

2 0 2 1

REINICIO

Para la
selección
y uso de
ensayos
de aptitud



2021

Guía para la selección y uso de ensayos de aptitud

Instituto Nacional de Metrología

Edwin Arvey Cristancho Pinilla
Director General

Rodolfo Manuel Gómez Rodríguez
Secretario General

Erika Bibiana Pedraza Guevara
Subdirectora de Servicios
Metrológicos y Relación
con el Ciudadano

Álvaro Bermúdez Coronel
Subdirector de Metrología Física.

Diego Alejandro Ahumada Forigua
Subdirector de Metrología
Química y Biología

Adriana Montenegro Bernal
Jefe de Oficina Asesora de Planeación

Organismo Nacional de Acreditación de Colombia

Alejandro Giraldo López
Director Ejecutivo

Mauricio Rodríguez Rodríguez
Director Técnico Internacional

Diana Milena Jácome M.
Coordinadora Sectorial de
Laboratorios de Ensayos y Clínicos.

Para mayor información y solicitud de copias, contacte a:

Instituto Nacional de Metrología
Av. Cra. 50 No. 26 - 55 Int. 2 CAN
Bogotá D.C - Colombia
Tel: +57 1 254 22 22
www.inm.gov.co

Organización de las Naciones
Unidas para el Desarrollo
Industrial - ONUDI, Colombia
Calle 115 No. 5-50, Bogotá
Tel: +57 1 477 98 88
www.gqspcolombia.org

GQSP Colombia - programa de calidad para la cadena química

Juan Pablo Díaz-Castillo
Gerente de Proyecto y Oficial de
Desarrollo Industrial de la ONUDI

Helen Jhoana Mier Giraldo
Coordinadora Técnica Nacional

Javier Francisco Fernández Rodríguez
Especialista Nacional de Calidad

Edición y redacción

Diego A. Ahumada F., Juliana C. Barrios G., Diego A. Garzón Z., Katherin Holguín A., Andrés F. León T., Laura V. Morales E., Mauricio Rodríguez R. y Diana M. Jácome.

Diagramación

Agencia Central

ISBN

978-958-52871-8-1

Agradecimientos

Esta guía es el resultado del trabajo conjunto entre el Instituto Nacional de Metrología de Colombia – INM y el Organismo Nacional de Acreditación de Colombia – ONAC.

La elaboración de este documento, no hubiera sido posible sin el apoyo de la Organización de las Naciones Unidas para el Desarrollo Industrial - ONUDI en Colombia, la cual a partir de su Programa de Calidad para la Cadena de Químicos- GQSP Colombia, ha buscado la consolidación y el trabajo articulado entre el Organismo Nacional de Acreditación de Colombia - ONAC, el Instituto Nacional de Metrología - INM y todos los actores del Subsistema Nacional de Calidad de Colombia- SICAL.

Finalmente, queremos extender un especial agradecimiento a Helen Jhoana Mier Giraldo y Javier Francisco Fernández Rodríguez del GQSP Colombia - Programa de Calidad para la Cadena Química liderado por la ONUDI; Mauricio Rodríguez, Lía Carrillo y Diana M. Jácome M. del Organismo Nacional de Acreditación-ONAC; Luisa Juana Bernal, María Mercedes Arias, John Emerson Leguizamón, Juliana Serna Saiz, Silvia Ramírez, Antonio García, y Sandra Mendoza del Instituto Nacional de Metrología. El equipo de trabajo extiende sus sinceros agradecimientos a Colombia Productiva, al Ministerio de Comercio, Industria y Turismo, a la Cooperación Económica y Desarrollo (SECO) de la Embajada de Suiza en Colombia así como todas las entidades, instituciones y empresas que participaron y cuyo grano de arena, permite hoy entregar el presente documento.

CONTENIDO

Abreviaturas y símbolos	09
Introducción	10

Generalidades de los ensayos de aptitud

1 Conceptos asociados

1.1	Comparabilidad metrológica de los resultados de medición	14
1.2	Comparación interlaboratorio	15
1.3	Compatibilidad metrológica de resultados de medición	16
1.4	Ensayo de aptitud	17
1.5	Homogeneidad	17
1.6	Indicador de desempeño	18
1.7	Intervalo de medición	18
1.8	Mensurando	19
1.9	Procedimiento de medición de referencia	19
1.10	Programa de ensayo de aptitud	20
1.11	Proveedor de ensayos de aptitud	20
1.12	Incertidumbre de medición	21
1.13	Valor asignado	21
1.14	Valor atípico	22

2 Propósito de los ensayos de aptitud

3 Tipos de ensayos de aptitud

3.1	De acuerdo a su propósito	25
3.1.1	Cuantitativos	25
3.1.2	Cualitativos	25
3.1.3	Interpretativos	26
3.2	De acuerdo a su ejecución	26
3.2.1	Secuenciales	26
3.2.2	Simultáneos	30
3.3	De acuerdo con la manera como se asigna el valor referencia	31

4 Indicadores de desempeño (ISO 13528)

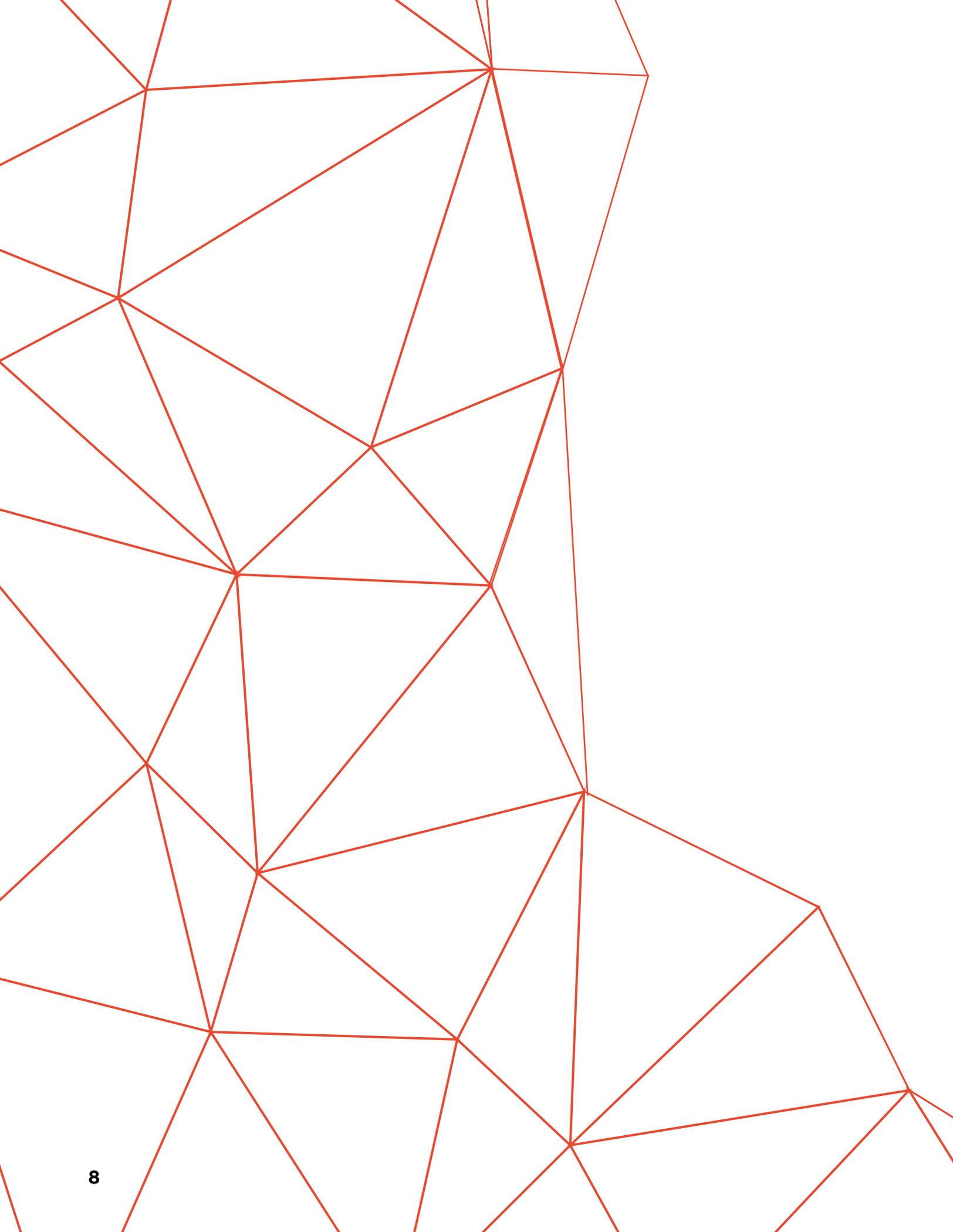
Referencias bibliográficas	37
----------------------------	----

Recomendaciones para la selección de los ensayos de aptitud

1 Generalidades en la selección	40
2 Funcionamiento de los ensayos de aptitud	43
3 Equivalencia u homologación de ensayos de aptitud	46
3.1 Ítem de comparación vs muestras de rutina	47
3.2 Equivalencia en función de la matriz y el analito	48
3.3 Sector de aplicación y técnicas analíticas empleadas	53
3.4 Recomendaciones de la guía Eurachem	57
4 Acerca del proceso de evaluación: indicadores de desempeño	58
4.1 Consideraciones acerca de los valores de referencia	58
4.2 Consideraciones acerca del z_1 score	59
5 Frecuencia de participación	63
6 Disponibilidad de los ensayos de aptitud	64
7 Guía práctica para la selección de EA	67
Referencias bibliográficas	67

Análisis de los resultados y mejora de los sistemas de medición

1 Introducción	72
2 Revisión de los resultados e interpretación	72
2.1 Revisión de los resultados por ronda de EA	74
2.2 Monitoreo del desempeño de los EA a lo largo del tiempo	74
3 Investigación de causas para establecer actividades de mejora	75
4 Posibles fuentes de error	77
5 Analisis de multiples ensayos de aptitud	82
5.1 Puntajes resumidos	82
5.2 Métodos gráficos	84
Referencias bibliográficas	87



Abreviaturas y símbolos

BIPM	Oficina internacional de pesas y medidas
CIL	Comparación interlaboratorio
EA	Ensayos de aptitud
IEA	Ítem de ensayo de aptitud
INM	Instituto Nacional de Metrología
ISO	Organización Internacional de Normalización
ICP-MS	Espectrometría de Masas con Plasma Acoplado Inductivamente
ILAC	International Laboratory Accreditation Cooperation
MR	Material de referencia
MRA	Acuerdo de reconocimiento mutuo
MRC	Material de referencia certificado
ONAC	Organismo Nacional de Acreditación de Colombia
ONU DI	Organización de las Naciones Unidas para el Desarrollo Industrial
PCB	Bifenilos policlorados
PAH	Hidrocarburos aromáticos policíclicos
PTRV	Valor de referencia del ensayo de aptitud (Proficiency test reference value)
u	Incertidumbre estándar de medición
U	Incertidumbre expandida de medición
VIM	Vocabulario internacional de metrología: conceptos básicos y generales, y términos asociados



Introducción

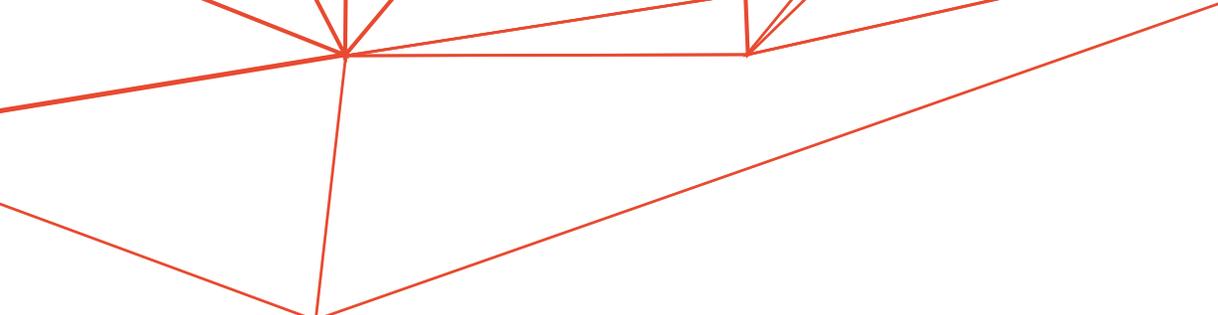
La globalización se ha convertido en uno de los principales componentes asociados al crecimiento de las economías más exitosas del mundo. El crecimiento económico, la generación de empleo y el bienestar social, requieren de políticas propicias que fomenten la inversión, el desarrollo y la innovación en procesos y la modernización de la infraestructura existente. Todo lo anterior, ha inducido un gran dinamismo en diferentes mercados.

El desarrollo de nuevos productos, la protección al consumidor, la eficiencia en la producción o el acceso a mercados más especializados, se convierte en un desafío para las empresas y gobiernos, pues exige el establecimiento de nuevos mecanismos de control de calidad, nuevas regulaciones, procesos estandarizados, entre otros factores que implican contar con una infraestructura de calidad robusta. La infraestructura de la calidad ofrece una amplia variedad de servicios que contribuyen a mejorar la competi-

tividad y la productividad, y facilitan el cumplimiento de la regulación y los estándares internacionales. En este sentido, resulta importante que estos servicios cuenten con un respaldo institucional que permita generar confianza sobre la correcta ejecución de un determinado procedimiento de medición, para lo cual es indispensable contar con servicios acreditados.

La norma ISO/IEC 17025 establece los requisitos, que deben cumplir los laboratorios de ensayo y calibración, para garantizar la competencia técnica y la fiabilidad de los resultados de medición. Dentro de estos requisitos, se encuentra la participación en ensayos de aptitud, los cuales son una herramienta para evaluar la competencia técnica de los laboratorios y empresas en la implementación de sistemas de medición específicos.

Los primeros ensayos de aptitud iniciaron como una actividad de algunos sectores o grupos de laboratorios con el propósito de lograr obtener



información respecto a la capacidad de medición, métodos de medición, mejoras en los procesos de medición, entre otros. Posteriormente, fueron objeto de estudio de diferentes organizaciones como la asociación de Asociación de Químicos Agrícolas Oficiales, la Unión Internacional de Química Pura y Aplicada (IUPAC), la Organización Internacional de Normalización (ISO), entre otros; las cuales establecieron diferentes protocolos que posicionaron a los ensayos de aptitud como la mejor herramienta para la evaluación del desempeño de un laboratorio. Esto ha derivado en que los ensayos de aptitud se constituyan en un pequeño sector industrial, en el cual participan la mayoría de laboratorios analíticos.

El propósito principal de los ensayos de aptitud es demostrar la capacidad técnica del laboratorio en el desarrollo de un ensayo o calibración. Por lo tanto, a través de la participación en ensayos de aptitud el laboratorio puede identificar oportunidades de mejora y establecer las acciones pertinentes para lograr resultados acordes con su propósito. Para los laboratorios de ensayo y calibración la participación en un ensayo de aptitud tiene las siguientes etapas: (i) selección del ensayo de aptitud, (ii) inscripción, (iii) medición del ítem de comparación, (iv) reporte de los resultados, (v) análisis de la información e (vi) implementación de mejoras de los procesos de medición, siendo estas dos últimas etapas muy relevantes para los laboratorios que tienen desempeños no satisfactorios o cuestionables.

En este contexto, a través de esta publicación queremos brindar una guía para los laboratorios de ensayo químico con el propósito de facilitar

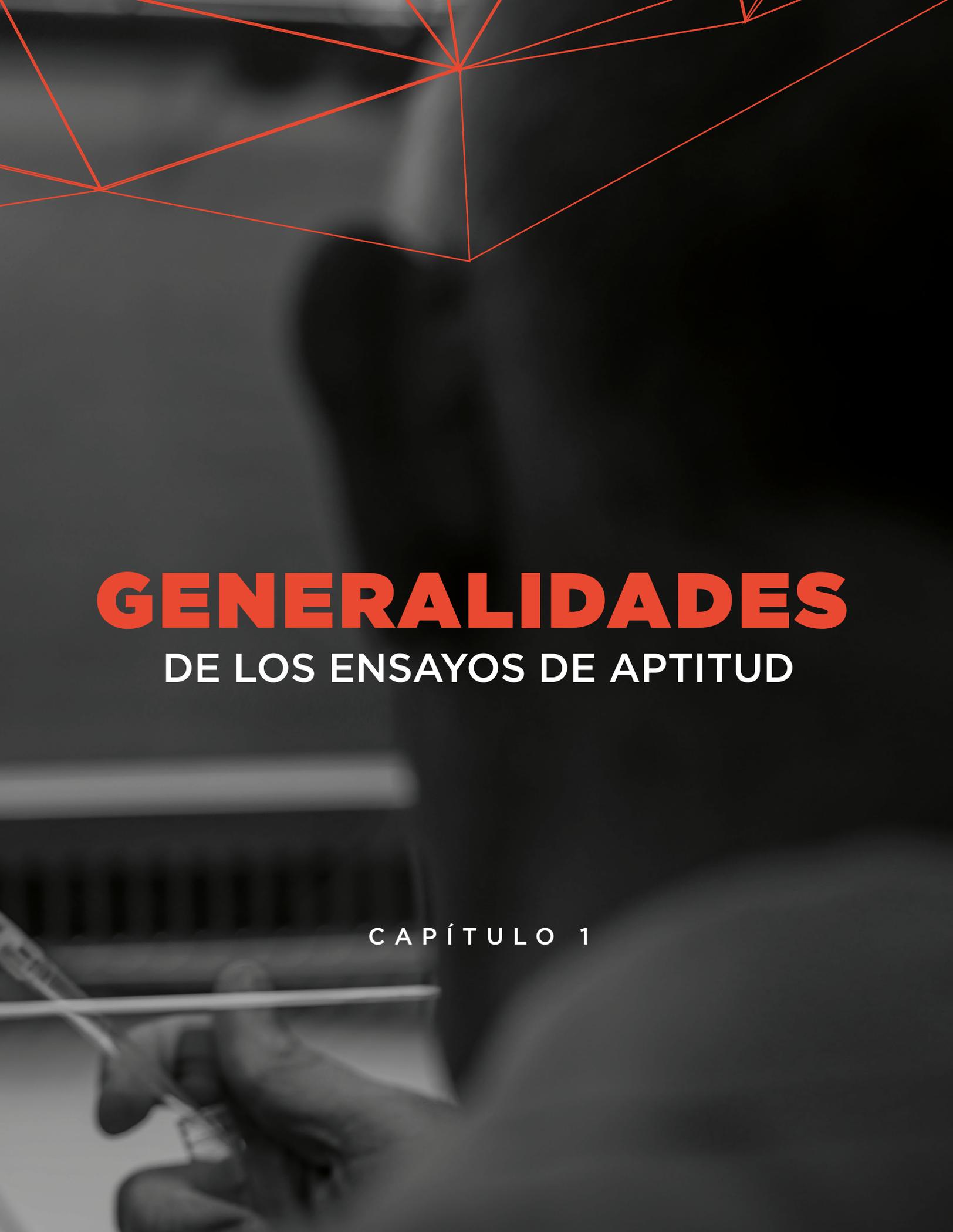
los procesos de selección, participación y análisis de la información de estos ensayos de aptitud.

El presente documento se encuentra estructurado en tres capítulos. En el primer capítulo se busca proporcionar una herramienta que permita a los laboratorios familiarizarse con los términos y generalidades de los ensayos de aptitud y las pruebas estadísticas de desempeño que son usualmente aplicadas; por último, se relacionan las directrices del Organismo Nacional de Acreditación de Colombia (ONAC) para la participación en estos ensayos de aptitud.

En el segundo capítulo, se brindan algunas recomendaciones para la selección de los ensayos de aptitud, dentro de las cuales se hace énfasis en algunos aspectos críticos para la selección como (i) el alcance del ensayo, (ii) las consideraciones prácticas acerca de los principales indicadores de desempeño, (iii) la equivalencia de ensayos de aptitud, (iv) la frecuencia de participación y, (v) se brinda un listado de posibilidades para la búsqueda de estos ensayos de aptitud.

Finalmente, en el tercer capítulo se tiene por objetivo guiar a los laboratorios en la interpretación y análisis de los resultados de los ensayos de aptitud, así como la toma de acciones relacionadas a un resultado no satisfactorio, cuestionable o un bajo rendimiento. Para ello, se abordan temas relacionados con: (i) la revisión e interpretación de los resultados en los diferentes esquemas de ensayos de aptitud, (ii) la investigación para establecer la causa de resultados no satisfactorios, indicando las posibles fuentes de error y, (iii) finalmente se dan recomendaciones para el análisis de los resultados derivados de múltiples ensayos de aptitud.





GENERALIDADES

DE LOS ENSAYOS DE APTITUD

CAPÍTULO 1

1. Conceptos asociados

1.1. Comparabilidad metrológica de los resultados de medición

Comparabilidad de resultados de medición, para magnitudes de una naturaleza dada, que son metrológicamente trazables a la misma referencia. (VIM 2012) [2]

Esta característica permite que si, por ejemplo, un laboratorio en Colombia mide residuos de un plaguicida presente en fresas obteniendo un valor de $1.8 \mu\text{g kg}^{-1} \pm 0.5 \mu\text{g kg}^{-1}$ y otro laboratorio en Alemania determina que el contenido del plaguicida es de $2.1 \mu\text{g kg}^{-1} \pm 0.4 \mu\text{g kg}^{-1}$ los resultados puedan ser equivalentes.

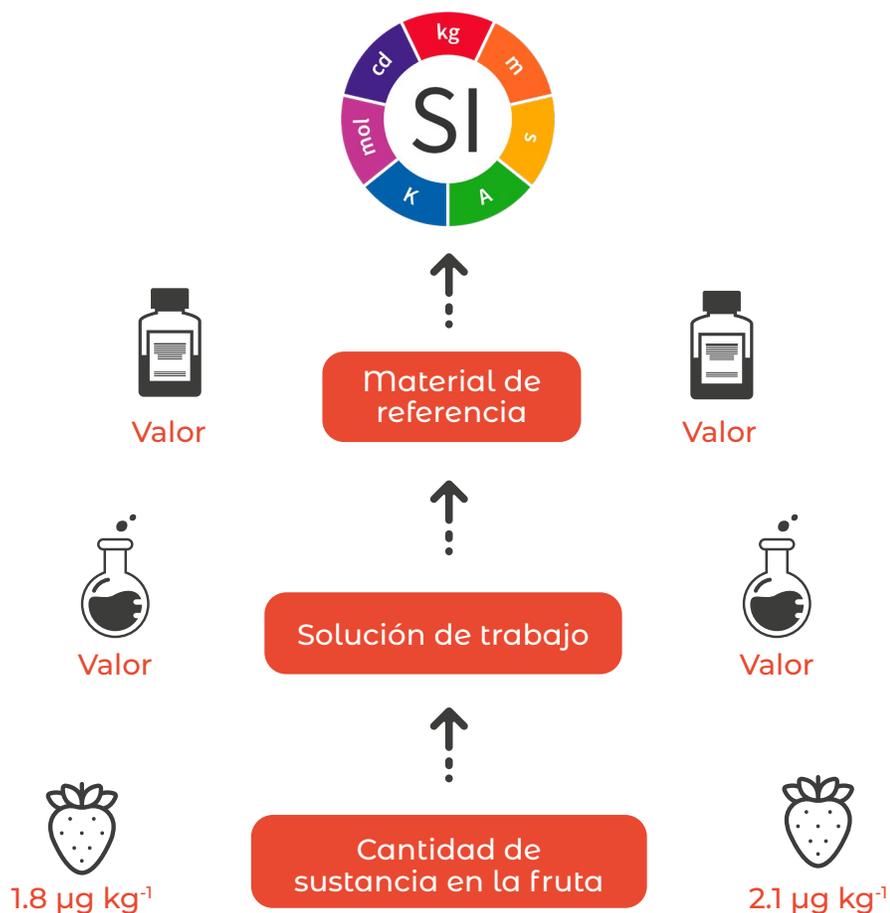
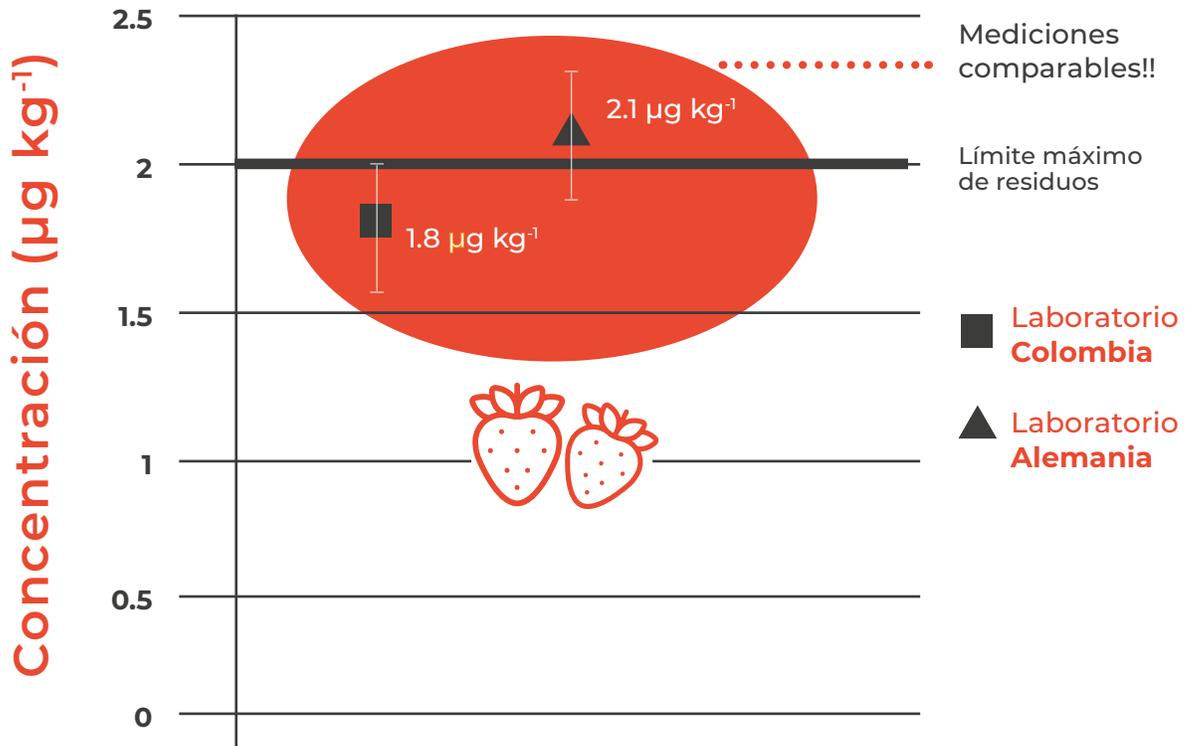


FIGURA 1.1
Ejemplo de comparabilidad metrológica

NOTA

la comparabilidad metrológica se da siempre y cuando las incertidumbres no se superpongan, y siempre y cuando los resultados sean trazables a la misma referencia.

**FIGURA 1.2**

Ejemplo de comparabilidad metrológica

1.2. Comparación Interlaboratorio

Organización, realización y evaluación de mediciones o ensayos sobre el mismo ítem o ítems similares por dos o más laboratorios de acuerdo con condiciones predeterminadas (ISO/IEC 17043:2010) [1]

Es un término general para diferentes ejercicios en los que participan varios laboratorios con el fin de comparar y mejorar la calidad de sus mediciones. Existen diferentes tipos de comparación de laboratorio, las cuales se muestran en la siguiente **Figura 1.3**

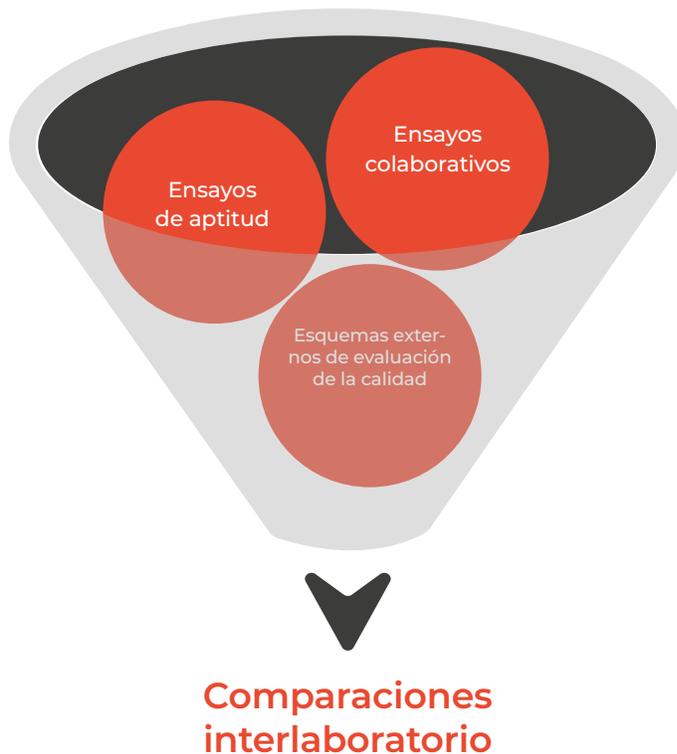


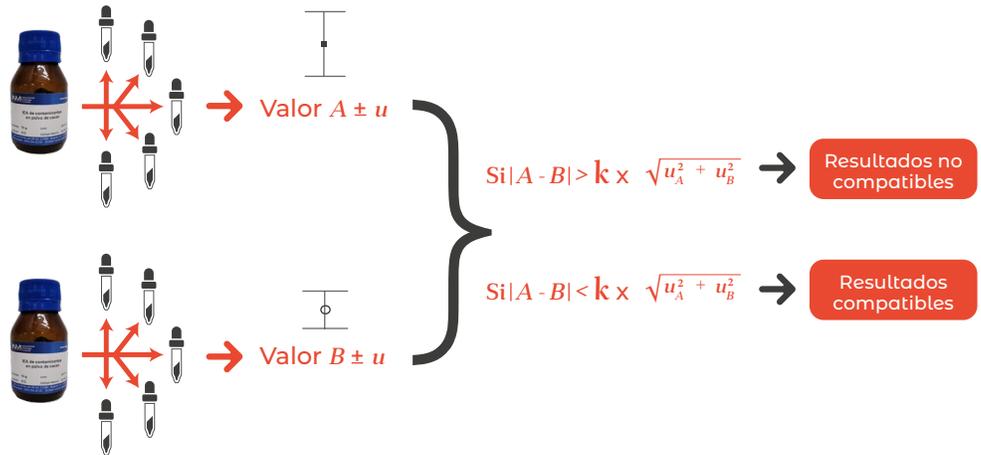
FIGURA 1.3

Comparaciones interlaboratorio

1.3. Compatibilidad metrológica del resultado de medición

Propiedad de un conjunto de resultados de medición de un mensurando específico, tal que el valor absoluto de la diferencia de cualquier par de valores medidos a partir de dos resultados de medición distintos sea inferior a un cierto múltiplo seleccionado de la incertidumbre estándar de esta diferencia. (VIM 2012) [2]

Esta característica permite que si, por ejemplo, se tienen dos resultados independientes para el contenido de cadmio (mg kg^{-1}) en muestras de cacao con sus respectivas incertidumbres, ambos resultados se encuentren acordes.

FIGURA 1.4Compatibilidad
metrológica

1.4. Ensayo de aptitud

Evaluación del desempeño de los participantes con respecto a criterios previamente establecidos mediante comparaciones interlaboratorios (ISO/IEC 17043:2010) [1]

Es un ejercicio que proporciona una evaluación objetiva, independiente y continua del desempeño de un laboratorio para llevar a cabo un determinado procedimiento de medición. Como resultado del ejercicio, cada laboratorio obtiene un indicador numérico que señala si el laboratorio tuvo o no un desempeño satisfactorio. Un resultado satisfactorio permite al laboratorio demostrar y asegurar la validez de sus resultados.

1.5. Homogeneidad

Uniformidad en el valor de una propiedad especificada, en una porción definida de un material de referencia. (Guía ISO 30:2015) [3]

Es uno de los atributos esenciales de los materiales de referencia, su evaluación permite asegurar que cada unidad o porción de una unidad, presenta el valor de la propiedad especificada. Las muestras empleadas en una **comparación interlaboratorio** deben tener determinado su grado de

homogeneidad, de modo que sea posible identificar las diferencias entre los laboratorios participantes.

! **Ejemplo 1:** una muestra de suelo en la cual el analito se encuentra distribuido o concentrado en una fracción específica del suelo, puede considerarse una muestra heterógena para la medición de elementos disponibles en suelos.

! **Ejemplo 2:** una muestra de acero, medida en diferentes puntos, para la que no se encontraron diferencias significativas entre los diferentes valores obtenidos, puede llegar a considerarse una muestra homogénea.

NOTA: es importante mencionar que, aunque los valores medidos sean estadísticamente iguales, no implica que la incertidumbre asociada a la heterogeneidad en la muestra sea igual a cero. Para más detalles consulte ISO guía 35 [4].

1.6. Indicador de desempeño

Puntajes de rendimiento que permiten una interpretación consistente en una gama de medidas y pueden permitir comparar los resultados de diferentes medidas en una base igual. (ISO 13528) [5]

Es un criterio cuantitativo a partir del cual se evalúa el desempeño de los participantes en las **comparaciones interlaboratorio**.

! **Ejemplo:** un laboratorio que participa en un ensayo de aptitud, obtiene un puntaje zeta (puntaje en una base igual) de 0.9, lo cual se considera un desempeño satisfactorio.

1.7. Intervalo de medición

Conjunto de los valores de magnitudes de la misma naturaleza que un instrumento o sistema de medición dado puede medir con una incertidumbre instrumental especificada bajo determinadas condiciones. (VIM 2012) [2]

! **Ejemplo 1:** el intervalo de medición de un laboratorio es de $1 \mu\text{g kg}^{-1}$ a $100 \mu\text{g kg}^{-1}$ de calcio en sangre con incertidumbres estándar de 10% y 2%, respectivamente.

! **Ejemplo 2:** en un método multianalito los intervalos de medición son:

Analito	Intervalo	Máxima incertidumbre Estándar (%)
HCB*	5 $\mu\text{g kg}^{-1}$ a 100 $\mu\text{g kg}^{-1}$	22%
Clorpirifos	1 $\mu\text{g kg}^{-1}$ a 500 $\mu\text{g kg}^{-1}$	11%
Atrazina	1 $\mu\text{g kg}^{-1}$ a 100 $\mu\text{g kg}^{-1}$	17%

*HCB: Hexaclorobenceno

1.8. Mensurando

Magnitud que se desea medir (VIM 2012) [1]

Es la descripción suficientemente detallada e inequívoca de lo que se pretende medir. Para el caso de mediciones químicas, esta descripción debe incluir: **el analito, las unidades y la matriz.**

! **Ejemplo 1:** en la medición de metales tóxicos en pescado el mensurando puede describirse como:

fracción másica (mg kg^{-1}) de mercurio en pescado; en la anterior descripción se identifican claramente lo que se quiere medir o el analito (mercurio), las unidades (mg kg^{-1}) y la matriz (pescado).

! **Ejemplo 2:** en la determinación del contenido de calcio en leche en polvo, el mensurando puede describirse como: contenido de calcio (mg g^{-1}) en leche; en la anterior descripción se identifican claramente lo que se quiere medir o el analito (calcio), la unidad (mg g^{-1}) y la matriz (leche)

1.9. Procedimiento de medición de referencia

Procedimiento de medición aceptado para producir resultados de medición apropiados para su uso previsto en la evaluación de la veracidad de los valores medidos obtenidos a partir de otros procedimientos de medición, para magnitudes de la misma naturaleza, para una calibración o para la caracterización de materiales de referencia. (VIM 2012) [2]

Son procedimientos que han sido muy bien caracterizados y que, normalmente tienen asociados bajos valores de **incertidumbre de medición**. Estos procedimientos, pueden establecer criterios de aceptación de parámetros de validación (p.e. precisión, veracidad, linealidad), facilitando su implementación y uso en los laboratorios. Normalmente son empleados para asignar valor a materiales de referencia, evaluar el desempeño de otros sistemas de medición, entre otros

- ! **Ejemplo 1:** el Comité Conjunto de Trazabilidad en Medicina de Laboratorio o JCTLM ha definido la espectrofotometría como el método de referencia para la determinación de la actividad enzimática de la fosfatasa alcalina.
- ! **Ejemplo 2:** el método Kjeldahl para la medición de proteínas.

1.10. Programa de ensayo de aptitud

Ensayos de aptitud diseñados y operados en una o más rondas para un área específica de ensayo, medida, calibración o inspección (ISO/IEC 17043:2010) [1]

Son herramientas que permiten a los laboratorios participantes demostrar su competencia en la aplicación de sus métodos de ensayo o calibración y, garantizar resultados de calidad; además permiten identificar problemas relacionados con sus procedimientos de medición o calibración e implementar herramientas de mejora.

- ! **Ejemplo:** se participa en un programa de ensayos de aptitud para laboratorios de alimentos, en los cuales se realizan diferentes comparaciones interlaboratorio empleando muestras de diferentes características como matrices grasas, matrices acuosas, matrices ácidas a diferentes concentraciones de los analitos de interés.

1.11. Proveedor de ensayos de aptitud

Organización que es responsable de todas las tareas relacionadas con el desarrollo y la operación de un programa de ensayos de aptitud (ISO/IEC 17043:2010) [1]

Los organizadores de este tipo de ensayos desempeñan un papel importante en la cadena de valor para garantizar productos y servicios. Sus programas, deben estar diseñados adecuadamente y ser estadísticamente robustos, de modo que se asegure que la evaluación del desempeño de los participantes se realiza de manera idónea, permitiendo a los laboratorios el uso de los resultados para identificar problemas en sus procedimientos de medición, implementar acciones de mejora y brindar confianza en sus resultados.

1.12. Incertidumbre de medición

Parámetro no negativo que caracteriza la dispersión de los valores atribuidos a un mensurando, a partir de la información que se utiliza. (VIM 2012) [2]

Es una indicación cuantitativa de la dispersión del resultado de medición y la certeza que se tiene acerca del mismo. La expresión correcta de la **incertidumbre de medición** debe incluir un intervalo, un nivel de confianza, indicar el tipo de distribución y la unidad.

- ❗ **Ejemplo:** el resultado de una determinación de contenido de glucosa (mg dL^{-1}) en sangre se puede expresar como: $154.5 \text{ mg dL}^{-1} \pm 11.7 \text{ mg dL}^{-1}$ a un nivel de confianza del 95%, empleando como factor de cobertura ($k=1.97$), asumiendo una distribución normal.

1.13. Valor asignado

Valor atribuido a una propiedad particular de un ítem de ensayo de aptitud (ISO/IEC 17043:2010) [2]

Es la mejor estimación del “valor verdadero” de la propiedad en prueba. Este valor se establece comúnmente a partir de los resultados de los laboratorios participantes (valor por consenso, cuando hay método dependencia), o puede ser asignado por el organismo coordinador del ensayo de aptitud a partir de mediciones previas (valor de referencia); en cualquier

caso, este valor se emplea para evaluar el desempeño de los participantes.

- ! **Ejemplo 1:** la asignación de la fracción másica de Cd en agua se realiza a través de doble dilución isotópica, el cual es un método primario reconocido por el comité consultivo de cantidad de sustancia.
- ! **Ejemplo 2:** un valor de referencia puede ser establecido a través de una red de laboratorios reconocidos y competentes, por ejemplo, la asignación de un valor a un material entre Institutos Nacionales de Metrología.

1.14. Valor atípico

Observación en un conjunto de datos que parece no concordar con los restantes datos de dicho conjunto. (ISO/IEC 17043:2010) [2]

Son casos en los que se presentan observaciones de características diferentes a las demás, usualmente son resultados que no representan la población y, pueden llegar a distorsionar los comportamientos estadísticos generales.

- ! **Ejemplo:** en un conjunto de datos de un interlaboratorio todos los participantes reportaron valores entre $10 \mu\text{g kg}^{-1}$ y $14 \mu\text{g kg}^{-1}$, a excepción de un laboratorio que reportó $381 \mu\text{g kg}^{-1}$, lo anterior parece indicar que es un valor atípico.

2. Propósitos de los ensayos de aptitud

Los ensayos de aptitud son una de las herramientas de control de calidad que, garantizan el aseguramiento de la validez de los resultados de los laboratorios (ISO/IEC 17025:2017) [6] y, a su vez, permiten verificar que los procedimientos de medición aplicados por los laboratorios son adecuados para el uso previsto. Este tipo de ensayos, permiten a los participantes:

La evaluación objetiva, continua e independiente

Los proveedores de ensayos de aptitud deben cumplir con los requisitos de competencia establecidos en las diferentes normativas, de modo que garanticen la confiabilidad en los resultados de los EA, de esta manera se generan evidencias objetivas del desempeño técnico de los participantes. Por otro lado, una participación continua de los laboratorios permite evaluar periódicamente su desempeño, proporcionando un registro de cualquier mejora que se haya logrado o, por el contrario, puede alertar a un laboratorio sobre la disminución del rendimiento y, por lo tanto, impulsar la introducción de soluciones. [7]

La identificación de problemas analíticos

Si los resultados de un laboratorio en un ensayo de aptitud no cumplen con los criterios de desempeño, es probable que los laboratorios investiguen e identifiquen las posibles fuentes de error o las circunstancias que derivaron en un desempeño no satisfactorio o cuestionable y tomar las acciones relevantes, evitando la prestación de un mal servicio a sus clientes. En este sentido, los EA generan una retroalimentación que estimula la mejora de la competencia técnica, brindando una herramienta de gestión de riesgos y mejora continua de la calidad.

Demostrar calidad de las mediciones

A partir de un desempeño satisfactorio en un ensayo de aptitud, se demuestra la capacidad del laboratorio para emitir resultados confiables y de calidad, ya que son obtenidos a partir de la adecuada ejecución de los procedimientos de medición o calibración de un laboratorio.

Mejora continua en los procesos de medición

En caso de un resultado no satisfactorio o cuestionable en un ensayo de aptitud se pueden establecer planes de mejora con las recomendaciones realizadas por el organizador. De igual manera se pueden plantear estrategias y evaluaciones de los métodos de medición aplicados; por ejemplo, si todos los participantes en el ensayo de aptitud emplearon el mismo método de medición y el mismo instrumento y solo uno de los participantes obtuvo un resultado significativamente diferente a los demás se pueden plantear acciones que permitan identificar problemas en alguna de las etapas de medición.

Por otro lado, en el caso de tener resultados positivos, se pueden plantear igualdades y categorías entre los participantes proporcionando así una comparabilidad entre distintos o el mismo método de medición aplicado.

Validar la incertidumbre declarada

Los ensayos de aptitud permiten la verificación periódica del desempeño de un laboratorio. En consecuencia, los resultados del laboratorio obtenidos a través de diferentes participaciones en EA pueden utilizarse para validar la incertidumbre de medición. [8]

La confiabilidad de los resultados para los clientes

Demostrar la competencia técnica de los laboratorios, a través de la participación en EA, puede generar confianza tanto al interior del laboratorio como a los usuarios externos incluyendo clientes y otras partes afectadas por los resultados de los ensayos y calibraciones.

Acceder a esquemas de acreditación

El desempeño satisfactorio de un laboratorio en un EA, o la implementación de acciones efectivas frente a un desempeño no exitoso, puede proporcionar a los organismos reguladores y de acreditación confianza en los resultados generados por el laboratorio y posicionarlos como organizaciones competentes.

Estado actual de la calidad analítica

A partir de los resultados de los EA, es posible obtener información acerca de la calidad de mediciones específicas realizadas por los laboratorios de un sector o región, y de esta manera identificar sus necesidades, fortalezas y debilidades en materia de medición [9]

3. Tipos de ensayos de aptitud

Existen diversas clasificaciones de los ensayos de aptitud. Para los fines de esta guía se han clasificado en tres grandes grupos: (i) según el propósito del ensayo de aptitud o lo que se pretende evaluar, (ii) según la manera en que se adelante la ejecución del ensayo de aptitud o (iii) según la manera en la que se asigne el valor de referencia.

3.1. De acuerdo a su propósito

3.1.1. Cuantitativos

En este tipo de ejercicio, se tiene por objetivo evaluar una o varias magnitudes relacionadas con la medición de un ítem de ensayo en donde el resultado es un valor numérico con su respectiva unidad. (ISO/IEC 17043:2010) [2]

Este es el tipo de ensayo de aptitud más común en los laboratorios de análisis químicos. El establecimiento del valor de referencia se puede lograr a través de un valor por consenso o un valor previamente determinado por el proveedor del ensayo de aptitud

- ❗ **Ejemplo 1:** se envía un ítem de comparación (Material de referencia de arroz) a todos los participantes, los cuales cuantifican la fracción másica (mg kg^{-1}) de arsénico en arroz, reportan su valor y posteriormente a través de un indicador de desempeño se evalúa su competencia respecto al valor asignado, que ha sido previamente establecido mediante un método primario de medición.
- ❗ **Ejemplo 2:** en la medición de grasa total en un alimento, se envía un alimento como ítem de comparación a diferentes laboratorios para la cuantificación de grasa total, posteriormente se realiza un tratamiento de los datos para encontrar el mejor valor de referencia a partir de los resultados de los participantes.

3.1.2. Cualitativos

Estos ensayos de aptitud, tienen como propósito describir una o más características de un ítem y se comunican en una escala categórica u ordinal. (ISO/IEC 17043:2010) [2]

Este es el tipo de ensayo de aptitud más común en los laboratorios de análisis químicos cualitativos y análisis biológicos, pues a través de este se busca evaluar la competencia del laboratorio para detectar la presencia o ausencia de un analito o microorganismo de interés. De igual manera, puede ser aplicado para la evaluación de algunas propiedades morfológicas o características que permiten clasificar o aceptar/rechazar un producto.

- ! **Ejemplo 1:** se envían 10 muestras (este es el ítem de comparación) que pueden o no contener SARS-CoV-2, el laboratorio debe reportar la presencia/ausencia en cada una de las muestras. Posteriormente, de acuerdo a su grado de aciertos/desaciertos se evalúa la competencia del laboratorio.

3.1.3. Interpretativos

A través de estos ensayos de aptitud se evalúa una característica descrita por parte del participante. (ISO/IEC 17043:2010) [2]

- ! **Ejemplo 1:** descripción de la morfología de un IEA y posterior clasificación taxonómica
- ! **Ejemplo 2:** interpretación de imagen diagnóstica en búsqueda de una patología sobre un cuerpo humano.

3.2. De acuerdo a su ejecución

3.2.1 Secuenciales

El proveedor del ensayo de aptitud asigna el valor al ítem de comparación o envía este ítem a un laboratorio de referencia para su caracterización y asignación del valor. Luego, el organismo coordinador envía el artículo a cada laboratorio participante para su posterior análisis [11]. Dentro de esta modalidad, existen varias modalidades:

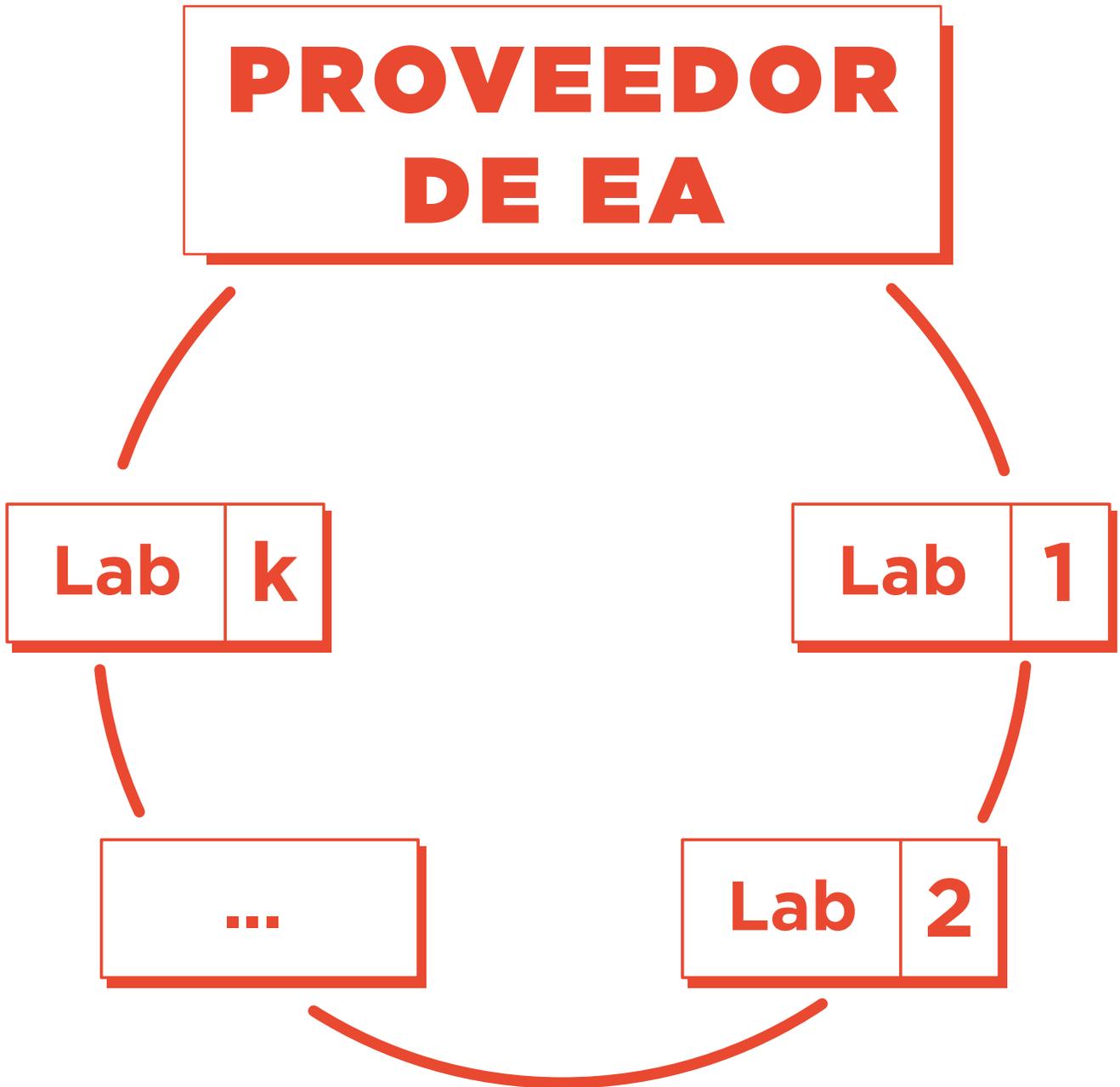
Secuencial clásico en anillo:

en este tipo de EA, el IEA se envía secuencialmente a varios participantes iniciando y finalizando en el proveedor del EA. Cada laboratorio participante mide el ítem de comparación de forma independiente, envía sus resultados al proveedor del EA y el IEA al siguiente laboratorio (ver **Figura 1.5**). Esta variante se emplea cuando:

- El ítem de comparación es lo suficientemente estable.
- El ítem de comparación se puede medir varias veces.

FIGURA 1.5

Ensayo de aptitud
secuencial forma 1



Secuencial “Round Robin”:

Una vez que cada laboratorio participante ha completado la evaluación del ítem, este se devuelve al proveedor de los ensayos de aptitud (ver **Figura 1.6**), para realizar verificaciones en intervalos de tiempo definidos. El proveedor de ensayos de aptitud evaluará los resultados de las pruebas y emitirá un informe de desempeño a cada laboratorio participante. Esta variante se emplea cuando:

- El ítem de comparación no es lo suficientemente estable o requiere realizar un control de calidad más exhaustivo por riesgos asociados al transporte o la manipulación que puedan afectar el valor asignado.
- El ítem de comparación se mide varias veces.

FIGURA 1.6

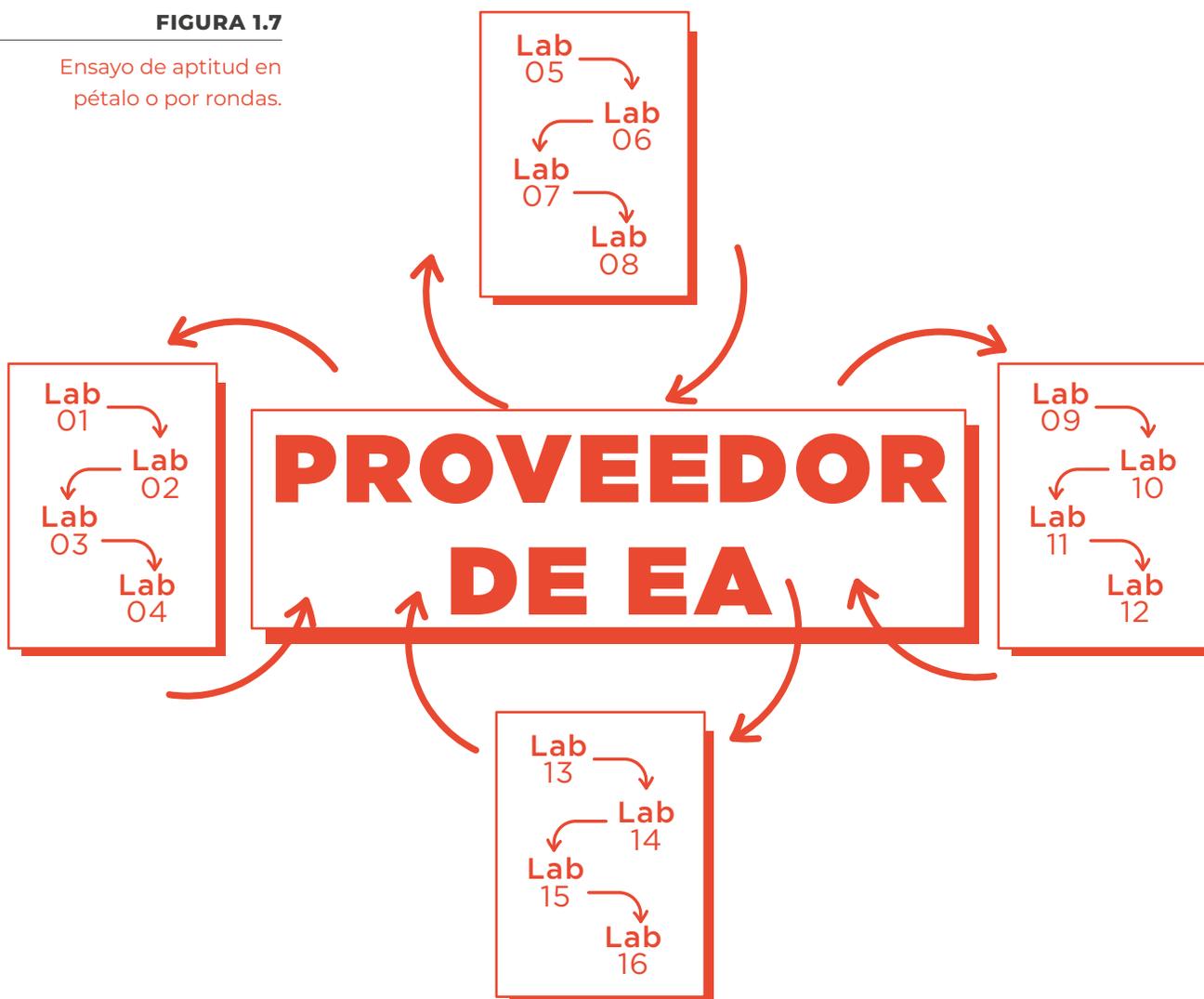
Ensayo de aptitud
secuencial forma 2



Secuencial en pétalo o por rondas:

Es un esquema de ensayos de aptitud secuencial, en el que cada ciclo o ronda se integra por un conjunto de laboratorios, una vez culmine las mediciones de este conjunto de laboratorios se devuelve al proveedor de ensayos de aptitud para verificaciones en intervalos de tiempo definidos. Posteriormente, se envía a otra ronda o a otros laboratorios diferentes.

FIGURA 1.7
 Ensayo de aptitud en
 pétalo o por rondas.



3.2.2. Simultáneos

Los ítems de ensayo se distribuyen en el mismo momento para que la medición se realice dentro de un periodo de tiempo definido. Normalmente involucran sub-muestras seleccionadas aleatoriamente de una fuente de material de referencia lo suficientemente homogéneo y estable (ISO/IEC 17043:2010) [1].

Los esquemas de participación simultánea son los más comunes en mediciones químicas y generalmente se usan para IEA de un sólo uso que se destruyen o descartan después de la medición. En general, los tres diseños o modalidades más comunes son:

Prueba de multinivel:

Es un esquema de ensayos de aptitud en los cuales se evalúa no sólo una concentración, sino un intervalo de concentraciones. Para esto, el ítem de comparación se encuentra compuesto de dos o más muestras, de diferentes concentraciones.

Prueba de muestra dividida:

Es un esquema de ensayos simultáneos en los que el participante recibe una muestra que ha sido dividida en dos o más ítems. El participante debe reportar tanto el promedio como la variación entre las muestras. Lo anterior permite no sólo evaluar el sesgo del valor medido, sino que permite evaluar su precisión.

Prueba de proceso parcial:

Es un esquema de ensayos de aptitud simultáneos en los cuales los participantes deben realizar sólo partes específicas del proceso de medición. Por ejemplo, se distribuyen extractos o soluciones que deben ser cuantificadas, pero el participante no debe realizar dicha extracción.

3.3. De acuerdo con la manera como se asigna el valor de referencia

La asignación del valor de referencia es una de las actividades que mayores retos implica para los proveedores de ensayos de aptitud. En este sentido y en el marco de esta modalidad, se cuentan tres variantes.

✓ **Valor asignado por un laboratorio de referencia (ISO 13528) [5]:**

Es un esquema de ensayos de aptitud en el cual un laboratorio de referencia asigna el valor al material de referencia, a través de métodos primarios, uso de diferentes métodos de medición, métodos de referencia entre otros. Posteriormente, los laboratorios participantes miden y, a partir de los resultados reportados por los participantes, se realiza la evaluación. (Figura 1.8).

FIGURA 1.8
Ensayo de aptitud
simultáneo con
laboratorio de referencia
asignando el valor al ítem.



✓ **Valor asignado por una red de laboratorios competentes (ISO 13528) [5]:**

Es un esquema de ensayos de aptitud en el cual una red de laboratorios competentes asigna el valor al material de referencia, a través de un único método o diferentes métodos de medición. Posteriormente, se realiza la evaluación contra el valor asignado por esta red de laboratorios (ver **Figura 1.9**).

FIGURA 1.9

Asignación del valor a un material a través de una red de laboratorios competentes



☑ **Valor asignado por consenso (ISO 13528) [5]:**

Es un esquema de ensayos de aptitud en el cual se asigna el valor al IEA, de los mismos resultados de los laboratorios participantes. Posteriormente, se realiza la evaluación contra el valor asignado por el laboratorio de referencia (ver **Figura 1.10**).

FIGURA 1.10

Ensayo de aptitud
simultáneo con
asignación del valor
por consenso.



4. Indicadores de desempeño (ISO 13528) [5]:

Las pruebas estadísticas evalúan el resultado reportado por el laboratorio frente al valor de referencia de una manera imparcial y por lo tanto son un indicador de la competencia técnica del laboratorio. Existen diversos indicadores de desempeño, sin embargo, el uso de cada uno de ellos está determinado por el tipo de ensayo de aptitud, el tipo de medición, entre otros factores. En este sentido, no se puede establecer un único estadístico de desempeño, sino que el proveedor debe ofrecer el más adecuado para el propósito previsto.

Normalmente, esta selección depende exclusivamente del proveedor de ensayos de aptitud, sin embargo, es recomendable que los participantes conozcan el sistema de evaluación del proveedor, con el propósito de seleccionar e interpretar de la mejor manera el ensayo de aptitud. Por lo anterior, la siguiente tabla presenta algunos de los principales indicadores empleados en ensayos de aptitud.

Tabla 1. Indicadores de desempeño empleados en ensayos de aptitud

Indicador	Fórmula	Criterio	Ventajas	Desventaja
Diferencia D	$D_i = x_i - x_{pt}$	Satisfactorio, si:	Permite realizar la estimación del sesgo del participante con el valor de referencia, lo cual hace que sea fácilmente interpretable	Debe verificarse la trazabilidad del valor asignado.
Diferencia Porcentual %D	$\%D_i = \frac{x_i - x_{pt}}{x_{pt}} * 100$	$-\delta_E < D_i < \delta_E$ ó $-3\sigma_{pt} < D_i < 3\sigma_{pt}$		Los puntajes entre diferentes rangos de la misma determinación pueden no ser comparables
(ISO 13528, numeral 9.3) [5]		No Satisfactorio, si: $D_i \leq -\delta_E$ ó $D_i \geq \delta_E$ ó $D_i \leq -3\sigma_{pt}$ ó $D_i \geq 3\sigma_{pt}$	$\%D_i$ es independiente de la magnitud del valor asignado	Si x_{pt} se obtiene con base a los resultados de los participantes para este caso, puede ser difícil comparar resultados sobre una base grupal ya que puede haber relativamente pocos participantes que tengan capacidades de medida comparables.

<p>Porcentaje de Diferencia permitida (ISO 13528, numeral 9.3) [5]</p>	$\%P_{Ai} = \frac{D_i}{\delta_E} * 100$	<p>Satisfactorio, Si:</p> $ \%P_{Ai} < 100\%$ <p>No Satisfactorio. si:</p> $ \%P_{Ai} \geq 100\%$	<p>Fácilmente interpretable</p> <p>Depende y se ajusta de acuerdo a una especificación técnica</p>	<p>Debe verificarse la confiabilidad del valor asignado.</p>
<p>Puntaje z (ISO 13528, numeral 9.4) [5]</p>	$z_i = \left(\frac{x_i - x_{pt}}{\sigma_{pt}} \right)$	<p>Aceptable</p> $ z_i < 2.0$ <p>Señal de advertencia</p> $2.0 \leq z_i < 3.0$ <p>Inaceptable</p> $ z_i \geq 3.0$	<p>Se tiene en cuenta la variabilidad del ensayo de aptitud y es fácilmente interpretable desde la distribución normal.</p>	<p>Si σ_{pt} se calcula erróneamente se puede afectar el cálculo del indicador. Si σ_{pt} se calcula haciendo uso de los resultados de los participantes se debe contar con una cantidad adecuada de participantes que permita obtener cálculos estadísticamente apropiados.</p> <p>Falta de certeza en el valor del indicador obtenido puesto que no se introduce la incertidumbre de la medida del laboratorio participante y del valor asignado.</p>
<p>Puntaje z' (ISO 13528, numeral 9.5) [5]</p>	$z'_i = \left(\frac{x_i - x_{pt}}{\sqrt{\sigma_{pt}^2 + u^2(x_{pt})}} \right)$	<p>Aceptable</p> $ z'_i < 2.0$ <p>Señal de advertencia</p> $2.0 \leq z'_i < 3.0$ <p>Inaceptable</p> $ z'_i \geq 3.0$	<p>Considera la incertidumbre del valor asignado dentro del cálculo, dado que es significativa</p> $(u(x_{pt}) < 0.3 \sigma_{pt})$	<p>Falta de certeza en el valor del indicador obtenido puesto que no se introduce la incertidumbre de la medida del laboratorio participante</p>

<p>Puntaje zeta (ζ_i)</p> <p>(ISO 13528, numeral 9.6) [5]</p>	$\zeta_i = \left(\frac{x_i - x_{pt}}{\sqrt{u^2(x_i) + u^2(x_{pt})}} \right)$	<p>Aceptable</p> $ \zeta_i < 2.0$ <p>Señal de advertencia</p> $2.0 \leq \zeta_i < 3.0$ <p>Inaceptable</p> $ \zeta_i \geq 3.0$	<p>Existe mayor certeza en el valor del indicador obtenido</p> <p>Debido a que se usan las incertidumbres estándar esto permite la misma interpretación que para los valores de z tradicionales.</p>	<p>El denominador es obtenido a través de una combinación de incertidumbres que puede no ser comparables; si las incertidumbres no son comparables, no sería adecuado interpretar un resultado como aceptable.</p>
<p>Error normalizado (E_n)</p> <p>(ISO 13528, numeral 9.7) [5]</p>	$(E_n)_i = \left(\frac{x_i - x_{pt}}{\sqrt{U^2(x_i) + U^2(x_{pt})}} \right)$	<p>Satisfactorio</p> $-1.0 < (E_n)_i < 1.0$ <p>No satisfactorio</p> $(E_n)_i \geq 1.0$ <p style="text-align: center;">ó</p> $(E_n)_i \leq -1.0$	<p>Existe mayor certeza en el valor del indicador obtenido.</p>	<p>El denominador es obtenido a través de una combinación de incertidumbres que puede no ser comparables; si las incertidumbres no son comparables, no sería adecuado interpretar un resultado como satisfactorio.</p>

Referencias bibliográficas

- [1] Joint committee for guides in Metrology (JCGM), International vocabulary of Metrology (VIM). 2012.
- [2] ISO, "ISO/IEC 17043:2010. Evaluación de la conformidad - Requisitos generales para los ensayos de aptitud. Traducción oficial.," vol. 2010, p. 51, 2010.
- [3] International Organization for Standardization, ISO/Guide 30:2015(en) Reference materials — Selected terms and definitions. 2015.
- [4] International Organization for Standardization, ISO/Guide 35:2017(en) Reference materials — Guidance for characterization and assessment of homogeneity and stability. 2017.
- [5] International Organization for Standardization, ISO 13528:2015(en) Statistical methods for use in proficiency testing by interlaboratory comparison.
- [6] Organización Internacional de Estandarización, ISO. ISO/IEC 17025, Requisitos Generales para la Competencia de los Laboratorios de Ensayo y Calibración. 2017
- [7] R. Society and B. House, "Proficiency testing of analytical laboratories: Organization and statistical assessment," Analyst, vol. 117, no. 1, pp. 97–104, 1992.
- [8] M. Désenfant and M. Priel, "Road map for measurement uncertainty evaluation," Meas. J. Int. Meas. Confed., vol. 39, no. 9, pp. 841–848, 2006.
- [9] F. M. de Albano and C. S. ten Caten, "Proficiency tests for laboratories: A systematic review," Accred. Qual. Assur., vol. 19, no. 4, pp. 245–257, 2014.
- [10] onac, "politica para la participación en ensayos de aptitud (ea) en laboratorios cea-3.0-04 (Antes cea-04) versión 03 nivel 1: 3.0 Prestación del servicio," 2018.
- [11] EEE-PT WG, Eurachem guide: selection, use and interpretation of proficiency testing (PT) schemes, 2nd ed. 2011.





RECOMENDACIONES

PARA LA SELECCIÓN DE LOS
ENSAYOS DE APTITUD

CAPÍTULO 2

Recomendaciones para la selección de los ensayos de aptitud

1. Generalidades en la selección

La participación en un ensayo de aptitud (EA) permite que un laboratorio acceda a una evaluación externa de su desempeño, y con ello, asegure la validez de sus resultados y consolide su estrategia interna de calidad.

Para que la participación en un EA sea efectiva, deberá hacerse una adecuada selección de éste tomando en cuenta diferentes aspectos: pertinencia del ensayo, frecuencia establecida, nivel de riesgo del laboratorio, entre otros. En la pertinencia del ensayo, es importante verificar la similitud o “equivalencia” entre las propiedades medidas en el laboratorio y las ofrecidas en el esquema interlaboratorio, evaluando: la naturaleza de la muestra a medir, los analitos y los niveles de concentración. Por ser la equivalencia un aspecto fundamental en la selección de un EA, se presenta información útil al respecto en el numeral 3 de este capítulo.

Un ejercicio correcto de EA implica varios pasos encadenados entre sí, la **Figura 2.1** presenta un esquema general de estos pasos.

FIGURA 2.1

Esquema general de un ensayo de aptitud



☑ **Selección e inscripción:**

Es necesario realizar una evaluación rigurosa de los documentos e información que provee el organizador del EA, para determinar si el ensayo es adecuado para el laboratorio y se ajusta a su realidad (propiedades y tipo de muestras). Participar en un EA requiere una inversión de recursos y personal, por lo tanto, es importante evaluar la calidad del proveedor de EA, éste deberá estar acreditado bajo la norma ISO/IEC 17043 para considerarlo competente y en lo posible, deberá ser independiente, es decir, que no tenga alianzas o intereses comerciales (marcas de reactivos, materiales o equipos) o demostrar imparcialidad. Si se trata de la primera participación es deseable escoger un EA que permita seleccionar una sola ronda.

Por otro lado, se debe considerar el sector en el que opera el laboratorio o la metodología empleada, por ejemplo: importancia y uso final del resultado en sector salud, forense, judicial, entre otros. De igual manera, se deben considerar aspectos como: estabilidad de las muestras, técnica de medición (si aplica), condiciones de envío de muestras, aduanas o restricciones legales en el ingreso o movilidad de las muestras en el país (si aplica).

Por último, se debe conocer cómo serán tratados los datos que resulten del ejercicio, las aproximaciones estadísticas y el estimador de desempeño que serán empleados para dicho análisis, puesto que al entender esto, el laboratorio cuenta con otro criterio para establecer si su participación en un EA es pertinente o no.

☑ **Medición y reporte:**

Durante la etapa de medición, esta deberá hacerse naturalmente, casi sin notar que el ítem de ensayo es una muestra diferente, lo ideal es incluirla con las muestras de rutina del laboratorio. En el caso de ser posible, el ítem se podría dividir en submuestras, de tal manera que se puedan evaluar los analistas o el efecto del tiempo en el laboratorio.

☑ **Análisis de resultados:**

Cuando se tienen los resultados finales del ensayo de aptitud deberá analizarse el desempeño de manera tal que, si no son satisfactorios o son cuestionables (o están dentro del intervalo de “señal de advertencia”), empiece un proceso de diagnóstico que permita encontrar las causas del bajo desempeño y diseñar acciones correctivas de carácter técnico. Por otro lado, en el caso en que sean satisfactorios los resultados es importante analizar la información relacionada con el sesgo y la repetibilidad obtenida con el propósito de establecer procesos de mejora continua o criterios de control de calidad.

Teniendo en cuenta lo anterior, en las siguientes secciones se desarrollan algunos de estos aspectos importantes para una adecuada selección de un EA. En la primera sección titulada: “*Funcionamiento de los ensayos de aptitud*” se darán algunas consideraciones acerca de la operación de los ensayos de aptitud, la cual tiene como objetivo brindar las herramientas a los laboratorios para lograr una mayor claridad en lo referente a los protocolos o documentación que brinda el proveedor.

En la segunda sección titulada: “*Equivalencia y homologación de ensayos de aptitud*”, se busca brindar conceptos para que los laboratorios encuentren alternativas de participación en ensayos de aptitud, y participen de una manera más estratégica, en especial cuando los realizan de manera frecuente.

En la sección: “*Consideraciones acerca del z_i score*”, se busca dar un complemento práctico al capítulo I y dar elementos que brinden conocimientos para que el lector genere sus propios criterios en lo relacionado con este indicador, el cual, es quizá, el más empleado en mediciones químicas.

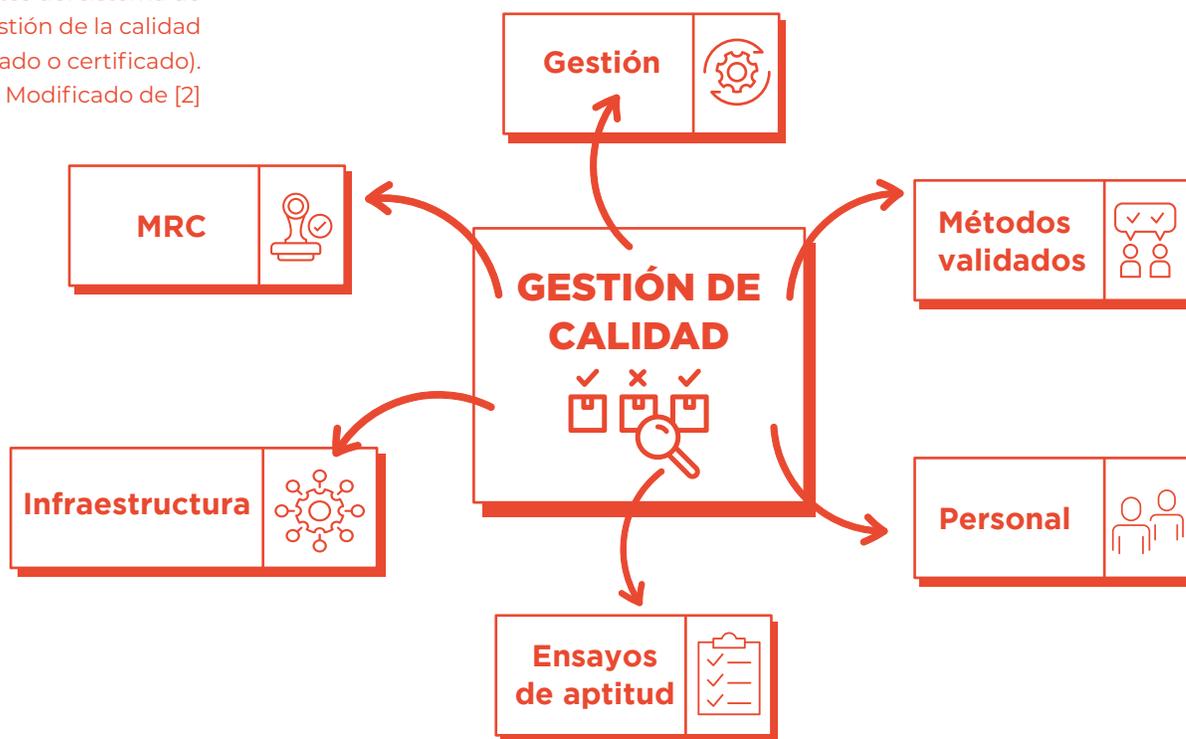
Por último, se presentan las secciones “*Frecuencia de participación*” y “*Disponibilidad de ensayos de aptitud*”, las cuales buscan dar recomendaciones para alinear la participación a los requerimientos de los organismos de acreditación.

2. Funcionamiento de los ensayos de aptitud

La participación en ensayos de aptitud es una parte fundamental del sistema de gestión de la calidad puesto que permite la evaluación externa, y en el caso en que el EA sea desarrollado con el empleo de un MRC, la evaluación del sesgo del laboratorio. Sin embargo, su uso por sí solo, desacoplado de un buen sistema de gestión de la calidad, carece de validez. En la **Figura 2.2**, se muestran las actividades complementarias que permiten el aseguramiento de la validez de los resultados y un adecuado sistema de gestión de la calidad [1].

FIGURA 2.2

Partes del sistema de gestión de la calidad (acreditado o certificado).
Modificado de [2]



Un EA se divide en etapas, en una primera, el organizador del EA distribuye un protocolo preliminar a los participantes, éste es un documento donde se definen los términos en los que se realizará el EA y se mencionan todos los aspectos de cómo se llevara a cabo el ejercicio, la descripción del IEA, la forma de obtención del valor de referencia de la propiedad medida en el IEA, los métodos a emplear (si aplica), los criterios estadísticos para

el análisis de los resultados, los criterios técnicos generales, la distribución de los IEA, entre otros. Por lo anterior, el laboratorio deberá leer y asegurarse de entender muy bien el protocolo y sus implicaciones, porque su aceptación es fundamental para la participación y bajo esas condiciones se registrará el EA.

Posteriormente, el organizador se encarga de preparar o contratar la preparación de los IEA. En los ensayos químicos, es usual que se preparen varias unidades del material de estudio, lo cual se denomina lote, y estas unidades posteriormente se distribuyen entre los laboratorios participantes. Generalmente, el material preparado se ha estudiado previamente, a través de estudios de homogeneidad y estabilidad, lo cual asegura la idoneidad del material para el ensayo de aptitud. El estudio de la homogeneidad asegura que cada unidad que se produjo es comparable o no presenta variaciones entre las diferentes unidades; el estudio de la estabilidad, asegura que el material permanece inalterado o que sus posibles alteraciones no afectan la(s) propiedad(es) de interés, durante el tiempo de transporte y durante el periodo previo a la medición de ítem comparación del EA.

El IEA se envía a los participantes bajo las condiciones que se establezcan para su preservación y es responsabilidad del laboratorio asegurar que esas condiciones se mantengan hasta que realice la medición. Los resultados son reportados por cada laboratorio al organizador, empleando los medios dispuestos para tal fin. El organizador deberá informar en el protocolo como protegerá la confidencialidad de la información, que generalmente se lleva a cabo mediante la codificación de los participantes. El organizador realiza el análisis estadístico de los datos y comunica los resultados a cada laboratorio mediante unos indicadores de desempeño predefinidos en el protocolo preliminar. Es importante entender cómo se obtienen estos indicadores y su implicación en la calificación del desempeño. La descripción de éstos se encuentra en el numeral 4 del capítulo I. En general, los resultados deberían estar diagramados de tal forma que sean fáciles de interpretar, adicionalmente, en el informe o reporte realizado por el organizador deberá existir un espacio donde se presenten los datos crudos reportados por los laboratorios, de tal manera que los participantes tengan la oportunidad de revisarlos y si existe algún error p.e. de transcripción, solicitar la revisión respectiva. En la **Figura 2.3**, se muestra el esquema mediante el cual opera un ensayo de aptitud.

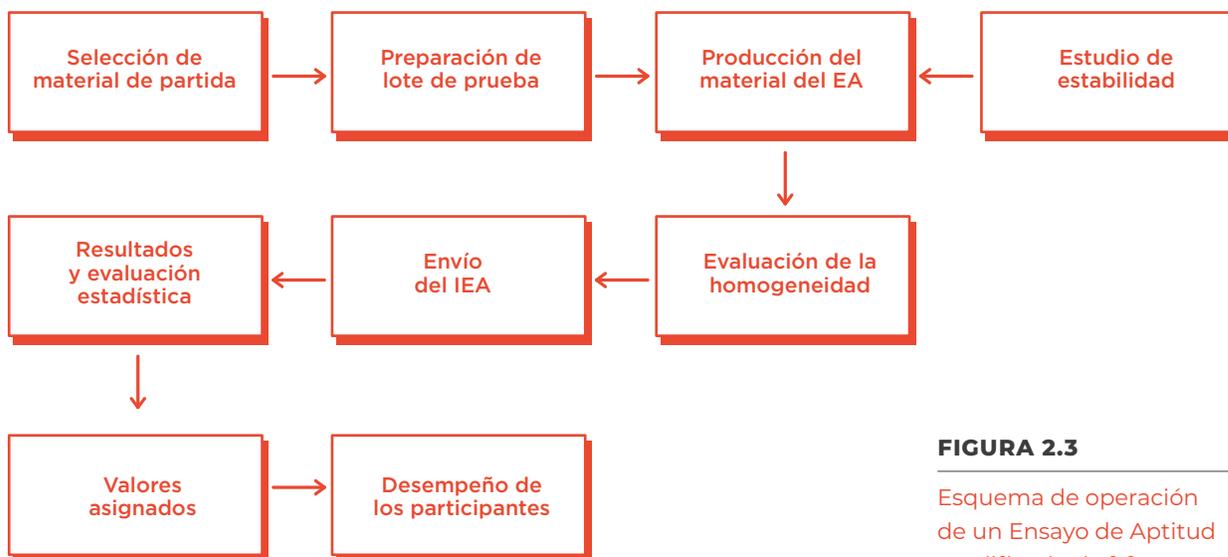


FIGURA 2.3

Esquema de operación de un Ensayo de Aptitud modificado de [3]

Es importante resaltar que, aunque en el marco de los sistemas de gestión de calidad basados en ISO 17025 es esencial la participación en EA, esta constituye una herramienta complementaria del proceso de aseguramiento de la calidad (**Figura 2.2**), puesto que al entender el funcionamiento de un EA, se hace evidente que existen serias limitaciones en el alcance de la información obtenida de estos ejercicios, entre las que pueden mencionarse [4]:

- I. La interpretación de los datos obtenidos del EA está sujeta a una incertidumbre estadística y algunas decisiones en el análisis de la información pueden ser arbitrarias.
- II. El proceso de preparación del ítem de comparación no necesariamente refleja el proceso de medición completo del laboratorio.
- III. Dependiendo del indicador de desempeño del laboratorio, puede existir dependencia con otros laboratorios (véase numeral 4 del capítulo I)
- IV. La variación en la fiabilidad de la información recopilada en el EA por prácticas inadecuadas de algunos participantes, fuertemente presionados por la obtención de buenos resultados.

- V. El alcance de la participación en EA está limitada por los costos, de tal manera que, generalmente solo algunas de las determinaciones realizadas por los laboratorios se someten a evaluación por EA. Generalmente, se selecciona un EA cuyo desafío de medición se considere representativo de los productos o las metodologías analíticas que realiza el laboratorio de ensayo. Sin embargo, en la mayoría de los casos, los resultados obtenidos no son representativos del comportamiento del laboratorio en todo su portafolio de actividades.

El resultado de la participación en un EA se verá reflejado en un criterio usualmente cuantitativo que permite juzgar el ejercicio llamado indicador de desempeño. Existen múltiples *indicadores de desempeño*, de los cuales, el proveedor del EA selecciona cual empleará. Debido a su relevancia, es importante conocer cómo se obtienen matemáticamente y las implicaciones que tiene su selección para los participantes. En la sección 4 del Capítulo I, se encuentra una breve descripción de estos aspectos. No obstante, en el numeral 4, se describirán algunas consideraciones adicionales.

3. Equivalencia u homologación de ensayos de aptitud

En el universo de las mediciones químicas, existe una infinidad de combinaciones de analitos, matrices o productos y niveles de concentración, lo cual lleva a que, en muchos casos, encontrar un ensayo de aptitud que coincida con las mediciones realizadas por los laboratorios de ensayo se convierta en un reto. Por lo anterior, es indispensable disponer de criterios para la selección de ensayos de aptitud equivalentes u homologables, por medio de los cuales, el laboratorio pueda evidenciar su competencia técnica y soportar su proceso de acreditación. Por lo anterior, en los siguientes numerales se dan algunas pautas para la selección de los ensayos de aptitud equivalentes u homologables, en los cuales se consideran:

- Diferencias entre el ítem de comparación y las muestras de rutina.
- Tipo de matrices y de analitos.
- Sector de aplicación y técnicas analíticas empleadas.

3.1. Ítem de comparación vs muestras de rutina

Detrás de la medición de uno o varios analitos químicos en una muestra, existe una cadena de eventos como son: la definición del problema, el muestreo, el pretratamiento y el análisis, en cada uno de los cuales existen factores que determinan el resultado de la medición y por tanto, deben evaluarse para determinar la pertinencia en la participación en un EA. Por ejemplo, si uno de los pasos decisivos de un análisis es el pretratamiento de la muestra, como en el caso de la digestión de un cosmético para la determinación de metales por absorción atómica, sería impropio participar de un EA en el que la matriz de la muestra sea acuosa. Tampoco sería procedente participar de un EA de tetraciclinas en miel de abejas si en el laboratorio de ensayo se validó una metodología para medir tetraciclinas en carne. El nivel de concentración del analito también debe tomarse muy en cuenta, por ejemplo, es un desafío diferente el medir colesterol en sangre que medir colesterol en leche bovina.

De lo anterior, es evidente la importancia de la naturaleza química de la matriz y del analito objeto de análisis para verificar la equivalencia. Al seleccionar un ensayo de aptitud, el material de referencia deberá ser similar a las muestras que se trabajan en el laboratorio, puesto que, de esta similitud, se desprenderá que los errores cometidos en el análisis sean comparables y puedan ser detectados en el resultado de la participación del ensayo de aptitud. La similitud del material, entendida como semejanza en estado físico del material, la composición de la matriz, la concentración del analito de interés, la forma como éste se enlaza a la matriz, y de ser viable, los posibles interferentes [5].

Por lo anterior, y con el propósito de considerar algunos de los aspectos anteriormente descritos relacionados con la naturaleza de la muestra, se deben plantear tres preguntas que pueden ayudar a seleccionar un EA que permita evaluar el desempeño de una metodología analítica de un laboratorio, sin que necesariamente el ítem a evaluar sea idéntico a las muestras que el laboratorio analiza rutinariamente

- * ¿El material de referencia ofrecido en el EA se comportará de manera similar que las muestras que se analizan con mi método de medición?
- * ¿Las diferencias entre el material de referencia del EA y las muestras que usualmente analizo influirán significativamente en el resultado de la medición?
- * ¿Puedo evaluar mi método de medición con una muestra diferente, pero garantizando que la diferencia entre el resultado de mi medición y el valor asignado en el EA se atribuye al sesgo de mi laboratorio y no a la diferencia entre la naturaleza de las muestras?.

3.2. Equivalencia en función de la matriz y el analito

Como se mencionó previamente, para establecer equivalencia u homologación entre ensayos de aptitud, es indispensable considerar la propiedad que se requiere medir, así como el material o la matriz sobre la cual se debe realizar esta medición. Para este último caso, y debido a la gran cantidad de matrices posibles, se presenta a continuación una clasificación que puede ser útil [6] para identificar el ensayo de aptitud en función de las muestras que se miden de manera rutinaria:

- Matrices líquidas de baja interferencia como agua potable, superficiales, residuales, bebidas, entre otras.
- Materiales clínicos (suero, orina, plasma, etc.)
- Matrices alimenticias (**Figura 2.4**) Subclasificadas según el triángulo de composición, de acuerdo a su porcentaje de grasas, carbohidratos o proteínas en: > 60% grasa, > 60% proteína, > 60% carbohidratos, ningún componente > 60%) [7]
- Matrices abióticas (suelos, lodos, otros como plásticos, polímeros, cosméticos)

FIGURA 2.4

Triángulo de matrices alimentarias AOAC [25].
Tomado de [8].



En cuanto a la propiedad medida (el analito de interés), la primera clasificación es su naturaleza química: orgánica, inorgánica o mixta. En el caso de los compuestos orgánicos, es importante su tamaño molecular y su polaridad, puesto que son propiedades que determinan diferencias significativas en los procedimientos de preparación de muestra, extracción, limpieza y análisis [24]. En el primer caso, se separan las moléculas de tamaños mayores a 500 uma (macromoléculas) de las de inferior peso molecular. En estas últimas, se pueden diferenciar, además, las polares ($pK_{ow} > -2$) de las no polares ($pK_{ow} < -2$).

3.2.1. Analitos orgánicos

A continuación, se proponen unos ejemplos de análisis de compuestos orgánicos en diferentes matrices y se ubican dentro de una tabla de clasificación (**ver Tabla 2.1**) con los conceptos mencionados anteriormente [6]. La Tabla 2.1, permite ubicar en un paréntesis el nivel de concentración en que se presenta el analito. En principio los análisis que se clasifiquen en la misma celda pueden considerarse equivalentes

- a. *Contenido de vitaminas B6, B3 y B2 en cereales*
- b. *Contenido de α -tocoferol y retinoides en pescado*
- c. *Contenido de α -tocoferol y retinoides en leche en polvo*
- d. *Perfil lipídico en aceites (coco, palma)*
- e. *Acrilamida en panes, panela*
- f. *Metanol en bebidas alcohólicas*
- g. *Ácido benzoico y ácido sórbico en jugos*
- h. *Cloranfenicol en productos cárnicos*
- i. *Cafeína en bebidas*
- j. *Trihalometanos en aguas crudas y tratadas*
- k. *BTEX en aguas*
- l. *PHA en aguas crudas y tratadas*
- m. *PHA en suelos*
- n. *Colesterol en suero sanguíneo*
- o. *Colesterol en lácteos*
- p. *Perfil lipídico en productos cárnicos*
- q. *Colorante verde malaquita en pescado*
- r. *PCB en lodos y suelos*

Tabla 2.1.

Ejemplo de clasificación por matriz para analitos orgánicos de bajo peso molecular (<500 uma).

Analito / Matriz		Matrices líquidas de baja interferencia		Materiales clínicos	Matrices alimenticias				Abióticas
		acuosas	orgánicas		>60% grasa	>60% proteína	>60% carbohidratos	Mixto	
moléculas (<500 uma)	polar ($pK_{ow} > -2$)	f g i j					a e	h	
	no polar ($pK_{ow} < -2$)	k l	q	n	d	b q	c o	p	m

Clasificación de acuerdo al nivel de concentración:

- I. > 1 g/kg; < 1 kg/kg [> %]
- II. > 1 mg/kg; < 1 g/kg [% <, > ppm]
- III. > 1 µg/kg; < 1 mg/kg [ppm <, > ppb]
- IV. < 1 µg/kg [ppb <]

3.2.2. Analitos inorgánicos

Entre los analitos inorgánicos, existen sustancias elementales y compuestos de naturaleza mixta como los organometálicos. En el caso de sustancias elementales, es necesario tomar en cuenta su ubicación en la tabla periódica para entender sus propiedades químicas y, por tanto, establecer similitudes en los análisis, adicionalmente los desafíos que ofrecen las interferencias son diferentes a los que se presentan en las mediciones de analitos orgánicos. **Por ejemplo**, si son matrices con altos contenidos de sílice, sales o compuestos orgánicos, así como, si son metales cuya disolución sea difícil. A continuación, algunos ejemplos y una tabla propuesta (tabla 2.2) para su clasificación según su grupo y los tipos de matriz [9].

- a. Elementos en agua de mar (Pb, Hg, Ni, Zn, As, Cd, Cu)
- b. Elementos traza en agua potable (As, B, Cd, Ca, Cr, Hg, Mo)
- c. Calcio en suero humano
- d. Selenio en harina de trigo
- e. Impurezas en alúmina en polvo (Fe, Mg, Si)
- f. Mercurio total en suelo
- g. Calcio en arcillas
- h. Halógenos en productos domésticos
- i. Nitrógeno en leche en polvo
- j. Elementos traza en biodiesel (Na, Ca, K, Mg, P)

Tabla 2.2.

Ejemplo de clasificación por matriz para analitos inorgánicos.

Analito	Acuosas	Alto contenido de sílice (suelos, sedimentos, plantas)	Alto contenido de sales (Agua de mar, orina)	Alto contenido de orgánicos (alimentos, sangre, cosméticos)	Metales difíciles de disolver	Matrices altamente volátiles
Grupo I (alcalinos y alcalinotérreos) Li, Na, K, Ca, Mg, Be, Rb, Sr, Ba	b	g			e	j
Elementos de transición Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Y, Zr, Nb, Mo, Tc, Ag, Cd, Ta, W, Au, Hg, Al, Ga, In, Tl, Pb, Po	b	f	a		e	
Elementos del grupo del platino Ru, Rh, Pd, Os, Ir, Pt						
Metaloides, semimetales B, Si, Ge, As, Sb, Te, Se	b		a	d	e	
No metales P, N, S, C, O				i		j
Halógenos (F, Cl, Br, I)	h					
Tierras raras (lantánidos y actínidos)						

En los anteriores ejemplos de clasificación de análisis orgánicos e inorgánicos en diferentes matrices puede notarse que los que se encuentren clasificados en la misma celda y tienen un nivel de concentración en el mismo intervalo se consideran equivalentes, sin embargo, es importante tener en cuenta que, aunque esto se cumpla, es la información de la validación de los métodos la que al final indica que esta equivalencia existe. Por ejemplo, si los analitos son comparables y la matriz también, pero el comportamiento químico de los analitos es diferente frente a las variaciones del método, probablemente no sean equivalentes.

3.3. Sector de aplicación y técnicas analíticas empleadas

Dentro de las mediciones químicas cuantitativas, se pueden identificar tres principales componentes:

- (i) La técnica de medición (p.e. Espectrofotometría, titulación potenciométrica).
- (ii) La propiedad medida (p.e. Cadmio, acrilamida).
- (iii) La matriz o el producto analizado (p.e. Suelo, agua potable, maíz etc.).

El Organismo Nacional de Acreditación de Colombia (ONAC) tiene agrupados los diferentes tipos de análisis químicos por sectores generales y específicos, y dentro de ellos por familias de técnicas, propiedad medida y el tipo de producto donde se hace la medición [10], de tal manera que, por ejemplo: en el sector específico de aguas y sedimentos en el sector específico de aguas y sedimentos, en el sector general de análisis fisicoquímicos, se encuentran todos los analitos que pueden analizarse por medio de los diferentes tipos de cromatografía (líquida, gaseosa, iónica etc.) como son: órgano clorados, órgano fosforados, PCBs, fenoles, hidrocarburos aromáticos policíclicos, compuestos orgánicos volátiles, BTEX, aniones y carbamatos. Esta clasificación es útil para definir los grupos de mediciones que se realizan en el laboratorio de ensayo.

Las técnicas de medición que se emplean para el análisis de una propiedad química (analito) por lo general exigen unos requisitos en la preparación de las muestras (previo a su análisis), y para cumplirlos, se debe tener en cuenta la matriz de la muestra, entre más compleja sea, es posible

que se deban tener más pasos de extracción, limpieza por técnicas como extracción en fase sólida (EFS), micro EFS, líquido-líquido, sólido-líquido, digestión, espacio de cabeza, concentración, dilución, entre otras. En la **Figura 2.5**, se exponen dos ejemplos una para cromatografía líquida y otra para cromatografía de gases en los cuales se evidencia que la técnica de medición y la matriz de la muestra definen en gran medida la metodología de análisis.

FIGURA 2.5

Ejemplos de la influencia de las técnicas analíticas sobre los procesos de medición.

Vitamina D en matrices grasas

Para analizar un compuesto como vitamina D en una matriz altamente grasa, como el aguacate o como una mantequilla, se debe retirar la mayor cantidad de grasa posible antes de poder cuantificar la vitamina mediante cromatografía líquida de alta eficiencia en fase reversa.

Residuos Plaguicidas en alimentos

En la medición de residuos de plaguicidas por cromatografía de gases, la muestra a inyectar, idealmente, no debe contener compuestos no volátiles, termosensibles o de alto peso molecular, porque estos pueden contaminar el sistema, de tal forma que el pretratamiento de la muestra usualmente involucra procesos de limpieza como cromatografía en columna, extracción en fase sólida, entre otros métodos.

El primer reto, es entonces, agrupar los análisis que se realizan en el laboratorio de ensayo con el fin de determinar cuáles son equivalentes u homologables entre sí, y formar grupos, lo que será muy útil puesto que el laboratorio de ensayo deberá buscar un EA para cada grupo y no para cada análisis que realiza, de tal forma que los resultados de la participación en el EA serán extensibles a los análisis que pertenezcan al grupo. El organismo de acreditación europeo sugiere definir estos grupos de análisis y los llama *subdisciplinas*¹.

¹Para los fines de este documento el término subdisciplinas es equivalente a grupos de análisis.

Lo primero que se debe tener en cuenta para definir las subdisciplinas es que los análisis que se realicen con el mismo método normalizado o estándar (ISO, AOAC, EN, DIN, entre otros) pertenecen a la misma subdisciplina. Por su parte, cuando se tienen métodos de análisis desarrollados y validados en el laboratorio habrá que revisar detenidamente las particularidades químicas de la matriz, del analito, su intervalo de concentración y el desempeño del método como tal.

Para aclarar el panorama se presentará un ejemplo tomado del organismo de acreditación europea que pretende facilitar el proceso de definición de subdisciplinas dentro de un laboratorio de ensayo químico [11].

El laboratorio de ensayo Chemassay, realiza análisis ambientales y desea acreditar las siguientes mediciones:

- * Bifenilos policlorados (PCB) por GC-MS en suelos y lodos
- * Hidrocarburos poliaromáticos (PHA) por GC-MS en suelos y lodos
- * Compuestos Orgánicos volátiles (VOC) por Purga y trampa- GC-MS en aguas
- * Elementos tóxicos por ICP-MS en suelos, lodos y aguas
- * pH en suelos, lodos y aguas

Para definir si existen subdisciplinas, lo primero es precisar si se emplea un método normalizado para la medición. En este caso, para la medición de pH, a pesar de que se mide en matrices particularmente diferentes, se emplea el mismo método ISO, por lo tanto, estas 3 mediciones se consideran en la misma subdisciplina, las demás mediciones que lleva a cabo el laboratorio se realizan mediante un método in house.

Compuestos volátiles orgánicos (VOC) en suelos y lodos

Las mediciones de VOC se realizan en una misma matriz, agua, y se emplea el mismo método validado in house de purga y trampa. Sin embargo, en este caso es fundamental observar el comportamiento de los diferentes analitos a medir al realizar la validación del método, si estos se comportan de manera similar ante variaciones en la metodología podría considerarse que son una misma subdisciplina, pero si variaciones en el análisis afectan de manera ostensible el resultado de la medición de uno(s) pero no de otro(s), no deberían estar en el mismo grupo de análisis, a menos que los parámetros que afectan la medición estén muy bien identificados y controlados por el laboratorio.

En este caso, supongamos que el laboratorio define que los diferentes VOC que analiza se comportan similarmente y que los parámetros que afectan la medición se encuentran muy bien controlados, por tanto, decide que están en la misma subdisciplina

PCB y PHA en suelos y lodos

El laboratorio los mide por la misma técnica, usando un método in house y en las mismas matrices, sin embargo, después de evaluar la validación, el laboratorio establece que definitivamente PCB y PAH se ven afectados de diferentes maneras por variaciones en la metodología y, por lo tanto, un desempeño aceptable en un EA para la medición de PCB no implica lo mismo para PAH y viceversa. Por lo tanto, el laboratorio decide no unirlos en la misma subdisciplina

Elementos tóxicos por ICP-MS

Debido a que la preparación de la muestra de agua es claramente diferente de la empleada para suelos y lodos, como se evidenció en la tabla de clasificación de matrices para analitos inorgánicos (Tabla 2. 2) y, debido a la implicación de esto en el análisis se definen dos subdisciplinas, una para aguas y otra para suelos y lodos.

De este ejercicio el laboratorio del ejemplo definió 6 subdisciplinas con lo que reduce eficientemente las mediciones para las que debe buscar ensayos de aptitud y demostrar la validez de todas sus mediciones a través de éstas.

3.4. Recomendaciones de la guía Eurachem

Una herramienta útil que presenta la guía Eurachem [12] para la selección, uso e interpretación de ensayos de aptitud, se presenta a través de una tabla ejemplo (Tabla 2.3), en la que es posible evaluar la relevancia de los diferentes aspectos y así valorar, si un EA es relevante para el laboratorio de ensayo. Para este caso se presenta el ejemplo de un laboratorio que analiza algunos metales en comida para gatos y pretende participar en un ensayo de aptitud de metales en trazas en comida para animales.

La elaboración y uso de esta tabla puede realizarse para una metodología de cada grupo de análisis subdisciplina que se quiera acreditar, aprovechando los conceptos de equivalencia en temas de matriz, analitos y técnicas de medición que se expusieron con anterioridad.

Tabla 2.3.

Matriz para la evaluación de la relevancia de la participación en un Ensayo de Aptitud con base en la equivalencia del IEA.

Analisis	Elementos traza en comida para animales		Referencia cruzada con los numerales del capítulo
Ensayo de aptitud	Metales en comida para animales (organiza; UP, Paises bajos)		
Situación en el laboratorio	Situación en el EA	Son las diferencias aceptables?	
Compuesto(s) analizado(s) usualmente	Compuesto(s) ofrecido(s) en el item	Sí	3.2
Cu, Ag, Hg	Ag, Al, Mg, Hg, Zn		
Matriz de trabajo usual	Matriz del item de ensayo	Sí	3.1 y 3.2
Comida para gatos	Comida para animales		
Intervalo de concentración usual	Intervalo de concentración del item	Sí	3.2
0-200 mg/kg	50-500 mg/kg		
Método empleado	Método estipulado (si aplica)	Sí	3.3
Solubilización ácida de cenizas	Ninguno		
Tipo de laboratorio	Tipo de laboratorios a los que se dirige el EA	Sí	2
Producto de comida para animales	Todo tipo de laboratorios de ensayo nacionales		
Se pueden adicionar otros aspectos que el laboratorio considere			1.5
Conclusión final: ¿EA es relevante?		Sí / No	

4. Acerca del proceso de evaluación: indicador de desempeño.

4.1. Consideraciones acerca de los valores de referencia

Como se mencionó en la sección 3.3 del capítulo I, los programas de EA pueden clasificarse según como se obtenga el valor de la propiedad asignada, también conocido como PTRV por sus siglas en inglés (*Proficiency Test Reference Value*), al Ítem de Ensayo de Aptitud (IEA), este valor es muy importante porque con él se realiza el primer cálculo encaminado a obtener la calificación del desempeño, el sesgo. El sesgo se define como:

Ecuación 1

$$\text{Sesgo} = x - X$$

Donde x es el valor del resultado obtenido y X es el valor de referencia o X_{pt} (PTRV), el cual se puede estimar de diferentes maneras, de acuerdo con lo mostrado en la sección 3.3 del Capítulo I.

Dentro de las diferentes modalidades, la asignación del valor por consenso es una de las prácticas más empleadas por los proveedores de ensayos de aptitud, tal vez debido a la dificultad de aplicar las otras dos modalidades (ver sección 3.3, Capítulo I). La asignación del valor por consenso resulta ser la más económica, sin embargo, tiene la desventaja de que la calidad del PTRV dependerá de la calidad de las mediciones, la experiencia de los laboratorios, las técnicas analíticas empleadas, entre otros factores que pueden afectar el PTRV, lo cual implica que el desempeño de los laboratorios dependerá de la calidad de diversos factores [4]. En este sentido, el laboratorio participante solo podrá obtener una idea de cómo está su medición con respecto a la de los demás laboratorios, pero no tiene información acerca del sesgo de su medición. Es importante tener clara la modalidad de asignación del PTRV (X_{pt}) al IEA para la elección de un EA, para mayor detalle revisar sección 3.3 del capítulo I.

Por otro lado, la participación frecuente en EA empleando un PTRV asignado por un laboratorio de referencia o una red de laboratorios de referencia, permite, no solo que un laboratorio pruebe su desempeño mediante una referencia externa apropiada, sino la revisión de este desempeño a través del tiempo, lo que hace posible la evaluación de los efectos que han tenido los cambios ocurridos en el laboratorio y que puedan afectar la metodología, por ejemplo: personal, equipos, consumibles, reactivos etc. También permite la comparación de su desempeño con sus pares en un tiempo determinado, esto siempre y cuando el σ_{pt} empleado para el cálculo del z' permanezca constante.

4.2. Consideraciones acerca del z_i score

Entre los indicadores más empleados, se destaca el z score (ecuación 2), que es la relación entre el sesgo en la medición del laboratorio (ecuación 1) y el valor de la desviación estándar objetivo de los resultados σ_{pt} . El z -score implica un resultado cuestionable por fuera del intervalo de [2] y demanda una acción si se encuentra por fuera del intervalo [3]. Este indicador asume que la incertidumbre del valor asignado x_{pt} es insignificante comparada con la incertidumbre del resultado del laboratorio participante.

Ecuación 2

$$z_i = \left(\frac{x_i - x_{pt}}{\sigma_{pt}} \right)$$

Como puede observarse, la incertidumbre de medición de los laboratorios no se tiene en cuenta en este indicador, lo que representa una limitación de su alcance, adicionalmente a que, como se mencionó en el numeral 2 del presente capítulo, su comparabilidad entre rondas y entre ejercicios diferentes de EA, dependerá de los estimadores empleados, el valor asignado al IEA, x_{pt} , y la desviación estándar objetivo, σ_{pt} . La dependencia del Z score de estos estimadores hace que en algunos casos no sea tan representativo de buen desempeño como en otros, p.ej. una σ_{pt} laxa ocultaría fácilmente un desempeño pobre. En otras palabras, el indicador es muy proclive a fallas desde su definición, por lo que el participante podría estar

sobreestimando o subestimando su desempeño real, e incluso confrontando resultados no comparables entre rondas, si no tiene claridad acerca de los estimadores empleados por el proveedor del EA. El indicador z-score además permite que, con una incertidumbre del laboratorio del doble de la σ_{pt} , el laboratorio aún tenga una probabilidad del 67% de estar en el intervalo de $|2|$ y un 83% de probabilidad de estar en ese mismo intervalo con un sesgo equivalente a σ_{pt} , lo que hace dudar de su capacidad para demostrar competencia. En todo caso, se reconoce que el z score es un buen indicador para la comparación del desempeño entre laboratorios en una misma ronda o incluso en el tiempo, si se mantienen sus estimadores [13-15].

La interpretación de los z score debe hacerse con extrema precaución puesto que debe evitarse la sobreestimación de su alcance. Por ejemplo, si el valor de referencia es asignado por un consenso en el ensayo de aptitud, un laboratorio que obtiene en una ronda un valor de 0.4 y en la siguiente un valor de 1.6, no debe asumir que su desempeño ha empeorado, esto puede ocasionarse por la variación natural, por lo tanto, tampoco es posible juzgar los laboratorios entre sí por sus resultados obtenidos dentro del intervalo de aceptación [16].

El z' score (ecuación 3) se recomienda cuando la incertidumbre del valor asignado al IEA, $u^2(x_{pt})$, es significativa para afectar considerablemente el indicador, esta situación se considera cuando $u^2(x_{pt}) < 0.3 \sigma_{pt}^2$. Sin embargo, el indicador será incapaz de diferenciar un desempeño deficiente de un valor asignado al IEA que no satisface las condiciones por lo que debe evaluarse con precaución [13].

Ecuación 3

$$z'_i = \left(\frac{x_i - x_{pt}}{\sqrt{\sigma_{pt}^2 + u^2(x_{pt})}} \right)$$

El indicador zeta (zeta-score) se calcula mediante la ecuación 4, para emplearlo se requiere que los laboratorios informen la incertidumbre de su resultado y que el IEA sea un material de referencia certificado. Como puede observarse, el indicador aumenta cuando el numerador (sesgo del participante) es muy grande o, cuando la incertidumbre estimada por el laboratorio es muy pequeña, $u^2(x_i)$, así, la subestimación de ésta es pena-

lizada, sin embargo, el laboratorio puede manipular el indicador sobreestimando la incertidumbre lo que afectaría la objetividad del indicador.

Ecuación 4

$$\zeta_i = \left(\frac{x_i - x_{pt}}{\sqrt{u^2(x_i) + u^2(x_{pt})}} \right)$$

El estimador E_n o error normalizado, se obtiene mediante la ecuación 5 y emplea la misma ecuación que el zeta score con la diferencia de que emplea las incertidumbres expandidas.

Ecuación 5

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

En general, aunque los indicadores zeta no incluyen la incertidumbre de los laboratorios y por tanto no pueden considerarse perfectos, los indicadores que emplean las incertidumbres reportadas por los laboratorios tampoco han sido exitosos, debido principalmente a los problemas asociados a su correcta estimación, en algunos casos subestimada y en otros sobreestimada [17].

Por otro lado, tanto z_i como z_i' son indicadores de desempeño que dependen del σ_{pt} , el cual es un valor que representa la desviación estándar del EA y puede obtenerse de las siguientes formas:

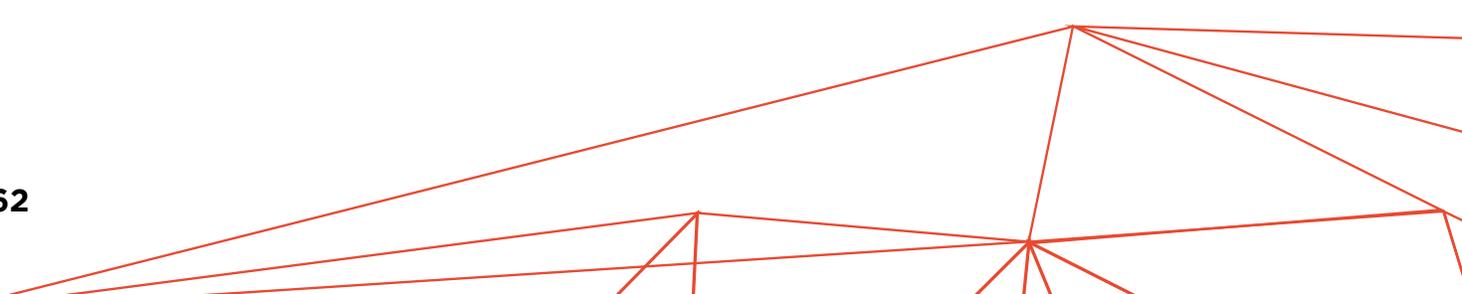
- a. Percepción, se define considerando cómo debería ser el desempeño esperado de los laboratorios, es arbitrario.
- b. De la variación actual en una ronda de participación después del retiro de datos anómalos o mediante el uso de métodos robustos,
- c. La σ_R del método normalizado
- d. Un modelo de precisión esperada (p.e. Horwitz).
- e. Precisión requerida definida por el propósito de la medición, conocida como σ_p .
- f. Regulación.

La manera como se obtiene el σ_{pt} es determinante en la utilidad del indicador de desempeño que lo emplea. Las opciones a y b son susceptibles de cambios en el tiempo y por ende los resultados no serán comparables de ronda a ronda si este estimador ha cambiado. Adicionalmente, cuando se usa la variación de la ronda actual para la obtención del σ_{pt} (opción b), existe un 95% de probabilidad de que un laboratorio presente un resultado satisfactorio, por la misma definición estadística, por lo que solo es posible la comparación del desempeño de los laboratorios entre sí y, al no existir una referencia externa válida, tampoco se puede inferir si el desempeño es realmente útil para el propósito [18].

La opción de obtener el estimador según un modelo (opción d) presenta el inconveniente de que seguir un modelo p.e. Horwitz, puede no ser indicado en todos los casos, por ejemplo, está demostrado que cuando se trabajan analitos a nivel de trazas, por debajo de 10 ug kg⁻¹ en fracción mássica, la función predice precisiones Inter laboratorio tan bajas (RSD=32%) que en la práctica podría estar en el límite de detección y se podría dudar acerca de la presencia o ausencia del analito [19].

El estimador actualmente reconocido como el más adecuado [20] se define mediante el propósito de la medición (opción e), definiendo el nivel de incertidumbre que se acepta como apropiado según el sector donde se emplean los resultados, de tal forma que la incertidumbre sea suficientemente pequeña para garantizar que las decisiones tomadas con base en los resultados sean raramente incorrectas, pero no tan pequeña que aumente innecesariamente el costo de los análisis [21] en todo caso, debe entenderse que este estimador no representa el desempeño de los laboratorios sino el desempeño que deberían tener para ofrecer los servicios de medición a sus clientes con confianza.

La IUPAC recomienda el uso de z scores siempre aplicando el criterio de aptitud para el propósito a la hora de elegir el σ_{pt} , de tal forma que un laboratorio puede conocer como esta su desempeño tomando en cuenta las necesidades del sector donde se aplica la medición tal como se discutió en la sección anterior.



5. Frecuencia de participación

Por último, la frecuencia de la participación deberá ajustarse a los lineamientos de las organismos de acreditación y a la pertinencia de estas participaciones para el laboratorio, lo cual dependerá del número de servicios que preste, los resultados previamente obtenidos, la disponibilidad del EA, los costos, entre otros factores. La frecuencia de participación también se establece según las herramientas de aseguramiento de calidad que tenga, pero sobre todo las destinadas a establecer la ausencia de sesgo en las mediciones, por ejemplo: uso de MRC, empleo de diferentes técnicas para establecer el valor de la misma propiedad, análisis de muestras ciegas externas, entre otros.

La frecuencia de participación corresponde a que tan seguido un laboratorio deberá participar en un EA para un determinado grupo de análisis (ver numeral 3.3). Dependiendo del sector donde se realicen los ensayos, la frecuencia en la participación varía ostensiblemente, ILAC recomienda una participación para la obtención de la acreditación y de ahí en adelante [22], una al menos cada 4 años, no obstante, en muchos casos es evidente la necesidad de una frecuencia de participación mucho mayor, de incluso varias veces al año. La determinación de la frecuencia de participación deberá ser establecida por el laboratorio, convertirse en parte fundamental de la política de calidad del laboratorio y justificarse ante el ente acreditador. Esta frecuencia puede ser determinada con base en un análisis detallado de algunos factores que se esquematizan en la **Figura 2.6** [21,23-25].



FIGURA 2.6

Algunos aspectos útiles en la evaluación para el establecimiento de la frecuencia de participación en EA por laboratorios de ensayo.

6. Disponibilidad de los ensayos de aptitud

Aunque la disponibilidad de EA se ha incrementado con el tiempo, es claro que hay un déficit de proveedores de estos servicios, sobre todo en países en vías en desarrollo, siendo claro además, que existen necesidades para el análisis de productos alimentarios particulares de determinadas regiones, que los proveedores internacionales no pueden suplir [31]. La siguiente tabla, tiene como objetivo facilitar la búsqueda por parte de los laboratorios de sus ensayos de aptitud.

Tabla 2.4

Algunas fuentes para la búsqueda de ensayos de aptitud

¿Quiénes?	¿Dónde buscar?	Recomendación
Proveedores de Ensayos de Aptitud acreditados por ONAC	https://onac.org.co/	Buscar dentro del directorio de organismos evaluadores de la conformidad acreditados, en el esquema de Proveedores de Ensayos de Aptitud - PEA
Organismos de acreditación que acreditan ensayos de aptitud	En la página web de cada organismo	ONAC tiene una compilación de estos organismos, la cual está disponible aquí. https://onac.org.co/
Instituto Nacional de Metrología de Colombia	www.inm.gov.co	Buscar dentro del servicio de ensayos de aptitud.
Institutos Nacionales de Metrología de otros países	https://www.bipm.org/kcdb/	En la página web del BIPM buscar institutos que realicen diseminación de la trazabilidad metrológica a través de ensayos de aptitud ONAC tiene una lista de algunos Institutos Nacionales de Metrología que son proveedores de ensayos de aptitud, la cual puede ser consultada aquí. https://onac.org.co/
Proveedores internacionales	https://www.eptis.org/	Es una base de datos internacional que contiene información acerca de los programas de ensayos de aptitud. Sin embargo, es importante confirmar en la página del organismo acreditador que el ensayo de aptitud se encuentra acreditado o dentro del alcance de la acreditación

En el caso puntual de Colombia, el Decreto Único Reglamentario del Sector Comercio, Industria y Turismo – DURCIT (Decreto No. 1074 de 2015) emitido por el Ministerio de Comercio, Industria y Turismo, establece quienes son los proveedores de los servicios de ensayos de aptitud/comparación interlaboratorios, los cuales corresponden a:

- ☑ Los institutos nacionales de metrología, como laboratorios primarios de otros países que sean firmantes del Acuerdo de Reconocimiento Mutuo MRA- en el ámbito del Comité Internacional de Pesas y Medidas – CIPM- de la Oficina Internacional de Pesas y Medidas – BIPM.
- ☑ Los organismos proveedores legalmente constituidos y que demuestren su competencia técnica mediante un certificado de acreditación vigente con la norma ISO/IEC 17043 (NTC-ISO/IEC 17043) o la que la modifique, sustituya o adicione y que su alcance cubra el servicio ofrecido.
- ☑ Las organizaciones internacionales reconocidas de desarrollo de estándares internacionales que ofrezcan servicios de ensayos de aptitud/comparación interlaboratorios y las organizaciones que ofrezcan servicios de ensayos de aptitud/comparación interlaboratorios aceptadas por el Organismo Nacional de Acreditación de Colombia, siempre y cuando no exista ningún proveedor de ensayos de aptitud/comparación interlaboratorios acreditado con la norma ISO/IEC 17043 (NTC-ISO/IEC 17043) a nivel nacional o internacional que su alcance cubra el servicio requerido.
- ☑ El Instituto Nacional de Metrología de Colombia-INM.

Cuando definitivamente no es posible encontrar un esquema relevante puede optarse por participar en alguno que permita evaluar las técnicas analíticas o parte de la metodología, siempre justificando esta alternativa [10].

El Organismo Nacional de Acreditación-ONAC, ha establecido una política para la participación en ensayos de aptitud en laboratorios, la cual tiene como objetivo establecer los criterios que deben cumplir los laboratorios de ensayo, laboratorios clínicos y laboratorios de calibración, con respecto a la participación y resultados de los ensayos de aptitud. Esta política da cumplimiento al “Acuerdo de Reconocimiento Mutuo de la Cooperación Internacional de Acreditación de Laboratorios (ILAC MRA), y se constituye en una parte fundamental para asegurar y mantener la condición de signatario de ONAC”.

Para mayor información acerca de esta política, visite <https://onac.org.co/>

Referencias bibliográficas

- [1] EEE-PT WG. Eurachem guide: selection, use and interpretation of proficiency testing (PT) schemes [Internet]. 2nd ed. Mann, Ian; Brookman B, editor. 2011. Available from: https://www.eurachem.org/images/stories/Guides/pdf/Eurachem_PT_Guide_2011.pdf
- [2] Eurachem Nederland PTS group. Eurachem guide: Selection, use and interpretation of proficiency testing (PT) schemes by laboratories-2000. 1.0. Boley N, editor. 2000.
- [3] Sorbo A, Ciprotti M, Colabucchi A, D'Amato M, Gregorio M Di, Luswergh GF, et al. Proficiency testing as a tool to assess quality of data: the experience of the EU Reference Laboratory for chemical elements in food of animal origin. Pure Appl Chem [Internet]. 2020;92(3):383–90. Available from: <https://www.degruyter.com/view/journals/pac/92/3/article-p383.xml>
- [4] RSC Analytical Methods Committee. Proficiency testing of analytical laboratories: organization and statistical assessment. Analyst [Internet]. 1992;117(1):97–104. Available from: <http://dx.doi.org/10.1039/AN9921700097>
- [5] Maier EA, Quevauviller P, Griepink B. Interlaboratory studies as a tool for many purposes: proficiency testing, learning exercises, quality control and certification of matrix materials. Anal Chim Acta [Internet]. 1993;283(1):590–9. Available from: <http://www.sciencedirect.com/science/article/pii/000326709385272L>
- [6] Document G. CCQM ORGANIC ANALYSIS WORKING GROUP (OAWG) Practices and Guidelines Document Contents. 2018;1–26.
- [7] Wolf WR, Andrews KW. A system for defining reference materials applicable to all food

matrices. Fresenius J Anal Chem. 1995;352(1-2):73-6.

[8] Measurements and Standards to Support Nutrition Labeling _ NIST [Internet]. Available from: <https://www.nist.gov/programs-projects/measurements-and-standards-support-nutrition-labeling>

[9] Fisicaro P, Infante HG. The new IAWG CC approach : objectives. 2019.

[10] Onac, Política para la participación en ensayos de aptitud (EA) en Laboratorios, CEA-3.0-04 versión 03.

[11] European Accreditation Commity. EA-4 / 18 INF : 2010 Guidance on the level and frequency of proficiency expression testing participation. 2010;(June):15.

[12] Eurachem's Proficiency Testing Working Group. Proficiency testing – How much and how often ? 1st ed. 2016. 1-2 p.

[13] Kisets D. Performance indication improvement for a proficiency testing. Accred Qual Assur [Internet]. 2006;10(9):461-5. Available from: <https://doi.org/10.1007/s00769-005-0021-8>

[14] Hund E, Massart DL, Smeyers-Verbeke J. Inter-laboratory studies in analytical chemistry. Anal Chim Acta [Internet]. 2000;423(2):145-65. Available from: <http://www.sciencedirect.com/science/article/pii/S0003267000011156>

[15] Committee AM, No A. Z -Scores and other scores in chemical proficiency testing - Their meanings, and some common misconceptions. Anal Methods [Internet]. 2016;8(28):5553-5. Available from: <http://dx.doi.org/10.1039/C6AY90078J>

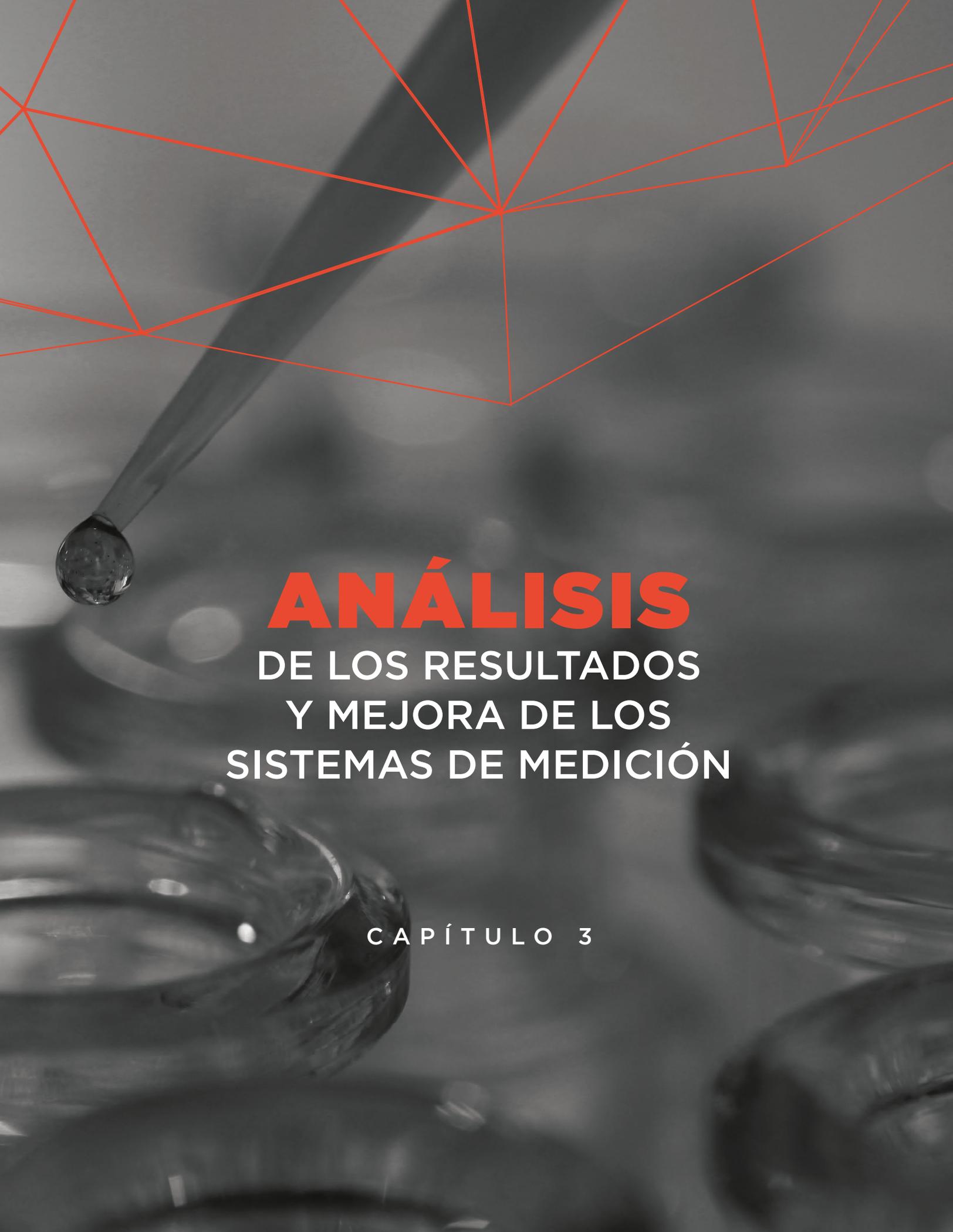
[16] Analytical Methods Committee. What is proficiency testing? Guide for end-users of chemical data. R Soc Chem. 2010;7(2):1-2.

[17] Mäkinen I. Use of uncertainty estimates as reported by participants in proficiency testing for the evaluation of their results: pros and cons. Accred Qual Assur [Internet]. 2009;14(1):35-41. Available from: <https://doi.org/10.1007/s00769-008-0462-y>

[18] Analytical Methods Committee. Understanding and acting on scores obtained in proficiency testing schemes. Anal Methods Comm Tech Br. 2002;(11):4 pp.

- [19] Thompson M. Recent trends in inter-laboratory precision at ppb and sub-ppb concentrations in relation to fitness for purpose criteria in proficiency testing. 2000;385–6.
- [20] Thompson M, Ellison S, Wood R. The International Harmonized Protocol for the Proficiency Testing of Analytical Chemistry Laboratories: (IUPAC Technical Report). Pure Appl Chem - PURE APPL CHEM. 2006 Jan 31;78:145–96.
- [21] Fearn T, Fisher SA, Thompson M, Ellison SLR. A decision theory approach to fitness for purpose in analytical measurement. Analyst [Internet]. 2002;127(6):818–24. Available from: <http://dx.doi.org/10.1039/B111465D>
- [22] Örnemark U, Fostel H, Straub R, van de Kreeke J. Policies, requirements and surveys concerning frequency for participation in proficiency testing schemes. Accred Qual Assur [Internet]. 2004;9(11):729–32. Available from: <https://doi.org/10.1007/s00769-004-0858-2>
- [23] Brookman B, Butler O, Ciaralli L, Horsky M, Laurent C, Noblett T, et al. Proficiency testing in analytical chemistry, microbiology and laboratory medicine: discussions on current practice and future directions. Accred Qual Assur [Internet]. 2019;24(1):93–101. Available from: <https://doi.org/10.1007/s00769-018-1343-7>
- [24] Brookman B, Evans S, Ellison S, Boley N, Schmidt A, Berkel E Van, et al. Proficiency testing in analytical chemistry, microbiology and laboratory medicine : working group discussions on current practice and future directions. Accred Qual Assur. 2004;635–41





ANÁLISIS

DE LOS RESULTADOS
Y MEJORA DE LOS
SISTEMAS DE MEDICIÓN

CAPÍTULO 3

Análisis de los resultados y mejora de los sistemas de medición

1. Introducción

La información de los resultados obtenidos en la participación de un EA proporciona a los laboratorios una confirmación del desempeño satisfactorio o una indicación de que hay oportunidades de mejora en las que se deben trabajar. Además, proporciona información sobre aspectos del sistema de gestión de calidad, como recepción, almacenamiento, tratamiento de la muestra, el tratamiento de los datos, el informe de resultados, entre otros.

Dada la importancia de los resultados de los EA, el presente capítulo tiene por objetivo guiar a los laboratorios en la interpretación y análisis de los mismos, así como la formulación de acciones relacionadas con un resultado no satisfactorio o un bajo desempeño. Para ello, en las secciones del capítulo se abordan temas asociados con la revisión e interpretación de los resultados en los diferentes esquemas de EA, se muestra cómo abordar una investigación para establecer la causa de resultados no satisfactorios o cuestionables, indicando las posibles fuentes de error que se pueden presentar, y facilitar la identificación de las actividades de mejora. Finalmente, se presenta el análisis de los resultados derivados de múltiples EA.

2. Revisión de los resultados e interpretación

La participación en un EA tiene un aporte limitado salvo que el laboratorio aproveche su evaluación de desempeño y la información que se propor-

ciona en el informe. Es importante que el laboratorio no solo reconozca el desempeño obtenido, sino que evalúe e interprete los resultados, evitando interpretaciones erróneas o excesivas. Esta evaluación debe hacerse después de cada ronda y, para esquemas continuos, a lo largo del tiempo. Existen algunos puntos básicos sobre la interpretación de los resultados de los EA, ya que estos ensayos no solo se tratan de “aprobar” o “fallar” una medición; sino de aprender de los resultados. Cuando un laboratorio recibe un informe de EA, primero observa los indicadores de desempeño calculados por el proveedor del EA, con respecto a su laboratorio. Cuando se encuentra un mal resultado y se corrobora que el dato del valor reportado por parte del laboratorio es correcto, la justificación habitual es que el bajo desempeño ha sido un incidente, posiblemente un error humano, o no se puede identificar. Como resultado, no se toman más medidas, rechazando así una oportunidad para mejorar la calidad de los resultados del laboratorio.

La información del desempeño debe analizarse con cuidado. Se puede dar el caso de un resultado satisfactorio en una ronda para un laboratorio, donde todos los participantes tienen un desempeño satisfactorio, pero la desviación estándar para la evaluación de aptitud podría ser demasiado grande, y el resultado no necesariamente indica un alto nivel de competencia por parte del laboratorio. Por otro lado, un desempeño no satisfactorio en una ronda, no necesariamente indica falta de competencia, ya que la evaluación del desempeño del participante además de su competencia técnica puede estar asociada a los estadísticos de desempeño seleccionados por el proveedor del EA, las características del ítem de EA, entre otros. No obstante, el bajo desempeño constante indica problemas importantes con el proceso de medición del laboratorio y cuando esto ocurre, se deben tomar acciones para evitar que esta situación se continúe presentando.

La interpretación del desempeño en los EA concierne a todos los niveles de gestión del laboratorio, desde el operador hasta la alta dirección. El personal responsable de la medición puede estar familiarizado con el funcionamiento del esquema del EA y normalmente debería proceder con la evaluación inicial de los resultados. Es posible que la alta gerencia no siempre esté familiarizada con el desempeño del EA, y es recomendable que obtengan un nivel adecuado de comprensión para la toma de decisiones.

2.1. Revisión de los resultados por ronda de EA

Los resultados de cada ronda de un EA deben evaluarse utilizando toda la información disponible en el informe, independientemente del desempeño obtenido, ya que un resultado satisfactorio no necesariamente significa un buen desempeño. Además, el desempeño no satisfactorio en el contexto de una ronda donde la mayoría de los participantes se desempeñó a un nivel satisfactorio debe contrastarse con el desempeño no satisfactorio donde un número significativo de participantes tuvo un desempeño no satisfactorio. Ambas situaciones deben considerarse, ya que estas indican problemas relacionados con el proceso de medición.

Como parte de la revisión de los resultados, el personal del laboratorio siempre debe verificar que los datos usados en el informe del EA sean los presentados por el laboratorio y, en particular, que el sistema de calificación de desempeño utilizado en el esquema es adecuado para el propósito establecido. Si es necesario, se debe contactar al proveedor del ensayo de aptitud para evitar cualquier interpretación errónea del desempeño.

Si después de un análisis exhaustivo, el laboratorio concluye que el resultado es realmente no satisfactorio, entonces se debe realizar la investigación para determinar las causas y si es necesario implementar acciones correctivas.

Los resultados de la participación en el ensayo de aptitud también se pueden utilizar para verificar la validez de la incertidumbre de medición del laboratorio. Además, los resultados pueden ser útiles para declarar la incertidumbre del laboratorio para un ensayo en particular en relación con sus capacidades de medición y calibración.

2.2. Monitoreo del desempeño de los EA a lo largo del tiempo

Después de una evaluación de los resultados de una sola ronda, se debe hacer un seguimiento del desempeño del EA a lo largo del tiempo, la base de la evaluación debe ser consistente de una ronda a otra, de modo que los puntajes en rondas sucesivas sean comparables; solo de esta manera puede un participante ver tendencias a largo plazo en su desempeño e identificar posibles problemas relacionados con la falta de precisión, el error sistemático o el error humano.

Para monitorear el desempeño del laboratorio a lo largo de los EA, es muy

útil construir una gráfica que relacione los indicadores de desempeño y las rondas. A menudo, esto lo proporciona el proveedor de EA en el informe, o el participante puede hacerlo. Este enfoque permite resaltar resultados inusuales o inesperados, además de ayudar a identificar tendencias. Normalmente, se esperaría que los procedimientos de control de calidad interno del laboratorio identifiquen tendencias asociadas con la medición, calibración instrumental, mantenimiento instrumental, uso de reactivos, entre otros; por lo tanto, el monitoreo del desempeño del EA a lo largo del tiempo actúa como un sistema complementario de aseguramiento de la calidad de los resultados de medición.

Para decidir si el desempeño mejora o disminuye con el tiempo, los datos de las rondas continuas deben ser comparables. Sin embargo, el conjunto de datos de la misma medición de dos rondas diferentes puede tener una desviación estándar diferente para la evaluación de aptitud y, por lo tanto, los indicadores de desempeño se calculan de manera diferente, dependiendo del grupo de laboratorios que participaron y la influencia de otras variables derivadas de la complejidad de las muestras. Los participantes pueden calcular sus propios z-score (u otros indicadores de desempeño) utilizando una desviación estándar adecuada para la evaluación de aptitud, si el valor utilizado en rondas consecutivas difiere sustancialmente. Si un laboratorio decide recalculer sus propios indicadores de desempeño, debería justificar y documentar su elección.

3. Investigación de causas para establecer actividades de mejora

Cualquier laboratorio ocasionalmente puede tener resultados no satisfactorios o cuestionables en los EA. Cuando esto ocurre, el laboratorio debe identificarlos y documentarlos claramente. Si se deben realizar investigaciones, deben tratarse como una no conformidad del sistema de gestión de calidad del laboratorio.

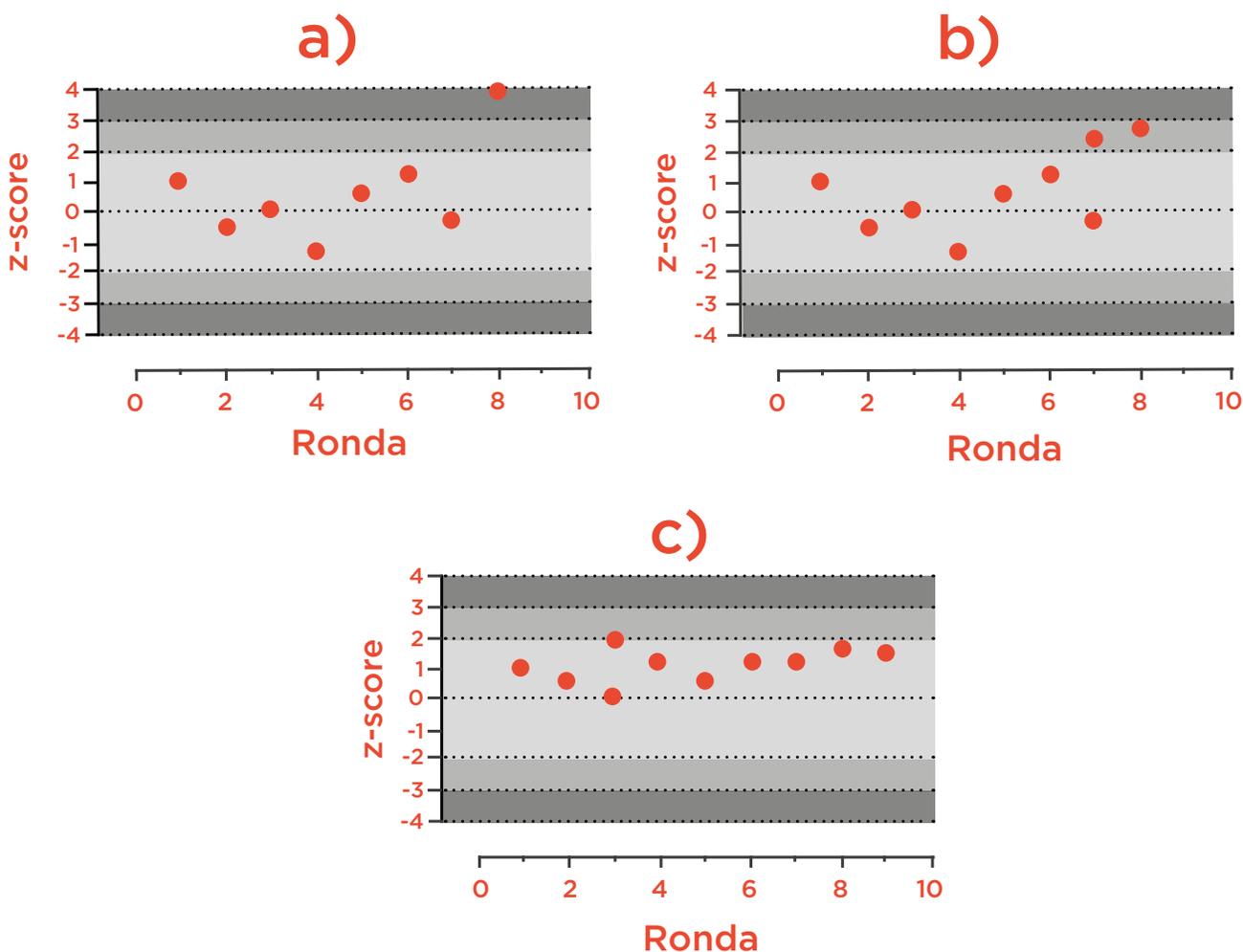
La profundidad de la investigación dependerá de una serie de factores, que pueden incluir, el procedimiento de medición, la frecuencia de resultados no satisfactorios y la evidencia de un sesgo. En todos los casos, el laboratorio deberá documentar la evaluación de los resultados, incluso si decide no tomar ninguna medida específica. Como el laboratorio debe utilizar métodos validados junto con controles de calidad internos, que

aseguran la validez de los resultados de medición, cualquier desempeño deficiente debe tomarse en serio, ya que puede indicar que hay un problema en el procedimiento de medición.

Cuando se participa en un esquema continuo de ensayo de aptitud, con varias rondas, el laboratorio puede elegir los siguientes criterios para hacer una investigación, por ejemplo (Figura 3.1): a) una puntuación no satisfactoria; b) 2 puntuaciones de rendimiento cuestionables consecutivas para el mismo parámetro; c) 9 puntuaciones de rendimiento consecutivas, para el mismo parámetro, que tienen el mismo signo de sesgo contra el valor asignado.

FIGURA 3.1

Ejemplos de resultados cuestionables en participación en un esquema continuo de ensayo de aptitud.



Es importante tener en cuenta que es responsabilidad del laboratorio establecer sus propios criterios para iniciar una investigación, de acuerdo con los factores mencionados anteriormente y teniendo en cuenta que la clave del análisis está en examinar las tendencias de los resultados.

Cuando se considera necesaria una investigación completa, se prefiere un enfoque gradual, para maximizar las posibilidades de determinar la causa raíz del problema. Un procedimiento adecuado de investigación debería consistir en los siguientes pasos e involucrar al personal que realizó el análisis, y cuando sea apropiado, a las directivas del laboratorio:

- a. Analizar el problema en función de los datos sin procesar, datos de control de calidad interno, el resultado general de los participantes en la ronda y tendencias con los resultados de rondas anteriores.
- b. Definir las actividades de mejora y hacer un plan.
- c. Ejecutar las actividades de mejora y registrar sus resultados.
- d. Verificar que las actividades de mejora fueron efectivas.

4. Posibles fuentes de error

Existen varias razones para obtener un bajo desempeño en un EA, resultando en una investigación compleja que consume tiempo; sin embargo, como las investigaciones resultan en un mejoramiento en el desempeño del laboratorio, es necesario hacer este esfuerzo. Con el propósito de identificar la causa del bajo desempeño, es importante centrarse en potenciales fuentes de error, tales como:

Error no analítico:

Puede incluir errores de transcripción, unidades, decimales, o etiquetado incorrecto. Estos errores pueden indicar que el laboratorio tiene un problema potencial al reportar los resultados al cliente. La identificación de estos errores es un primer paso en una investigación. Si los errores no analíticos son la causa frecuente de resultados no satisfactorios, entonces la investigación se debe enfocar en aspectos del sistema de gestión de calidad. Este tipo de error no está relacionado con la competencia técnica del laboratorio.

☑ **Problema técnico:**

Puede ocurrir en cualquier nivel del procedimiento analítico y cada uno de los pasos deben ser evaluados: Almacenamiento / pre-tratamiento de la muestra, método / controles de calidad, equipos / reactivos / calibración, condiciones ambientales, y procesamiento de datos; además se debe considerar la formación del personal. En el caso de no identificar la fuente de error asociada a un problema en la ejecución del método de medición, puede ser necesario evaluar la validación del método.

☑ **Problema relacionado con el esquema del EA:**

Un bajo desempeño puede deberse a la selección inadecuada del esquema de EA o un problema con el ítem de ensayo. Estos problemas se pueden relacionar con diferencias entre el ítem de EA y las muestras de rutina, deterioro potencial del ítem de EA, inadecuados niveles de concentración, pérdida de estabilidad u homogeneidad, problemas de almacenamiento del ítem de EA, instrucciones inapropiadas del proveedor a los participantes, inadecuada agrupación de participantes, valor asignado inapropiado, desviación estándar inapropiada, y entrada incorrecta de datos por parte del proveedor de EA.

Para facilitar las investigaciones es útil tener en mente las principales causas de un bajo desempeño, y así tener un mejor enfoque. A continuación, se relacionan las fuentes de error que ocurren con frecuencia [5].

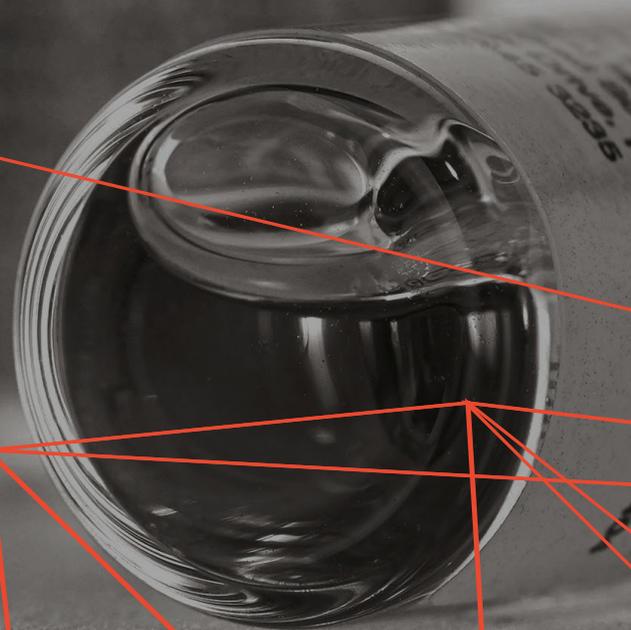
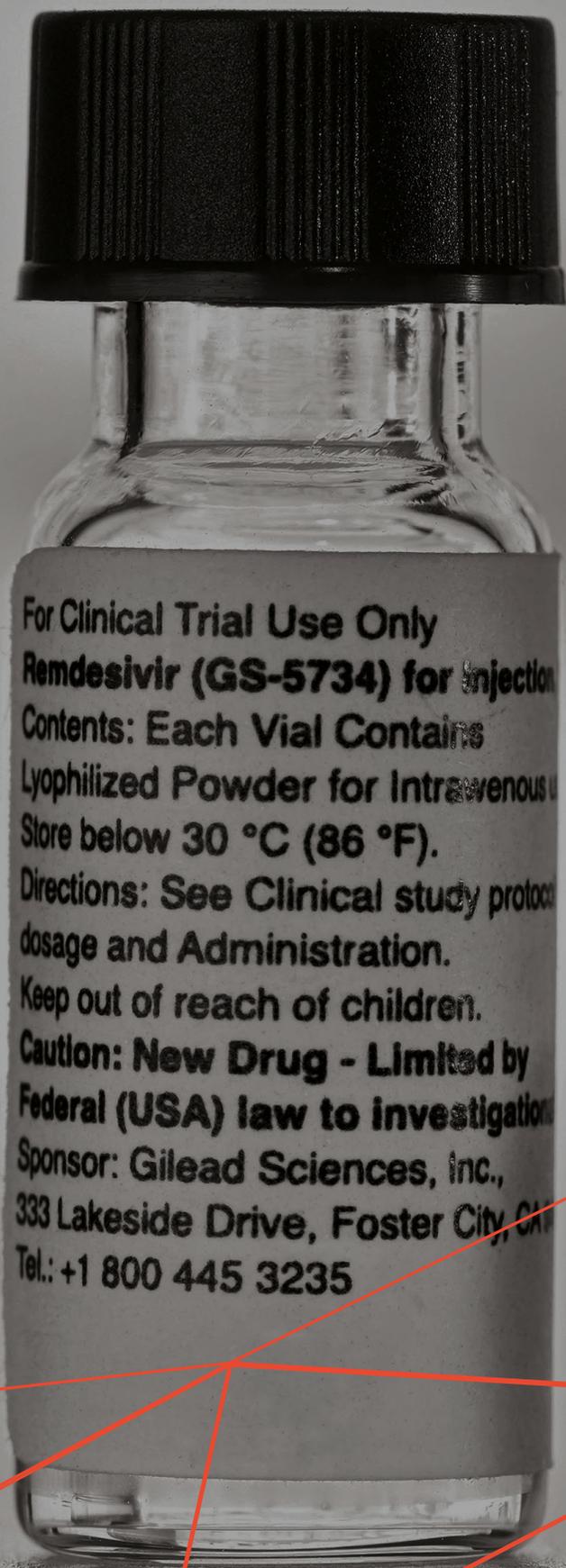


Tabla 3.1

Fuentes frecuentes de error [5].

	Fuente de error	Ejemplo
Error no analítico	Error humano	Falta de entrenamiento o experiencia, error de transcripción, instrucciones de reporte inadecuadas / seguimiento incorrecto, error en la lectura del instrumento, y error de interpretación
	Problema en el reporte (formato, unidades, interpretación)	Valor reportado correctamente pero no con las unidades usadas por el proveedor del EA, unidades declaradas incorrectamente, error en transcripción del valor, reportar como “no detectado” o “menor que x” cuando se solicita un valor numérico, y declarar incertidumbre de medición que no cumple con los requisitos del proveedor de EA.
	Seguimiento de la muestra (etiquetado y cadena de custodia)	Resultados atribuidos a la muestra incorrecta, mal etiquetado de la muestra durante o posterior a la recepción.
Problema técnico	Preparación de la muestra	Error en la dilución de las muestras, problemas de extracción y recuperación, fallas en la digestión de la muestra, limpieza inadecuada, problemas en el proceso de pesaje, secado, y molienda.
	Falla en el sistema de medición	Fallas del instrumento de medición, problemas en la configuración del instrumento, equipo fuera del intervalo de mantenimiento, el instrumento no pasa las verificaciones de rutina.
	Calibración	Procedimiento de calibración incorrecto (p.e. Calibración con un solo punto vs multipunto, forma de curva incorrecta), MR no disponible, MR en mal estado, intervalo de calibración insuficiente, matriz inadecuada, condiciones de medición no controladas (p.e. temperatura, presión, etc.).
	Error de cálculo	Valores de entrada incorrectos, problemas en la hoja de cálculo donde el usuario introduce la fórmula, errores aritméticos en cálculos hechos a mano, software usado incorrectamente, error de software del instrumento, problemas causados en el software de hoja de cálculo.

Problema técnico	Selección del método de medición	Precisión inadecuada, límite de detección inadecuado, incertidumbre del método inaceptable, diferencias en el desempeño entre métodos de referencia, material o aplicación fuera del alcance del método elegido
	Fallas en el método de medición	Problemas en la validación o verificación del método, falta de controles de medición, falta de trazabilidad metrológica en las mediciones, error en el procedimiento de medición, o en su ejecución, medición fuera de los criterios de calidad establecidos para el método, controles internos de calidad no cumplen con los parámetros establecidos.
	Almacenamiento de la muestra	Condiciones de almacenamiento incorrectas, contenedor inapropiado o en mal estado, tiempos largos de almacenamiento.
	Muestreo primario	Muestra no representativa, problema en la estabilización de la muestra, falla en el protocolo de muestreo, uso de contenedor incorrecto, contaminación cruzada durante muestreo.
	Condiciones ambientales del laboratorio	Inadecuado control de condiciones ambientales, y contaminación cruzada dentro de laboratorio.
Problema relacionado con el esquema del EA	Problema en el ítem de ensayo	Insuficiente cantidad de la muestra, matriz inesperada o presencia de interferencias que afectan al método de medición, material de ensayo fuera del alcance del método.
	Problema del proveedor del EA	Los puntajes no tienen en cuenta diferencias entre métodos o diferentes instrumentos, muestra incorrectamente empacada, condiciones de transporte inapropiadas, resultado atribuido a la muestra o el laboratorio incorrecto, valor asignado incorrecto, error de etiquetado, poca distribución de los datos, tratamiento de datos de límite de detección incorrectos.

Es posible que después de una investigación exhaustiva, no se identifique el origen del bajo desempeño. Un resultado no satisfactorio podría atribuirse a un error aleatorio, si se trata de un resultado repetido, entonces debe ser cuestionado el proceso analítico.

5. Análisis de múltiples ensayos de aptitud

La participación en ensayos de aptitud permite evaluar el desempeño de un laboratorio cronológicamente, el desempeño de diferentes analistas, desempeño por métodos, desempeño por niveles de concentración y/o el desempeño por grupos de analitos.

La puntuación o el valor del indicador de desempeño obtenido como resultado en un ensayo de aptitud proporciona una indicación valiosa del desempeño del laboratorio, un conjunto o secuencia de puntuaciones proporciona una visión más profunda. La evaluación de una secuencia de puntuaciones se puede hacer apropiadamente por métodos gráficos o numéricos que resuman el conjunto de valores obtenidos; sin embargo, se requiere la debida precaución con los métodos numéricos para evitar conclusiones incorrectas. No se recomienda el uso de un puntaje resumido derivado de los puntajes relacionados con diferentes analitos, ya que tiene un rango muy limitado de aplicaciones válidas y tiende a ocultar problemas esporádicos o persistentes con analitos individuales; además, es propenso al mal uso.

5.1. Puntajes resumidos

Los puntajes resumidos permiten evaluar una secuencia de resultados para un mismo analito, material de ensayo y método; estos tienen como utilidad hacer un seguimiento continuo de los resultados contemplando todos los puntajes que hacen parte de la secuencia. A continuación, se presentan dos tipos de puntajes resumidos que tienen una base estadística sólida y pueden ser útiles para evaluar una secuencia de z-score ($z_1, z_2, \dots, z_i, \dots, z_n$).

Suma reescalada de z-score

$$S_{z,sr} = \sum_i z_i / \sqrt{n}$$

El resultado de la suma reescalada puede interpretarse sobre la misma base de un z-score. Este estadístico tiene una propiedad útil para demostrar sesgo o tendencia, como es el caso de tener una secuencia de resultados de cuatro z-score con valores de (1.5, 1.5, 1.5, 1.5) que proporcionaría un $S_{z,sr}$ de 3.0, que resulta en un desempeño no satisfactorio, aun cuando cada resultado individual de z-score este dentro del intervalo de un resultado satisfactorio. Sin embargo, este estadístico podría ocultar el efecto de dos z-score grandes con signo opuesto, como puede ocurrir en la secuencia (1.5, 4.5, -3.6, 0.6) que resulta en un $S_{z,sr}$ de 1.5, considerado como un desempeño satisfactorio [6].

Suma de cuadrados de z-score

$$S_{zz} = \sum_i z_i^2$$

La suma de los cuadrados de z-score tiene que ser comparada contra el valor critico chi-cuadrado (χ^2), cuando el valor de S_{zz} es mayor se considera un desempeño no satisfactorio [7]. Este estadístico tiene la ventaja de evitar la cancelación de z-score grandes de signo opuesto, pero es menos sensible a pequeños sesgos.

Ambos estadísticos deben protegerse contra puntajes atípicos previos, que podrían afectar a largo plazo el resultado. S_{zz} es especialmente sensible a valores atípicos. Ambos estadísticos pueden relacionarse con la incertidumbre de medición en la siguiente forma, si los z-score se basan en la aptitud para el propósito y, por lo tanto, se consideran aleatorios $N(0,1)$, valores grandes en estos estadísticos indican que la incertidumbre de los participantes es mayor a la indicada en los esquemas con el criterio de aptos para el propósito [6].

5.2. Métodos gráficos

Los métodos gráficos se deberían emplear cuando se obtienen resultados para más de un ítem de ensayo de aptitud o para varios mensurandos, siempre que estén estrechamente relacionados y/o se obtengan mediante el mismo método. Estos procedimientos combinan resultados de indicadores de desempeño, de tal forma que no ocultan valores elevados de resultados de indicadores individuales y pueden revelar información adicional sobre el desempeño, como la correlación entre los resultados para diferentes analitos, que no es evidente en tablas de indicadores individuales.

Los métodos gráficos para analizar un conjunto de z-scores pueden ser tan informativos como los puntajes resumidos y puede ser menos propensos a malas interpretaciones. Los gráficos de Shewhart (con límites de advertencia y acción en $z=\pm 2$ y $z=\pm 3$, respectivamente) se pueden aplicar. Múltiples gráficos simbólicos univariados, como los que se muestran a continuación, brindan una visión general y son especialmente útiles cuando los puntajes de un grupo de analitos son determinados por un método común.

El gráfico de control de z-scores para diferentes analitos (**Figura 3.2**), muestra símbolos apuntando hacia arriba para indicar z-scores mayores que cero y símbolos apuntado hacia abajo para aquellos valores menores que cero. Los símbolos pequeños representan valores de $2 \leq |z| < 3$, y símbolos grandes valores de $|z| \geq 3$. Los datos ilustrados muestran algunas características notables. En el caso de la **Figura 3.3**, se encuentra que en la ronda 11 los resultados en su mayoría son bajos, lo que demuestra un problema en el procedimiento, mientras que el analito 7 da resultados altos frecuentemente entre las rondas, lo que demuestra un problema persistente para este analito específico.

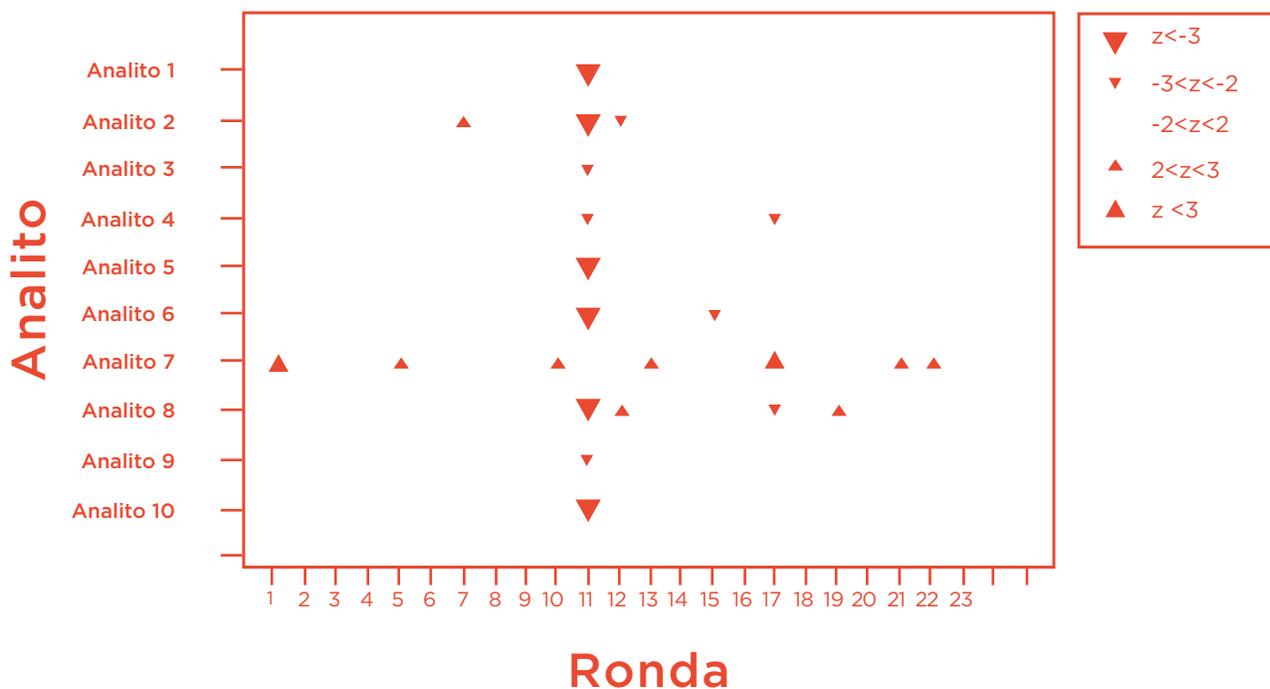


FIGURA 3.2

Gráfico J (gráfico de zona) para diferentes analitos [6].

Los gráficos J, también conocidos como gráficos de zona, son más informativos. Este gráfico permite detectar sesgos menores que son persistentes, así como cambios abruptos en el sistema analítico, ya que estos gráficos acumulan valores de J-score atribuidos a resultados sucesivos en uno de los lados de la línea del cero. Para construir los gráficos es necesario hacer la conversión del z-score a J-score y sumar el acumulado con resultados anteriores. Las equivalencias para convertir los z-score a J-score son:

Tabla 3.2

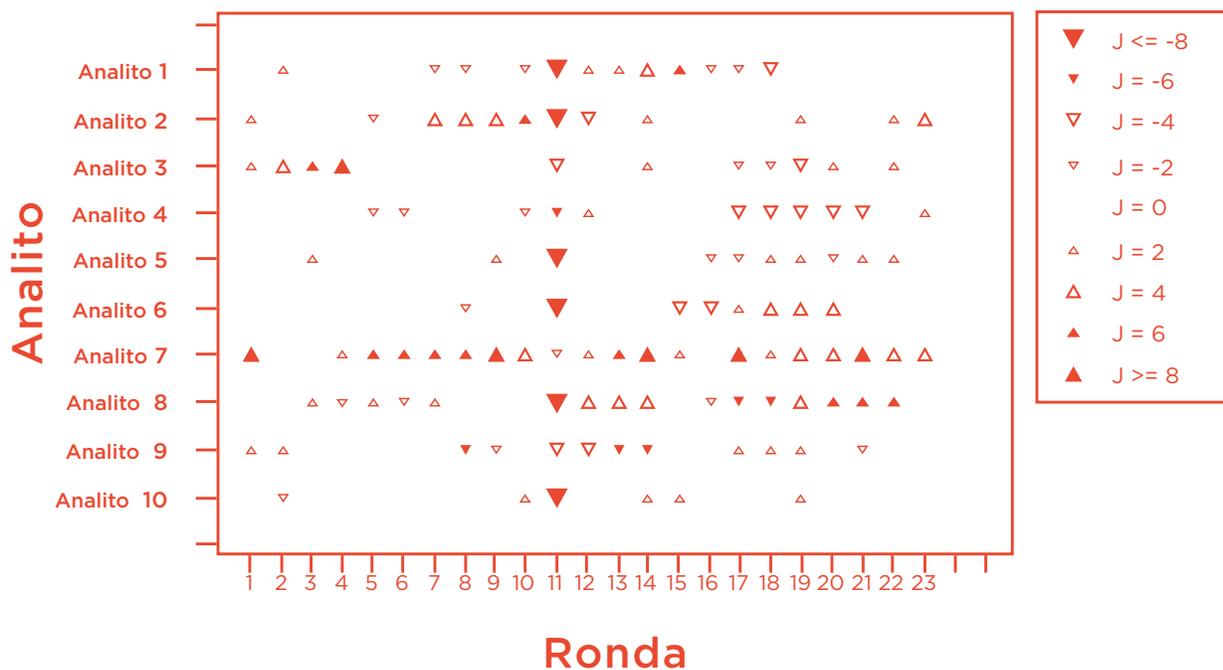
Equivalencias entre J-score y z-score.

z-score	≥ 3	$2 \leq z < 3$	$1 \leq z < 2$	$-1 < z < 1$	$-2 < z \leq -1$	$-3 < z \leq -2$	≤ -3
J-score	8	4	2	0	-2	-4	-8

Cuando los J-score de rondas sucesivas acumulan valores por encima de 8 o por debajo de -8, se llega a los límites de acción y se deben establecer procedimientos de investigación. Hechas las investigaciones e implementadas las acciones de mejora, se reestablece el acumulador del J-score a cero. En la **Figura 3.3** se muestran varios ejemplos del sesgo acumulativo que son visibles (estos ilustran los mismos resultados de la **Figura 3.2**). Por ejemplo, el analito 3 en las rondas 1 a la 4 recibe z-scores de 1.5, 1.2, 1.5 y 1.1 respectivamente, que equivalen a valores de J-score de 2, 2, 2 y 2, que se acumulan a 8 en la cuarta ronda y conllevan a una investigación de causas. Un ejemplo similar se puede observar en el analito 7.

FIGURA 3.3

Gráfico J (gráfico de zona) para diferentes analitos [6].



Referencias bibliográficas

- [1] EURACHEM/CITAC and C. of E.-P. W. Group, "Selection, Use and Interpretation of Proficiency Testing (PT) Schemes," Eurachem, pp. 1-52, 2011.
- [2] R. G. Visser, "Interpretation of interlaboratory comparison results to evaluate laboratory proficiency," *Accredit. Qual. Assur.*, vol. 10, no. 10, pp. 521-526, 2006, doi: 10.1007/s00769-005-0051-2.
- [3] Analytical Methods Committee, "Understanding and acting on scores obtained in proficiency testing schemes," *Anal. Methods Comm. Tech. Br.*, no. 11, p. 4 pp., 2002.
- [4] R. W. Jenny and K. Y. Jackson-Tarentino, "Causes of unsatisfactory performance in proficiency testing," *Clin. Chem.*, vol. 46, no. 1, pp. 89-99, 2000, doi: 10.1093/clinchem/46.1.89.
- [5] S. L. R. Ellison and W. A. Hardcastle, "Causes of error in analytical chemistry: Results of a web-based survey of proficiency testing participants," *Accredit. Qual. Assur.*, vol. 17, no. 4, pp. 453-464, 2012, doi: 10.1007/s00769-012-0894-2.
- [6] M. Thompson, S. L. R. Ellison, and R. Wood, "The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories: (IUPAC technical report)," *Pure Appl. Chem.*, vol. 78, no. 1, pp. 145-196, 2006, doi: 10.1351/pac200678010145.
- [7] D. A. Lacher, "Evaluation of a sum of squared Z scores quality control rule," *Clinica Chimica Acta*, vol. 164, no. 1, pp. 109-110, Apr. 15, 1987, doi: 10.1016/0009-8981(87)90112-4.
- [8] International Organization for Standardization, ISO 13528:2015(en) Statistical methods for use in proficiency testing by interlaboratory comparison-[org/10.1039/B111465D](https://www.iso.org/10.1039/B111465D)

- [9] RSC Analytical Methods Committee. Proficiency testing of analytical laboratories: organization and statistical assessment. *Analyst* [Internet]. 1992;117(1):97–104. Available from: <http://dx.doi.org/10.1039/AN9921700097>
- [10] Kisets D. Performance indication improvement for a proficiency testing. *Accredit Qual Assur* [Internet]. 2006;10(9):461–5. Available from: <https://doi.org/10.1007/s00769-005-0021-8>
- [11] Hund E, Massart DL, Smeyers-Verbeke J. Interlaboratory studies in analytical chemistry. *Anal Chim Acta* [Internet]. 2000;423(2):145–65. Available from: <http://www.sciencedirect.com/science/article/pii/S0003267000011156>
- [12] Committee AM, No A. Z -Scores and other scores in chemical proficiency testing - Their meanings, and some common misconceptions. *Anal Methods* [Internet]. 2016;8(28):5553–5. Available from: <http://dx.doi.org/10.1039/C6AY90078J>
- [13] Analytical Methods Committee. What is proficiency testing? Guide for end-users of chemical data. *R Soc Chem*. 2010;7(2):1–2.
- [14] Mäkinen I. Use of uncertainty estimates as reported by participants in proficiency testing for the evaluation of their results: pros and cons. *Accredit Qual Assur* [Internet]. 2009;14(1):35–41. Available from: <https://doi.org/10.1007/s00769-008-0462-y>
- [15] Maier EA, Quevauviller P, Griepink B. Interlaboratory studies as a tool for many purposes: proficiency testing, learning exercises, quality control and certification of matrix materials. *Anal Chim Acta* [Internet]. 1993;283(1):590–9. Available from: <http://www.sciencedirect.com/science/article/pii/S000326709385272L>
- [16] ONAC. POLITICA PARA LA PARTICIPACIÓN EN ENSAYOS DE APTITUD (EA) EN LABORATORIOS CEA-3.0-04 (Antes CEA-04) Versión 03 NIVEL 1: 3.0 PRESTACIÓN DEL SERVICIO. Vol. 04. 2018.
- [17] Document G. CCQM ORGANIC ANALYSIS WORKING GROUP(OAWG) Practices and Guidelines Document Contents. 2018;1–26.
- [18] Wolf WR, Andrews KW. A system for defining reference materials applicable to all food matrices. *Fresenius J Anal Chem*. 1995;352(1–2):73–6.

- [19] Measurements and Standards to Support Nutrition Labeling _ NIST [Internet]. Available from: <https://www.nist.gov/programs-projects/measurements-and-standards-support-nutrition-labeling>
- [20] Fiscaro P, Infante HG. The new IAWG CC approach : objectives. 2019.
- [21] European Accreditation Commity. EA-4 / 18 INF : 2010 Guidance on the level and frequency of proficiency expression testing participation. 2010;(June):15.
- [22] Örnemark U, Fostel H, Straub R, van de Kreeke J. Policies, requirements and surveys concerning frequency for participation in proficiency testing schemes. *Accredit Qual Assur* [Internet]. 2004;9(11):729–32. Available from: <https://doi.org/10.1007/s00769-004-0858-2>
- [23] Brookman B, Butler O, Ciaralli L, Horsky M, Laurent C, Noblett T, et al. Proficiency testing in analytical chemistry, microbiology and laboratory medicine: discussions on current practice and future directions. *Accredit Qual Assur* [Internet]. 2019;24(1):93–101. Available from: <https://doi.org/10.1007/s00769-018-1343-7>
- [24] Brookman B, Evans S, Ellison S, Boley N, Schmidt A, Berkel E Van, et al. Proficiency testing in analytical chemistry , microbiology and laboratory medicine : working group discussions on current practice and future directions. *Accredit Qual Assur*. 2004;635–41.
- [25] Eurachem’s Proficiency Testing Working Group. Proficiency testing – How much and how often ? 1st ed. 2016. 1–2 p.

GQSP COLOMBIA

Calle 115 No. 5-50, Bogotá
Tel: +57 1 477 98 88
www.gqspcolombia.org

ISBN

978-958-52871-8-1



Con el apoyo de:



ORGANIZACIÓN DE LAS NACIONES UNIDAS
PARA EL DESARROLLO INDUSTRIAL



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Confederación Suiza

Departamento Federal de Economía,
Formación e Investigación DEFI
Secretaría de Estado para Asuntos Económicos SECO



El progreso
es de todos

Mincomercio

