

EVALUATION D'APTITUDE PAR COMPARAISONS INTERLABORATOIRES

PROFICIENCY ASSESSMENT USING INTERLABORATORY COMPARISONS

Michèle Désenfant, Soraya Amarouche

LNE

Laboratoire national de métrologie et d'essais
1, rue Gaston Boissier 75724 Paris cedex 15

Résumé

La mise en œuvre d'une comparaison interlaboratoires peut répondre à différents objectifs dont les trois principaux sont : attribuer une valeur consensuelle à une caractéristique d'un objet (Matériau de référence), estimer l'exactitude d'une méthode (Performance d'une méthode), évaluer l'aptitude des laboratoires (Essais d'aptitude). Les différents documents normatifs associés à chaque objectif sont présentés, puis l'article se focalise sur les essais d'aptitude qui sont en plein essor. Plusieurs documents sont en cours d'évolution dont la future norme ISO/CEI 17043. Les comparaisons interlaboratoires en essai et en étalonnage sont ensuite étudiées à partir d'une sélection de points clés dans leurs processus.

Abstract

Interlaboratory comparisons could achieve different goals of which the three main ones are: to assign a consensus value to a characteristic (reference material), to estimate the accuracy of a measurement method (proficiency method), to evaluate laboratory's proficiency (proficiency testing). The different standards associated to each goal are presented. Then, the text focuses on proficiency assessments which are progressing. Several documents are being updated and a new international standard ISO/IEC 17043 is being prepared. Then, interlaboratory comparisons in testing and calibration are studied from selected criteria.

Les différentes comparaisons interlaboratoires

Les comparaisons interlaboratoires sont définies comme l'organisation, l'exécution et l'évaluation d'essais ou de mesurages sur des échantillons identiques ou semblables

par au moins deux laboratoires différents dans des conditions prédéterminées.

La mise en œuvre d'une comparaison interlaboratoires peut répondre à différents objectifs dont les trois principaux sont les suivants :

- Attribuer une valeur consensuelle à une caractéristique d'un objet
Matériau de référence
- Estimer l'exactitude (justesse et fidélité) d'une méthode de mesure ou d'essai
Performance méthode
- Evaluer l'aptitude des laboratoires
Essais d'aptitude

On les représente classiquement par le triangle de la figure 1.

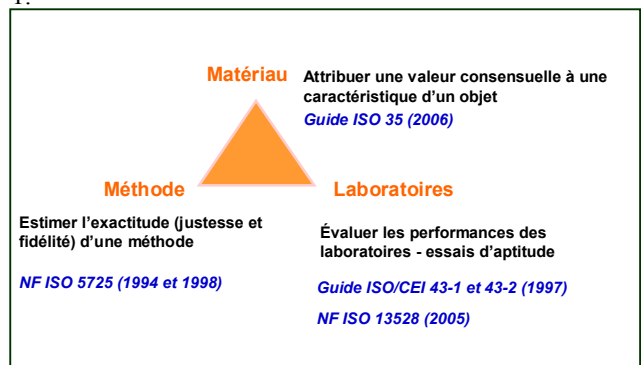


Figure 1 : les principaux objectifs possibles d'une comparaison interlaboratoires avec les référentiels associés.

Les référentiels normatifs associés et les outils statistiques diffèrent selon l'objectif de la comparaison. Il est important d'en faire la distinction afin d'utiliser le document qui correspond le mieux au problème posé. Voici un descriptif des principaux documents, sachant que nous allons détailler plus spécialement les essais d'aptitude.

Les comparaisons interlaboratoires organisées dans le but d'assigner une valeur consensuelle à une caractéristique

d'un objet se référeront au *guide ISO 35 Certification des matériaux de référence – Principes généraux et statistiques* [1]. Notons que ce document n'est pas spécifique aux comparaisons interlaboratoires, à l'instar des autres références citées ci après.

Les comparaisons interlaboratoires organisées dans le but d'estimer la performance d'une méthode ont pour objet d'évaluer la justesse et la fidélité d'une méthode de mesure ou d'essai produisant des résultats sur une échelle continue de valeurs. Tous les participants doivent bien sûr mettre en œuvre la même méthode, le même protocole. On trouve régulièrement publiée dans les normes d'essais, une annexe sur la fidélité de la méthode. Le référentiel normatif est la norme *NF ISO 5725* déclinée en six parties dont la plus utilisée est la partie 2 *Méthode de base pour la détermination de la répétabilité et de la reproductibilité d'une méthode de mesure normalisée* [2] décrivant les estimations des écarts-types de répétabilité et de reproductibilité associés à une méthode ainsi que différents tests de valeurs aberrantes (Grubbs et Cochran). Ce document est l'une des normes statistiques les plus utilisées et il devrait faire l'objet d'une révision prochainement par l'ISO TC69/SC6. Citons également le FD ISO/TR 22971 [3], paru en 2005, qui est un guide sur l'utilisation pratique de la NF ISO 5725-2.

Les comparaisons interlaboratoires organisées dans le but d'évaluer l'aptitude des laboratoires sont appelées généralement essais d'aptitude ou encore en anglais proficiency testing (PT). Dans ce document, nous employons aussi le terme plus général d'évaluation d'aptitude afin de couvrir aussi bien les essais que les étalonnages. Les laboratoires s'inscrivent pour participer à une ou plusieurs campagnes, proches de leurs propres essais de routine. Ils mettent en œuvre leur méthode habituelle ce qui leur permet d'avoir une évaluation externe de la qualité de leur essai ou mesurage et de comparer leurs résultats aux autres laboratoires participants.

Le référentiel normatif sur le sujet est le *guide ISO/CEI 43-1 Essais d'aptitude des laboratoires par intercomparaison - partie 1 : Développement et mise en œuvre de système d'essais d'aptitude* [4], dont l'annexe A traite des exemples de méthodes statistiques pour le traitement des données d'essais d'aptitude. Cette annexe, renvoie à la norme internationale *NF ISO 13528 Méthodes statistiques utilisées dans les essais d'aptitude par comparaisons interlaboratoires* [5] qui présente essentiellement l'homogénéité, la stabilité des échantillons, ainsi que les différentes déterminations possibles de la valeur assignée, l'incertitude associée, et les différents critères de performance possibles.

Les évaluations d'aptitude ou essais d'aptitude

Le développement des essais d'aptitude

Les essais d'aptitude ou évaluations d'aptitude prennent de plus en plus d'importance et l'une des causes est certainement les recommandations du chapitre 5.9 de la

norme *NF EN ISO/CEI 17025* [6] qui cite « la participation à des programmes de comparaisons entre laboratoires ».

Les comparaisons interlaboratoires (notées ILC dans la suite du texte) ont pour objectif d'évaluer et de démontrer les performances des participants à réaliser des essais ou des mesures et constituent également un véritable outil de progrès pour les participants - si l'examen des résultats est convenablement mené pour déclencher si nécessaire des actions. Les ILC complètent les procédures internes de maîtrise de la qualité en fournissant une mesure externe supplémentaire de leur compétence en matière d'essai ou d'étalonnage. Les ILC complètent également la technique d'évaluation des laboratoires in situ par les experts techniques.

Ce développement et cette importance à démontrer la qualité des résultats a impliqué le développement de l'accréditation des organisateurs d'essai d'aptitude.

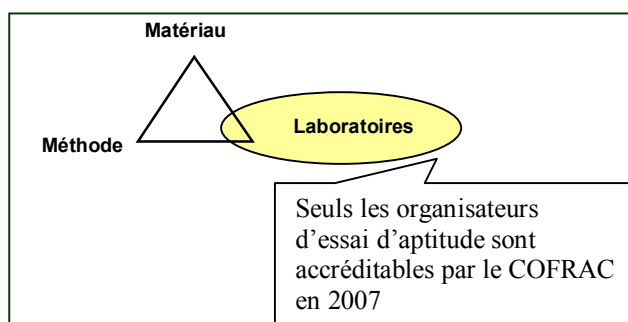


Figure 2 : comparaisons interlaboratoires et accréditation

Evolution des référentiels autour des essais d'aptitude

En 2006, l'ILAC International Laboratory Accreditation cooperation vient de réviser le document *ILAC G13 lignes directrices aux exigences en matière de compétence des organisateurs d'essais d'aptitude* [7].

Les principales modifications de l'ILAC G13 ont porté sur :

- l'alignement avec la nouvelle version de l'ISO/IEC 17025 :2005 et de l'ISO 9001 :2000
- la prise en compte des expériences d'accréditation des organisateurs d'essais d'aptitude ces dernières années
- la couverture des comparaisons dans le secteur clinique pour éviter la parution de guide sectoriel.

A l'initiative de l'ILAC, le CASCO/WG 28 travaille sur l'évolution du guide ISO 43 en norme ISO/CEI 17043. La cohérence avec la série 17000 est de toute importance. La future norme ISO/CEI 17043 sera destinée en priorité aux organisateurs d'essai d'aptitude, aux laboratoires participants, ainsi qu'aux accréditeurs.

Comparaisons interlaboratoires en essai et en étalonnage

Les ILC s'appliquent aussi bien aux essais qu'aux étalonnages, et il ne faut surtout pas les opposer, simplement leur mise en œuvre implique différentes organisations. En fait le principal critère de différenciation n'est pas essai ou étalonnage mais le support de comparaison comme décrit ci-dessous.

- Support de comparaison

Dans le domaine des essais, le support de la comparaison est souvent un échantillon soumis à essai envoyé simultanément à tous les participants et dont on a vérifié la stabilité et l'homogénéité. Il peut aussi quelquefois s'agir d'un essai circulaire "round robin", le même produit est essayé dans tous les laboratoires successivement.

Dans le domaine des étalonnages, le support de comparaison est un instrument ou une grandeur matérialisée soumis à étalonnage, qui circule d'un participant à l'autre. L'homogénéité n'a pas lieu d'être vérifiée mais la stabilité est d'autant plus importante que la comparaison est étalée dans le temps. voir figure 3 différentes organisations schématisées

- Durée de la comparaison

Dans le domaine des essais, la comparaison se déroule généralement sur une courte période pouvant aller d'une semaine à deux mois (peut être plus long si essais circulaires de type round robin).

Concernant la circulation d'étalons, la comparaison peut s'étendre sur une période pouvant aller de 6 mois à un an, voir plus.

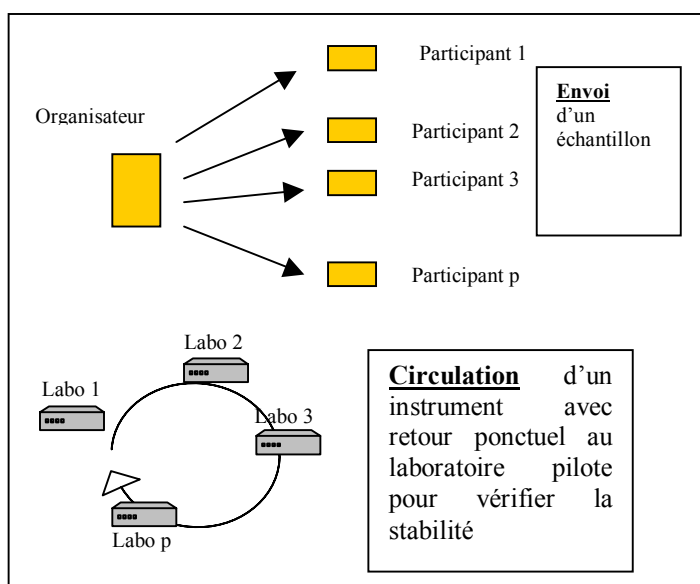


Figure 3 : Schéma de deux organisations possibles de comparaison (envoi simultané ou circulation).

- Confidentialité

C'est un critère très important pour les participants, et l'organisateur doit veiller à l'assurer. Cependant, il faut bien réaliser que les contraintes (aussi bien techniques que financières) ne sont pas les mêmes pour un envoi unique en parallèle (comparaison dans le domaine des essais) que pour une circulation en série dans les comparaisons en étalonnage. Sauf précautions onéreuses et riches en temps, en étalonnage chaque participant est amené à connaître le laboratoire précédent (expéditeur du support de comparaison) et le laboratoire suivant (destinataire de l'envoi de l'étalon). Plusieurs solutions d'organisation sont possibles, il faut que les participants, et éventuellement l'organisme d'accréditation, en soient informés et y adhèrent.

- Détermination de la valeur assignée et de son incertitude

Le terme *valeur assignée* est celui utilisé dans la norme NF ISO 13528.

Le critère fondamental dans la détermination de la valeur assignée consiste à savoir si la valeur est déterminée indépendamment des résultats des participants ou à partir des résultats des participants. C'est la première question de l'organigramme de la figure 1 de la norme 13528.

Dans certains domaines comme en métrologie, les deux méthodes sont envisageables car il est possible d'avoir des valeurs de référence indépendantes des résultats des participants, dans d'autres domaines comme en biologie médicale ou en œnologie, seule une valeur consensuelle de tous (ou d'une partie) des participants est possible.

Dans le cas de valeur consensuelle, il faut déterminer le mode de calcul de l'estimation de cette valeur assignée : soit moyenne arithmétique après élimination de possibles valeurs aberrantes, soit moyenne robuste, soit autre méthode justifiée, sachant que la norme préconise l'usage de la moyenne robuste. Dans tous les cas, les participants doivent être informés du mode de détermination avant participation.

- Statistique de performance

Plusieurs critères de performance sont calculables (la norme en liste huit), en fonction de la détermination de la valeur assignée, de l'estimation ou non de l'incertitude de chaque participant, de la détermination de l'écart-type pour l'évaluation d'aptitude.

En schématisant, dans le domaine des essais, quand la valeur assignée est obtenue par la moyenne robuste des participants et l'écart-type pour l'évaluation d'aptitude par l'écart-type robuste, et que l'on ne dispose pas de l'incertitude associée à chacun des résultats des participants, le critère classiquement calculé est le *score z*.

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

où x est le résultat rapporté par un participant

X est la valeur assignée

$\hat{\sigma}$ est l'écart-type pour l'évaluation de l'aptitude

Les critères courants du score z sont :

$$|z| \leq 2 \text{ satisfaisant}$$

$$2 < |z| < 3 \text{ discutable}$$

$$|z| \geq 3 \text{ insatisfaisant}$$

Dans le domaine des étalonnages et plus généralement quand la valeur assignée est obtenue en dehors des résultats des participants, par une valeur de référence et que chaque participant fournit une incertitude associée à son résultat, le critère classique est le nombre E_n .

$$E_n = \frac{x - X}{\sqrt{U_{lab}^2 + U_{ref}^2}}$$

où la valeur X est déterminée dans le laboratoire de référence,

U_{ref} est l'incertitude élargie de X

U_{lab} est l'incertitude élargie du résultat x d'un participant

Les critères courants du nombre E_n sont :

$$|E_n| \leq 1 \text{ satisfaisant}$$

$$|E_n| > 1 \text{ insatisfaisant}$$

- Rapport

Il convient que les rapports soient livrés rapidement et contiennent dans le meilleur des cas toutes les données des participants, dans certains cas, seulement une synthèse de tous les résultats. Ceci est donc possible une fois que tous les participants ont renvoyé leurs résultats. Dans le cas des essais circulaires, des rapports intermédiaires sont nécessaires. Ils sont non définitifs, sans les résultats des participants bien sur, sans la valeur de référence mais doivent donner une information sur la qualité des résultats afin de ne pas attendre par exemple huit mois avant d'engager une action corrective si un résultat était très éloigné de la valeur assignée.

Conclusion

Les évaluations d'aptitude par comparaisons interlaboratoires sont en développement dans de nombreux domaines. Elles ont pour objectif d'évaluer et de démontrer les performances des participants à réaliser des essais ou des mesures. Elles constituent également un outil de progrès quand les résultats sont réellement analysés. Il faut bien les distinguer des essais interlaboratoires destinés à

évaluer la performance d'une méthode d'essai ou de mesure.

Les référentiels évoluent pour tenir compte à la fois des retours d'expérience dans ce domaine, mais aussi pour suivre les modifications des documents qualité. Enfin, chaque comparaison interlaboratoires doit s'adapter aux contraintes techniques et économiques de son domaine et il faut bien être conscient qu'une comparaison par circulation d'un instrument de mesure ou d'essai ne peut pas être analysé comme celle par envoi d'échantillons, ce que les référentiels s'efforcent de prendre en compte.

Références

[1] Guide ISO 35 (2006) Certification des matériaux de référence – Principes généraux et statistiques.

[2] NF ISO 5725-2 (1994) Exactitude (justesse et fidélité) des résultats et méthodes de mesure. Partie 2 Méthode de base pour la détermination de la répétabilité et de la reproductibilité d'une méthode de mesure normalisée

[3] FD ISO/TR 22971 (2005) Lignes directrices pratiques pour l'utilisation de l'ISO 5725-2 :1994 pour la conception, la mise en œuvre et l'analyse statistique des résultats de répétabilité et de reproductibilité interlaboratoires.

[4] Guide ISO/CEI 43-1 (1997) Essais d'aptitude des laboratoires par intercomparaison - partie 1 : Développement et mise en œuvre de système d'essais d'aptitude.

[5] NF ISO 13528 (déc. 2005) Méthodes statistiques utilisées dans les essais d'aptitude par comparaisons interlaboratoires.

[6] NF EN ISO/CEI 17025 (sept.2005) Exigences générales concernant la compétence des laboratoires d'étalonnages et d'essais.

[7] ILAC G13 (2006) lignes directrices aux exigences en matière de compétence des organisateurs d'essais d'aptitude.

Nous remercions Christophe Perruchet, président de l'ISO/TC 69 "Application des méthodes statistiques", pour la relecture attentive de ce texte.