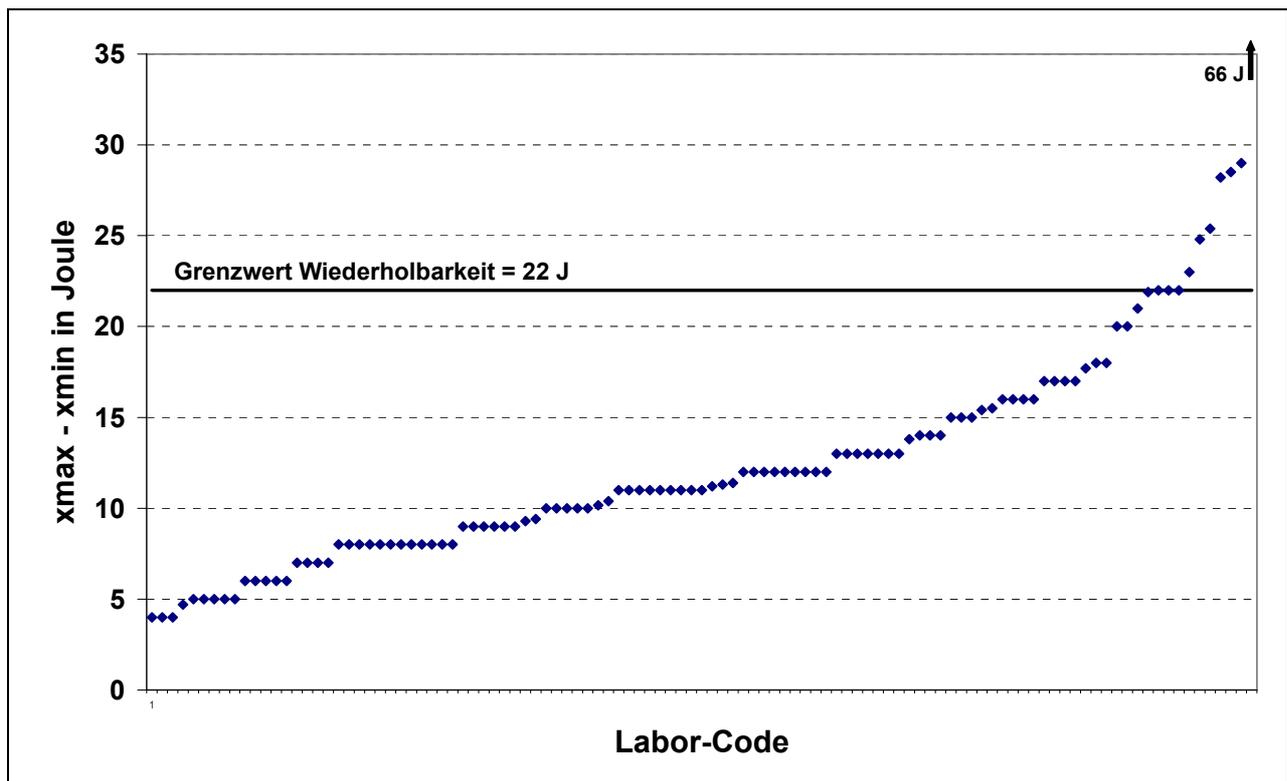


Bild 5.4.10: CIT 2001, Wiederholbarkeit, Übersicht

5.4.6 Genauigkeit der Probenfertigung

Die Teilnehmer der Eignungsprüfung CIT 2001 hatten ausreichend Probematerial erhalten, um eine zusätzliche (Referenz-) Probe mit V-Kerb nach DIN EN ISO 10045-1, Tabelle 2 [118] herzustellen, die sie anschließend vermessen sollten. Die Laboratorien sollten eine Aussage (ja / nein) darüber machen, ob die von ihnen hergestellten Proben normkonform sind. Diese Referenzprobe wurde anschließend in einem akkreditierten Messlabor des Materialprüfungsamtes des Landes Brandenburg nachgemessen. Im Laufe der Eignungsprüfung bzw. nach Auswertung der Messungen des Materialprüfungsamt des Landes Brandenburg ergab sich folgendes Bild:

- Einige Laboratorien forderten neues Probenmaterial an, nachdem sie erkannten, dass die von ihnen gefertigten Proben nicht normkonform waren.
- Neunzehn Laboratorien bestätigen die Normkonformität ihrer Proben und prüften diese, obwohl bereits ihre eigenen Messungen Abweichungen von der Norm ergeben hatten.
- Sechszwanzig Laboratorien weichen bei einem oder mehreren Maßen von der Normvorgabe ab; die Laboratorien konnten die Abweichung(en) selbst nicht

feststellen, daher bestätigten sie die Normkonformität der Proben. Die Abweichungen wurden bei der Nachmessung durch das akkreditierte Messlabor des Landes Brandenburg entdeckt.

- Am schwierigsten einzuhalten sind die Maße „Höhe im Kerbgrund“ und „Kerbradius“. Die Nachprüfung des Materialprüfungsamt des Landes Brandenburg zeigt außerdem, dass das Bestimmen dieser Maße nicht von allen Laboratorien beherrscht wird [29].
- Ein Laboratorium hatte statt der vorgeschriebenen V-Kerb-Probe eine U-Kerb-Probe gefertigt. Die Prüfergebnisse dieses Laboratoriums unterscheiden sich signifikant von allen anderen Prüfergebnissen.

Daraus folgt: die Vorgaben des vereinheitlichten Messverfahrens zur Probenherstellung sind entweder nicht ausreichend bekannt oder werden von den Prüflaboratorien nicht hinreichend beachtet. Allerdings scheint der Einfluss einer nicht norm-konformen Probe auf das Prüfergebnis (und damit auf das Ergebnis der Eignungsprüfung) von vernachlässigbarer Größe zu sein, solange es sich um die gleiche Probenform handelt und die Kerbtiefe nicht zu einer signifikanten Verminderung der Querschnittsfläche unter dem Kerb führt.

Zusammenfassend kann man sagen, dass für das Gelingen einer Eignungsprüfung „Kerbschlagbiegeversuch nach DIN EN 10045 [118] die Homogenität des Probematerials und die sorgfältige Beurteilung der rechnerisch ermittelten Größen „Stoffwert“ und „Standardabweichung der Eignungsprüfung“ im Hinblick auf ihre technische Relevanz von ausschlaggebender Bedeutung sind.

5.5 Eignungsprüfung Emissionsspektroskopie

Die Atomemissionsspektroskopie AES (auch optische Emissionsspektroskopie OES genannt) ermöglicht es, in Proben unterschiedlichster Art über einen weiten Konzentrationsbereich hinweg Multielementbestimmungen durchzuführen. In der OES wird die Probensubstanz in eine Strahlungsquelle gebracht, dort atomisiert und teilweise ionisiert. Ionen und Atome werden dann zu ihrer spezifischen Linienstrahlung angeregt, die mittels Spektralapparat zerlegt wird. Grundlage der qualitativen Spektralanalyse ist das Auftreten einer elementcharakteristischen Linie, die auf die Anwesenheit dieses Elementes in der Probe hinweist. Die OES ist eine Relativmethode. Elementkonzentrationen können nach einer Kalibrierung durch den Vergleich der gemessenen Lichtintensitäten mit den Lichtintensitäten von Standardproben bekannter Zusammensetzung bestimmt werden. Dies ist die Grundlage der quantitativen OES.

Bei der Wareneingangsprüfung, der Produktions- und Endproduktkontrolle wird in der Metallverarbeitenden Industrie die Funkenspektroskopie routinemäßig verwendet. Mittelspannungsfunken in Argon als Strahlungsquelle werden heute in Verbindung mit Simultanspektrometern für die schnelle Bestimmung vieler Elemente (15 und mehr, darunter auch Kohlenstoff, Phosphor, Schwefel und Bor) eingesetzt. Die gesamte Analysendauer liegt bei etwa 20s. Die Nachweisgrenzen liegen für die verschiedenen Metallmatrizes im $\mu\text{g/g}$ -Bereich. Auch leichte Elemente wie Beryllium, Bor oder Kohlenstoff können gut bestimmt werden. Entsprechende interne Standards vorausgesetzt, kann eine relative Standardabweichung unterhalb 1% erreicht werden. Die Wahl der Kalibrierproben ist für das Analyseergebnis sehr wichtig, da durch Störungen in der Verdampfung oder Ionisations- und Temperaturbeeinflussung im Plasma große Matrixeffekte verursacht werden können, gegebenenfalls wird ein rechnerisches Korrekturverfahren notwendig. [123]

Die Emissionsspektrometrie ist in einem mechanisch-technologisch ausgerichteten Prüflaboratorium ein ergänzendes Arbeitsgebiet. Es dient in vielen Laboratorien zur schnellen Identifizierung und Analyse von Probenmaterial. Die Emissionsspektrometrie ist kein genormtes Prüfverfahren. Die Analysen werden auf der Basis validierter Hausverfahren durchgeführt, die in der Regel durch den Einsatz von zertifizierten Referenzmaterialien rückverfolgbar sind.

5.5.1 Untersuchungsprogramm

Das Leistungskriterium, nach dem die teilnehmenden Laboratorien in der Eignungsprüfung Emissionsspektroskopie EMS 2001 [27] beurteilt werden, ist die Abweichung des Labor-mittelwertes vom Stoffwert. Die Bewertung wird für jedes Element einzeln vorgenommen.

Auf die im Rahmen der Eignungsprüfung gewonnenen Daten wurden die in den Tabellen 5.5.1 bis 5.5.3 gekennzeichneten Methoden angewendet, kritisch untersucht und bewertet:

Tabelle 5.5.1: Bestimmung des Stoffwertes X

Methode (nach Kapitel 4.2)	
Bekannter Wert (Methode A)	-
Zertifiziertes Referenzmaterial (Methode B)	-
Referenzmaterial (Methode C)	X
Konsenswert Experten (Methode D)	-
Konsenswert Teilnehmer (Methode E)	X

Tabelle 5.5.2: Berechnung der Standardabweichung der Eignungsprüfung $\hat{\sigma}$

Methode (nach Kapitel 4.4.1)	
Vorgeschriebener Wert	-
Festlegung eines erwünschten Leistungsniveaus	-
Präzisionsexperiment nach DIN ISO 5725-2	X
Konsenswert, robuste Datenauswertung	X

Tabelle 5.5.3: Berechnung der Leistungsstatistik

Methode (nach Kapitel 4.4.2)	
Schätzung der systematischen Abweichungskomponente D	-
Prozentuale Differenz $D_{\%}$	-
z-Scores	X
Histogramm	-
Säulendiagramm	X

5.5.2 Programm der Eignungsprüfung

Die Teilnehmer der Eignungsprüfung erhielten eine pulvermetallurgisch hergestellte Probe unbekannter Zusammensetzung, die an der Bundesanstalt für Materialforschung und –prüfung, Berlin (BAM) hergestellt wurde. An diesen Proben sollte mittels einer fünfmal zu wiederholenden Analyse der Gehalt von insgesamt vierzehn Elementen bestimmt werden. Den Laboratorien war freigestellt, für welche Element eine Analyse durchgeführt und zur Bewertung eingereicht wird. Insgesamt nahmen 68 Laboratorien an der Eignungsprüfung teil. Drei Laboratorien haben sich mit zwei Geräten beteiligt, so dass insgesamt bis zu 71 Analysen je Element vorlagen.

5.5.3 Materialauswahl und Homogenitätsprüfung

Bei der Auswahl des Probenmaterials wurden erhöhte Anforderungen an dessen Homogenität gestellt.

Die Versuche zur Homogenitätsprüfung wurden von der Bundesanstalt für Materialforschung und –prüfung (BAM), Berlin, Labor I.11 mit dem Funkenspektrometer OBLF QSL 150 durchgeführt.

Homogenität ist ein Zustand, in dem eine einheitliche Zusammensetzung für eine oder mehrere festgelegte Eigenschaften vorliegt. Die Forderung nach Homogenität ist gebunden an die „kleinste Probeneinheit“, die analysiert werden kann. Nur zwischen diesen „kleinsten Probeneinheiten“ ist Homogenität erforderlich, da eine weitere Unterscheidung analytisch nicht möglich ist. Die „kleinste Probeneinheit“ eines Versuchsmaterials ist die Menge des Versuchsmaterials, die für eine Einzelmessung eingesetzt wird bzw. effektiv zum Einsatz gelangt. Bei der Emissionsspektroskopie ist dies die Masse des verdampften Materials. Da diese sehr gering ist, sind die Anforderungen an die Homogenität des Probenmaterials besonders hoch [68].

Bei der Herstellung und Homogenitätsprüfung des Materials für diese Eignungsprüfung wurde wie folgt vorgegangen:

Das pulverförmige Ausgangsmaterial wurde durch heißisostatisches Pressen kompaktiert. Es wurden insgesamt 18 Stangen Rund-Material hergestellt. Im ersten Schritt wurde eine Stange in 18 Probeabschnitte unterteilt. Aus jedem dieser Probeabschnitte wurde jeweils eine Probe analysiert. Es wurde außen, im Zentrum der Probe und ungefähr in der Mitte des Probenradius analysiert, um die Homogenität über der Probenfläche zu prüfen. Im zweiten Schritt wurden jeweils zwei Proben aus der Mitte weiterer vier Stangen analysiert, um die Homogenität zwischen den Stangen zu testen. Für die Eignungsprüfung wurden ausschließlich Proben verwendet, die aus dem mittleren Abschnitt der Stangen stammen.

Die Größe der Standardabweichungen für die einzelnen Elemente wird zur Beurteilung der Homogenität des Materials herangezogen, siehe Tabelle 5.5.4.

Die Standardabweichung als Maß für die Streuung setzt sich prinzipiell aus einer verfahrensbedingten Streuung und einer werkstoffbedingten Streuung (Inhomogenität) zusammen. Für den vorliegenden Fall folgt daher, dass die werkstoffbedingte Streuung (Inhomogenität) kleiner oder gleich den in Tabelle 5.5.4 angegebenen Werten sein muss.

Tabelle 5.5.4: Homogenitätsprüfung: mittlere Standardabweichungen (STABW); alle Angaben in Massen-%

Element	STABW	Element	STABW	Element	STABW
<i>C</i>	0,01008	<i>N</i>	0,00792	<i>W</i>	0,00499
<i>Si</i>	0,00419	<i>Cr</i>	0,00387	<i>Cu</i>	0,00349
<i>Mn</i>	0,00629	<i>Mo</i>	0,00884	<i>Co</i>	0,00166
<i>P</i>	0,00024	<i>Ni</i>	0,02253	<i>Nb</i>	0,00152
<i>S</i>	0,00047	<i>V</i>	0,00343		

Das Material ist daher für eine Verwendung in der Eignungsprüfung ausreichend homogen.

5.5.4 Abweichung des Laborwertes vom Stoffwert

Im Folgenden werden die in Kapitel 4 vorgestellten Modelle zur Ermittlung des Stoffwertes X und zur Berechnung der Standardabweichung der Eignungsprüfung $\hat{\sigma}$ wie in den Tabellen 5.5.1 bis 5.5.3 dargestellt auf die Daten der Eignungsprüfung EMS 2001 systematisch angewendet.

Variante 1: Bestimmung der Stoffwerte X und der Standardabweichungen der Eignungsprüfung $\hat{\sigma}$, Referenzmaterial (Methode C, Kap 4.2)

Das Labor I.11 der BAM, Berlin stellt üblicherweise zertifizierte Referenzmaterialien für die Emissionsspektrometrie her. Daher kann davon ausgegangen werden, dass entsprechende rückführbare Normale vorhanden sind und die Analyseneinrichtungen und ihr Bedienpersonal entsprechend hohen Anforderungen genügen. Betrachtet man das Versuchsmaterial der Eignungsprüfung als Referenzmaterial, so gibt das Labor I.11 der BAM, Berlin für die einzelnen Elemente die in Tabelle 5.5.5 zusammengestellten Werte für den jeweiligen Stoffwert X_{VI} an. Als Standardabweichungen der Eignungsprüfung $\hat{\sigma}_{VI}$ werden die jeweiligen Standardabweichungen der bei Homogenitätsprüfung für jedes Element durchgeführten Messungen verwendet. Die Beurteilung der eingereichten Ergebnisse erfolgt mittels z-Score (s. Kap. 4.4.2). In Tabelle 5.5.5 sind für jedes Element die Zahl der Laboratorien aufgeführt, die ein Warn- bzw. Aktionssignal erhalten.

Tabelle 5.5.5: Variante 1; Abweichung vom Stoffwert, Warn- und Aktionssignale, Methode C

Element	Stoffwert X_{VI}	STABW $\hat{\sigma}_{VI}$	Anzahl der Teilnehmer		
	in Massen-%		Gesamt	Warnsignal $2 < z < 3$	Aktionssignal $ z \geq 3$
C	0,871	0,01008	68	11	38
Si	0,828	0,00419	70	1	65
Mn	1,061	0,00629	71	6	55
P	0,020	0,00024	68	1	59
S	0,018	0,00047	67	3	56
N	0,019	0,00792	24	0	0
Cr	1,676	0,00387	71	2	67
Mo	0,880	0,00884	71	4	45
Ni	2,866	0,02253	71	12	45
V	0,552	0,00343	70	5	58
W	0,699	0,00499	61	6	45
Cu	0,742	0,00349	68	2	59
Co	0,310	0,00166	59	3	42
Nb	0,530	0,00152	55	0	54

Variante 2: Bestimmung der Stoffwerte X und der Standardabweichungen der Eignungsprüfung $\hat{\sigma}$, Konsenswert Teilnehmer (Methode E, Kap 4.2), klassische statistische Auswertung nach DIN ISO 5725

Die Ausreißerprüfung nach Cochran und nach Grubbs ergaben die in Tabelle 5.5.6 aufgeführte Anzahl von Ausreißern. Für das Element Niob wurde nur der Cochran Test durchgeführt, der abgebrochen wurde, nachdem bereits 35% der Ergebnisse als Ausreißer zurückgewiesen worden waren.

Für das Element Niob wurde eine grafische Ausreißerprüfung mittels Mandels k-Statistik für Ausreißer mit zu großer laborinterner Streuung bzw. Mandels h-Statistik für Ausreißer zwischen den Laboratorien erstellt. In den Bildern 5.5.1 und 5.5.2 sind die Ergebnisse aufgetragen. Bild 5.5.1 zeigt vier Ausreißer aufgrund zu hoher Streuung zwischen den Laboratorien. Bild 5.5.2 zeigt fünf Ausreißer und einen Fastausreißer bezüglich der laborinternen Streuung. Die Ausreißer aus Bild 5.5.1 sind nicht identisch mit denen aus Bild 5.5.2.

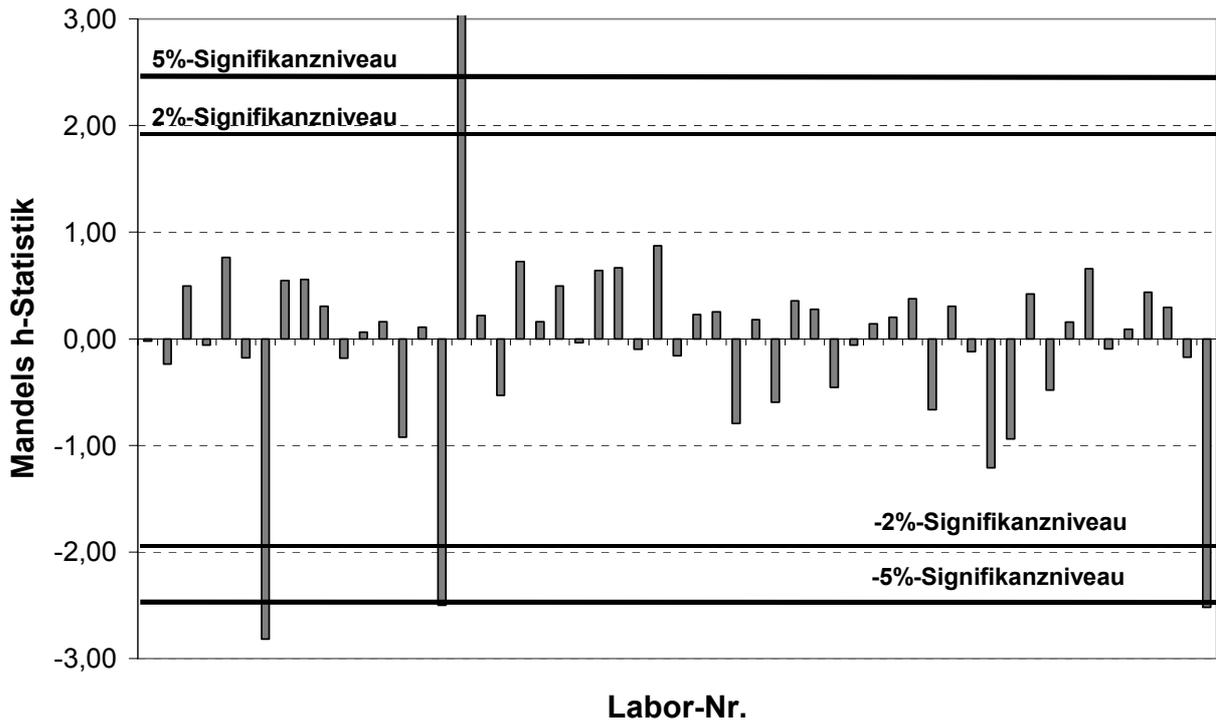


Bild 5.5.1: Element Niob, grafische Ausreißerprüfung, Mandels h-Statistik
Ausreißer zwischen den Laboratorien

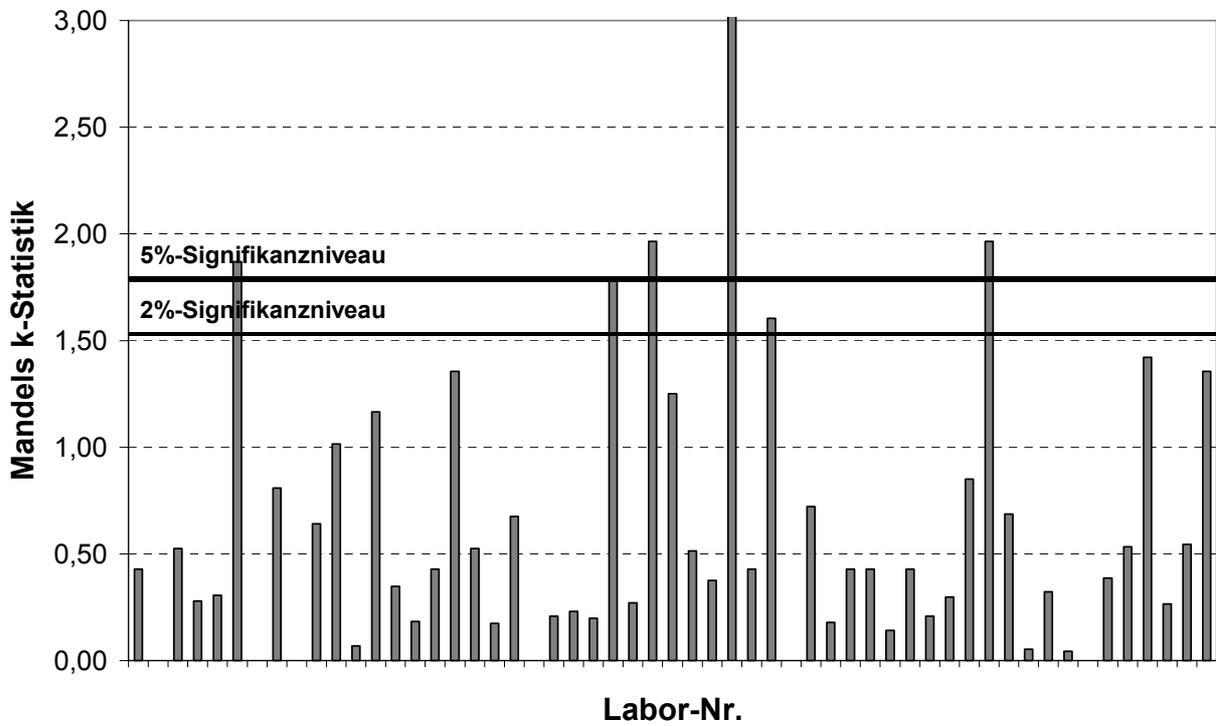


Bild 5.5.2: Element Niob, grafische Ausreißerprüfung, Mandels k-Statistik
Ausreißer mit laborinterner Streuung

Tabelle 5.5.6: Ergebnisse der Ausreißertests nach Cochran und Grubbs

Element	Ausreißertest nach Cochran		Ausreißertest nach Grubbs	
	Fastausreißer	Ausreißer	Fastausreißer	Ausreißer
C	1	4	0	0
Si	1	7	1	0
Mn	2	8	0	0
P	0	13	0	1
S	0	6	0	1
N	0	1	0	0
Cr	5	8	1	0
Mo	1	6	0	0
Ni	4	8	0	2
V	0	5	0	2
W	0	3	1	3
Cu	1	3	1	1
Co	0	8	2	1
Nb	---	---	---	---

Tabelle 5.5.5: Variante 2; Abweichung vom Stoffwert, Warn- und Aktionssignale, Methode E

Element	Stoffwert X_{V2}	STABW $\hat{\sigma}_{V2}$	Anzahl der Teilnehmer		
	in Massen-%		Gesamt	Warnsignal $2 < z < 3$	Aktionssignal $ z \geq 3$
C	0,908	0,0453	68	3	3
Si	0,797	0,0418	70	6	1
Mn	1,086	0,0315	71	4	2
P	0,0175	0,0040	68	5	4
S	0,0155	0,0029	67	4	3
N	0,0151	0,0023	24	1	0
Cr	1,722	0,0779	71	4	3
Mo	0,898	0,0434	71	5	3
Ni	2,949	0,1343	71	6	3
V	0,591	0,0280	70	4	3
W	0,717	0,1006	61	4	4
Cu	0,712	0,0464	68	8	2
Co	0,321	0,0209	59	2	4
Nb	---	---	55	---	---

Die Daten der Ausreißer nach Grubbs und Cochran wurden zur Berechnung der Präzisionskenngrößen aus dem Datensatz entfernt.

Die Berechnung erfolgt nach den Gleichungen (4.3.20) bis (4.3.25) sowie Gleichung (4.4.4). In Tabelle 5.5.7 sind für jedes Element der jeweilige Stoffwert X_{V2} und die zugehörige Standardabweichung der Eignungsprüfung $\hat{\sigma}_{V2}$ eingetragen. Die Beurteilung der Ergebnisse erfolgt mittels z-Score (s. Kap. 4.4.2). In Tabelle 5.5.7 sind für jedes Element die Zahl der Laboratorien, die ein Warn- bzw. Aktionssignal erhalten, aufgeführt.

Variante 3: Bestimmung der Stoffwerte X und der Standardabweichungen der Eignungsprüfung $\hat{\sigma}$, Konsenswert Teilnehmer (Methode E, Kap 4.2), einfache robuste statistische Auswertung

Die Ausreißerdatensätze werden nicht entfernt. Für jedes Element ist der Stoffwert X_{V3} der Median der Labormittelwerte. Die Standardabweichungen der Eignungsprüfung $\hat{\sigma}_{V3}$ werden für jedes Element durch den normalisierten Quartilsabstand nIQR nach Gleichung 5.5.1 berechnet:

$$\hat{\sigma}_{V3} = nIQR = 0,7413 \cdot IQR \quad \text{mit } IQR=Q_3-Q_1 \quad (5.5.1).$$

Q_3 ist der Wert unterhalb dessen 75 % aller Werte liegen, Q_1 ist der Wert unterhalb dessen 25 % aller Werte liegen. Der Faktor 0,7413 ergibt sich aus der Standardnormalverteilung, die einen Mittelwert von Null und eine Standardabweichung von Eins hat. Die Breite von (Q_3-Q_1) beträgt in solcher Verteilung 1,34898 und daher ergibt sich $1/1,34898=0,7413$ [87]. Multipliziert man (Q_3-Q_1) mit diesem Faktor, so kann man ihn mit einer Standardabweichung vergleichen.

Tabelle 5.5.8: Variante 3; Abweichung vom Stoffwert, Warn- und Aktionssignale, Methode E

Element	Stoffwert X_{V3}	STABW $\hat{\sigma}_{V3}$	Anzahl der Teilnehmer		
	in Massen-%		Gesamt	Warnsignal $2 < z < 3$	Aktionssignal $ z \geq 3$
C	0,908	0,0404	68	4	3
Si	0,791	0,0286	70	6	7
Mn	1,086	0,0302	71	6	2
P	0,0178	0,00357	68	4	5
S	0,0154	0,00245	67	6	4
N	0,0152	0,00169	24	4	1
Cr	1,732	0,0480	71	5	7
Mo	0,906	0,0343	71	1	8
Ni	2,928	0,1601	71	2	3
V	0,520	0,0266	70	4	4
W	0,711	0,0538	61	4	9
Cu	0,712	0,0280	68	2	10
Co	0,314	0,0159	59	1	6
Nb	0,596	0,0409	55	3	5

In Tabelle 5.5.8 sind für jedes Element der Stoffwert X_{V3} und die Standardabweichung der Eignungsprüfung $\hat{\sigma}_{V3}$ eingetragen. Die Beurteilung der eingereichten Ergebnisse erfolgt mittels z-Score (s. Kap. 4.4.2). In Tabelle 5.5.8 sind für jedes Element die Anzahl der Laboratorien, die ein Warn- bzw. Aktionssignal erhalten, aufgeführt.

Die Variante 1 ist die am leichtesten auszuwertende Variante, da man die Ergebnisse des Referenzlabors übernehmen kann. Eine Auswertung nach dieser Variante stellt jedoch eindeutig eine Überforderung der Laboratorien dar, da diese offensichtlich nicht die Wiederholgenauigkeit des Referenzlaboratoriums erreichen.

Variante 2 ist die am wenigsten scharfe Auswertung, d. h. die Zahl der Laboratorien, die ein Aktionssignal oder ein Warnsignal erhalten, ist am niedrigsten. Der Rechenaufwand ist vergleichsweise groß, da für jedes Element eine einfaktorielle Varianzanalyse durchgeführt werden muss. Außerdem setzt diese Form der Auswertung eine Normalverteilung der Ergebnisse voraus. Wie Tabelle 5.5.6 zu entnehmen, dürfen eine ganze Reihe von Daten bei der Berechnung der Kennwerte nicht berücksichtigt werden, da sie auf Grund des Cochran Tests als Ausreißer bezüglich der laborinternen Streuung zurückgewiesen wurden. Findet man auf unterschiedlichen Merkmalsniveaus bei demselben Laboratorium mehrere Fastausreißer bzw. statistische Ausreißer, kann dies nach DIN ISO 5725-2 [54] ein ernster Hinweis sein, dass bei diesem Laboratorium die Varianz innerhalb des Labors außergewöhnlich hoch ist.

Variante 3 urteilt etwas schärfer als Variante 2. Der Rechenaufwand zur Bestimmung der Kennwerte ist vergleichsweise niedrig, und der Rechenweg ist für die Teilnehmer einfach nachzuvollziehen. Es wird keine Normalverteilung der Ergebnisse vorausgesetzt. Die Eignungsprüfung EMS 2001 [27] wurde auf Basis von Variante 3 ausgewertet.

Bild 5.5.3 gibt in Form eines Säulendiagramms für sechzehn Laboratorien eine Übersicht über die Gesamtleistung der Laboratorien. Die z-Scores für die Elemente sind in der Reihenfolge (von links nach rechts) C, Si, Mn, P, S, N, Cr, Mo, Ni, V, W, Cu, Co, Nb nebeneinander über der Labor-Nr. aufgetragen. Die Grenzen für $z=\pm 3$ und $z=\pm 2$ sind parallel zur x-Achse in das Diagramm eingetragen. Zur Bewertung der Gesamtleistung wurde die Anzahl der mit einem z-Score $|z|\geq 3$ bewerteten, analysierten Elemente auf die Anzahl der von einem Laboratorium analysierten Elemente bezogen. Wenn mehr als 20 % der Analysen eines Labors mit „unzureichend“ bewertet wurden, so wird die gesamte Eignungsprüfung für dieses Labor mit „nicht erfüllt“ bewertet (siehe Kapitel 4.4.2, VIII). In Bild 5.5.1 ist Laboratorium Nr. 12 ein Beispiel für ein Labor, das die Eignungsprüfung nicht bestanden hat. Das Laboratorium hat 11 Elemente analysiert, die Analyse von 5 Elementen (45%) wurde mit „unzureichend“ bewertet (z-Score $|z|\geq 3$). Als Gegenbeispiel sei Labor Nr. 1 genannt: dort

wurden 13 Elemente analysiert, alle Analysen wurden mit „ausreichend“ (z-Score $|z| \leq 2$) bewertet.

Die Auftrugung der Ergebnisse in einem Säulendiagramm ist sehr gut geeignet, dem Auswertenden einen schnellen Überblick über die Gesamtleistung der Laboratorien zu geben, wenn mehrere Kennwerte, die gemeinsam bewertet werden, zu bestimmen sind. Zwölf von 72 Laboratorien haben die Eignungsprüfung mit dem Gesamtergebnis „nicht erfüllt“ abgeschlossen [27].

Wie sich in dieser Eignungsprüfung herausstellte, muss die Verwendung eines Stoffwertes bzw. einer Standardabweichung der Eignungsprüfung, welche von einem Expertenlabor bestimmt wurden, sorgfältig erwogen werden muss. Ist das Expertenlaborium den Teilnehmern der Eignungsprüfung deutlich in der Güte Ausrüstung und der Erfahrung bzw. Qualifikation der Mitarbeiter voraus, kann dies dazu führen, dass die überwältigende Mehrzahl der Teilnehmer die Eignungsprüfung mit einem schlechten („nicht erfüllt“) Ergebnis abschließt. In diesem Fall muss sorgfältig erwogen werden, ob z. B. die sehr hohe Wiederholgenauigkeit, welche das Expertenlabor angibt, für im industriellen Maßstab durchgeführte Prüfungen (Wareneingangsprüfung, Verwechslungsprüfung) relevant bzw. notwendig ist.

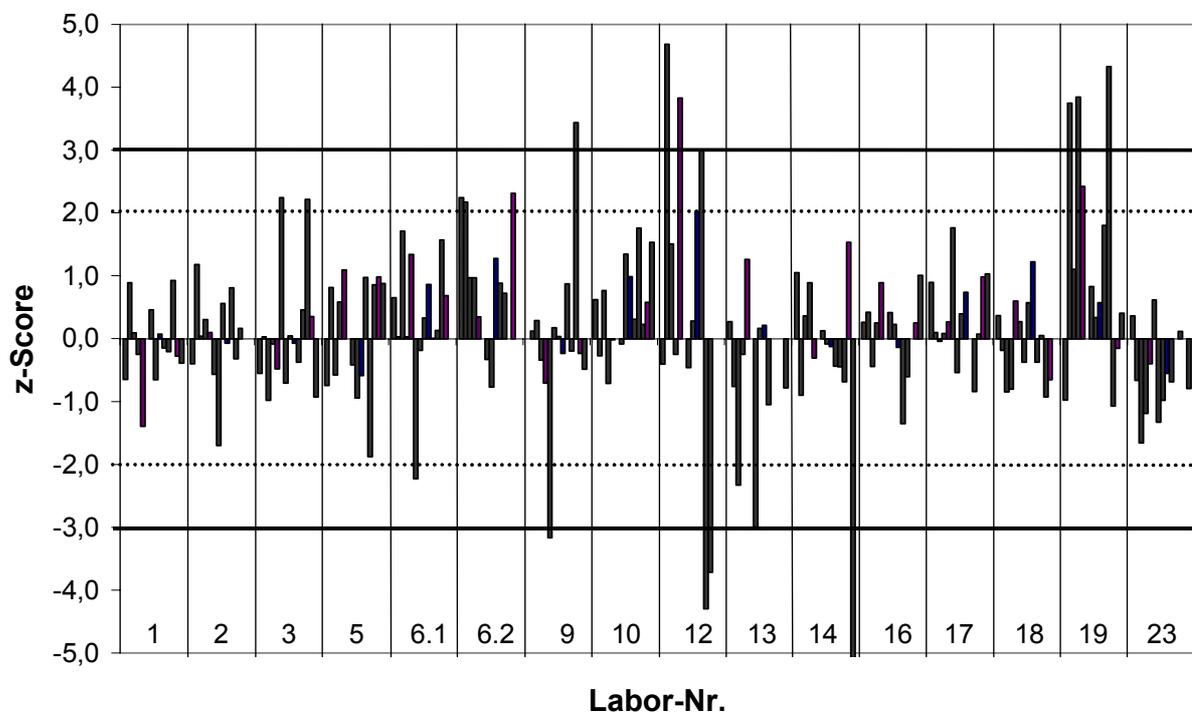


Bild 5.5.3: Eignungsprüfung Emissionsspektroskopie 2001, z-Scores der Ergebnisse von 16 Teilnehmern

6 Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurden erstmalig Eignungsprüfungen speziell für verschiedene mechanisch-technologische Werkstoffprüfverfahren konzipiert, mit insgesamt 600 Teilnehmern aus 25 Nationen durchgeführt und systematisch ausgewertet. Ziel der Untersuchung war die Entwicklung von Methoden zur Durchführung und Auswertung von Eignungsprüfungen, die einen prüftechnisch angemessenen und wissenschaftlich begründbaren Kompetenznachweis ermöglichen.

Die aus den Ergebnissen dieser Untersuchung resultierenden Eignungsprüfungen lassen sich für folgende Aufgaben verwenden:

- Akkreditierte Prüflaboratorien können bei erfolgreicher Teilnahme einen belastbaren Kompetenznachweis erbringen und die in ISO/IEC 17025 gestellte Anforderung erfüllen, an „einem geeigneten Programm für Vergleichsmessungen zwischen Laboratorien“ teilzunehmen.
- Nicht akkreditierten Laboratorien ermöglichen diese Eignungsprüfungen erstmalig die Beurteilung ihrer Prüfleistungen im Vergleich mit führenden europäischen Laboratorien auf Grundlage belastbarer Kriterien.
- Die aus den Daten der Eignungsprüfungen gewonnenen Ergebnisse können zusätzlich zur Weiterentwicklung und Normung der mechanisch-technologischen Prüfverfahren, z. B. zur Bestimmung der Messunsicherheit von Härtemessungen, verwendet werden.

Die Ergebnisse der Untersuchungen lassen sich wie folgt zusammenfassen:

- Die günstigste Voraussetzung für die Durchführung einer Eignungsprüfung auf dem Gebiet der mechanisch-technologischen Werkstoffprüfung liegt vor, wenn ein vereinheitlichtes (genormtes) Prüfverfahren existiert, das bereits Grenzwerte für die Anwendbarkeit des Prüfverfahrens enthält, wie z. B. die Normen für die gängigen Härteprüfverfahren nach Vickers oder Brinell oder die Kerbschlagbiegeprüfung. Im Rahmen dieser Arbeit konnte nachgewiesen werden, dass die in den zugehörigen Normen festgelegten Grenzwerte zur Prüfung der Prüfmaschinen, wie die Größen „Abweichung vom Stoffwert“ und die „Wiederholpräzision“ technisch relevant sind und von den Laboratorien eingehalten werden können. Da die Normfestlegungen konservativ sind, besteht hier noch ein Verbesserungspotential. Die Auswertung der

Eignungsprüfung wird deutlich vereinfacht und gewinnt an Aussagesicherheit, wenn ein zertifiziertes Referenzmaterial als Probenmaterial eingesetzt werden kann.

- Muss zur Bestimmung des Stoffwertes auf einen Konsenswert aus den Ergebnissen aller Teilnehmer zurückgegriffen werden, erfolgt die Berechnung des Konsenswertes am besten auf Basis eines robusten statistischen Verfahrens. Die robusten statistischen Verfahren setzen anders als die klassische Auswertung nach DIN ISO 5725 keine Normalverteilung der Daten und keine Homogenität der Varianzen voraus. Das Rechenverfahren auf Basis eines robusten statistischen Verfahrens ist für die Teilnehmer der Eignungsprüfung einfacher nachzuvollziehen.
- Die Berechnung eines robusten Mittelwertes bzw. einer robusten Standardabweichung nach ISO/DIS 13528 führte im Rahmen dieser Untersuchung bei deutlich höherem Rechenaufwand zu keinem signifikanten Unterschied gegenüber der Berechnung mittels Median und normalem Quartilsabstand.
- Erfolgt die Bewertung einer Eignungsprüfung mittels z-Score, hängt das Ergebnis essentiell von der Standardabweichung der Eignungsprüfung $\hat{\sigma}$ ab. Daher ist es unbedingt notwendig, dass bei der Bestimmung dieser Größe die technische Relevanz des Wertes berücksichtigt wird. Es ist gegebenenfalls notwendig, einen rechnerisch ermittelten Wert für $\hat{\sigma}$ mittels Expertenfestlegung zu korrigieren.
- Die Ergebnisse von entsprechend angelegten Eignungsprüfungen sind sehr gut geeignet, die notwendigen Daten zur Ermittlung der Messunsicherheit der mechanisch-technologischen Prüfverfahren zu erfassen und für die weitere Auswertung bereitzustellen. Die im Rahmen dieser Arbeit durchgeführte Eignungsprüfung Härteprüfung nach Brinell ist die erste speziell unter Berücksichtigung dieses Aspektes gestaltete Eignungsprüfung.
- Die in der chemischen Analytik übliche Bewertung der Eignungsprüfung mittels E_n -Zahlen, z^2 -Scores oder Zeta-Scores ist nicht möglich. Diese Bewertung setzt voraus, dass die Messunsicherheit der Verfahren bekannt sind. Bislang lag zur Bestimmung der Messunsicherheit für die Verfahren der mechanisch-technologischen Werkstoffprüfung keine ausreichende Datenbasis vor. Diese Datenbasis wurde im Rahmen dieser Arbeit erstmalig für die Härteprüfung nach Brinell erarbeitet.
- Die Eignungsprüfung ist ein hochwirksames Mittel zur Überprüfung und Bestätigung der Kompetenz von Prüflaboratorien der mechanisch-technologischen Werkstoffprüfung. Mit Hilfe der Eignungsprüfung wurden folgende weit verbreitete Probleme erkannt:

- Prüfergebnisse, besonders solche, die von EDV-gestützten Prüfsystemen stammen, werden kritiklos akzeptiert und (an den Kunden!) weitergegeben.
- Die internen Qualitätssicherungsverfahren funktionieren nicht ausreichend, da offensichtlich falsche Prüfergebnisse weitergegeben werden.
- Die Prüfnormen sind nicht ausreichend bekannt oder werden nicht exakt angewendet.
- Die Laboratorien verwenden fast ausschließlich regelmäßig kalibrierte Prüfmaschinen. Unzureichende Eignungsprüfungsergebnisse aufgrund nicht ausreichender Kalibrierung konnten nicht festgestellt werden.
- Die Ergebnisse von Eignungsprüfungen sind geeignet, die Arbeit eines Fachbegutachters im Akkreditierungsprozess deutlich zu erleichtern, da die Laborleistung auf Basis belastbarer und technisch relevanter Kriterien durch Dritte überprüft wurde.

Die europaweit durchgeführten und von vielen europäischen Akkreditierungsstellen anerkannten Eignungsprüfungen geben einen guten Überblick über die Qualität der Prüfleistungen auf dem Gebiet der mechanisch-technologischen Werkstoffprüfung. Es hat sich gezeigt, dass es trotz vereinheitlichter Prüfverfahren, der weiten Verbreitung von labor-internen Qualitätsmanagementsystemen, der regelmäßigen Kalibrierung der Prüfmaschinen und der regelmäßigen Schulung der Mitarbeiter weiterhin Laboratorien gibt, deren Leistung signifikant schwächer ist. Weitere Erkenntnisse, in wie weit die Teilnahme an Eignungsprüfungen zu einer Verbesserung bzw. Vereinheitlichung der Prüfleistungen führt, werden von der regelmäßigen Wiederholung der Eignungsprüfungen erwartet. Über einen längeren Zeitraum betrachtet, können gute Ergebnisse in den Eignungsprüfungen dann zu Erleichterungen im Akkreditierungsprozess genutzt werden.

Verbesserte Prüfleistungen sind die Voraussetzung zu einer konsequenteren Ausnutzung der mechanischen Eigenschaften von Konstruktionswerkstoffen und können zu einer besseren Werkstoffausnutzung (Leichtbau) und einer erhöhten Bauteilsicherheit führen. Die Abwicklung von Warenannahmeprüfungen wird erleichtert, wenn die Vergleichbarkeit von Prüfergebnissen unterschiedlicher Prüflaboratorien sichergestellt ist, so wird die Senkung von Qualitäts- und Prüfkosten möglich.

7 Literaturverzeichnis

- [1] EA-1/02: EA Brochure, Feb 1998.
- [2] ILAC-I2:1994: Testing, Quality Assurance, Certification and Accreditation, 1996.
- [3] EA-1/04 (rev.01): EA Memorandum of Understanding, Feb 2001.
- [4] Deutscher Akkreditierungsrat (DAR): [http:// www.dar.bam.de](http://www.dar.bam.de), Anzahl der zur Zeit akkreditierten Stellen vom 02. Mai 2002.
- [5] JUNIPER, I.R.: Quality issues in proficiency testing. *Accred Qual Assur* (1999) 4: pp. 366-341, 1999.
- [6] DIN EN 45001:1990, Allgemeine Kriterien zum Betreiben von Prüflaboratorien. Beuth Verlag, Berlin, Mai 1990.
- [7] DIN EN ISO/IEC 17025:2000, Allgemeine Anforderungen an die Kompetenz von Prüf- und Kalibrierlaboratorien. Beuth Verlag, Berlin, April 2000.
- [8] ANDERSSON, H.O.F.: The rational use of proficiency tests and intercomparisons. *Accred Qual Assur* (1998) 3: pp. 224-226, 1998.
- [9] BOLEY, N.: Benefits arising from proficiency testing schemes: The comparison of analytical methods. *Accred Qual Assur* (1998) 3: pp. 459-461, 1998.
- [10] ISO/IEC Guide 43-1:1997, Proficiency Testing by Interlaboratory Comparisons. Part 1: Development and operation of proficiency testing schemes. International Organisation for Standardization, Genève, First Edition 1997.
- [11] DIN V 55394-1:2000, Eignungsprüfung durch Vergleiche zwischen Laboratorien. Teil 1: Entwicklung und Durchführung von Programmen für die Eignungsprüfung. Beuth Verlag, Berlin, 2000.
- [12] GÖBEL, U.; STEFFEN, B.: Studie „Ringversuche in Deutschland“, (Bericht für BCR), Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin 1993.
- [13] DINKLOH, L. et al.: Ringversuche in der Wasseranalytik. Teil 1: Planung und Organisation. In: *Z. Wasser Abwasser Forsch.* 13 (1980) Nr. 2, S. 54-62. Teil 2: Auswertung. In: *Z. Wasser Abwasser Forsch.* 13 (1980) Nr. 5, S. 174-187. Teil 3: Beispiel einer Ringversuchsauswertung und Vergleich mit dem Entwurf DIN ISO 5725. In: *Z. Wasser Abwasser Forsch.* 13 (1980) Nr. 6, S. 209-212.
- [14] ROSS, J. W.: A Theoretical Basis for Clinically Relevant Proficiency Testing Evaluation Limits. In: *Arch. Pathol. Lab. Med* – Vol. 112, April 1988, pp. 421-434.
- [15] ROSS, J. W.: Evolution of Evaluation Criteria in the College of American Pathologists Surveys. In: *Arch. Pathol. Lab. Med* – Vol. 112, April 1988, pp. 334-339.
- [16] THOLEN, D.: Reference Values and Participant Means as Targets in Proficiency Testing. In: *Arch. Pathol. Lab. Med* – Vol. 117, September 1993, pp. 885-889.
- [17] TIETZ, N. W.; RODGERSON, D. O.; LAESSIG, R. H.: Are Clinical Laboratory Proficiency Tests as Good as They Can Be? In: *Clinical Chemistry*, Vol. 38. No. 4, 1992, pp.473-475.

- [18] DAR-EM17: Leitfaden zu Eignungsprüfungen als Instrumente der Akkreditierung im Prüfwesen. Version 1.0, 04/1998, DAR, Berlin 1998.
- [19] DAR-EM17/1: Anhang I zu DAR-EM17: Empfehlungen zur Begutachtung von Laboratorien. Version 3.0 12/97, DAR, Berlin 1997.
- [20] SIEBEL, E.: Grundlagen und Entwicklung der Werkstoffprüfung. In: E. Siebel (Hrg.), Handbuch der Werkstoffprüfung, 1. Band, Prüf- und Messeinrichtungen. 2. Auflage, S. 1-14. Springer Verlag Berlin, Göttingen, Heidelberg 1958.
- [21] FRENZ, H.; KÜHN, H.-J.; FRITZ, T.: Der Zugversuch. Ergebnisse eines Vergleichsversuches mit 91 europäischen Prüflaboratorien. Materialprüfung 40 (1998) 11-12, S. 480-485, Carl Hanser Verlag, München 1998.
- [22] TRIEBEL, J.; FRENZ, H.: Begutachtung von Prüflaboratorien auf der Grundlage der neuen DIN EN ISO/IEC 17025. DVM-Tagung Werkstoffprüfung 2000, Bad Nauheim 7. und 8.12.2000, Tagungsband S. 201-206, DVM 2000.
- [23] RZ DAP-1: Final Report „Tensile Test according to DIN EN 10002 part 1; Results of a Proficiency Testing Programme for Mechanical Testing Laboratories.“ BAM-V.21-report, Berlin, Juli 1998.
- [24] VOGT, C.; LANDGREBE, R.; LEHMANN, C.; TRIEBEL, J.; FRENZ, H.: Härteprüfung nach Vickers – Ergebnisse einer Eignungsprüfung mit 108 europäischen Prüflaboratorien. In: Materialprüfung 43 (2001) 5, S. 164-171. Carl Hanser Verlag, München 2001.
- [25] Eignungsprüfung Härteprüfung nach Vickers, EPHV-1 HV 10/30: Abschlußbericht. FH Gelsenkirchen, Fachbereich Chemie und Materialtechnik, Recklinghausen, November 1999.
- [26] Eignungsprüfung Metallographie MET-1/00: Abschlussbericht. FH Gelsenkirchen Fachbereich Chemie und Materialtechnik, November 2000.
- [27] Eignungsprüfung Emissionsspektrometrie EMS 2001, Abschlußbericht. Institut für Eignungsprüfung, Recklinghausen, November 2001.
- [28] Eignungsprüfung Härteprüfung nach Brinell HBW 2001, Abschlußbericht. Institut für Eignungsprüfung, Recklinghausen, Juni 2002.
- [29] Eignungsprüfung Kerbschlagbiegeversuch CIT 2001, Abschlußbericht. Institut für Eignungsprüfung, Recklinghausen, Dezember 2001.
- [30] BAUER, C.O.: Notwendiges Wissen vom Recht für Prüffingenieure. Sonderdruck Fachbeitrag aus den ZfP-Zeitungen, Ausgaben 66-68, 1999.
- [31] FORSTÉN, J.: Are we on the right track towards improving the quality of test results? In: Accred Qual Assur (1998) 3: pp. 222-223, 1998.
- [32] HENSELER, P.: Rechtliche Aspekte von Prüfbescheinigungen. In: Tagungsband „Sicherheit durch Prüfbescheinigungen“ 5.2.02 in Kassel, Deutsches Institut für Normung, Berlin 2002.
- [33] HACH, H. F.: Bleche und Band sowie gewalzte Stäbe für den Druckbehälterbau komplett europäisch genormt. In: Jahrbuch Stahl 2001, Band 1, S. 336 – 339, Verlag Stahleisen GmbH, Düsseldorf 2000.
- [34] FRENZ, H.; BOLEY, N. ; VAN DER VEEN, A.: Training Module: How to conduct proficiency testing, Second draft, Recklinghausen, Oktober 1998.

- [35] EA-03/04, Rev. 1: Use of Proficiency Testing as a Tool for Accreditation in Testing. EA-EUROLAB-EURACHEM Group, August 2001.
- [36] DAR-EM32: Anleitung zur Akkreditierung nach ISO/IEC 17025. Version 2.0, 03/2001, DAR, Berlin 2001.
- [37] ILAC-G13:2000, Guidelines for the Requirements for the Competence of Providers of Proficiency Testing Schemes. ILAC Technical Accreditation Issues Committee, 2000.
- [38] AQS Baden-Württemberg am ISWA:
http:// www.uni-stuttgart.de/siwa/ch/ags/allgemeines.htm, Stand April 2002.
- [39] Land Baden-Württemberg: WassR 2.3.9: Verordnung des Ministeriums für Umwelt und Verkehr über sachverständige Stellen in der Wasserwirtschaft, vom 02. Mai 2001 (GBl. S. 399).
- [40] AQS Baden-Württemberg am ISWA:
http:// www.uni-stuttgart.de/siwa/ch/aqs/aqsbw.htm , Stand April 2002.
- [41] WAGNER, R. J.: Ringversuche im Rahmen der Analytischen Qualitätssicherung. In: tm-Technisches Messen 59 (1992), Heft 4, S. 167-172.
- [42] Analytische Qualitätssicherung Baden-Württemberg: 6. Länderübergreifender Abwasser-Ringversuch (AOX, TOC, CSB und Phenol-Index sowie optional MKW), Stuttgart, Dezember 2001.
- [43] ISO/IEC Guide 2:1996, General terms and their definitions concerning standardization and related activities. International Organisation for Standardization, Genève, 1996.
- [44] ISO 3534-1:1993, Statistics -- Vocabulary and symbols -- Part 1: Probability and general statistical terms. International Organisation for Standardization, Genève, 1993.
- [45] DIN EN 45003:1995, Akkreditierungssysteme für Prüf- und Kalibrierlaboratorien. Allgemeine Anforderungen für Betrieb und Anerkennung. Beuth Verlag, Berlin, 1995.
- [46] International vocabulary of basic and general terms in metrology: 1993 issued by BIPM; IEC; IFCC; ISO; IUPAC: IUPAP; OIML.
- [47] DIN 55350-13:1987, Begriffe der Qualitätssicherung und Statistik. Teil 13: Begriffe zur Genauigkeit von Ermittlungsverfahren und Ermittlungsergebnissen. Beuth Verlag, Berlin, 1987.
- [48] DIN ISO 5725-1:1997, Genauigkeit (Richtigkeit und Präzision) von Meßverfahren und Meßergebnissen. Teil 1: Allgemeine Grundlagen und Begriffe. Beuth Verlag, Berlin 1997.
- [49] MÄDER, M.: Auswertung von Ringversuchen mit Hilfe robuster Statistik. Dissertation, Würzburg 1996.
- [50] Forschungsgesellschaft für das Straßenwesen: Merkblatt über die statistische Auswertung von Prüfergebnissen, Teil 3: Planung und Durchführung von Ringversuchen, Ausgabe 1979.
- [51] ISO/IEC Guide 43-2:1997, Proficiency Testing by Interlaboratory Comparisons. Part 2: Selection and use of proficiency testing schemes by laboratory accreditation bodies. International Organisation for Standardization, Genève, First Edition 1997.

- [52] DIN V 55394-2:2000, Eignungsprüfung durch Vergleiche zwischen Laboratorien. Teil 2: Auswahl und Anwendung von Programmen zur Eignungsprüfung durch Stellen, die Laboratorien akkreditieren. Beuth Verlag, Berlin, 2000.
- [53] CORTEZ, L. et al.: How to interpret information from proficiency test exercises concerning the relative performance of accredited laboratories. *EEE/pt(01)33*.
- [54] E DIN ISO 5725-2:2000, Genauigkeit (Richtigkeit und Präzision) von Meßverfahren und Meßergebnissen. Teil 2: Grundlegende Methode für die Ermittlung der Wiederhol- und Vergleichspräzision eines vereinheitlichten Messverfahrens. Beuth Verlag, Berlin 2000.
- [55] EPTIS 2002, www.eptis.bam.de; "European Information System on Proficiency Testing Schemes", European Community Concerted Action SMT4-CT98-8002
- [56] DACH Deutsche Akkreditierungsstelle Chemie GmbH: 1. DACH-Ringversuch Bestimmung der Globalmigration aus Kunststoffen in wässrigen Prüfsimulantien
- [57] KOCH, M.: Ringversuche als Werkzeug zur Analytischen Qualitätssicherung. In: *Ges. Deutscher Chemiker (Hg.), Tagung Umwelt und Chemie vom 7.10. bis 10.10. 1996 in Ulm, 1996*.
- [58] EA Interlaboratory Comparison MT 1: Tensile Testing of Metals, Final Report, Czech Testing Laboratories Association (CTLA) Czech Accreditation Institute (CAI), March 2000
- [59] BAM, RZ DAP-1: Zugversuch nach DIN EN 10002 Teil 1: Ergebnisse eines Vergleichsversuches mit 50 deutschen Prüflaboratorien. Berlin, 1998.
- [60] BOYER, K.; HORWITZ, W.; ALBERT, R.: Interlaboratory Variability in Trace Element Analysis. In: *Anal. Chem.* 1985, 57, pp. 454-459.
- [61] DÜRR, W.; DAMMANN, V.; DINKLOH, L.; FUNK, W.; KRUTZ, H.; VONDERHEID, C.: Vergleich von Auswerteverfahren für Ringversuche anhand realer Daten. In: *Vom Wasser* 1981, 57, S.183-188.
- [62] BLANDFORT, H.; BUTTSTEDT, R.; HEMMICH, H.-J.; JUNG, H.; RINNE, D.; TRISSLER, K.; WALIGORSKI, F.: Erfahrungsbericht über Ringversuche in der Wasserwirtschaftsverwaltung von Rheinland-Pfalz. In: *Z. Wasser-, Abwasser-Forsch.* 20, S. 168-173, VCH Verlagsges. mbH, Weinheim, 1987.
- [63] ASTM C 802: 1996, Standard Practise for Conducting an Interlaboratory Test Program to Determine the Precision of Test Methods for Construction Materials.
- [64] VEEN, A.H.M. van der; ALINK, A.: A practitioner's approach to the assessment of the quality of sampling, sample preparation, and sub sampling. In: *Accreditation and quality assurance* 3 (1998), pp. 20-26.
- [65] ISO Guide 34: 1996, Quality system guidelines for the production of reference materials. International Organization for Standardization, Geneva, 1996.
- [66] ISO Guide 35: 1989, Certification of reference materials – General and statistical principles. International Organization for Standardization, Geneva, 1989.
- [67] PAUWELS, J.; LAMBERTY, A.; SCHIMMEL, H.: Homogeneity testing of reference materials. In: *Accreditation and Quality Assurance* 3 (1998), pp. 51-55.
- [68] NOACK, S.: Homogenitätsprüfungen an Ausgangsmaterialien zur Herstellung von zertifiziertem Referenzmaterial. In: *QMH – I.13 UE – StAA-7-ANA 7.7*, Bundesanstalt für Materialforschung und –prüfung (BAM), Berlin, 1998.

- [69] ZÖFEL, P.: Statistik verstehen. Ein Begleitbuch zur computergestützten Anwendung. Addison-Wesley Verlag München, 2001.
- [70] MONKA, M.; VOB, W.: Statistik am PC, Lösungen mit Excel. 2. vollst. überarbeitete Aufl., Carl Hanser Verlag München, Wien, 1999.
- [71] BÜHL, A.; ZÖFEL, P.: SPSS Version 10. Einführung in die moderne Datenanalyse unter Windows. 7. überarb. Aufl., Addison-Wesley Verlag München, 2000.
- [72] DIN V ENV 13005:1999, Leitfaden zur Angabe der Unsicherheit beim Messen. Beuth Verlag, Berlin, Juni 1999.
- [73] ISO/DIS 13528:2002 (E), Statistical methods for use in proficiency testing by interlaboratory comparisons. International Organization for Standardization, Geneva, Feb. 2002.
- [74] VOGT, C.: How to Determine the Assigned Value in Proficiency Tests for Accredited Material-testing Laboratories. In: MAT 2001 Proceedings, MAT 2001 International Conference Material Testing and Research, Nuremberg, May 8-10, 2001. AMA Service GmbH, Wunstorf, 2001.
- [75] SENONER, M.: Auswertung von Ringversuchen zur Verfahrensvalidierung. In: Materialprüfung 43 (2001) 7-8, S. 298-303. Carl Hanser Verlag, München 2001.
- [76] GRUBBS, F.: Procedures for Detecting Outlying Observations in Samples. In: Technometrics, Vol. 11, No. 1, February 1969, pp. 1-21.
- [77] GRUBBS, F.; BECK, G.: Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations. In: Technometrics, Vol. 14, No. 4, November 1972, pp. 847-854
- [78] STREULI, H.: Fehlerhafte Interpretation und Anwendung von Ausreißer-Tests, insbesondere bei Ringversuchen zur Überprüfung analytisch-chemischer Untersuchungsmethoden. In: Fresenius Z. Anal. Chem. 303, 406-408 (1980).
- [79] DAVIES, P. L.: Statistical evaluation of interlaboratory tests. In: Fresenius Z. Anal. Chem. 331, pp. 513-519 (1988).
- [80] ANSCOMBE, F.J.: Rejection of Outliers. In: Technometrics 2, 1960, pp. 123-147.
- [81] E DIN ISO 5725-5, 1996, Genauigkeit (Richtigkeit und Präzision) von Meßverfahren und Meßergebnissen. Teil 5: Alternative Methoden für die Ermittlung der Richtigkeit eines vereinheitlichten Messverfahrens. Beuth Verlag, Berlin 1996.
- [82] HORWITZ, W.: Evaluation of Analytical Methods Used for Regulation of Foods and Drugs. In: Analytical Chemistry, Vol. 54, No.1, January 1982, pp. 67A-76A.
- [83] LÜCK, D.: Practical aspects concerning planning and realisation of proficiency tests; statistical evaluation. Bundesanstalt für Materialforschung und –prüfung (BAM), Berlin 1996.
- [84] 2. BAM-Ringversuch “Altlasten”: Polyzyklische aromatische Kohlenwasserstoffe (PAK) im Boden, Abschlussbericht. Bundesanstalt für Materialforschung und –prüfung, Berlin, Dezember 1997.
- [85] YOUNDEN, W. J.: Graphical Diagnosis of Interlaboratory Test Results. In: Journal of Quality Technology Vol. 4, No. 1, January 1972, pp. 29-33.
- [86] SACHS, L.: Angewandte Statistik. Anwendung statistischer Methoden. Neunte, überarbeitete Auflage, Springer Verlag, Berlin, Heidelberg 1999.

- [87] APLAC T002: Tensile Testing of Metal, Proficiency Testing Program, Final Report, Rhodes, NSW, Australia 1997.
- [88] WAWRZINIÖK, O.: Handbuch des Materialprüfungswesens für Maschinen- und Bauingenieure. Zweite vermehrte und vollständig umgearbeitete Auflage. Julius Springer Verlag, Berlin, 1923.
- [89] Schrader, A.: Ätzheft, Berlin 1957.
- [90] DIN 17100:1980, Allgemeine Baustähle. Beuth Verlag Berlin, Januar 1980.
- [91] DIN EN ISO 9001:2000, Qualitätsmanagementsysteme Anforderungen. Beuth Verlag, Berlin, Dezember 2000.
- [92] VDA 6 Qualitätsmanagement in der Automobilindustrie, Teil 1: QM-Systemaudit, 4. überarbeitete Auflage, Verband der Automobilindustrie e. V. (VDA), Frankfurt a. M., 1998.
- [93] DIN EN 10002-1:2001, Zugversuch. Teil 1: Prüfverfahren bei Raumtemperatur. Beuth Verlag, Berlin, Dezember 2001.
- [94] MERKEL, M.; THOMAS, K.-H.: Taschenbuch der Werkstoffe. 5., neubearbeitete Auflage. Fachbuchverlag Leipzig, 2000.
- [95] EA Interlaboratory Comparison MT 1, Tensile Test of Metals: Final Report. Czech Testing Laboratories Association (CTLA), Czech Accreditation Institute (CAI), Prague, March 2000.
- [96] VOGT, C.; FRENZ, H.: „Die Eignungsprüfung als Instrument zur Kompetenzbewertung von Prüflaboratorien für mechanisch-technologische Werkstoffprüfungen“, Tagungsband DVM-Tagung Werkstoffprüfung 2000, DVM e. V., Berlin, S. 207-212, 2000.
- [97] AEGERTER, J.; BOCHING, H.; SONNE, H.-M.: Influence of Testing Speed on the Yield/Proof Strength. Materialprüfung, 43 (2001) (10), S. 393-403, Carl Hanser Verlag, München 2001.
- [98] GRAU, P.; MEINHARD, H.: Härtemessungen auf rauen Oberflächen. In: Härtereitechnische Mitteilungen (2001) Heft 4.
- [99] DIN EN ISO 6507-1:1998, Härteprüfung nach Vickers. Teil 1: Prüfverfahren. Beuth Verlag, Berlin, Januar 1998.
- [100] BARGEL, H.-J.; CARDINAL, P.; HILBRANS, H.; HÜBNER, K.-H.; SCHULZE, G.; WURZEL, G.; Hrsg. von H.-J. Bargel und G. Schulze: Werkstoffkunde. 5. Aufl. – Düsseldorf, VDI-Verlag, 1988.
- [101] DIN EN ISO 6507-2:1998, Härteprüfung nach Vickers. Teil 2: Prüfung der Prüfmaschinen. Beuth Verlag, Berlin, Januar 1998.
- [102] DIN EN ISO 6506-1:1999, Härteprüfung nach Brinell. Teil 1: Prüfverfahren. Beuth Verlag, Berlin, Oktober 1999.
- [103] GABAUER, W., Manual of Codes of Practice for the Determination of Uncertainties in Mechanical Tests on Metallic Materials, The Estimation of Uncertainties in Hardness Measurements, Project, No. SMT4-CT97-2165, UNCERT COP 14: 2000
- [104] KÜHLMEYER, M.: Statistische Auswertungsmethoden für Ingenieure. Springer Verlag Berlin, Heidelberg, New York, 2001.

- [105] DENGEL, D.: Genauigkeit von Werkstoffprüfergebnissen. In: Materialprüfung (39) 1997 (5), S. 190-195. Carl Hanser Verlag, München 1997.
- [106] POLZIN, T., SCHWENK D., Estimation of uncertainty of hardness testing; PC file for the determination, Materialprüfung (44) 2002 (3), S. 64-71. Carl Hanser Verlag, München 2002.
- [107] PAPULA, L., Mathematik für Ingenieure und Naturwissenschaftler, Band 3: Vektoranalysis, Wahrscheinlichkeitsrechnung, Mathematische Statistik, Fehler- und Ausgleichsrechnung. Friedrich Vieweg & Sohn Verlagsgesellschaft, Braunschweig/Wiesbaden, 3. verb. Auflage 1999.
- [108] DIN EN ISO 6506-2:1999, Härteprüfung nach Brinell. Teil 2: Prüfung und Kalibrierung der Härteprüfmaschinen. Beuth Verlag, Berlin, Oktober 1999.
- [109] DOMKE, W.: Werkstoffkunde und Werkstoffprüfung. 9. verbesserte Auflage, Verlag W. Giradet, Essen, 1981.
- [110] SCHRAMM, J.: Einrichtungen und Verfahren der metallographischen Prüfung. In: E. Siebel (Hrg.), Handbuch der Werkstoffprüfung, 1. Band, Prüf- und Messeinrichtungen. 2. Auflage, S. 677-778. Springer Verlag Berlin, Göttingen, Heidelberg 1958.
- [111] DIN 50601:1985, Ermittlung der Ferrit- oder Austenitkorngröße von Stahl und Eisenwerkstoffen. Beuth-Verlag Berlin, August 1985.
- [112] ASTM E 562-99, Standard Test Method for Determining Volume Fraction by Systematic Manual Point, ASTM Philadelphia, 1999.
- [113] ISO 9042-88 (E), Steels-Manual point counting method for statistically estimating the volume fraction of a constituent with a point grid International Organization for Standardization, Geneva, 1988.
- [114] ISO 643-83, Stähle - mikrographische Bestimmung der ferritischen oder austenitischen Korngröße. International Organization for Standardization, Geneva Sep. 1983.
- [115] DIN 17440:1996, Nichtrostende Stähle. Technische Lieferbedingungen für Blech, Warmband und gewalzte Stäbe für Druckbehälter, gezogenen Draht und Schmiedestücke. Beuth-Verlag Berlin, September 1996.
- [116] DIN EN 10025:1994, Warmgewalzte Erzeugnisse aus unlegierten Baustählen. Technische Lieferbedingungen. Beuth-Verlag Berlin, März 1994.
- [117] HARTUNG, J.; ELPELT, B., KLÖSENER, K.-H.: Statistik. Lehr- und Handbuch der angewandten Statistik. 12., unwesentlich veränderte Auflage – München; Wien, Oldenbourg, 1999.
- [118] DIN EN 10045-1:1991, Kerbschlagbiegeversuch nach Charpy. Teil 1: Prüfverfahren. Beuth Verlag, Berlin, April 1991.
- [119] LASKA, R.; FELSCH, C.: Werkstoffkunde für Ingenieure. 2. durchgesehene Auflage, Fr. Vieweg & Sohn Verlagsgesellschaft mbH Braunschweig, 1987.
- [120] DIN EN 10028-2:1993, Flacherzeugnisse aus Druckbehälterstählen; Teil 2: Unlegierte und legierte warmfeste Stähle. Beuth Verlag, Berlin, April 1993.
- [121] DIN EN 10204:1995, Metallische Erzeugnisse - Arten von Prüfbescheinigungen. Beuth Verlag, Berlin, August 1995.

- [122] DIN EN 10045-2:1993, Kerbschlagbiegeversuch nach Charpy. Teil 2: Prüfung der Prüfmaschine (Pendelschlagwerk). Beuth Verlag, Berlin, Januar 1993.
- [123] BROEKAERT, J. A. C.: Atomemissionsspektroskopie. In: Naumer, H., Heller, W. (Hrsg): Untersuchungsmethoden in der Chemie, S. 293-307, Stuttgart, NewYork, 1990.

8 Definitionen und Erläuterungen verwendeter Begriffe

Im Rahmen dieser Arbeit gelten die folgenden Definitionen und Begriffsbestimmungen:

Prüfung – technischer Vorgang, der aus dem Ermitteln eines oder mehrerer Merkmale eines Produktes, Prozesses oder einer Dienstleistung besteht und nach einer vorgeschriebenen Verfahrensweise durchgeführt wird [43].

Prüflaboratorium – Laboratorium, das Prüfungen (und / oder Kalibrierungen) durchführt [43].

Prüfgegenstand (synonym: Probe) – Werkstoff oder Gegenstand, der dem teilnehmenden Laboratorium für die Eignungsprüfung übergeben wird [10].

Prüfverfahren – vorgegebene technische Verfahrensweise zur Durchführung einer Prüfung [43].

Ermittlungsergebnis (synonym: Prüfergebnis) – Wert eines Merkmals, gewonnen durch die Anwendung eines vereinheitlichten Messverfahrens [44].

Akkreditierung – Verfahren, durch welches eine maßgebliche Stelle formell anerkennt, dass eine Organisation bzw. Person kompetent ist, bestimmte Aufgaben auszuführen [45].

Inspektion – Überprüfung eines Produktmusters, Produktes, einer Dienstleistung, eines Prozesses bzw. einer Anlage und Feststellung seiner Übereinstimmung (Konformität) mit spezifischen Anforderungen bzw. – auf Grundlage eines fachlichen Urteils – mit allgemeinen Anforderungen [45]. Die Inspektion von Prozessen schließt Personal, Einrichtungen, Technologie und Methodik ein. Die Ergebnisse einer Inspektion können zur Unterstützung der Zertifizierung verwendet werden [34].

Eignungsprüfung (eines Prüflaboratoriums) – Bestimmung der Leistungsfähigkeit eines Prüflaboratoriums mittels Vergleichsprüfungen [43].

Vergleichsprüfungen durch Prüflaboratorien – Organisation, Durchführung und Bewertung von Prüfungen gleicher oder gleichartiger Gegenstände oder Stoffe durch zwei oder mehrere Prüflaboratorien unter vorgegebenen Bedingungen [10].

Referenzmaterial (RM) – Material oder Substanz von ausreichender Homogenität, von dem bzw. der ein oder mehrere Merkmalswerte so genau festgelegt sind, dass sie zum Kalibrieren von Messgeräten, zur Beurteilung von Messverfahren oder Zuweisung von Stoffwerten verwendet werden können [46].

Zertifiziertes Referenzmaterial (ZRM) – Referenzmaterial mit einem Zertifikat, in dem unter Angabe der Unsicherheit und des zugehörigen Vertrauensniveaus ein oder mehrere Merk-

malswerte aufgrund eines Ermittlungsverfahrens zertifiziert sind, mit dem die Rückverfolgbarkeit der Werte auf eine genaue Realisierung der Einheit erreicht wird [46].

Referenzlaboratorium – Laboratorium, das einem Prüfgegenstand Referenzwerte zuordnet.

Stoffwert (engl. assigned value) – kennzeichnender Wert für eine bestimmte Größe, der, manchmal durch Übereinkunft, mit einer Unsicherheit angenommen wird, die für einen bestimmten Zweck geeignet ist [10, 11, 46].

Rückverfolgbarkeit – Eigenschaft eines Messergebnisses oder des Wertes eines Normals, durch eine ununterbrochene Kette von Vergleichsmessungen mit angegebenen Messunsicherheiten auf geeignete Normale, im allgemeinen internationale oder nationale, bezogen zu sein [46].

Koordinator – Organisation oder Person mit Verantwortlichkeit für die Koordinierung sämtlicher Tätigkeiten, die mit der Durchführung eines Programms der Eignungsprüfung verbunden sind [10].

Wahrer Wert – tatsächlicher Merkmalswert unter den bei der Ermittlung herrschenden Bedingungen [47].

Richtiger Wert – Wert für Vergleichszwecke, dessen Abweichung vom wahren Wert für den Vergleichszweck als vernachlässigbar betrachtet wird [47].

Erwartungswert – das mittlere Ermittlungsergebnis, welches aus der unablässig wiederholten Anwendung des unter vorgegebenen Bedingungen angewendeten Ermittlungsverfahrens gewonnen werden könnte [47].

Genauigkeit – qualitative Bezeichnung für das Ausmaß der Annäherung von Ermittlungsergebnissen an den Bezugswert, wobei dieser je nach Festlegung oder Vereinbarung der wahre, der richtige oder der Erwartungswert sein kann [47].

Richtigkeit – qualitative Bezeichnung für das Ausmaß der Annäherung des Erwartungswertes des Ermittlungsergebnisses an den Bezugswert, wobei dieser je nach Festlegung oder Vereinbarung der wahre oder der richtige Wert sein kann [47].

Präzision – qualitative Bezeichnung für das Ausmaß der gegenseitigen Annäherung voneinander unabhängiger Ermittlungsergebnisse bei mehrfacher Anwendung eines festgelegten Ermittlungsverfahrens unter vorgegebenen Bedingungen [47].

Wiederholbedingungen – Bedingungen bei der Gewinnung von voneinander unabhängigen Ermittlungsergebnissen, bestehend in der Anwendung desselben Verfahrens am identischen Untersuchungsobjekt im selben Labor durch denselben Bearbeiter mit derselben Geräteausstattung in kurzen Zeitabstände [48].

Vergleichsbedingungen – Bedingungen bei der Gewinnung von voneinander unabhängigen Ermittlungsergebnissen, bestehend in der Anwendung desselben Verfahrens an identischen Untersuchungsobjekten in verschiedenen Laboratorien durch verschiedene Bearbeiter mit verschiedenen Geräteausrüstungen [48].

(Untersuchtes) Merkmalsniveau in einer Eignungsprüfung – Der Gesamtmittelwert der Ermittlungsergebnisse von allen Laboratorien für ein untersuchte einzelnes Material oder eine einzelne untersuchte Probe [48].

Ausreißer – ein Wert unter einer Serie von Werten, der mit den anderen Werten dieser Serie nicht verträglich ist [48].

Extremwerte – Ausreißer und sonstige Werte, die nicht zu den übrigen Elementen eines Datensatzes passen. Diese Ergebnisse können einen starken Einfluss auf statistische Berechnungen mit Summenbildung und Standardabweichung ausüben [10, 11].

Stabile (robuste) statistische Verfahrensweisen – Verfahrensweisen, mit denen der Einfluss minimiert wird, den Extremwerte auf Schätzungen von Mittelwert und Standardabweichung haben können [10, 11].

Messunsicherheit – den Messergebnissen zugeordneter Parameter, der die Streuung der Werte kennzeichnet, die vernünftigerweise der Messgröße zugeordnet werden können [46].

9 Aufstellung der verwendeten Abkürzungen

AES	Atomemissionsspektroskopie
AQS	Analytische Qualitätssicherung
ANOVA	Analysis of Variances (einfaktorielle Varianzanalyse)
APLAC	Asia Pacific Laboratory Accreditation Co-operation
ASTM	American Society for Testing and Materials
BAM	Bundesanstalt für Materialforschung und -prüfung
CIT	Charpy Impact Test (Kerbschlagbiegeversuch nach Charpy)
COP	UNCERT Code of Practice
CRM	Certified Reference Material
DAR	Deutscher Akkreditierungsrat
DAP	Deutsches Akkreditierungssystem Prüfwesen GmbH
DIN	Deutsches Institut für Normung e. V.
EA	European co-operation for Accreditation
EPTIS	European Proficiency Testing Information System
IEC	International Electrotechnical Commission
IfEP	Institut für Eignungsprüfung GbR, Recklinghausen
ILAC	International Laboratory Accreditation Cooperation
IQR	Interquartile Range (Quartilabstand)
nIQR	Normalised Interquartile Range (normalisierter Quartilabstand)
ISO	International Organization for Standardization
ISWA	Institut für Siedlungswasserbau, Wassergüte- und Abfallwirtschaft der Universität Stuttgart
LMPA	Landesmaterialprüfamt
MPA	Staatliche Materialprüfungsanstalt
MTP	Mechanisch-technologische (Werkstoff)Prüfung
OES	Optische Emissionsspektroskopie
STABW	Standardabweichung
TÜV	Technischer Überwachungsverein
TU	Technische Universität
ZRM	Zertifiziertes Referenzmaterial

10 Verzeichnis der verwendeten Formelzeichen

A	Bruchdehnung in %
A_V	Kerbschlagarbeit in Joule
B	systematische Abweichungskomponente unter Wiederholbedingungen
C	Prüfgröße nach Cochran
D	Schätzwert für die systematische Laborabweichung
$D_{\%}$	relative Differenz
E	Elastizitätsmodul in GPa
F	Prüfgröße
G	Prüfgröße nach Grubbs
MED	Median
MED_{Stabw}	Median der Standardabweichungen
MW	arithmetischer Mittelwert
MQ	mittlere Quadrate
OG	oberer Grenzwert
U	erweiterte Messunsicherheit
\tilde{U}	relative erweiterte Messunsicherheit
UG	unterer Grenzwert
R	Spannweite
R_{eH}	obere Streckgrenze in MPa
R_m	Zugfestigkeit in MPa
$R_{p0,2}$	0,2-%-Dehngrenze in MPa
SAQ	Summe der Abweichungsquadrate
T	Intervallgrenze
VAR	Gesamtvarianz
X	Stoffwert
Z	Brucheinschnürung in %
e	zufällige Abweichung
g	Größe einer entnommenen Stichprobe
$h_{i,j}$	Vereinbarkeitsprüfgröße nach Mandels h-Statistik
k	Anzahl der Stichproben (Probemengen)
$k_{i,j}$	Vereinbarkeitsprüfgröße nach Mandels k-Statistik
m	Erwartungswert
n	Anzahl der Einzelmessungen
p	Anzahl der Laboratorien, Datensätze
s	Schätzwert für die Standardabweichung σ
u_X	Standardmessunsicherheit des Stoffwertes

u_E	Standardmessunsicherheit aus der Grenzabweichung (1σ) der Härteprüfmaschine
$u_{\bar{x}_{CRM}}$	Messunsicherheit (1σ) der Härteprüfmaschine zum Kalibrieren der Härtevergleichsplatte
u_{CRM}	Standardmessunsicherheit der Härtevergleichsplatte
$u_{\bar{H}}$	Standardmessunsicherheit der Härteprüfmaschine bei Messung auf der Härtevergleichsplatte
u_{MS}	Standardmessunsicherheit durch den kleinsten Ableseschritt
$u_{\bar{x}}$	Standardmessunsicherheit bei der Prüfung eines Probematerials
s_s	Standardabweichung zwischen zwei Stichproben
s_w	Standardabweichung innerhalb einer Stichprobe
s^*	robuste Standardabweichung
t	Nummer der Stichprobe
w_t	Spannweite zwischen den Probemengen
x	(Einzel-) Messwert
\bar{x}	arithmetischer Mittelwert
\bar{x}	Gesamtmittelwert
x^*	robuster Mittelwert
\bar{y}	Gesamtmittelwert der Homogenitätsprüfung
z	z-Score
z'	z'-Score
α	statistische Irrtumswahrscheinlichkeit
ε_t	Gesamtdehnung in %
ε_p	Plastische Dehnung in %
σ	Spannung in MPa
$\hat{\sigma}$	Standardabweichung einer Eignungsprüfung
σ_L	Standardabweichung zwischen den Laboratorien
σ_R	Vergleichsstandardabweichung
σ_r	Wiederholstandardabweichung
$\hat{\rho}$	Korrelationskoeffizient
ζ	Zeta-Score