CNAS-GL032

能力验证的选择核查与利用指南

Guidance on the Selection，Review and Use of Proficiency Testing

中国合格评定国家认可委员会

2018 年 3 月 1 日 发布　　2018 年 3 月 1 日 实施

目 录

前 言

能力验证是利用实验室间比对，按照预先制定的准则评价参加者的能力。参加能力验证是实验室质量保证的重要手段，有助于实验室评价和证明其测量数据可靠性，发现自身存在的问题，改进实验室的技术能力和管理水平。能力验证的结果可作为证明实验室技术能力的有效证明，为管理部门、认可机构、客户和其他利益相关方选择、评价、认可有能力的实验室提供依据。

实验室作为参加能力验证的主体，需基于自身需求和外部对能力验证的要求，在综合考虑内部质控水平、人员能力、设备状况、风险、运行成本等因素的基础上，合理策划并积极寻求适当的能力验证计划。

实验室可获得的能力验证除认可的能力验证提供者（PTP）组织的能力验证计划外，还有大量非认可的PTP包括行业组织运作的能力验证计划。对于认可的PTP在其认可范围内开展的能力验证计划，实验室可根据需求选择参加并直接利用能力验证结果。对于其他的能力验证计划，实验室可参照本指南和其他相关标准、指南等文件对其进行核查，以确认其是否满足要求。

关于能力验证的组织运作和统计评价有两个通用标准：GB/T 27043《合格评定 能力验证的通用要求》（ISO/IEC 17043:2010）和ISO 13528:2015《利用实验室间比对进行能力验证的统计方法 》（Statistical methods for use in proficiency testing by interlaboratory comparison ）。本指南文件中有关能力验证核查部分（包括结果分析、指定值、能力评定标准差和能力评定等）主要参考ISO 13528:2015中相关内容而制定。

能力验证的选择核查与利用指南

1 目的范围

本文为能力验证参加者和其他利益相关方（如认可机构、监管机构或实验室的客户）选择、核查和利用能力验证提供指南，也可为实验室开展质量控制提供指导。

2 引用文件

下列文件中的条款通过引用而成为本文件的条款。以下引用的文件，注明日期的，仅引用的版本适用；未注明日期的，引用文件的最新版本（包括任何修订）适用。

GB/T 27043 合格评定 能力验证的通用要求（ISO/IEC 17043，IDT）

ISO 13528 利用实验室间比对进行能力验证的统计方法 （Statistical methods for use in proficiency testing by interlaboratory comparison）

GB/T 15000.3 标准样品 定值的一般原则和统计方法

JJF 1059.1 测量不确定度评定与表示

CNAS-RL02 能力验证规则

CNAS-GL002 能力验证结果的统计处理和能力评价指南

CNAS-GL003 能力验证样品均匀性和稳定性评价指南

3 术语和定义

GB/T 27043 和 ISO 13528 界定的术语和定义适用于本文件。为方便使用，重复列出以下术语和定义：

3.1 能力验证 proficiency testing

利用实验室间比对，按照预先制定的准则评价参加者的能力。

注 1：在本指南中，术语“能力验证”具有极为广泛的含义，包括但不限于以下类型：

a) 定量计划 quantitative scheme

该类计划是确定能力验证物品的一个或多个被测量的量。

b) 定性计划 qualitative scheme

该类计划是对能力验证物品的一个或多个特性进行鉴别或描述。

c) 顺序计划 sequential scheme

该类计划是将检测或测量的一个或多个能力验证物品按顺序分发，并按期返回能力验证提供者。

d) 同步计划 simultaneous scheme

该类计划中，分发能力验证物品，在规定期限内同时进行检测或测量。

e) 单次计划 single occasion exercise

该类计划中，为单个需求提供能力验证物品。

f) 连续计划 continuous scheme

该类计划中，按规定间隔提供能力验证物品。

g) 抽样 sampling

该类计划中，为后续的分析抽取样品。

h) 数据转换和解释 data transformation and interpretation

该类计划中，提供成组的数据或其他信息，要求对信息进行处理以给出解释（或其他结论）。

注2：在医学领域的某些能力验证提供者，利用术语“外部质量评价(EQA)”表示其能力验证计划和/或更广义的计划。

3.2 能力验证计划 proficiency testing scheme

在检测、测量、校准或检验的某个特定领域，设计和运作的一轮或多轮次能力验证。

注：一项能力验证计划可以包含对能力验证物品的一种或多种特定类型的检测、校准或检查。对一个参加者进行“一对一”能力评价的能力验证计划有时也被称为测量审核。

3.3 指定值 assigned value

对能力验证物品的特定性质赋予的值。

3.4 离群值 outlier

一组数据中被认为与该组其他数据不一致的观测值。

注1：离群值可能来源于不同的总体，或是不正确记录或其他粗大误差的结果。

注2：多数能力验证计划采用离群值这一术语表示产生行动信号的结果，但这并非这一术语的预期用途。尽管离群值通常会产生行动信号，但行动信号亦可能由非离群值产生。

3.5 能力评定标准差 standard deviation for proficiency assessment

用于评价能力验证结果的离散性度量。

注1：该定义可以理解为由严格按要求操作的实验室所组成的一个假设总体，其测试结果的总体标准差。

注2：能力评定标准差仅适用于定比尺度和定距尺度的结果。

注3：并非所有的能力验证计划都根据结果的离散性进行评价。

3.6 最大允许误差 maximum permissible error criterion for differences

由规范或标准所允许的测量，相对于已知参考量值的测量误差的极限值。

3.7 稳健统计方法 robust statistical method

对给定概率模型假定条件的微小偏离不敏感的统计方法。

3.8 测量不确定度 measurement uncertainty

根据所用到的信息，表征赋予被测量的量值分散性的非负参数。

4 能力验证的选择

4.1 制定能力验证参加计划

实验室需制定能力验证参加计划，这是国际实验室认可合作组织（ILAC）的明确要求。获得或准备申请CNAS认可的实验室，参加的能力验证计划的领域、频次至少应满足CNAS能力验证政策的要求。实验室需根据自身需求制定适宜的能力验证参加计划，该计划需基于管理和技术方面的风险、实验室质量控制、管理部门和认可机构的要求和能力验证的可获得性等进行制定。

4.1.1 管理和技术方面的风险

制定的能力验证参加计划需重点考虑实验室可能存在的风险，包括（但不限于）以下内容：

1) 日常开展的检测或测量任务量多少；

2) 技术人员流动情况；

3) 溯源是否得到保证，如标准物质/标准样品是否可以获得；

4) 测量技术的稳定性；

5) 测量结果的重要程度，如司法鉴定结果要求较高的可信度；

6) 环境设施、仪器设备的变化情况。

4.1.2 实验室质量控制

参加能力验证是一种有效的外部质量控制方式，实验室制定的能力验证参加计划需考虑并结合开展的质控手段，如（但不限于）：

1) 定期使用（有证）标准物质/标准样品；

2) 不同技术方法间的比较；

3) 与其他实验室间的比对；

4) 其他内部质量控制，如留样再测、人员比对和仪器比对等。

4.1.3 管理部门和认可机构的要求

CNAS-RL02《能力验证规则》对申请认可和已获认可的实验室参加能力验证有明确的要求。有关实验室的政府主管部门和行业管理部门都将能力验证作为有

效的质量控制手段，明确要求或鼓励其管理的实验室参加适宜的能力验证。因此实验室制定的能力验证参加计划需满足管理部门和认可机构的要求。

4.1.4 能力验证的可获得性

由于技术和成本的原因，不是所有的领域(包括项目、方法和物品)都可开展能力验证。实验室在制定能力验证参加计划时，需考虑能力验证是否可获得，只有可获得的项目才考虑列入参加计划。

4.2 选择途径

实验室可从 CNAS 网站能力验证专栏和其他途径获取能力验证信息。实验室优先选择按照 ISO/IEC 17043 运作的能力验证计划，并按照以下顺序选择参加：

1）CNAS认可的能力验证提供者（PTP）以及已签署PTP相互承认协议（MRA）的认可机构认可的PTP在其认可范围内运作的能力验证计划；

2）未签署PTP MRA的认可机构依据ISO/IEC 17043认可的PTP在其认可范围内运作的能力验证计划；

3）国际认可合作组织运作的能力验证计划，例如：亚太实验室认可合作组织（APLAC）等开展的能力验证计划；

4）国际权威组织实施的实验室间比对，例如：国际计量委员会（CIPM）、亚太计量规划组织（APMP）、世界反兴奋剂联盟（WADA）等开展的国际、区域实验室间比对；

5）依据ISO/IEC 17043获准认可的PTP在其认可范围外运作的能力验证计划；

6）行业主管部门或行业协会组织的实验室间比对；

7）其他机构组织的实验室间比对。

4.3 选择依据

能力验证的选择主要从以下几个方面进行考虑：

1）选择的能力验证需符合实验室的预期目标。实验室的预期目标最好能和实施机构组织能力验证的目的一致，即使不一致也需满足实验室的预期目标。组织能力验证的目的通常有：评价实验室从事特定测量能力及监测其持续能力，识别实验室间的差异，建立方法的等效性和可比性，增强实验室客户的信心，帮助实验室提高能力，确认实验室声称的不确定度等。

2）能力验证样品的特性参数需与参加者的日常测量类似。测量方法需尽量与实验室日常使用的方法一致或类似。

3）校准领域能力验证的指定值（通常称为参考值）需具有可溯源性，参考值的测量不确定度需优于参加者。

4）当能力验证计划方案可获得时，可核查实施机构的样品制备、均匀性和稳定性评价、指定值和能力评定标准差的确定、能力评定等设计是否合理。

5）参加者应能获取能力验证作业指导书，并在能力验证计划结束后能获取最终报告。参加者可从最终报告中能获知样品制备、均匀性和稳定性检验和统计评价等内容。

6）实施机构是否可以连续提供能力验证，是否可为实验室提供技术服务，帮助实验室识别问题并提升水平。

5 能力验证核查

通常参加者可依据 GB/T 27043 核查实施机构是否按其要求组织开展能力验证计划。当实验室选择经认可的能力验证提供者在其认可范围内组织开展的能力验证计划，或国际权威组织实施的实验室间比对时，可直接利用能力验证结果。当实验室按顺序选择 4.2 中 5）至 7）项中的能力验证计划时，需对实施机构能力验证相关文件和信息进行核查，核查其组织的能力验证计划是否合理并满足实验室自身的预期目标。核查的内容包括：能力验证设计方案（可获得时）、样品类型和检测方法、作业指导书、样品均匀性和稳定性、结果分析、指定值、能力评定标准差、能力统计量和能力验证报告等必要内容。可参考本指南相关章节内容进行核查，同时填写下表《能力验证适宜性核查表》。

表 1　能力验证适宜性核查表

序号	核查项目	参考本指南章节	核查情况	备注
1	能力验证活动（PT）所用的检测/校准/检验方法是否是本合格评定机构的日常检测/校准/检验方法？	5.2 样品类型和测量方法	□是□否 □不适用	
2	组织/实施机构是否提供了如何完成 PT 的说明文件，例如作业指导书？	5.3 能力验证作业指导书	□是□否 □不适用	
3	组织/实施机构是否提供了有关 PT 物品的必要说明，例如物品的处置方法和存储条件。	5.3 能力验证作业指导书 附录 A	□是□否 □不适用	
4	组织/实施机构是否提供了有关 PT 物品的均匀性和/或稳定性评估的必要细节？PT 物品是否均匀和/或稳定？	5.4 样品的均匀性和稳定性 附录 B	□是□否 □不适用	
5	PT 物品是否与本合格评定机构日常检测/校准/检验物品类型和测量范围相同或相似？	5.2 样品类型和测量方法	□是□否 □不适用	
6	组织/实施机构是否给出了指定值（参考值）及其确定方式？校准项	5.6 指定值 附录 C	□是□否 □不适用	

	目是否给出了指定值（参考值）的计量溯源性和测量不确定度信息？	附录 D		
7	组织/实施机构是否给出了能力评价的方法并对本合格评定机构能力给出了评价？	5.7 能力评定标准差 附录 E 5.8 能力评价 附录 F	□是□否 □不适用	
8	组织/实施机构是否提供了 PT 报告？	5.9 能力验证报告	□是□否 □不适用	
9	本次 PT 是否满足本合格评定机构的需要？		□是□否	

注 1：本表由能力验证参加者填写。参加者可在选择能力计划时核查序号 1 和 5 相应的内容，收到样品时可核查序号 2 和 3 相应的内容，收到最终报告时可核查序号 4 和序号 6 至 9 相应的内容。

注 2：核查情况为“不适用”时，需在备注中给出说明。如果核查项目有“否”时，参加者选择“本次 PT 满足本合格评定机构的需要”，需在备注中给出说明。

5.1 能力验证设计方案

能力验证方案中通常包括以下内容：实施机构的联系方式，参加者的数量及类型，参加条件，潜在的误差来源，样品的特性或预期的量值范围，样品制备和均匀性、稳定性检验（包括方法及程序），样品储存、运输和分发，样品丢失或损害时采取的措施，检测/校准方法，日程安排，数据处理和采用的统计分析，指定值及不确定度的确定方法，指定值的计量溯源性，结果评价标准，反馈给参加者的数据等的描述，对计划的结果及结论公布的范围，参加者反馈结果的标准化报告格式等必要内容。

5.2 样品类型和测量方法

实验室选择的能力验证计划中的样品需尽量与其日常测量物品相同或类似，且被测量在大致相似的浓度水平（适用时），使用的测量方法需尽量与其日常使用的方法一致。如果能力验证计划允许多种测量方法，或允许实验室选择测量方法，实施机构应对方法的差异进行比对、分析，确保方法不会对参加者的能力评定产生影响。

5.3 能力验证作业指导书

实施机构需提供详细的文件化的作业指导书。指导书需包括：

1）要求参加者按照日常检测样品的处理方式处理能力验证物品（除非能力验证计划有特定要求）；

2）能力验证物品检测或校准影响因素的详细说明，例如：能力验证物品的性质、存储条件、是否限定检测方法，以及测量的时间要求；

3）进行检测或校准之前，能力验证物品的准备和/或状态调节的详细要求；

4）处置能力验证物品的适当指导，包括安全要求；

5）参加者检测和/或校准时特定的环境条件，如适用，要求参加者报告测量期间相关环境条件；

6）测量结果及其不确定度记录和报告方式的详细的说明。如果指导书要求报告结果的测量不确定度，需包括包含因子和包含概率（适用时）；

7）结果反馈的格式和具体要求，如测量单位、有效数字或小数位数、结果报告的依据（如按干基重量计或湿重计）等内容；

8）实施机构接收用于分析的能力验证结果的截止日期；

9）实施机构的详细联络信息；

10）适用时，返回或传递能力验证物品的说明。

能力验证作业指导书示例见附录 A

5.4 样品的均匀性和稳定性

参加者应可获取样品制备及其均匀性和稳定性等必要信息，包括：样品制备方式、均匀性和稳定性检测数据、取样方式、取样数量、测量方法、评价方式。对于校准领域能力验证计划，由于样品一般为测量仪器，通常只需考察样品的稳定性。

均匀性检检验和稳定性检验的测量方法应具有充分小的重复性标准差（S_r），以便能够检测出任何明显的不均匀和不稳定。测量方法的重复性标准差与能力评定标准差之比，通常应小于 0.5，即S_r<0.5σ_{pt}。

当不能进行重复测量，如进行破坏性测试时，可将测量结果的标准差当样品间标准差S_s使用，使用的测量方法应具有较小的重复性标准差。

在某些情况下，需对能力验证样品进行全部检验以确认样品的均匀性，如在电气领域的能力验证计划中，往往需对样品实施全部检验。

当不能进行均匀性和稳定性检验时，应证明能力验证样品的收集、制备、包装和分发程序可以充分满足能力验证要求。

对于批量样品，通常样品间的不均匀和不稳定标准差均不应超过$0.3\sigma_{pt}$（σ_{pt}为能力评定标准差）。有时实施机构仅可获得不完全均匀或稳定的材料，如果在指定值的不确定度中或者结果评价时考虑到这一点，这些材料仍可用作能力验证样品。

对于样品的均匀性，可利用$S_s \le 0.3\sigma_{pt}$（S_s为样品间标准差）准则或F检验进行评价；稳定性可利用$|\bar{x}-\bar{y}| \le 0.3\sigma_{pt}$准则或$t$检验进行评价。适当时也可用线性回归分析的方法判断样品稳定性。

样品均匀性和稳定性检验详见附录 B

5.5 结果分析

能力验证结果的数据分布决定所用的统计方法。很多能力验证的统计分析方法均假定数据呈近似正态分布，或者至少是单峰分布、对称（必要时可在转换后符合）分布。当使用稳健统计方法时，通常无需验证参加者结果是否成正态分布，但至少需确认其近似对称，这一点尤为重要。若无法确认对称性，需采用适宜于不对称分布的稳健统计方法。

参加者可通过数据直方图或核密度图核查结果分布是否对称，统计假设是否合理，是否存在异常（如双峰分布、离群值比例较大或异常偏倚）等。

基于正态分布的经典统计方法在应用过程中受离群值的影响，可能导致经典统计失效，因此采用经典统计方法前通常需剔除离群值。为了减少离群值的影响，可根据能力验证的目的和参加者情况，采用不同的稳健统计方法。通常可计算中位值或稳健均值作为指定值，计算标准化四分位距、尺度化中位绝对差（scaled median absolute deviation）或稳健标准差作为能力评定标准差。不同的稳健统计方法具有不同的适用条件、统计效率（efficiency）和失效点（breakdown point）。

参加者结果统计分析方法见附录 C。

5.6 指定值

指定值的选择与计算是能力验证计划成功的关键，实施机构需采用适当的指定值确定方法。定性计划确定指定值的方法通常有：专家判定、利用标准物质/标准样品的参考值、已知物品来源和利用参加者结果众数或中位值；定量计划确定指定值的方法通常有配方法、有证参考值、独家定值、专家公议值和参加者公议值。

通常指定值的不确定度不应超过 $0.3\sigma_{pt}$，对于校准项目，实施机构需给出指定值的计量溯源性和测量不确定度。

注：GB/T 27043 附录 B 和 CNAS-GL002 中给出指定值的确定方法有：已知值、有证参考值、参考值、专家公议值和参加者公议值。本指南 5.6 中的配方法和 GB/T 27043 附录 B 中的已知值是同一方法；独家定值方法包含利用高度匹配的有证标准物质校准获得指定值的参考值方法，和由一家实验室利用基准方法进行定值的方法。

指定值的确定见附录 D。

5.7 能力评定标准差

确定能力评定标准差通常有 5 种方法：规定值、经验值、一般模型、测量方法精密度试验和由参加者结果确定。

能力评定标准差的确定见附录 E。

5.8 能力评价

定性能力验证的评价方式，与其性质和要求有关，不同的专业和行业有不同的评价要求和评价形式。通常只需根据参加者的检测结果与指定值是否准确一致，即可给出合格或不合格、满意或不满意的能力评价。但在某些能力验证计划中，要求对参加者进行多方面的综合评价，参加者的报告需提供给多位专家，最后协商给出公议的结论，并赋予参加者整体的评价或评分。

定量能力验证结果通常需要转化为能力统计量，以便进行解释和与其他确定的目标做比较。其目的是依据能力评定标准来度量与指定值的偏离。常用的能力统计量有：偏差或百分相对差（D或$D\%$）、z 值、z'值、ζ 值和E_n值。这几种能力统计量的计算方法如下:

1）偏差（测量误差）或百分相对差

$$D_i = x_i - x_{pt} \tag{1}$$

$$D_i\% = \frac{\left(x_i - x_{pt}\right)}{x_{pt}} \times 100 \tag{2}$$

式中 x_i 为参加者 i 报告的结果，x_{pt} 为指定值。如果用 δ_E 表示测量结果的最大允许误差，当 $-\delta_E \le D \le \delta_E$，则表示满意结果（无行动信号），否则为不满意结果（给予行动信号）。当 $-\delta_E / x_{pt}\% \le D\% \le \delta_E / x_{pt}\%$ 时，表示结果满意（无行动信号），否则为不满意（给予行动信号）。

2） z 值

$$z_i = \frac{x_i - x_{pt}}{\sigma_{pt}} \tag{3}$$

z 值的评定标准为：

当 $|z| \le 2.0$ 表示满意结果；

当 $2.0 < |z| < 3.0$ 则表示有问题结果，给予警告信号；

当 $|z| \ge 3.0$ 时为不满意结果，给予行动信号。

参加者收到警告信号后，应检查测量程序有无问题。若参加者收到行动信号，则应采取纠正措施，全面检查相关的测量步骤。

3） z'值

$$z'_i = \frac{x_i - x_{pt}}{\sqrt{\sigma^2_{pt} + u^2(x_{pt})}} \tag{4}$$

当指定值不确定度 $u(x_{pt}) > 0.3\sigma_{pt}$ 时，能力评定标准差中需考虑加入指定值不确定度分量。z 值的评定标准适用于 z'值。

4) ζ 值

$$\zeta_i = (x_i - x_{pt}) \Big/ \sqrt{u^2(x_i) + u^2(x_{pt})} \qquad (5)$$

式中 $u(x_i)$ 是参加者结果 x_i 的标准不确定度，$u(x_{pt})$ 是指定值 x_{pt} 的标准不确定度。z 值的评定标准适用于 ζ 值。

5) E_n 值

$$(E_n)_i = \frac{x_i - x_{pt}}{\sqrt{U^2(x_i) + U^2(x_{pt})}} \qquad (6)$$

式中 x_{pt} 为参考实验室确定的指定值，$U(x_i)$ 为参加者结果 x_i 的扩展不确定度，$U(x_{pt})$ 为 x_{pt} 的扩展不确定度。当 $|E_n| \leq 1.0$ 时表示结果满意，$|E_n| > 1.0$ 时则表示结果不满意。在校准能力验证计划中，常使用 E_n 值来评定参加者结果。E_n 值也经常被用于测量审核的结果评价。

能力统计量的计算见附录 F。

5.9 能力验证报告

参加者需有途径获得能力验证报告，能力验证报告需全面，包含所有参加者结果的信息，并对参加者的结果做出评价。除非不适用或实施机构有正当理由，否则报告需包括以下内容：

1) 实施机构的名称和详细联系信息；

2) 联系人的姓名和详细联系信息；

3) 报告批准人的姓名、职位、签名或等效标识；

4) 实施机构分包活动的说明；

5) 报告发布日期和状态（如初期的、中期的或最终的报告）；

6) 报告的页码和清晰的结束标记；

7) 结果保密程度的声明；

8) 能力验证计划报告的编号和清晰标识；

9) 对能力验证物品的清晰描述，包括能力验证物品制备、均匀性和稳定性评定的必要细节；

10) 参加者的结果；

11) 统计数据及结果统计量，包括指定值、可接受结果的范围；

12) 用于确定指定值的方法；

13) 指定值的计量溯源性和测量不确定度的详细信息（校准能力验证计划需提供，某些检测能力验证计划也可说明）；

14) 用于确定能力评定标准差或其它评定准则的方法；

15）对应每组参加者使用的检测方法/程序的指定值和结果统计量（如果不同组的参加者使用了不同的方法）；

16）实施机构对参加者的能力评述；

17）能力验证计划设计和实施的信息；

18）数据统计分析的方法；

19）对统计分析解释的建议；

20）基于本轮能力验证结果的评述或建议。

注：对于连续能力验证计划，提供较简单的报告即可，上述很多内容在连续计划常规报告中可以省略，但需包含在参加者可获得的能力验证计划协议或阶段性的汇总报告。测量审核报告可适当简化，但至少需提供测量审核物品来源、稳定性、指定值和能力评定标准差的确定方式等必要信息。

5.10 测量审核的核查

参加者若选择测量审核，即“一对一”的能力验证计划，需核查测量审核样品来源、稳定性、指定值和能力评定标准差（用 En 值评定时没有能力评定标准差）的确定方式。测量审核样品通常来源于有证标准物质（标准样品）、能力验证剩余样品或者定制样品等。

当测量审核样品为有证标准物质（标准样品）时，参加者需核查样品是否在证书规定的有效期内，指定值是否利用证书提供的参考值，是否利用附录 F 中的 E_n 值或其他合理方式评定参加者的能力。

当测量审核样品为能力验证剩余样品时，参加者需核查实施机构是否确认样品稳定，能力评定方式是否合理。如样品被测特性不稳定，参加者可要求实施机构需重新定值或重新选用特性值稳定可靠的样品。能力评定标准差可采用经验值，或上一轮能力验证计划确定的能力评定标准差。可利用附录 F 中的 z 值评定参加者的能力。

当测量审核样品为定制样品时，参加者需核查实施机构是否采用附录 D 中的配方法或其他合理方式确定指定值，能力评定标准差可采用经验值或其他合理的方式确定，比如利用 Horwitz 公式或方法精密度确定。可利用附录 F 中的 z 值评定参加者的能力。

6 能力验证结果利用

能力验证参加者和利益相关方需正确利用能力验证结果，不应只关心能力验证结果是否满意，需从能力验证报告中获取尽可能多的有用信息，如关注不同测量方法间的差异，实施机构提供的技术分析和建议等。

不管是参加者还是利益相关方不应过度关注结果满意与否。一次能力验证结果并不不一定证明参加者的能力或技术水平的优劣。即使在一个运作良好、有经验丰富工作人员的实验室，偶尔也会得到异常数据；同样，即使一个已由精密度试验验证为有效的标准测量方法，也有可能存在缺陷，而这个缺陷可能只在多轮能力验证后才能显现；并且能力验证计划本身也可能存在缺陷。因此能力验证的结果不宜作为处罚实验室的依据。

如果参加者总体水平参差不齐，结果比较离散，使用参加者的结果确定的能力评定标准差较大，即使参加者的结果并不理想（离指定值的偏差较大），也可能被评定为满意。如果参加者总体水平较高，结果比较集中，使用参加者的结果确定的能力评定标准差较小，即使参加者的结果离指定值的偏差很小，也可能被评定为不满意。因此需关注在连续能力验证计划中参加者能力随时间的变化情况。

6.1 单次能力验证结果的利用

根据 CNAS-RL02《能力验证规则》的要求，当实验室在参加能力验证中结果为不满意且已不能符合认可项目依据的标准或规范所规定的判定要求时，应自行暂停在相应项目的证书/报告中使用 CNAS 认可标识，并按照实验室体系文件的规定采取相应的纠正措施，验证措施的有效性。在验证纠正措施有效后，实验室自行恢复使用认可标识。实验室的纠正措施和验证活动（可行时）应在 180 天（自能力验证最终报告发布之日起计）内完成。实验室应保存上述记录以备评审组检查。纠正措施有效性的验证方式包括：再次参加能力验证计划（包括测量审核）或通过 CNAS 评审组的现场评价。

当实验室能力验证结果为可疑或有问题时，应对相应项目进行风险评估，必要时，采取预防或纠正措施。

造成结果不满意或可疑（有问题）的原因，主要有管理和技术上的原因。管理原因如：抄写错误、贴错标识、小数点错误等。技术上的原因如：物品的储存或前处理不当、测量方法或内部质控有问题、标准物质/标准样品异常、设备状态不佳、环境条件不适宜或数据处理出现问题等。

当利用公议值确定能力评定标准差时，如果参加者总体水平高度一致，或由于串通而使得能力评定标准差较小，导致参加者结果不满意（行动信号）时，参加者可联系实施机构了解结果离群原因。如实验室参加能力验证的结果虽为不满意，但仍符合认可项目依据的标准或规范所规定的判定要求，或当实验室参加能力验证结果为可疑或有问题时，实验室应对相应项目进行风险评估，必要时，采取预防或纠正措施。

当参加者结果不满意或可疑（有问题）时，通常采取的纠正措施有：核查相关人员是否理解并遵循测量程序；核查测量程序的所有细节是否正确；核查设备校准和试剂的成分；更换可疑的设备或试剂；与另一个实验室进行人员、设备和/或试剂的比对测试。

6.2 连续能力验证结果的利用

在利用单次能力验证结果的基础上，实验室可参加连续能力验证计划监测其工作质量随时间的变化情况。参加的连续能力验证需是同一参数在相同或近似水平范围内的多轮次能力验证计划，可以识别出与随机误差、系统误差或人为错误等相关的潜在问题。

连续能力验证结果的利用可采用统计或图示的方法进行，其中将多轮次的能力评定结果制成控制图，以此来监测实验室工作质量随时间的变化，识别实验结果的趋势和其他特征。

附录 A 能力验证作业指导书示例

各参加实验室：

“鱼肉中恩诺沙星和氧氟沙星药物残留量的测定”能力验证计划是某实施机构组织开展的能力验证计划。本次能力验证计划中，贵实验室的代码为_____。

本次能力验证计划的要求和相关信息如下：

1 样品信息

1.1 样品状态

本次能力验证计划提供一份约 30g 待测样品，样品为鱼肉糜样（固体），以密封袋包装。样品编号为_____。

1.2 样品分发

（1）本次能力验证以邮寄方式发放样品。由某实施机构直接寄送；

（2）参加实验室收到样品后，当场进行外观检查并填写《被测物品接收状态确认表》（见表 2），并于收件日当天把该表传真或 E-mail 至组织本次计划的某实施机构，逾期不报者视为样品状态正常。

2 测试程序

2.1 测试项目

本次能力验证测试项目为：恩诺沙星（Enrofloxacin）和氧氟沙星（Ofloxacin）。样品中所含的 2 种测试参数浓度范围在 20-200 μg/kg 之间。本次能力验证计划不限定检测方法，参加实验室可采用实验室日常检测方法，所使用检测方法的定量限应优于 10μg/kg。

2.2 注意事项

虽经验证，样品可在室温下保存至少两周，但建议将未开封的样品置于-18℃下保存。开封后请将样品置于 4℃冷藏保存。请平行 2 次检测并提供 2 次检测结果及平均值。

3 结果反馈

请将测试结果填入《测定结果报告单》（见表 3），结果保留整数，单位为μg/kg。

请于测试后 3 个工作日内将《测定结果报告单》传真或 E-mail 至组织该计划的某实施机构，同时将纸质文件寄回某实施机构。

4 联系方式

联 系 人：_____ 电子邮箱：_____

联系电话：_____ 传 真：_____

地 址：_____ 邮 编：_____

表 2 被测物品接收状态确认表

编号：

能力验证计划名称	鱼肉中恩诺沙星和氧氟沙星药物残留量的测定		
组 织 机 构			
发 送 机 构			
电 话 / 传 真		联 系 人	
发 送 日 期		运输单据号码	
发 送 状 态	完好 □ 不完好 □	发送人签名	
接收实验室名称： 联系地址：　　邮编： 联系电话/传真：　　联系人： 接收时间：　　接收人签名：			
接收时，被测物品状态是否良好：　　是 □　　否 □ 如需要，对接收状态的详细说明：			

表 3 测定结果报告单

实验室代码：

样品编号	测试项目	测试结果（μg/kg）			备注
		1	2	平均值	
	恩诺沙星				
	氧氟沙星				

（测试结果保留整数）

检测人（签名）：

测试日期：

实验室名称（盖章）：

附录 B 均匀性和稳定性检验

B.1 总则

对于制备批量样品的检测能力验证计划，需确保能力验证样品充分均匀和稳定。

采用的均匀性和稳定性评价标准，需确保样品的不均匀和不稳定不会对能力评定产生不良影响，可利用B.2和B.3中的均匀性和稳定性检验进行确认；也可用先前轮次使用相似的能力验证样品获得的经验，并在当前轮次中进行必要的验证。对于长期或多轮次能力验证计划，可能会根据积累的经验减少均匀性和稳定性检验。

仅在以下情况下才可利用先前轮次获得的经验：

1）制备能力验证样品的程序不会发生影响样品均匀性的变更；

2）制备能力验证样品的材料不会有影响样品均匀性的变化；

3）可通过均匀性检验或参加者回报结果确认均匀性；

4）可定期核查材料的均匀性，核查时考虑该材料的预期用途，以确保该制备程序下获得的均匀性仍能满足预期用途。

示例：先前轮次使用的能力验证样品，经证明具有充分的均匀性和稳定性。如果同一批参加者在当前轮次中的实验室间标准差不大于前几轮次，即表明当前轮次使用的能力验证物品具有充分的均匀性和稳定性。

定性物品的均匀性和稳定性检验取决于物品的性质特点和测量要求。有些物品的均匀性和稳定性可通过定性观察确定，有的则需用定量或半定量的测量方式确定。例如在动物、植物检验的病毒检测能力验证中，检验制备物品的均匀性和稳定性时，可从制备的物品中随机抽取一定数量的样品，然后用聚合酶链式反应（PCR）或酶联免疫吸附实验（ELISA）法，对抽取的样品进行定量或半定量测试，再根据测试的结果判定物品是否均匀或稳定。有时根据能力验证计划的设计要求，只需测试样品的Ct（cycle threshold, Ct值就是扩增曲线达到阈值时的循环圈数）值在临界值之上或之下，达到可以判别阳性或阴性即可。对于有害昆虫及物种鉴定的能力验证计划，物品的均匀性（或一致性）是指每种待检的昆虫或种子，包括干扰物种，都经专家鉴定，确保准确和一致性，即可证明样品均匀。

B.2 均匀性检验

通常从总体样品中随机抽取10个或10个以上样品进行均匀性检验，每个样品在重复条件下至少检测两次。可使用 $S_s \le 0.3\sigma_{pt}$ 准则（S_s 为样品间标准差，σ_{pt} 为能力评定标准差）和 F 检验进行评价。

当开展一项新的能力验证计划时，如果σ_{pt}未知，可用F检验初步评价样品的均匀性。通常在参加者结果回报后获得σ_{pt}时，需再次利用$S_s \le 0.3\sigma_{pt}$确认样品是否均匀。如σ_{pt}已知（比如由先前轮次获得），可直接利用$S_s \le 0.3\sigma_{pt}$进行评价。

B.2.1 F 检验

为检验样品的均匀性，抽取i个样品(i=1、2、……m)，每个样在重复条件下检测j次（j=1、2、……n）。

每个样品的检测平均值

$$\overline{x_i} = \sum_{j=1}^{n} x_{ij} / n_i \tag{B.1}$$

全部样品检测的总平均值

$$\overline{\overline{x}} = \sum_{i=1}^{m} \overline{x_i} / \ n \tag{B.2}$$

检测总次数

$$N = \sum_{i=1}^{m} n_i \tag{B.3}$$

样品间平方和

$$SS_1 = \sum_{i=1}^{m} n_i (\overline{x_i} - \overline{\overline{x}})^2 \tag{B.4}$$

样品间均方

$$MS_1 = \frac{SS_1}{f_1} \tag{B.5}$$

样品内平方和

$$SS_2 = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (x_{ij} - \overline{x_i})^2 \tag{B.6}$$

样品内均方

$$MS_2 = \frac{SS_2}{f_2} \tag{B.7}$$

自由度

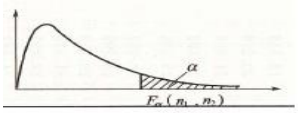
$$f_1 = m - 1 \qquad f_2 = N - n \tag{B.8}$$

统计量

$$F = \frac{MS_1}{MS_2} \tag{B.9}$$

若F＜自由度（f_1，f_2）及给定显著性水平α（通常α=0.05）的临界值$F_\alpha(f_1,f_2)$(查表4 F 分布表)，则表明样品内样品间无显著差异，样品是均匀的。

表4 F 分布表



$P\{F(f_1,f_2)>F_\alpha(f_1,f_2)\}=\alpha$ $\alpha=0.05$

f_2 \ f_1	8	9	10	11	12	13	14	15	16	17	18	19	20
9	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01	2.99	2.97	2.96	2.95	2.94
10	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85	2.83	2.81	2.80	2.79	2.77
11	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72	2.70	2.69	2.67	2.66	2.65
12	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62	2.60	2.58	2.57	2.56	2.54
13	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53	2.51	2.50	2.48	2.47	2.46
14	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46	2.44	2.43	2.41	2.40	2.39
15	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40	2.38	2.37	2.35	2.34	2.33
16	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35	2.33	2.32	2.30	2.29	2.28
17	2.55	2.49	2.45	2.41	2.38	2.35	2.33	2.31	2.29	2.27	2.26	2.24	2.23
18	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27	2.25	2.23	2.22	2.20	2.19
19	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.21	2.20	2.18	2.17	2.16
20	2.45	2.39	2.35	2.31	2.28	2.25	2.23	2.20	2.18	2.17	2.15	2.14	2.12

（摘自 GB 4086.4-1983 统计分布数值表 F 分布）[2]

注1：重复性条件是指在同一实验室，由同一操作员使用相同的设备，按相同的测试方法，在短时间内对同一被测对象相互独立进行的测试条件[1]。

注 2：F 检验是两个正态总体方差一致性检验，它是样品间方差与样品内方差的比较。为了确保F 检验的结果符合能力验证的要求，应在对参加者结果进行统计处理取得能力评定标准差σ_{pt}后，再将方差分析的所得的S_s（见公式 B.10），按$S_s \le 0.3\sigma_{pt}$准则进行检验。测量方法的重复性标准差与能力评定标准差之比，通常应小于 0.5，即S_r<0.5σ_{pt}。

注3：根据均匀性F 检验的统计假设，样品间的均方应大于或等于样品内的均方，即$MS_1 \ge MS_2$,所以计算的F 值应≥1。但在不均匀性很小时，由于测量数据的随机波动，F 值有可能出现＜1的现象。若F ＜1,但仍很接近1时，可将样品间标准差视为0。倘若F 值远小于1，这可能是由某种不正常的因素造成，如测量方法的精密度、样品内部不均匀、或样品的制备和测量不符合重复性条件等。这时不能以简单的F ＜$F_\alpha(f_1,f_2)$为依据判定样品是均匀的，而应查找问题的原因，并采取相应的措施。

B.2.2 $S_s \le 0.3\sigma_{pt}$ 准则

从能力验证计划制备的样品中随机抽取i个样品(i=1、2、……m)，每个样品在重复性条件下检测j次（j=1、2、……n）。按上述B.2.1中计算MS_1、MS_2。

若每个样品的重复检测次数均为n次。按下式计算样品之间的不均匀性标准偏差S_s：

$$S_s = \sqrt{(MS_1 - MS_2)/n} \qquad \text{(B.10)}$$

式中：MS_1—样品间均方；

MS_2—样品内均方；

n— 测量次数。

若$S_s \le 0.3\sigma_{pt}$，则使用的样品可以认为是均匀的。

B.3 稳定性检验

通常从能力验证样品总体中随机抽取足够具有代表性的样品进行稳定性检验，每个样品在重复条件下至少检测两次。稳定性检验的统计方法通常有$|\bar{x}-\bar{y}| \le 0.3\sigma_{pt}$准则、$t$检验法等。

B.3.1 $|\bar{x}-\bar{y}| \le 0.3\sigma_{pt}$ 准则

若$|\bar{x}-\bar{y}| \le 0.3\sigma_{pt}$成立，则认为样品是稳定的。

式中：$\bar{x}$—均匀性检验的总平均值

$\bar{y}$—稳定性检验的检测平均值。

注：取样数≥2，每次单独取样。每个样品重复测试3次以上，检测方法与均匀性检验相同。

B.3.2 t检验

由于许多能力验证计划的能力评定标准差来自参加者结果的公议值，因此在物品制备后无法立即应用$|\bar{x}-\bar{y}| \le 0.3\sigma_{pt}$准则来判定物品的稳定性。这时可采用$t$检验法检验物品的稳定性。

B.3.2.1 二个平均值之间的一致性

当将各时段的测量结果与第一次测量的结果进行比较（通常是均匀性检验结果），可按下式计算 t 值：

$$t=\frac{\left|\overline{x_2}-\overline{x_1}\right|}{\sqrt{\frac{(n_1-1)s_1^{\ 2}+(n_2-1)s_2^{\ 2}}{n_1+n_2-2}\times\frac{n_1+n_2}{n_1\times n_2}}} \tag{B.11}$$

式中：$\overline{x_1}$—第一次检测数据的平均值；

$\overline{x_2}$—第二次检测数据的平均值；

s_1—第一次检测数据的标准偏差；

s_2—第二次检测数据的标准偏差；

n_1—第一次检测总次数；n_2—第二次检测总次数。

若t＜显著性水平α（通常α=0.05）自由度为n_1+n_2-2的临界值$t_{1-\frac{\alpha}{2}(n_1+n_2)}$（查表5 t分布表），则两个平均值之间无显著差异，表明样品是稳定的。

注1：为了保证平均值和标准偏差的准确度，n_1和n_2均≥6。

注2：t检验是基于两个样品平均值差值的显著性检验，而$|\bar{x}-\bar{y}|\le$ σ_{pt}准则是以样品间的差异不影响能力评定为依据，二者的评价方法不同。所以为了确保能力验证样品满足稳定性的要求，应在对参加者结果进行统计处理取得能力评定标准差σ_{pt}后，再用$|\bar{x}-\bar{y}|\le 0.3\sigma_{pt}$准则进行核验。

注3：不稳定样品与不均匀样品的处理相似。首先应调查出现不稳定的原因，必要时在能力评定中对不稳定性产生的影响进行适当修正。若不稳定性是由系统误差造成，则应考虑修正指定值；若不稳定性是由随机误差造成，则考虑修正指定值的不确定度或能力评定标准差。

B.3.2.2 一系列测量的平均值与标准值/参考值的比较

对于已知指定值的能力验证物品，如标准物质/标准样品，或先前能力验证计划留存的物品，为了检验物品在存储或运输条件下是否保持稳定，可按下式计算t值：

$$t=\frac{\left|\bar{x}-\mu\right|\sqrt{n}}{S} \tag{B.12}$$

式中：$\bar{x}$ — n次测量的平均值；

μ — 标准值/参考值；

n — 测量次数；

S — n次测量结果的标准偏差。

注：为了保证平均值和标准偏差的准确度，n≥6。

若t＜显著性水平α（通常α=0.05）自由度为n-1 的临界值$t_{\alpha(n-1)}$（查表 5 t 分布表），则平均值与标准值/参考值之间无显著性差异，表明样品稳定。

表5 t 分布表

自由度	10	11	12	13	14	15	16	17	18	19
临界值	2.2281	2.2010	2.1788	2.1604	2.1448	2.1315	2.1199	2.1098	2.1009	2.0930
自由度	20	21	22	23	24	25	26	27	28	29
临界值	2.0860	2.0796	2.0739	2.0687	2.0639	2.0595	2.0555	2.0518	2.0484	2.0452
自由度	30	31	32	33	34	35	36	37	38	39
临界值	2.0423	2.0395	2.0369	2.0345	2.0322	2.0301	2.0281	2.0262	2.0244	2.0227

（摘自 GB 4086.3-1983 统计分布数值表 t 分布）[3]

附录 C 参加者结果统计分析方法

能力验证的统计方法需考虑数据特性（定量、定性、解释等）、统计假设和误差性质，以及预期参加者结果数量。同时统计设计应考虑参加者结果的评价方式。

基于不同的参加者结果评价方式，统计设计考虑也有所不同，常见如下：

1）若将参加者结果与预先确定的参考值和预先确定的限值（如最大允许误差，或法规规定值）进行比较，统计设计需考虑有途径能够获得参考值和限值，同时考虑能力评定方法。

2）若将参加者结果与公议值确定的指定值的差值和预先确定的限值进行比较，统计设计需要考虑如何利用公议值确定指定值，以及确定限值和能力评定方法。

3）若将参加者结果与公议值确定的指定值的差值和能力评定标准差比较，统计设计需考虑指定值和能力评定标准差的合理性，同时考虑能力评定方法。

4）若将参加者结果与指定值进行比较，同时考虑参加者的测量不确定度，统计设计时需考虑如何获得指定值及其不确定度，同时还需考虑将如何参加者的测量不确定度用于结果评定。

5）若比较不同测量方法的差异，统计设计需考虑相关的统计量及计算方法。

利用参加者结果计算指定值和能力评定标准差的统计方法，有经典统计方法和稳健统计方法。为减少离群值的影响，通常优先考虑采用稳健统计方法。

C.1 经典统计方法（classical statistical method）

当参加者结果呈正态分布时，可采用经典统计分析方法，计算样本平均值（$\bar{x}$）作为指定值（x_{pt}），计算样本标准差（s）作为能力评定标准差（σ_{pt}）。

假设有 p 个参加者参加测量得到 p 个测量数据，表示为 $x_1, x_2, \cdots x_i, \cdots, x_p$，可按下式计算样本均值（$\bar{x}$）和样本标准差（$s$）

$$\bar{x} = \sum_{i=1}^{p} x_i \,/\, p \qquad (C.1)$$

$$s = \sqrt{\sum_{i=1}^{p} (x_i - \bar{x}) \Big/ (p-1)} \qquad (C.2)$$

样本均值和标准差是总体均值和标准差最小无偏估计量，因此经典统计方法具有最高的统计效率（见表9）。但经典统计方法对离群值敏感，数据总体中即使只有一个离群值也会对经典统计方法产生很大影响。

当使用经典统计方法时，可用高置信水平的离群值检验剔除离群值后，计算平均值和标准差。若对离群值进行剔除，实施机构需：

1）记录剔除所用的检验方法及置信水平；

2）如采用连续离群值检验，需设定剔除数据的比例；

3）确认所产生的指定值及能力评定标准差满足能力验证计划的目的。

离群值剔除是数据处理程序的一部分，即使是离群剔除的结果，也要对剔除结果进行能力评定。

注1： GB/T 6379.2[4]中7.3.4给出了用格拉布斯（Grubbs）检验识别离群值的方法，此方法利用所有参加者结果的标准差（包括潜在离群值）进行检验。

注 2：当能力验证计划中使用相同的能力验证样品，并要求提交重复测量结果时，通常会针对重复性测量结果的离群值采用科克伦（Cochran）检验，详见 GB/T 6379.2 中 7.3.3 部分。

注 3：亦可利用稳健方法识别离群值，例如，若已计算出稳健平均值和标准差，则参加者结果与稳健平均值之间的差值超过 3 倍以上稳健标准差的结果可以视为离群值。

C.2 稳健统计方法 (robust statistical method)

由于经典统计方法对离群值敏感，因此通常优先采用对离群值相对不敏感的稳健统计方法。

注：采用剔除离群值后计算标准差的经典统计方法，通常会低估近似正态分布数据的离散性。因此，通常采用稳健统计方法，给出离散性的无偏估计。

中位值、尺度化中位绝对差（$MADe$）和标准化四分位距（$nIQR$）均是简易稳健统计量。算法A通过迭代方法转化原始数据，为近似正态分布提供均值和标准偏差的替代计算方法，这种方法在预期离群值比例低于20%的情况下非常有用。

C.2.1 对总体平均值和标准偏差的简单估计方法

C.2.1.1 中位值(median)

中位值是对称分布总体平均值的一种简单估计，该方法对离群值不敏感，可用 $med(x)$ 表示中位值。假设参加者提交的 p 个数据按递增顺序表示为：$x_1, x_2, \cdots x_i, \cdots, x_p$。当 p 为奇数时，中位值为第 $(p+1)/2$ 位的数值；当 p 为偶数时，中位值为第 $p/2$ 位和第 $(1+p)/2$ 位数值的平均值。可用下式表示：

$$med(x)=\begin{cases} x_{\{(p+1)/2\}} & p \text{ 为奇数} \\ \dfrac{\left[x_{\{p/2\}}+x_{\{1+p/2\}}\right]}{2} & p \text{ 为偶数} \end{cases} \qquad (C.3)$$

C.2.1.2 尺度化中位绝对差（scaled median absolute deviation, $MADe$）

尺度化中位绝对差（$MADe$）是正态分布数据的总体标准偏差的估计值。$MADe$ 计算方法对较高比例（50%）的离群值不敏感。假设参加者提交的 p 个数据按递增顺序排列表示为：$x_1, x_2, \cdots x_i, \cdots, x_p$。可先计算 p 个数据的中位值，然后计算 p 个数据中每个数据与中位值的绝对差 d_i（$i=1$ 到 p），再计算绝对差的中位值，将得到的中位值乘以 1.483 即可得到尺度化中位绝对差值。计算公式如下：

1)按下式计算绝对差 d_i（$i=1$ 到 p）：

$$d_i = \left| x_i - med(x) \right| \qquad \text{(C.4)}$$

2）计算 $MADe(x)$ 为：

$$MADe(x) = 1.483 med(d): \qquad \text{(C.5)}$$

如果 50%或者更多的参加者的结果是一致，那么中位值 $med(x)$ 将为零，可能有必要使用标准差化四分位距（见 C.2.1.3）或具有更高效率的统计方法（见表 9）计算样本标准差。

C.2.1.3 标准化四分位距(normalized interquartile range, $nIQR$)

标准化四分位距法是一种类似于尺度化中位绝对差的稳健统计方法，该方法相对简单并使用广泛。可将参加者结果按递增顺序排列，计算第 75 百分位（或第三个四分位）和第 25 百分位（或第一个四分位）参加者结果的差值，然后乘以系数 0.7413 即可得到标准化四分位距。可按下式计算得到：

$$nIQR(x) = 0.7413[Q_3(x) - Q_1(x)] \qquad \text{(C.6)}$$

$Q_1(x)$ 表示 x_i 的第25百分位数 $(i = 1,2,...,p)$ ，

$Q_3(x)$ 表示 x_i 的第 75 百分位数 $(i = 1,2,...,p)$。

如果第 75 百分位数和第 25 百分位数相同，则 $nIQR$ 将为零，需在剔除离群值后计算标准偏差作为总体标准差估计值。

注 1：与尺度化中位绝对差方法相比，标准化四分位距法只需要对数据进行一次排序，但标准化四分位距的失效点（breakdown point）为 25%（见表 8），而尺度化中位绝对差法的失效点为 50%，因此尺度化中位绝对差方法比标准化四分位距法能忍受更高比例的离群值。

注 2：标准化四分位距法和尺度化中位绝对差法这两种统计方法，在参加者结果数 $p < 30$ 时会导致分散性计算值明显偏小,可能会影响参加者结果的能力评定。

C.2.1.4 算法 A（Algorithm A）

C.2.1.4.1 算法 A 迭代程序

应用此算法计算可得到总体平均值和标准差的稳健值。

假设参加者提交的 p 个数据按递增顺序排列表示为：$x_1, x_2, \cdots x_i, \cdots, x_p$。这些数据的稳健平均值和稳健标准差记为 x^* 和 s^*。先计算 p 个数据的中位值作为初始稳健平均值（x^*），计算其绝对中位差作为初始稳健标准差（s^*）。计算公式如下：

计算 x^* 和 s^* 的初始值如下（med 表示中位值）：

$$x^* = \mathrm{med} x_i \qquad (i = 1,2,\cdots, p) \tag{C.7}$$

$$s^* = 1.483 \times \mathrm{med}\left|x_i - x^*\right| \qquad (i = 1,2,\cdots, p) \tag{C.8}$$

根据以下步骤更新 x^* 和 s^* 的值。计算：

$$\delta = 1.5 s^* \tag{C.9}$$

对每个 $x_i (i = 1,2,\cdots, p)$，计算

$$x_i^* = \begin{cases} x^* - \delta, & 若 x_i < x^* - \delta \\ x^* + \delta, & 若 x_i > x^* + \delta \\ x_i, & 其他 \end{cases} \tag{C.10}$$

再由下式计算 x^* 和 s^* 的新的取值：

$$x^* = \sum x_i^* / p \tag{C.11}$$

$$s^* = 1.134 \sqrt{\sum \left(x_i^* - x^*\right)^2 / (p-1)} \tag{C.12}$$

其中求和符号对 i 求和。

稳健估计值 x^* 和 s^* 可由迭代计算得出，例如用新取值数据更新 x^* 和 s^*，直至过程收敛。当稳健平均值和稳健标准差的第三位有效数字在连续两次迭代中不再变化时，即可认为过程是收敛的。这是一种可用计算机编程实现的简单方法。

算法 A 统计方法对一定比例的离群值不敏感，其失效点约为 25%（见表 8）。当 $MADe(x)$ 为 0，数据总体中有极端离群值时，初始 s^* 可能显著降低对离群值的耐受力。当数据总体中离群值比例预期超过 20%，或当初始 s^* 受极端离群值的不利影响，可考虑采用下面方法：

1）当 $MADe$=0，用 $\mathrm{med}(\left|x_i - \bar{x}\right|)$ 代替 $MADe$，或使用其他估计值如使用标准差（剔除离群值）。

2）如能力评定不用稳健标准差，使用 $MADe$ 并在迭代过程中不更新 s^*。如能力评定使用稳健标准差，可使用具有更高效率的统计方法（见表 9）得出的估计值代替 s*并在迭代过程中不更新 s*。

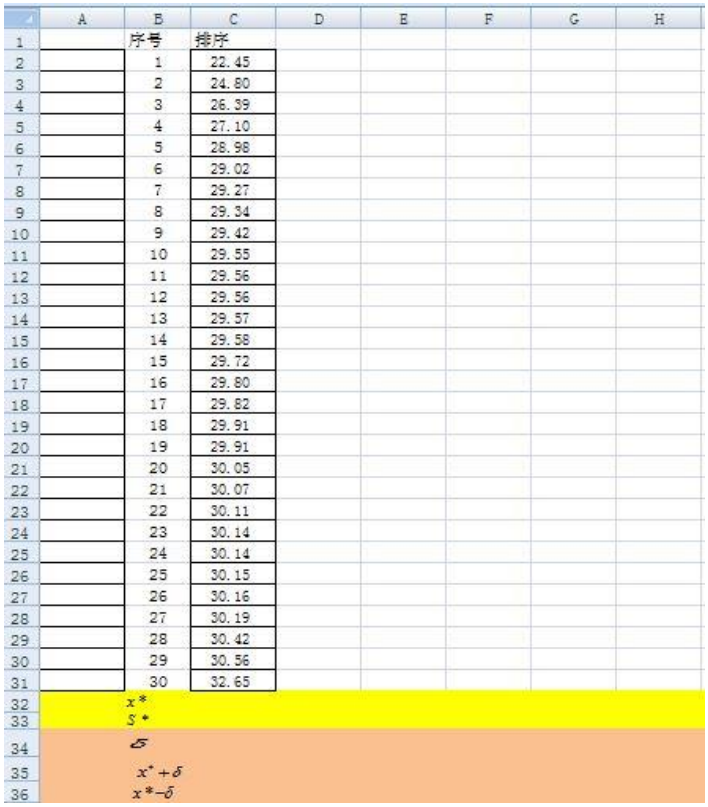
注：方法 2)将算法 A 的失效点提高到 50%，能处理较高比例的离群值。

C.2.1.4.2 算法 A EXCEL 实现过程

1）数据输入

假设某次能力验证计划有 30 个结果，在 B1 单元格输入“序号”，在 B2-B31 生成一个序列 1-30，在 C1 单元格输入“排序”，再将 30 个数据输入工作表的 C 列，将 C1-C31 选中，在“数据”菜单中选择“排序”，排序依据选择“以当前选定区域排序”，得到这组数据的非降序列。在 B32-B36 分别输入 x^*、S^*、δ、$x^*+\delta$、$x^*-\delta$（见表 6）。

表 6　数据表



	A	B	C	D	E	F	G	H
1		序号	排序					
2		1	22.45					
3		2	24.80					
4		3	26.39					
5		4	27.10					
6		5	28.98					
7		6	29.02					
8		7	29.27					
9		8	29.34					
10		9	29.42					
11		10	29.55					
12		11	29.56					
13		12	29.56					
14		13	29.57					
15		14	29.58					
16		15	29.72					
17		16	29.80					
18		17	29.82					
19		18	29.91					
20		19	29.91					
21		20	30.05					
22		21	30.07					
23		22	30.11					
24		23	30.14					
25		24	30.14					
26		25	30.15					
27		26	30.16					
28		27	30.19					
29		28	30.42					
30		29	30.56					
31		30	32.65					
32		x^*						
33		S^*						
34		δ						
35		$x^*+\delta$						
36		$x^*-\delta$						

2）参数计算

根据公式 C.7，在 C32 输入“=MEDIAN(C2:C31)”得到x^*。

根据公式 C.8，在 A2 输入“=ABS(C2-C$32)”，选中 A2 向下填充到 A31，在 A32 输入“=MEDIAN(A2:A31)”，得到$|x_i - x^*|$的中位值。在 C33 输入“=1.483*A32”得到s^*。

根据公式 C.9，在 C34 输入“=1.5*C33”，得到δ。在 C35 和 C36 分别输入“=C32+C34”和“=C32-C34”，得到$x^*+\delta$和$x^*-\delta$。

下面进行迭代计算。

根据公式 C.10，在 D1 输入“迭代 1”，在 D2 输入函数“=IF($C2<C$36,C$36,IF($C2>C$35,C$35,$C2))”，选中 D2 向下填充到 D31，可以得到x_i^*，$(i=1,2,\cdots,p)$。

根据公式 C.11，重新计算x^*，在 D32 输入“=AVERAGE(D2:D31)”得到新的x^*。

根据公式 C.12，重新计算s^*，在 D33 输入“=1.134*STDEV(D2:D31)” 得到新的s^*。

选中 C34、C35、C36 向右填充，在 D34、D35、D36 得到新的δ、$x^*+\delta$和$x^*-\delta$。

选中 D1~D36，向右填充至任意一列，即可得到多次迭代结果。（见表 7）

表 7　稳健均值和稳健标准偏差计算模板

	A	B	C	D	E	F	G	H	I	J	K
1		序号	排序	迭代1	迭代2	迭代3	迭代4	迭代5	迭代6	迭代7	迭代8
2	7.31	1	22.45	28.91469	28.83855	28.80171	28.7816	28.7703	28.76388	28.76022	28.75813
3	4.96	2	24.80	28.91469	28.83855	28.80171	28.7816	28.7703	28.76388	28.76022	28.75813
4	3.37	3	26.39	28.91469	28.83855	28.80171	28.7816	28.7703	28.76388	28.76022	28.75813
5	2.66	4	27.10	28.91469	28.83855	28.80171	28.7816	28.7703	28.76388	28.76022	28.75813
6	0.78	5	28.98	28.98	28.98	28.98	28.98	28.98	28.98	28.98	28.98
7	0.74	6	29.02	29.02	29.02	29.02	29.02	29.02	29.02	29.02	29.02
8	0.49	7	29.27	29.27	29.27	29.27	29.27	29.27	29.27	29.27	29.27
9	0.42	8	29.34	29.34	29.34	29.34	29.34	29.34	29.34	29.34	29.34
10	0.34	9	29.42	29.42	29.42	29.42	29.42	29.42	29.42	29.42	29.42
11	0.21	10	29.55	29.55	29.55	29.55	29.55	29.55	29.55	29.55	29.55
12	0.20	11	29.56	29.56	29.56	29.56	29.56	29.56	29.56	29.56	29.56
13	0.20	12	29.56	29.56	29.56	29.56	29.56	29.56	29.56	29.56	29.56
14	0.19	13	29.57	29.57	29.57	29.57	29.57	29.57	29.57	29.57	29.57
15	0.18	14	29.58	29.58	29.58	29.58	29.58	29.58	29.58	29.58	29.58
16	0.04	15	29.72	29.72	29.72	29.72	29.72	29.72	29.72	29.72	29.72
17	0.04	16	29.80	29.8	29.8	29.8	29.8	29.8	29.8	29.8	29.8
18	0.06	17	29.82	29.82	29.82	29.82	29.82	29.82	29.82	29.82	29.82
19	0.15	18	29.91	29.91	29.91	29.91	29.91	29.91	29.91	29.91	29.91
20	0.15	19	29.91	29.91	29.91	29.91	29.91	29.91	29.91	29.91	29.91
21	0.29	20	30.05	30.05	30.05	30.05	30.05	30.05	30.05	30.05	30.05
22	0.31	21	30.07	30.07	30.07	30.07	30.07	30.07	30.07	30.07	30.07
23	0.35	22	30.11	30.11	30.11	30.11	30.11	30.11	30.11	30.11	30.11
24	0.38	23	30.14	30.14	30.14	30.14	30.14	30.14	30.14	30.14	30.14
25	0.38	24	30.14	30.14	30.14	30.14	30.14	30.14	30.14	30.14	30.14
26	0.39	25	30.15	30.15	30.15	30.15	30.15	30.15	30.15	30.15	30.15
27	0.40	26	30.16	30.16	30.16	30.16	30.16	30.16	30.16	30.16	30.16
28	0.43	27	30.19	30.19	30.19	30.19	30.19	30.19	30.19	30.19	30.19
29	0.66	28	30.42	30.42	30.42	30.42	30.42	30.42	30.42	30.42	30.42
30	0.80	29	30.56	30.56	30.56	30.56	30.56	30.56	30.56	30.56	30.56
31	2.89	30	32.65	30.60531	30.57906	30.59384	30.60511	30.61181	30.61565	30.61786	30.61912
32	0.38	x^*	29.76	29.7088	29.69777	29.69336	29.69105	29.68977	29.68904	29.68862	29.68839
33		S^*	0.56354	0.58017	0.597375	0.607836	0.613837	0.617259	0.619213	0.620329	0.620967
34		δ	0.84531	0.870255	0.896063	0.911754	0.920755	0.925889	0.928819	0.930493	0.931451
35		$x^*+\delta$	30.61	30.58	30.59	30.61	30.61	30.62	30.62	30.62	30.62
36		$x^*-\delta$	28.91	28.84	28.80	28.78	28.77	28.76	28.76	28.76	28.76

3）模板应用

一组能力验证数据若有 n 个结果，例如 50 个，可以在模板（表 7）中，在 31 和 32 行中间插入 20 个空行。将数据重新输入到 C 列，然后选定输入的数据，按选定区域排序，将其他列的空白单元格分别向下填充，即可自动得到迭代结果。

C.3 效率与失效点（efficiency and breakdown point）

稳健统计方法稳健性的高低，与该方法对离群值的敏感性和稳健统计量本身的离散性质有关，有时也与对微小众数的敏感性（不敏感即耐受）这一特性有关。稳健统计方法可用三个特性量来描述，即失效点、效率和微小众数耐受力(resistance to minor modes)。

失效点— 数据组中允许离群值的最大比例，在该比例之下，估计量不受影响，反之将导致估计量失效。

效率— 某统计估计量的方差除以相应最小方差估计量的方差，其结果倒数即为该统计估计量的效率。

微小众数耐受力— 统计方法耐受小众偏离数据（通常小于样本数据量的20%）的能力。

C.3.1 失效点

失效点是数据组中允许离群值的最大比例，是一个统计估计量耐受离群值的度量，高失效点意味着耐受离群值的能力越强。表8给出了附录C中所涉及的不同统计方法的失效点和对微小众数的耐受力。

表 8 均值和标准偏差估计值的失效点

统计方法	总体统计参数	失效点	对微小众数的耐受力
样本均值	均值	0%	差
样本标准偏差	标准偏差	0%	差
样本中位值	均值	50%	好
$nIQR$	标准偏差	25%	中等
$MADe$	标准偏差	50%	中等到好
算法A	均值和标准偏差	25%	中等
Qn和Q/ Hampel	均值和标准偏差	50%	中等 (用于超过6S^*的微小众数效果好)

注1：这里使用的失效点的定义是： 正态分布数据总体中的一部分数据被替换成“+无穷大”,但估计值不会向无穷大移动的数据占总体数据的比例。例如，数据集里50%以下的数据被“+无穷大”替换，中位值依然在未被替换的数据里面。

注2：样本均值和标准偏差可能会因为1个离群值而受影响。中位值、$MADe$和Q/Hampel方法的稳健方法能耐受较大比例的离群值。使用迭代标准偏差的算法A和$nIQR$的失效点为25%。当离群值比例过大（如大于20%）时，经典或稳健方法都可能得到不合理的指定值和能力评定标准差估计值，需谨慎使用。

C.3.2 相对效率（relative efficiency）

所有统计估计量都有取样方差，即估计值会随不同轮次能力验证计划而变化。即使所有参加者都是有能力的，参加者结果分布不会出现离群和子集，从一轮至另一轮，各轮统计量的估计值也会发生变化，这种变化可用方差表示。而每一种统计估计量具有不同的方差。稳健统计方法基于理论假设，对偏离分布中心较远的数据进行修正，以降低其对估计量的影响。当总体数据为正态分布时，稳健统计估计量比经典统计估计量（样本均值和标准偏差）具有更大的方差。表9给出了附录C中不同统计估计量的相对效率。

表 9 稳健统计估计量的相对效率

统计方法	均值,n=50	均值,n=500	SD,n=50	SD,n=500
样本均值和标准偏差	100%	100%	100%	100%
中位值和$nIQR$	66%	65%	38%	37%
中位值和$MADe$	66%	65%	37%	37%
算法 A	97%	97%	74%	73%
Qn和Q/Hampel	96%	96%	73%	81%

表9结果表明没有任何一种统计方法对所有情况都是完美的。对正态分布数据，总体均值和标准偏差虽是最优，但当有离群值时会失效。简易稳健统计如中

位值、$MADe$或$nIQR$对于正态分布数据相对而言不是最优，但当有离群值或样本量较少时仍然是有效。

C.4 适用于参加者较少的能力验证统计方法

C.4.1 概述

通常样本估计值与总体参数实际值之间的差异会随样本量减少而增大。如果能力验证计划利用参加者的公议值评价参加者的能力，参加者数量通常应大于30；当参加者数量小于20时，由于公议值确定的指定值的不确定度相对较大，不可忽略，导致能力评定不可靠[5]。例如，当样本数为30时，样本离散性的估计值与总体标准差的差异，在95%置信水平时高达25%以上，而随着样本数的减少，其差异更大。所以对于参加者较少的能力验证计划，一般不推荐利用参加者结果分散性的公议值来进行能力评定，理想情况下可使用独立于参加者的、有明确溯源途径的方法确定指定值，如用配方法或由有证参考值给出。能力评定标准差也最好基于外部标准，如专家判断或适用性目标（fitness for purpose）。如果使用预先确定的指定值和评定标准，即使仅有一个参加者，能力验证也可进行。此种类型的实验室间比对被称为“一对一”能力验证或测量审核，在很多情况下非常有用，如校准。

当不能使用独立于参加者的、有明确溯源途径的方法确定指定值时，指定值或/和能力评定标准差可能需要利用参加者公议值得到。如果参加者数量过少，能力评定可能变得不可靠，需要考虑设定能力评定中参加者的最少数量。

C.4.2 识别离群值的方法

当参加者数据中有离群值时，宜优先采用稳健统计，但对于数据量较少的情况通常并不推荐使用。参加者较少（如少于18家）的能力验证计划，宜优先使用经典统计方法，可使用离群值检验，先剔除离群值，然后计算均值和标准偏差。

不同的数据量可使用不同的离群值检验方法。可用Grubbs检验对一个离群值和对同一方向的两个离群值同时进行剔除，Grubbs检验及其他检验需要预先规定可能的离群值个数，但可能由于有多个离群值而失效，这对于参加者结果数$p>10$时非常有用（取决于离群值的可能分布）。

注1：对分散性的估计在剔除离群值后需特别谨慎，因为估计值会偏低。当基于99%或以上置信水平进行剔除时，偏倚通常不严重。

注2：大多数指定值和分散性的单因子稳健估计量在$p\geq 12$可接受。

C.4.3 指定值计算方法

从少数参加者结果中获得的指定值，其不确定度$u(x_{pt})$需满足$u(x_{pt})<0.3\sigma_{pt}$。如果采用平均值作为指定值，当剔除离群值后，对于p $(p\leq 12)$家参加者结果，

用参加者结果标准偏差作为能力评定标准差，$u(x_{pt}) < 0.3\sigma_{pt}$ 准则不能满足。当使用中位值作为指定值（效率为64%），参加者结果数 $p \leq 18$ 时该准则不能满足。

其他稳健统计方法，如算法A有中等程度效率，当 $p > 12$ 时有可能满足 $u(x_{pt}) < 0.3\sigma_{pt}$ 准则。虽然有时可对更少参加者结果进行能力评定，某些指定值计算方法仍然要求有一定数量的参加者结果。如果参加者结果较少，很难有高度稳健的统计方法用于计算平均值。典型的数据量下限为 $p \geq 15$，中位值最低可应用于 $p = 2$（等同于均值），但当 $3 \leq p \leq 5$ 时，中位值没有均值好，除非有高风险离群值。

C.4.4 分散性计算方法

对于小数据量，由于分散性估计量的高变异性，不宜使用公议值作为能力评定标准差。

当分散性估计量用于其他目的（如作为指定值的分散性估计），或当能力验证计划可以承受较高变异的分散性估计量，处理小数据量宜选择可获得的最高效率的分散性计算方法。

注 1：“可获得”可理解为考虑合适的软件和专业知识的可获得性。

注 2： ISO13528 附录 C 给出的标准偏差的计算方法 Qn 比 $MADe$ 或 $nIQR$ 方法具有更高效率。

注 3：当数据量非常少时，对分散性进行计算有如下建议：

— $p = 2$，使用 $|x_1 - x_2|/\sqrt{2}$

— $p = 3$，如指定值和分散性未知，可使用 $MADe$ 以避免对标准偏差的过高估计，或使用平均值绝对偏差以避免对标准偏差的不当低估，如修约会导致两个相同的值。

— $p \geq 4$，推荐使用基于对数加权函数的标准偏差的 M—估计值[6]，近似于不用迭代计算指定值的算法 A，仅使用中位值作为指定值。

附录 D 指定值的确定

定性计划确定指定值的方法通常有：专家判定、利用标准物质/标准样品的参考值、已知物品来源和利用参加者结果众数或中位值。定量计划确定指定值的方法通常有：配方法、有证参考值、独家定值、专家公议值和参加者公议值等方法。

D.1 定性计划

D.1.1 专家判定

能力验证计划通常有一个专家组，它由若干具有适当资格的专家组成。指定值由专家组公议确定，如果专家组对某个物品的指定值不能达成公议值，则可考虑选用其他的定值方法，如果确实没有合适的其他方法，则该物品不宜用于能力验证。特殊情况下，也可由一个专家确定指定值。

D.1.2 利用标准物质/标准样品的参考值

当利用标准物质/标准样品（*CRM*）作为能力验证样品时，则以相应的参考值作为指定值。但*CRM*作为能力验证样品成本相对较高，而且*CRM*的信息是公开的，不利于能力验证实施过程中指定值的保密性。

D.1.3 已知物品来源

由已知来源的物质制备的能力验证物品，指定值可以依据物品来源确定。例如，鉴定昆虫或微生物（包括病毒）种类的能力验证计划，其样品制备可以用来自标准库或参考实验室经过鉴定的样品。在物品制备和处理过程中应尽力避免污染，以免影响参加者结果。

D.1.4 利用参加者结果众数或中位值

定性检测的指定值也可用参加者结果的统计量，如众数或中位值来确定。以众数作为指定值可用于类别和定序的结果，但中位值一般只适用于定序标尺的结果。如果利用众数或中位值作为指定值，在能力验证报告中应说明符合指定值的结果数及其所占的比例数。

D.2 定量计划

D.2.1 配方法（formulation）

将已知含量（或浓度）的被测物质或含有被测物的样品添加到空白基质中，得到的添加值作为指定值。当指定值由配方法来确定时，指定值的标准不确定度可根据 JJF 1059.1《测量不确定度评定与表示》，用不确定度分量来合成标准不确定度。

当采用配方法确定指定值时需注意：

1）基质需不受添加成分的影响，或者添加成分在基质中的比例是精确已知的；

2）必要时，所有成分需混合均匀；

3）所有显著误差来源是已被识别（例如，人们往往意识不到玻璃可以吸收汞化合物，实际上汞化合物水溶液的浓度会因容器材质而改变）；

4）添加成分和基质之间没有反应；

5）含有添加物的能力验证物品的性质需与实际测试样品类似。例如，人工配制样品与天然样品中相比，人工制品中添加成分通常更容易被提取。若人工配制样品与天然样品存在较大差异，实施机构需确保测试方法对人工配制样品适用。

D.2.2 有证参考值 （certified reference material value）

可直接用有证标准物质/标准样品（*CRM*）作为能力验证样品，它能提供一个独立于参加者测量结果的标准值，并能提供相应的溯源再现性。当使用*CRM*作为能力验证样品时，指定值及其不确定度由证书中给出。

这种方法的局限性在于：

1）为每个参加者都提供一份*CRM*会比较昂贵；

2）为确保长期稳定性，有时需对*CRM*进行进一步加工处理，这可能会对*CRM*的特性产生一定的影响。

3）参加者可能已经了解某种*CRM*，因此需隐藏*CRM*的识别标识。

D.2.3 独家定值 （results from one laboratory）

指定值x_{pt}可由一个实验室使用参考方法（比如基准方法）确定，该参考方法应易于理解并进行充分描述，包含完整的不确定度声明及计量溯源性，并适用于能力验证计划，该参考方法对参加者使用的所有测试方法需具有可替换性。利用参考方法确定的指定值是定值研究的平均值，使用多个能力验证物品或在不同测量条件下，进行多次重复测量得到，指定值的不确定度是当使用参考方法并在定值研究条件下的不确定度。

指定值x_{pt}也可由一个实验室采用合适的测量方法，通过使用高度匹配的*CRM*实施的校准获得。该方法假设*CRM*对于参加者所有测量方法具有互换性。这种定值方式需要利用同一测量方法在重复性条件下，在同一个实验室对能力验证物品和*CRM*实施一系列测试，选用同能力验证样品在基质、浓度和种类等方面具有相似性或可比性的*CRM*，在重复性条件下将*CRM*与能力验证样品一起，采用相同的分析方法，按照随机的顺序进行多次分析测量。对照*CRM*的参考值可以得到能力验证样品的指定值和标准不确定度。

$$x_{pt} = x_{CRM} + \bar{d} \tag{D.1}$$

$$u(x_{pt}) = \sqrt{u^2_{CRM} + u^2_{\bar{d}}} \tag{D.2}$$

式中：x_{pt}为能力验证样品的指定值；

$u(x_{pt})$为指定值的标准不确定度；

x_{CRM}为CRM的参考值；

u_{CRM}为CRM参考值的不确定度；

$\bar{d}$为能力验证样品测量结果（平均值）和CRM测量结果（平均值）之差的平均值，$u_{\bar{d}}$为其不确定度。

注：x_{CRM}与$\bar{d}$相互独立，除非定值实验室亦制备了该CRM。

如果在能力验证计划开始前样品已经有了参考值，且实施机构利用同一测量方法核查了该参考值，核查结果与参考值的差值应小于该差值的不确定度的两倍，即结果应具有计量兼容性（metrological compatible）。在此情况下，实施机构可选择带有适当不确定度的测量平均值作为指定值。如果结果不具有计量兼容性，则实施机构应查明原因，包括利用其他方法确定指定值及其不确定度，或放弃此轮能力验证计划。

D.2.4 专家公议值（consensus value from expert laboratories）

可采用标准物质/标准样品定值的方式由专家实验室进行实验室间比对确定指定值，定值方法详见GB/T 15000.3《标准样品 定值的一般原则和统计方法》。选取的专家实验室通常具有较高测量水平和较好的测试精度，能对测量条件进行严格控制。在分发能力验证物品之前，随机选取一部分能力验证物品，由一组专家实验室按规定的测试方案进行测试，该测试方案规定能力验证物品数和重复测试次数以及其它相关条件。每个专家实验室提供的结果需包含标准不确定度。

指定值也可由专家实验室报告结果的稳健平均值得到，具体计算可使用C.2.1.4部分中的算法A。也可使用其他计算方法代替算法A，只要该方法有可靠的统计学基础，并在能力验证报告中描述所使用方法即可。

如果专家实验室只报告了结果，且测试方案中并未要求专家实验室报告不确定度，或报告结果的不确定度不可靠，则通常将专家实验室分组然后计算组内参加者公议值，具体计算方法见D.2.5部分参加者结果公议值方法。如果专家实验室分别报告了一个以上的结果（例如重复测试），实施机构需建立具有统计意义的替代方法，以确定指定值及其不确定度，同时需考虑出现离群值或其它非预期结果的可能性。

如果专家实验室报告了带有不确定度的结果，利用专家公议值计算指定值相对比较复杂，目前有多种方法，包括：加权平均值、非加权平均值，对过度分散（over dispersion）结果进行处理的方法，以及对可能出现的异常结果或错误结果及不确定度评定进行处理和计算的方法。指定值的计算需：

1）核查所报告不确定度估计值的有效性，比如核查所报告的不确定度是否能与所观测到的结果的离散性一致；

2）使用适用于所报告不确定度的大小与可靠性的一种加权方法，具体包括：当所报告的不确定度相同，或不确定可靠性很差，或不确定未知的情况下实施同量加权；

3）考虑所报告的不确定度可能无法解释所观察到的分散性（过度分散），当过度分散时可增加额外分量；

4）考虑报告结果或其不确定度出现非预期离群值的可能性；

5）有可靠的理论基础；

6）有被证实能够满足能力验证计划目的。

D.2.5 参加者结果公议值(consensus value from participant results)

使用该方法得到的能力验证物品的指定值x_{pt}为中心位置（location）估计值（如稳健平均值、中位值或算术平均值），该估计值由某轮能力验证计划参加者报告的结果计算得出，具体见C.2稳健统计方法部分。

有时实施机构可能会使用通过预先设定的标准（如是否认可或先前的测量水平）来对参加者进行分组，这种情况可考虑使用参加者公议值。也可使用其他统计方法替代C.2中的稳健统计方法，只要该方法有可靠的统计学基础并在能力验证报告中说明即可。

使用参加者公议值的方法优点在于：

1）无需额外测试即可获得指定值；

2）该方法适用于标准化、程序化的被测量，因为通常无更可靠的方法获得等效结果。

该方法的局限性在于：

1）参加者的一致性可能不够充分；

2）公议值可能因参加者普遍使用了错误的方法而存在未知偏倚，而指定值标准不确定度不会包含该偏倚；

3）确定指定值的方法可能存在偏倚而使公议值发生偏倚。

4）可能难以确定公议值的计量溯源性。当结果溯源至各实验室时，只有在实施机构完全了解与公议值相关的所有参加者使用的校准标准，并掌握其他相关方法条件的信息时，才可以做出明确的溯源性声明。

指定值的标准不确定度取决于所使用的统计方法。当利用附录 C.2 中的稳健统计方法计算指定值时，指定值x_{pt}的标准不确定度可按下式计算：

$$u(x_{pt}) = 1.25 \times \frac{s^*}{\sqrt{p}} \qquad (D.3)$$

式中：$u(x_{pt})$为稳健平均值的标准不确定度；

s^*是结果的稳健标准差；

p为参加者数。

注 1：指定值和稳健标准差由参加者结果获得，可假定指定值的不确定度包含不均匀性、运输和不稳定性对不确定度的影响。

注 2：基于正态分布的大量数据的中位值的标准差，或是中位值作为平均值估计值的效率，确定校正系数为 1.25。较复杂的稳健统计方法的效率远大于中位值效率，因此校正系数小于 1.25，之所以选用该系数，是因为能力验证结果一般不严格服从正态分布，且包含未知比例的源于不同分布的结果（“污染结果”）。就可能存在的“污染结果”而言，1.25 是一个保守（高）估计值，实施机构可依据经验和所使用的稳健方法选择较小的校正系数或不同的公式。

D.2.6 对指定值不确定度的限定

当指定值的标准不确定度$u(x_{pt})$与能力评定标准差相比不可忽略时，会存在一种风险，即某些参加者将会因指定值的不准确而收到行动信号或警戒信号，而不是因为参加者本身的原因。因此需确定指定值的标准不确定度，并通知参加者。

当满足以下准则时，指定值的不确定度可忽略。

$$u(x_{pt}) < 0.3\sigma_{pt} \quad 或 \quad u(x_{pt}) < 0.1\delta_{\mathrm{E}} \tag{D.4}$$

式中$u(x_{pt})$为指定值的标准不确定度，σ_{pt} 为能力评定标准差，δ_{E}为最大允许测量误差。

注：$0.3\sigma_{pt} = 0.1\delta_{E}$，当$|z| \geq 3.0$，将产生行动信号。

当以上准则不满足时，实施机构需考虑采取以下措施：

1）选择另一种指定值确定方法，使其不确定度满足上述公式 D.4 所确定的准则。

2）在能力验证的结果解释中考虑不确定度（见z'值、ζ 值或E_n值的描述）。

3）如果指定值是由参加者结果得出，且较大不确定度是由可识别的各参加者子集（sub-population）间的差异所致，则分别报告各子集（如参加者使用不同测量方法）的指定值和不确定度。

4）通知参加者，指定值的不确定度不可忽略。

如果上述1)-4)均不适用，则需通知参加者，无法确定可靠指定值，也无法进行能力评定。

D.2.7 指定值和独立参考值的比较

当使用参加者公议值确定指定值(x_{pt})时，若同时可获得一个可靠、独立的估计值（表示为x_{ref}），如基于制备方法或基于参考值。需将公议值x_{pt}与x_{ref}比较。

如使用 D.2.1 至 D.2.4 中的方法确定指定值，每轮能力验证计划后，需将每轮计划得到的稳健平均值 x^* 与指定值比较。差值的计算公式为

$x_{diff}=(x_{ref}-x_{pt})$或（$x^*-x_{pt}$），差值的标准不确定度由下式估计：

$$u_{diff}=\sqrt{u^2(x_{ref})+u^2(x_{pt})} \tag{D.5}$$

式中 $u(x_{ref})$ 是参考值的不确定度，$u(x_{pt})$ 是指定值的不确定度。

如果该差值大于其标准不确定度的两倍，需查找原因，可能的原因如下：

1）参考测量方法存在偏倚；

2）参加者结果存在普遍偏倚；

3）使用配方法时未意识到测量方法的局限性；

4）使用独家定值或专家公议值时，专家实验室的结果存在偏倚；

5）独立参考值和指定值不能溯源到同一计量基准。

实施机构需依据查找的原因确定是否对结果进行评估，对于连续能力验证计划可考虑在后续能力验证计划中是否修正设计方案。若差值大到足以影响能力评定或表明参加者使用的测量方法间存在严重偏倚，需在能力验证计划的报告说明，如果出现这种情况，在后续设计能力验证计划时，需考虑该差值。

附录 E 能力评定标准差的确定

需根据能力验证的目的和相关规定（如管理部门的要求）确定合适的能力评定标准差（σ_{pt}）。确定能力评定标准差通常有五种方法：规定值、经验值、一般模型、测量方法精密度和由参加者结果确定。

E.1 由规定值确定

可根据管理部门、认可机构或实施机构认为参加者可实现的合理能力水平确定最大允许误差（δ_E）或能力评定标准差。将最大允许误差除以确定不满意结果（行动信号）时使用的σ_{pt}的倍数，即可获得能力评定标准差。同理，也可将σ_{pt}转化为最大允许误差（δ_E）。

如果为满足管理要求或适用性目标而给出标准差，则可直接将该标准差用于能力评定标准差σ_{pt}，如果该要求或目标给出的是最大允许误差（δ_E），则将δ_E除以行动限（3.0）可得到能力评定标准差（σ_{pt}）。所规定的最大允许误差（δ_E）可用于计算偏差或百分相对差（D或$D\%$）。

E.2 由经验值确定

能力评定标准差可设定为符合实验室能力水平的经验值，它是实施机构和实验室利益相关方希望实验室可达到的预期值。

可利用从先前轮次能力验证中获得的经验，确定能力评定标准差（σ_{pt}）和最大允许误差（δ_E）。当前轮次应与先前轮次具有相同被测量且特性值相当，同时参加者使用了相似的测量程序。在对适用性目标未达成一致意见时，该方法有用。该方法的优点在于：

1）依据合理能力预期开展能力评定；

2）评定标准不会因各轮次参加者的随机变化或参加者群体变化而改变；

3）当存在两个或多个实施机构时，评定标准将不会因实施机构的不同而变化。

评估先前开展的能力验证计划时，需考虑有能力的参加者可达到的水平。该水平不会受新参加者，或参加者数量较少或某特定轮次计划特有因素导致的随机变化的影响。可通过核查先前开展的能力验证计划确认其一致性，或依据平均值或适用于被测量值的回归模型确定。回归方程可能为直线，也可能为曲线。需考虑标准差和相对标准差，分析这两个统计量在被测量水平的适当范围内的适用性，选择适用性较好的一个，也可通过这种方式获得最大允许误差。

E.3 由一般模型确定

能力评定标准差可由测量方法再现性的一般模型得出，计算测量方法的再现性标准差作为能力评定标准差。该方法的优点在以实际经验为基础且对于被测量可保持客观和一致性。依据所使用的模型，该方法可视为能力评定标准差确定方法中使用适用性目标的特例。通过一般模型选择的预期标准差需具有合理性。如果大多数或极少数参加者为不满意或有疑问，实施机构需确保这种情况满足能力验证计划的目的。

对于一般模型方法，通常优先考虑测量特性问题。因此，在使用一般模型前，需先尝试使用专家经验、先前能力验证经验和测量方法精密度等方法确定能力评定标准差。

能力评定标准差可由检测方法再现性的一般模型得出。例如，Horwitz 公式给出了化学分析方法再现性标准差的一般模型，这个方法可得到以下再现性标准差的表达式：

$$\sigma_{pt}=\sigma_R=\begin{cases}0.22c & c<1.2\times10^{-7}\\ 0.02c^{0.8495} & 1.2\times10^{-7}\le c\le 0.138\\ 0.01c^{0.5} & c>0.138\end{cases} \quad \text{(E.1)}$$

式中c是以百分数表示的待测化学成分的浓度（质量分数）。

注：Horwitz 公式是经验公式，基于长时间内对多个参数的协作实验。当协作实验未出现明显问题时，σ_R值是实验室间变异的预期上限，在某些能力验证计划中可能不适合作为能力评定标准差。

E.4 由测量方法精密度确定

当能力验证计划中使用标准化的测量方法，且测量方法的重复性标准差（σ_r）和再现性标准差（σ_R）可获得时，能力评定标准差（σ_{pt}）计算如下：

$$\sigma_{pt}=\sqrt{\sigma_R^2-\sigma_r^2(1-1/m)} \quad \text{(E.2)}$$

式中，m是一轮能力验证计划中各参加者实施的重复测量次数。

E.5 由参加者结果确定

能力评定标准差σ_{pt}由同一轮能力验证参加者结果计算得出，使用该方法计算能力统计量时，适宜使用z值，通常使用稳健统计方法（见C.2稳健统计方法）计算所有参加者报告结果的稳健标准差（σ_{pt}）。

使用参加者结果得到的公议值可能不适宜作为能力评定标准差，原因如下：

1）如果稳健标准差极小，实施机构需给所使用的σ_{pt} 设定一个最低限，选定的最低限需满足在测量误差符合极端预期用途的情况下，$z<3.0$。

示例：在纺织行业，每厘米的线程数（number of threads）小于4线程/厘米的误差被认为不显著。在某些轮次的织物能力验证计划中稳健标准差很小，比如小于1线程/厘米。当能力验证稳健标准差不小于1.3线程/厘米时，实施机构将稳健标准差用作σ_{pt} 。而当稳健标准差小于1线程/厘米时，采用$\sigma_{pt}=1.3$，此时3σ_{pt} 为3.9接近4，

2）如果稳健标准差极大，实施机构需给使用的σ_{pt} 设定一个最高限，或对认为满意测量结果设定一个最高限。最高限为不满足适用性目标的结果将收到行动信号。如果对称区间包含不满足能力验证适用性目标的结果，实施机构可对被认为满意结果（无警戒信号或行动信号）的区间设定上限或下限。

示例：某非饮用水能力验证计划，法规规定结果须在参加者结果的稳健平均值的$3\sigma_{pt}$范围内，但由于有时满意结果的范围可能包括0μg/L，任何低于配制值10%的结果视为不满意结果。某能力验证物品中的限制物质配制含量为4.0μg/L，参加者稳健均值为3.2μg/L，σ_{pt} 为1.1μg/L，因此，如果参加者结果为0.0μg/L，仍在$3\sigma_{pt}$范围内；但由于低于配制值10%的结果视为不满意，所以任何低于0.4μg/L的结果将被认为是不满意结果。

该方法的主要优点在于简单和实用，可用于很多场合，且可能是唯一可行的方法。

该方法的缺点在于：

1）由于σ_{pt} 值可能在每轮计划都会有显著的变化，因而对于参加者而言，利用z值寻找多轮次计划中可能的趋势时会有一定的困难。

2）当能力验证计划中参加者数量较少或者参加者采用多种不同测量方法，标准差可能不可靠。例如：如果数$p=20$，正态分布的标准差与其真值的偏离在不同轮次能力验证中可能会有约±30%的变异。

3）将产生近似恒定比例的满意结果。通常较差的测量结果不能通过能力评分值体现，而优秀的参加者的较好测量结果却可能得到比较差的能力评分值。

4）由于该方法是基于统计假设，无法对结果应用的适用性提供有效说明，因此对结果分析解释需结合专业判断。

附录 F 能力统计量的计算

能力验证的结果通常需转化为能力统计量，以便于进行解释和与其他确定的目标进行比较。能力统计量通常有：偏差（D）、百分相对差（$D\%$）、z 值、z'值、ζ 值和E_n值等。

F.1 偏差（测量误差）

用x_i表示一轮能力验证计划中参加者i的测量结果（或重复测试结果的平均值）。则可用简单的方法通过计算结果x_i和指定值x_{pt}间的差值来评定参加者的结果如下：

$$D_i = x_i - x_{pt} \qquad (F.1)$$

如果指定值（x_{pt}）视作约定值或参考值，则D_i可视为报告结果的测量误差。D_i可与指定值的单位相同，也可以用以下公下式计算偏差百分比：

$$D_i\% = 100(x_i - x_{pt}) / x_{pt}\% \qquad (F.2)$$

D或者$D\%$通常与最大允许误差（δ_E）进行比较，δ_E可基于适用性目标或从前一轮能力验证计划获得的经验得出。如果$-\delta_E \le D \le \delta_E$，则表示结果满意，否则为不满意；当$-\delta_E / x_{pt}\% \le D\% \le \delta_E / x_{pt}\%$时，表示结果满意，否则为不满意。

δ_E与用于z值计算的能力评定标准差（σ_{pt}）紧密相关，δ_E和能力评定标准的关系由z值的评定标准确定，如果$z \ge 3.0$时为不满意，则$\delta_E = 3\sigma_{pt}$，或相当于$\sigma_{pt} = \delta_E / 3$。$\delta_E$常用于医学领域的能力验证以及测量方法和产品的性能规范。

用D和δ_E进行能力评定标准的优势在于：这些统计量与测量误差直接相关，且常用作确定适用性目标的标准，因而参加者能直观地理解这些统计量。

用百分相对差（$D\%$）的优势在于：$D\%$是对被测量水平的标准化度量，且与产生误差的常见原因相关（如校准错误或稀释偏倚），因而参加者能直观理解该统计量。不足之处在于：在很多国家或测量领域并不常用这些统计量；且在多个被分析物或不同水平的被测量拥有不同能力评定标准差的能力验证计划中，D没有标准化，不能实现简单判断以给出行动信号。

使用D和$D\%$时，通常假设参加者结果是对称分布，可接受范围为$-\delta_E \le D \le \delta_E$。

由于适用目标可能不同，为对不同被测量水平进行比较，或比较不同轮次和被测量，可将D和$D\%$转化为“允差百分比”（P_A）。计算公式如下：

$$P_{Ai} = (D_i / \delta_E) \times 100\% \qquad (F.3)$$

如果$P_A \ge 100\%$或$P_A \le -100\%$，则表示结果不满意，否则为不满意。

注 1：可比较不同被测量水平和不同轮次能力验证计划的P_A值，或在图形中标示P_A值。P_A值的使用和解释与z值相同（如$z \le -3.0$或$z \ge 3.0$时，产生行动信号）。

注 2：偏差（测量误差）的各种形式，常应用于能力验证频率高、被分析物量大的医学领域。

F.2 Z值

能力验证结果x_i的z值的计算公式如下：

$$z_i = \frac{(x_i - x_{pt})}{\sigma_{pt}} \qquad (F.4)$$

式中

x_{pt}为指定值，σ_{pt}表示能力评定标准差。

z值的解释如下：

— 当$|z| \le 2.0$，满意结果（无行动或警戒信号）。

— 当$2.0 < |z| < 3.0$，有疑问结果（警戒信号）。

— 当$|z| \ge 3.0$，不满意结果（行动信号）。

注 1：某些要求严格的应用中，$|z| > 2.0$视为不满意结果。

注 2： 能力评定标准差（σ_{pt}）的选择通常需满足上述z值解释，该解释广泛应用于能力评定，与控制图限值十分相似。

注 3：将z值的限值确定为 2.0 或 3.0 的原因如下：若正确开展的测量得到的结果（如有必要需对结果进行转化）服从正态分布，其均值为x_{pt}，标准差为σ_{pt}。那么z值也将呈正态分布，其均值为 0，标准差为 1.0。在此条件下，z值落在$-3.0 \le z \le 3.0$的区间之外的概率只有 0.3%，而在$-2.0 \le z \le 2.0$的区间之外的概率则有 5%。因为不满意发生概率很低，当没有真正的问题存在时很难出现不满意，所以若出现了不满意则认为有异常情况出现是合理的。

注 4：如果实际的实验室间差异小于σ_{pt}，则误判（misclassification）的概率会降低。

注 5：当能力评定标准差由专家经验或一般模型方法确定时，它可能与结果（稳健）标准差差异很大，此时结果落在±2.0 和±3.0 范围外的概率就不再是 5%和 0.3%。

需根据所报告z值的有效数字的位数来确定结果、指定值和能力评定标准差的有效数字。对于z值而言，在小数点后保留多于两位小数的意义不大。

当参加者数量很大且参加者的结果的标准差用作σ_{pt}时，建议实施机构检查参加者结果或z值分布的正态性。在极端情况下，当参加者数量很少时，可能不会给出行动信号。这种情况下，使用结合了多轮计划能力评定的图示法，将会提供比单轮计划结果更多的参加者能力的信息。

F.3 Z'值

如果考虑指定值x_{pt}的不确定度$u(x_{pt})$，比如当$u(x_{pt})>0.3\sigma_{pt}$时，需在能力统计量计算公式的分母增加不确定度分量。该统计量值称为z'值，计算公式如下：

$$Z_i' = \frac{x_i - x_{pt}}{\sqrt{\sigma^2_{pt} + u^2(x_{pt})}} \tag{F.5}$$

注：当利用参加者结果计算x_{pt}和/或σ_{pt}时，由于各参加者的结果可能会影响稳健均值和标准差，这时能力评定值与各个参加者的结果相关。能力评定值与各参加者结果的相关性取决于该参加者的结果在组合统计中的权重。因此，分母含指定值不确定度，不考虑相关性的能力评定值，低于分母考虑协方差时的能力评定值。例如，如果分母增加指定值不确定分量且$u(x_{pt})=0.3\sigma_{pt}$时，则z'值会减小约10%。因此，由参加者结果确定x_{pt}和/或σ_{pt}时，可用公式F.5来计算。

根据能力验证计划设计，z'值可与z值有相同的解释，并有相同的临界值（2.0和3.0）。

利用D和$D\%$评价能力时，若考虑指定值的不确定度，可按公式F.6进行修正。

$$\delta_E' = \sqrt{\delta_E^2 + U^2(x_{pt})} \tag{F.6}$$

式中，$U(x_{pt})$是利用包含因子k（k=2）计算得出的指定值x_{pt}的扩展不确定度。

当δ_E'为修正后最大允许测量误差，则$|D|<\delta_E'$被认为是满意的结果；用相对误差表示，即$|D\%|<\delta_E'/x_{pt}\%$为满意结果。

F.4 ζ 值

当核查参加者结果与指定值的差值是否在其所声称的不确定度范围内时，可使用ζ值。可用以下公式计算ζ值：

$$\zeta_i = \frac{x_i - x_{pt}}{\sqrt{u^2(x_i) + u^2(x_{pt})}} \tag{F.7}$$

式中

$u(x_i)$ 是参加者结果 x_i 的标准不确定度，

$u(x_{pt})$ 是指定值 x_{pt} 的标准不确定度。

注 1：当用参加者的公议值计算指定值（x_{pt}）时，x_{pt} 与各参加者的结果相关。指定值与各参加者的结果的相关性取决于该参加者的结果在指定值（甚至在指定值不确定度）中的权重。因此，分母含指定值不确定度，不考虑相关性的能力评定值，低于分母考虑协方差时的能力评定值。如果指定值的不确定度较小，则考虑相关性与不考虑相关性计算的 ζ 值差异不大；使用稳健统计方法时，对于结果与指定值偏离最大的参加者，该差异最小，因而可将公式 F.7 和公议值计算方法一起使用，无需因为相关性进行调整。

注 2：ζ 值与 E_n 值（见公式 F.8）的差别在于用标准不确定度 $u(x_i)$ 和 $u(x_{pt})$ 代替扩展不确定度 $U(x_i)$ 和 $U(x_{pt})$，当存在方法系统偏倚或参加者对测量不确度估计不可靠时，可能导致 ζ 值大于 2 或小于-2，因此，ζ 值表示对参加者提交的结果的严格评估。

使用ζ值时，无论实验室是否能报告准确的结果，可直接对其结果进行评估，即在测量不确定度范围之内参加者结果是否与 x_{pt} 一致(通常适用于校准实验室)。ζ值解释可采用与 z 值相同的临界值（2.0和3.0），或乘以估计扩展不确定度时使用的包含因子，但是超出临界值范围的ζ值可能表示 x_i 与 x_{pt} 的差异较大和（或）参加者对不确定度的估计值偏小。

ζ 值可以与 z 值联合使用，作为提升实验室能力的一种辅助手段。若某参加者所得z值多次超出临界值3.0，则有必要逐一检查测量过程，并识别测量过程中最大的不确定度来源以寻找改进空间。若ζ值反复超出临界值3.0，则表明不确定度评估未包含重要的不确定度分量（例如，遗漏了某些重要因素）。相反，如果某参加者 $z \geq 3.0$ 但ζ<2.0，则表明该参加者可能准确评估了结果的不确定度，但结果并不符合该能力验证计划中的预期能力。可能存在这种情况，如某参加者在测量程序中使用初筛法，而其他参加者使用准确定量法时会有这种情况发生。如参加者认为结果的不确定度能满足应用，则无需采取任何行动。

注：单独使用 ζ 值时，ζ 值仅可解释为评价参加者的不确定度是否符合测试偏差，而不可作为对特定参加者结果的适用性目标的表示。适用性目标可由参加者或认可机构通过评估偏差（$x - x_{pt}$）来确定，或将合成不确定度与目标不确定度进行比较来确定。

F.5 E_n 值

E_n 值可用于评估参加者结果与指定值的差异是否在所声称的扩展不确定度范围内。该统计方法常用于校准能力验证，也可用于其他类型的能力验证。

计算公式如下：

$$(E_n)_i = \frac{x_i - x_{pt}}{\sqrt{U^2(x_i) + U^2(x_{pt})}} \qquad (F.8)$$

式中

x_{pt} 是参考实验室中确定的指定值。

$U(x_{pt})$ 是指定值 x_{pt} 的扩展不确定度。

$U(x_i)$ 是参加者的结果 x_i 的扩展不确定度。

注：合并两扩展不确定度与ISO/IEC Guide98-3的要求不一致，合并不等同于计算合成扩展不确定度，除非 $U(x_i)$ 和 $U(x_{pt})$ 的包含因子与有效自由度相等。

E_n 值是两个不同（但有关联）能力度量的比值，解释时需谨慎。分子是结果与指定值的偏差，分母是合并的扩展不确定度，若参加者和实施机构能正确评估其不确定度，分母大于分子。因此，如果 $E_n \geqslant 1.0$ 或 $E_n \leqslant -1.0$，则有必要核查不确定度估计值，或纠正其测量；同样，如果 $-1.0 < E_n < 1.0$，则表明结果满意，前提条件是不确定度需正确评估，且偏差 $(x - x_{pt})$ 小于实验室客户的要求。

注：尽管对 E_n 值解释比较困难，但并不妨碍其应用，将不确定度信息与能力验证结果的解释相结合，将极大提高参加者对测量不确定度及其能力评定的理解。

F.6 参加者结果不确定度评估

校准能力验证计划已要求参加者报告不确定度，但在检测能力验证计划中并不常用。

即使能力评定时未使用不确定度，参加者报告能力验证结果的不确定度仍然有用。收集该信息的目的如下：

1) 认可机构可确定参加者报告的不确定度是否在认可范围内（通常适用于校准实验室）；

2) 参加者可核查自身和其他参加者报告的不确定度，从而评估其一致性，确定评估不确定度时是否未考虑所有相关因素、或是否考虑了过多因素；

3) 能力验证可用来证实所声称的不确定度，在报告结果时报告不确定度是最简便的办法。

当利用配方法、有证参考值、独家定值和专家公议值方法确定指定值（见D.2.1至D.2.4中所述方法），且$u(x_{pt})$满足$u(x_{pt}) < 0.3\sigma_{pt}$条件时，参加者结果的标准不确定度不可能小于$u(x_{pt})$，因此可将$u(x_{pt})$用作筛选下限，称为$u_{\min}$。如果用参加者结果公议值确定指定值（见D.2.5），则实施机构需确定$u_{\min}$的实际限值。

注：如果$u(x_{pt})$包含不均匀性或不稳定性造成的变异，则参加者的$u(x_i)$可能小于$u_{\min}$。

参加者报告的标准不确定度不应大于所有参加者稳健标准差（s^*）的1.5倍，因此稳健标准偏差的1.5倍可用作核查报告的不确定度的实际上限，称为$u_{\max}$。

注：基于F分布的百分位数值的平方根，系数1.5是对10个或10个以上结果的标准差的预期变异的上限，当核查参加者报告的不确定度时实施机构可使用不同系数。

如果利用$u_{\min}$或$u_{\max}$或其它标准识别异常不确定度，实施机构需向参加者说明，并说明即使$u(x_i)$小于$u_{\min}$或大于$u_{\max}$，该$u(x_i)$可能仍然有效，如果出现这种情况，参加者和利益相关方需检查结果或不确定度估计值。同样，如果报告的不确定度大于$u_{\min}$且小于$u_{\max}$，该不确定度可能仍然无效。$u_{\min}$和$u_{\max}$仅提供指示信息。

参考资料

[1] GB/T 6379.1-2004, 测量方法与结果的准确度（正确度与精密度）第1部分：总则与定义

[2] GB 4086.4-1983 统计分布数值表 F 分布

[3] GB 4086.3-1983 统计分布数值表 t 分布

[4] GB/T 6379.2-2004, 测量方法与结果的准确度（正确度与精密度）第2部分：确定标准测量方法重复性与再现性的基本方法。

[5] IUPAC/CITAC Guide: Selection and use of proficiency testing schemes for a limited number of participants-chemical analytic laboratories(IUPAC TR). Pure Appl Chem.2010,82(5)

[6] Rousseeuw P.J., & Verboven S. Comput. Stat. Data Anal. 2002, 40 pp. 741－758