

انتخاب، استفاده و تفسیر طرح‌های آزمون مهارت (PT)

ویرایش دوم – ۲۰۱۱



انتخاب، استفاده و تفسیر طرح‌های آزمون مهارت (PT)

ویرایش دوم - ۲۰۱۱

Editors

Ian Mann (SAS, Switzerland)
Brian Brookman (LGC Standards, UK)

Composition of EEE-PT Working Group

Brian Brookman (Chair), *LGC Standards, UK*
Natalie Stephenson (Secretary), *LGC Standards, UK*
Frank Baumeister, *AQS Baden-Württemberg, Germany*
Greg Ewbank, *UKAS, UK*
Holger Frenz, *Institut für Eignungsprüfung (IfEP), Germany**
Maria Grazia del Monte, *ACCREDIA, Italy*
Magnus Holmgren, *SP Technical Research Institute, Sweden*
Eva Klokocnikova, *CAI, Czech Republic*
Christian Lehmann, *DAkkS, Germany*
Ulrich Leist, *DRRR, Germany*
Mirja Leivuori, *SYKE, Finland*
Minna Loikkanen, *Labquality, Finland*
Ian Mann, *SAS, Switzerland**
Piet Meijer, *ECAT Foundation, The Netherlands**
Marina Patriarca, *Istituto Superiore di Sanità (ISS), Italy*
Poul Erik Poulsen, *DANAK, Denmark*
Piotr Robouch, *IRMM, European Commission**
Pedro Rosario Ruiz, *RPS-QUALITAS, Spain**
Annette Thomas, *WEQAS, UK*
Johannes van de Kreeke, *BAM, Germany*

قدردانی

بازنگری این سند، از سوی گروه کاری EEE-PT (EA) - یورولب - یوراکم) انجام شد. سازمان EQALM (سازمان اروپایی برنامه‌های تضمین کیفیت خارجی در آزمایشگاه‌های طبی) نیز به عنوان عضو وابسته در این بازنگری نقش داشته است.

ترجمه به فارسی: محمد رحمانی

E-mail: momohumer@yahoo.com

ترجمه این اثر به معلم علم و اخلاق، دکتر سید مهدی گلابی تقدیم می‌شود.

*Members of the drafting group led by Ian Mann

انتخاب، استفاده و تفسیر طرح‌های آزمون مهارت (PT)

ویرایش انگلیسی

ویرایش دوم - ۲۰۱۱

کپی‌رایت © ۲۰۱۱

حق معنوی این سند در انحصار سازمان‌های معرفی شده از سوی اعضای EEE-PT WG می‌باشد.

هرگونه درخواست ترجمه، تولید و توزیع ویرایش‌های جدید این سند می‌بایست با دبیرخانه‌ی یوراکم در میان گذاشته شود.

فهرست مطالب

۱	 مقدمه
۲	 دامنه کاربرد این راهنما
۲	 تعاریف
۴	 مقدمه‌ای بر آزمون مهارت
۴	 ۱.۴ نقش PT در سیستم مدیریت
۵	 ۲.۴ انواع طرح‌های PT
۶	 ۵. انتخاب طرح‌های PT
۷	 ۱.۵ استراتژی مشارکت در PT
۸	 ۲.۵ دسترس‌پذیری طرح‌های PT
۸	 ۳.۵ چگونگی تصمیم‌گیری در باره اینکه آیا طرح PT انتخاب شده برازنده هدف می‌باشد یا خیر؟
۹	 ۶. استفاده از PT از سوی آزمایشگاه‌ها
۱۲	 ۷. چطور برگزارکننده PT عملکرد آزمایشگاه را برآورد می‌کند؟
۱۲	 ۱.۷ مقدمه
۱۲	 ۲.۷ عناصر اصلی در برآورد نتایج PT
۱۲	 ۱.۲.۷ مقدار اسمی
۱۳	 ۲.۲.۷ انحراف استاندارد برای ارزیابی مهارت
۱۳	 ۳.۲.۷ ارزیابی عملکرد
۱۵	 ۴.۲.۷ اثر عدم قطعیت مقدار اسمی
۱۵	 ۵.۲.۷ طرح‌های PT کیفی
۱۵	 ۶.۲.۷ داده‌های انحرافی
۱۶	 ۸. تفسیر آزمایشگاهی نتایج
۱۶	 ۱.۸ ارزیابی عملکرد از سوی آزمایشگاه
۱۶	 ۱.۱.۸ بازبینی نتایج یک دور PT
۱۷	 ۲.۱.۸ پایش عملکرد PT با گذشت زمان
۱۷	 ۲.۸ بررسی نتایج PT سؤال‌برانگیز یا نارضایت‌بخش
۱۷	 ۱.۲.۸ ضرورت بررسی
۱۸	 ۲.۲.۸ بررسی علت اصلی
۱۸	 ۳.۲.۸ علل عملکرد ضعیف
۲۱	 پیوست A: فرم انتخاب طرح PT مرتبط
۲۲	 پیوست B: فرم مستندسازی بررسی‌های PT
۲۴	 پیوست C: تفسیر داده‌های PT از سوی کاربران
۲۶	 پیوست D: جنبه‌های آماری PT
۲۹	 پیوست E: برآورد عملکرد PT
۳۲	 پیوست F: مثال‌هایی از رویکردهای مختلف در برآورد عملکرد
۳۷	 پیوست G: مثال ارزیابی عملکرد طولانی‌مدت
۴۰	 پیوست H: مثال استفاده از عدم قطعیت اندازه‌گیری

۱. مقدمه

ارزیابی مستقل و منظم از عملکرد فنی آزمایشگاه برای کسب اطمینان از اعتبار اندازه‌گیری‌ها یا آزمون‌ها (در این سند واژه‌ی «اندازه‌گیری» استفاده می‌شود که هم اندازه‌گیری و هم آزمون‌ها را شامل می‌شود) ضروری است و می‌بایست به عنوان بخشی از استراتژی کیفیت کلی لحاظ شود. یک رویکرد مشترک برای این ارزیابی مستقل، استفاده از طرح‌های آزمون مهارت (PT) مستقل است. طرح PT، سیستمی برای ارزیابی عینی عملکرد یک آزمایشگاه با استفاده از ابزارهای خارجی است که شامل مقایسه نظام‌مند نتایج یک آزمایشگاه با نتایج دیگر آزمایشگاه‌ها می‌باشد. این امر با توزیع اقلام PT پایدار و همگن از سوی برگزارکننده PT، بین مشارکت‌کنندگان جهت آنالیز و گزارش نتایج امکان‌پذیر می‌شود. به هر بار توزیع اقلام PT یک "دور" می‌گویند. هدف اصلی طرح PT کمک به مشارکت‌کنندگان جهت ارزیابی صحت اندازه‌گیری‌هایشان می‌باشد [1]. به علاوه مشارکت در یک طرح PT مناسب برای آزمایشگاه‌هایی که به دنبال اثبات شایستگی خود از طریق تأیید صلاحیت در برابر استانداردهای ISO/IEC 17025 یا ISO 15189 می‌باشند، لازم است [2],[3]. در بعضی از بخش‌ها، مشارکت در طرح‌های خاص اجباری می‌باشد.

طرح‌های PT به خاطر نفع مشارکت‌کنندگان اجرا می‌شوند. با این حال، دیگر طرف‌ها نیز از نفع قانونی در طرح‌های PT برخوردار می‌باشند. این گروه‌ها عبارتند از: مشتری‌های خدمات آزمایشگاهی تجزیه، مجموعه‌های تأیید صلاحیت، مراجع صلاحیت‌دار نظارتی و دیگر کاربران نتایج آزمایشگاهی. مهم است که برگزارکنندگان طرح PT، نیازهای این سازمان‌ها را جهت استفاده از نتایج طرح‌های PT مدنظر داشته باشند تا بتوانند به درک آنها از قابلیت‌ها و شایستگی آزمایشگاه‌ها کمک کنند (پیوست C را ببینید). برای آزمایشگاه‌ها مهم است که در مورد دسترسی‌پذیری و دامنه کاربرد طرح‌های PT در حیطه‌ی کاری خود اطلاعات جامعی داشته باشند. این موضوع، آنها را قادر خواهد ساخت تا در مورد اینکه در کدام طرح(ها) مشارکت کنند، تصمیم‌های مناسب اتخاذ کنند. مهم است که این نوع اطلاعات به طور گسترده در دسترس باشند تا انتخاب مناسب‌ترین طرح برای مشارکت‌کننده‌ها امکان‌پذیر باشد.

آزمایشگاه‌ها لازم است تا درک کاری مناسبی از PT داشته باشند، بدین معنی که نسبت به اهداف طرح‌های PT، چگونگی ارزیابی داده‌ها از سوی برگزارکننده PT و چگونگی ارزیابی داده‌ها به شکل داخلی و نحوه‌ی استفاده از آنها مطلع باشند.

اصول کلیدی زیادی در این سند گنجانده شده که به کمک آنها گروه‌های ذینفع می‌توانند از مناسبت مشارکت در طرح‌های PT اطمینان حاصل نمایند. بعضی از این اصول کلیدی عبارتند از:

(a) طرح PT که در آن آزمایشگاه شرکت می‌کند می‌بایست تا حد امکان از لحاظ نمونه‌های اندازه‌گیری، آنالیت‌ها و مقادیر آن، شبیه کار روتین آن آزمایشگاه باشد. هر تفاوتی می‌بایست در نظر گرفته شده و لحاظ شود.

(b) آزمایشگاه‌ها می‌بایست اقلام PT را مثل نمونه‌های روتین تیمار کنند، یعنی اینکه تیمار خاصی را برای آنها لحاظ نکنند.

(c) ارزیابی و تفسیر عملکرد در طرح PT می‌بایست خطر مرتبط با اندازه‌گیری را لحاظ کند.

(d) نتایج سؤال‌برانگیز تکراری و نارضایت‌بخش می‌بایست به طور جامع بررسی شده تا آزمایشگاه به دلایل عملکرد ضعیف خود واقف شده و در صورت لزوم آن را تصحیح کند.

(e) عملکرد یک آزمایشگاه در چند دور از یک طرح PT و آنالیز روندهای آن در تعیین میزان موفقیت مشارکت در درجه اول اهمیت قرار دارد.

(f) به منظور اینکه تمام گروه‌ها از چگونگی اجرای طرح مطلع باشند، می‌بایست اطلاعات روشن و آشکاری در مستندسازی طرح PT و پروتوکول‌ها ارائه شده باشند.

g) جهت درک صحیح طرح و چگونگی اجرای آن برگزارکننده PT می‌بایست این آمادگی را داشته باشد تا صادقانه با طرف‌های ذی‌نفع بحث و گفتگو کند.

h) آزمایشگاه‌ها می‌بایست به مشارکت در PT به شکل یک ابزار آموزشی نگاه کرده و با استفاده از نتایج طرح بازخورد لازم را به پرسنل و بهبود فرایند ارائه دهند.

۲. دامنه کاربرد این راهنما

هدف این سند ارائه راهنمایی لازم به آزمایشگاه‌ها درباره موارد ذیل می‌باشد:

a) اهداف و مزایای شرکت در طرح‌های PT

b) انتخاب مناسب‌ترین طرح PT

c) درک آمار پایه و نحوه امتیازدهی عملکرد از سوی برگزارکنندگان PT

d) استفاده و تفسیر نتایج PT به منظور بهبود عملکرد کل آزمایشگاه

این سند عمدتاً بر طرح‌های PT کمی تمرکز دارد، چرا که اکثر موارد این طرح PT از سوی آزمایشگاه‌ها استفاده می‌شود. با این حال، راهنمایی‌هایی در رابطه با طرح‌های تفسیری و کیفی نیز وجود دارند. هر چند، بسیاری از اصول عمومی برای تمام انواع طرح‌های PT قابل کاربرد می‌باشند. گرچه هدف این سند عمدتاً آزمایشگاه‌های آزمون می‌باشد، با این حال بعضی از اصول ذکر شده در این سند ممکن است برای آزمایشگاه‌های کالیبراسیون نیز به کار روند. اطلاعات همچنین می‌توانند برای دیگر مشارکت‌کنندگان در طرح‌های PT نظیر افراد، سازمان‌ها یا مجموعه‌های بازرسی نیز به کار گرفته شوند. این سند به بررسی آن دسته از مقایسه‌های بین آزمایشگاهی (ILC) که به ارزیابی ویژگی‌های عملکردی متد یا تعیین و استناد مقادیر به مواد مرجع می‌پردازند، همینطور به بررسی «مقایسه‌های کلیدی» که وظیفه انستیتوهای اندازه‌شناسی ملی هستند، نمی‌پردازد.

اطلاعات به دست آمده برای دیگر طرف‌ها نظیر نهادهای تأیید صلاحیت، سازمان‌های نظارتی یا مشتری‌های آزمایشگاهی نیز بسیار مفید است.

۳. تعاریف

۱.۳ مقایسه بین آزمایشگاهی (ILC)

سازماندهی، اجرا و ارزیابی اندازه‌گیری‌ها یا آزمون‌ها بر روی اقلام یکسان یا مشابه از سوی دو یا تعداد بیشتری آزمایشگاه، مطابق با شرایط از قبل تعیین شده.

[4] [3.4 تعریف, ISO/IEC 17043]

۲.۳ آزمون مهارت (PT)

ارزیابی عملکرد مشارکت‌کننده در برابر معیارهای از قبل تعیین شده به وسیله مقایسه‌های بین آزمایشگاهی.

نکته: بعضی از برگزارکنندگان آزمون مهارت در بخش پزشکی برای طرح‌های آزمون مهارت و یا برای برنامه‌های گسترده‌تر از اصطلاح ارزیابی کیفیت خارجی (ISO/IEC 17043 را ببینید) استفاده می‌کنند.

[4] [3.7 تعریف, ISO/IEC 17043]

۳.۳ طرح آزمون مهارت (طرح PT)

آزمون مهارت در یک یا چند دور و برای یک زمینه خاص آزمایشی، اندازه‌گیری، کالیبراسیون یا بازرسی طراحی و اجرا می‌شود.

[4] [3.11 تعریف, ISO/IEC 17043]

۴.۳ مشارکت‌کننده

آزمایشگاه، سازمان یا فرد که اقلام آزمون مهارت را دریافت کرده و نتایج را جهت بازنگری به برگزارکننده آزمون مهارت ارائه می‌کند.

نکته: در بعضی موارد مشارکت‌کننده می‌تواند یک نهاد بازرسی باشد.

[4] [3.6 تعریف, ISO/IEC 17043]

۵.۳ مقدار اسمی

مقدار نسبت داده شده به یک ویژگی خاص از یک آیتم آزمون مهارت.

[4] [3.1 تعریف, ISO/IEC 17043]

۶.۳ انحراف استاندارد برای ارزیابی مهارت

معیار پراکندگی استفاده شده در ارزیابی نتایج آزمون مهارت، بر پایه اطلاعات در دسترس.

نکته ۱: انحراف استاندارد فقط برای نسبت و نتایج مقیاس تفاضلی به کار می‌رود.

نکته ۲: همه طرح‌های آزمون مهارت، مهارت را بر اساس پراکندگی نتایج ارزیابی نمی‌کنند.

[4] [3.3 تعریف, ISO/IEC 17043]

۷.۳ اندازه‌گیری

فرایند به دست آوردن یک یا تعدادی از مقادیر کمی به شکل آزمایشگاهی که به طور منطقی به یک کمیت نسبت داده می‌شود.

نکته ۱: اندازه‌گیری برای خواص اسمی به کار نمی‌رود.

نکته ۲: اندازه‌گیری اشاره ضمنی به مقایسه کمیت‌ها داشته و شامل شمارش ذوات است.

نکته ۳: اندازه‌گیری، توصیفی از کمیت متناسب با هدف مربوط به یک نتیجه اندازه‌گیری، روش اجرایی اندازه‌گیری و سیستم اندازه‌گیری کالیبره شده را که بر اساس روش اندازه‌گیری خاص عمل می‌کند، همینطور شرایط اندازه‌گیری را مسلم فرض می‌کند.

[5] [2.1 تعریف, JCGM 200:2008 / 99:2007 راهنمای ISO/IEC]

۸.۳ عدم قطعیت اندازه‌گیری

پارامتر نامنفی تعیین‌کننده پراکندگی مقادیر کمی که می‌توانند بر اساس اطلاعات استفاده شده به یک اندازه ده نسبت داده شوند.

نکته ۱: عدم قطعیت اندازه‌گیری شامل مؤلفه‌هایی است که از اثرات سیستماتیک نظیر مؤلفه‌های مرتبط با تصحیح‌ها و مقادیر کمیت اسمی استانداردهای اندازه‌گیری، همینطور عدم قطعیت تعریفی ناشی می‌شوند. بعضی اوقات اثرات سیستماتیک تخمینی تصحیح نمی‌شوند، و به جای آن مؤلفه‌های عدم قطعیت اندازه‌گیری مرتبط لحاظ می‌شوند.

نکته ۲: پارامتر ممکن است برای مثال یک انحراف استاندارد باشد که عدم قطعیت اندازه‌گیری استاندارد (یا مضرب خاصی از آن) نامیده شده و یا نیمه پهنای یک بازه با احتمال پوشش بیان شده باشد.

نکته ۳: به طور کلی عدم قطعیت اندازه‌گیری از مؤلفه‌های زیادی تشکیل می‌شود. بعضی از این مؤلفه‌ها از طریق برآورد تیپ

A عدم قطعیت اندازه‌گیری و از توزیع آماری مقادیر کمی مجموعه‌ای از اندازه‌گیری‌ها ارزیابی شده و با انحراف استانداردها تعیین می‌شوند. دیگر مؤلفه‌ها از طریق برآورد تیپ B عدم قطعیت اندازه‌گیری ارزیابی و به کمک انحراف استانداردهایی تعیین می‌شوند که از توابع دانسیته احتمال و بر اساس تجربه یا دیگر اطلاعات به دست آمده‌اند.

نکته ۴: به طور کلی آشکار است که برای یک مجموعه از اطلاعات، عدم قطعیت اندازه‌گیری با مقدار کمیت نسبت داده شده به اندازه ده مرتب است، اصلاح این مقدار به اصلاح عدم قطعیت مرتبط منجر خواهد شد.

[5] [2.26 تعریف, JCGM 200:2008 / 5] [99:2007 راهنما ISO/IEC]

۹.۳ تکنیک اندازه‌گیری

فرایند تست / کالیبراسیون / شناسایی ویژگی، شامل هر گونه پیش‌تیمار لازم برای معرفی نمونه، از دریافت نمونه توسط آزمایشگاه گرفته تا وسیله اندازه‌گیری (برای مثال ICP-MS، سختی راکول Rockwell، PCR، میکروسکوپی، اندازه‌گیری نیرو) [EA-4/18] [6]

۱۰.۳ ویژگی

کمیتی که اندازه‌گیری می‌شود (برای مثال ارسنیک، کراتینین، طول، سختی، نیرو) [EA-4/18] [6]

نکته: در این راهنما، ویژگی می‌تواند بیشتر گسترش یافته و شامل دیگر ویژگی‌ها مثل نظرات، رنگ، طعم، حضور/عدم حضور و... نیز باشد.

۱۱.۳ محصول

آیتمی که تکنیک اندازه‌گیری برای آن به کار می‌رود (مثل خاک، سبزیجات، سرم، پلی‌استیرن، بتون) [EA-4/18] [6]

۱۲.۳ سطح مشارکت

تعداد حوزه‌های فرعی که یک سازمان درون محدوده کاربردی‌اش مشخص می‌کند، و در نتیجه تعداد آزمون‌های مهارت خاصی که می‌بایست برای مشارکت لحاظ شوند. [EA-4/18] [6]

۱۳.۳ دفعات مشارکت

اینکه آزمایشگاه چه تعداد مشارکت در PT را بر اساس نیازهای خود و برای یک حوزه فرعی معین تصمیم‌گیری کند، ممکن است دفعات مشارکت از یک حوزه فرعی به حوزه‌ی دیگر درون یک آزمایشگاه و همینطور بین آزمایشگاه‌های با حوزه‌های فرعی یکسان تغییر کند. [EA-4/18] [6]

۱۴.۳ حوزه فرعی (زمینه شایستگی فنی)

زمینه‌ای از شایستگی فنی که توسط حداقل یک تکنیک اندازه‌گیری، ویژگی و محصول که با یکدیگر مرتبط هستند (برای مثال تعیین ارسنیک در خاک با ICP-MS) تعریف می‌شود. [EA-4/18] [6]

۴. مقدمه‌ای بر آزمون مهارت

۱.۴ نقش PT در سیستم مدیریت

به کارگیری روش‌های کنترل کیفیت جهت پایش قابلیت اطمینان اندازه‌گیری‌ها برای آزمایشگاه مهم است. برای آزمایشگاه‌های دارای تأیید صلاحیت، یا آزمایشگاه‌هایی که به دنبال تأیید صلاحیت می‌باشند، این اقدامات جنبه مهم الزامات می‌باشند. شرکت در PT و همینطور برای مثال، استفاده از مواد مرجع (گواهی‌دار) یکی از این اقدامات می‌باشند.

PT با فراهم آوردن یک شاهد عینی برای صلاحیت و شایستگی مشارکت‌کننده نقش بسیار ارزشمندی را در این زمینه ایفا می‌کند. این دلیل و یا شاهد عینی می‌تواند برای بهبود عملکرد مشارکت‌کنندگان و یا اطمینان‌دهی به توانایی مشارکت‌کنندگان جهت اجرای یک اندازه‌گیری خاص مورد استفاده قرار گیرد.

به علاوه، مشارکت در طرح‌های PT نه تنها اطلاعاتی را در رابطه با عملکرد سیستم تجزیه‌ای ارائه می‌کند، بلکه در رابطه با دیگر جنبه‌های سیستم مدیریت نظیر دریافت/ تیمار نمونه، تیمار داده‌ها، گزارش‌دهی نتیجه و غیره نیز اطلاعاتی را عرضه می‌کند. خیلی مهم است که آزمایشگاه یک استراتژی مناسب برای مشارکت در طرح‌های PT داشته باشد (بخش ۱.۵ را ببینید).

PT فرصتی را برای مقایسه‌های بین داده‌های مشارکت‌کننده‌ها و مقایسه نتایج آنها با مقادیر اسمی (یا دیگر معیارهای عملکرد) و یا عملکرد آزمایشگاه‌های همکارشان فراهم می‌کند. نتایج یک چنین مشارکتی ممکن است به تأییدی مبنی بر رضایت‌بخش بودن عملکرد یا هشدار برای بررسی مشکلات بالقوه در این زمینه، منجر شود.

منابع مالی و انسانی لازم برای مشارکت PT می‌تواند مهم باشد، اما نمی‌بایست بیش از اندازه محدود شود، چرا که خیلی از خطرات مربوط به داده‌های مشارکت‌کننده‌ها که ممکن است دارای خطاها، بایاس‌ها یا اختلاف‌های معنادار در مقایسه با همتایانشان باشند به احتمال زیاد بدون تشخیص باقی خواهند ماند.

مهم است به این نکته تأکید شود که هدف PT صرفاً اجرای خوب یا بد آزمایش‌ها نیست، هدف اصلی آن است که مشارکت‌کننده‌ها از حضورشان در طرح‌های PT مطالب تازه‌ای آموخته و از این اطلاعات برای بهبود کیفیت اندازه‌گیری‌شان استفاده کنند.

گرچه هدف اصلی طرح PT، ارزیابی مشارکت‌کننده‌ها است، با این حال مزایای زیاد دیگری هم وجود دارند که به طور جامع در فصل ۶ بحث می‌شوند.

۲.۴ انواع طرح‌های PT

انواع مختلف طرح‌های PT در دسترس هستند که هر کدام از آنها حداقل بر یکی از عناصر چهار مقوله زیر پایه‌گذاری می‌شوند:

۱. (a) کیفی: نتایج آزمون‌های کیفی توصیفی بوده و بر اساس یک مقیاس اسمی یا ترتیبی گزارش می‌شوند.
نکته ۱: مقیاس اسمی یک سیستم عددی استفاده شده برای طبقه‌بندی و گروه‌بندی داده‌هاست، برای مثال جنسیت (مذکر/ مؤنث). یک مقیاس اسمی بعضی اوقات مقیاس مقوله‌ای نیز نامیده می‌شود.
نکته ۲: مقیاس ترتیبی نوعی مقیاس اندازه‌گیری است که مقادیری را به اشیا بر اساس رتبه‌بندی و نسبت به یکدیگر نسبت می‌دهد، برای مثال +، ++، +++.

- (b) کمی: نتایج اندازه‌گیری‌های کمی عددی بوده و بر اساس یک بازه یا مقیاس نسبت گزارش می‌شوند.
نکته ۳: مقیاس بازه نوعی مقیاس اندازه‌گیری است که در آن فاصله معینی در امتداد مقیاس به معنای مفهومی یکسان بوده و مهم نیست که شما کجای مقیاس باشید. با این حال، عدد "0" روی مقیاس، بیانگر غیاب شیء اندازه‌گیری‌شونده نمی‌باشد، برای مثال مقیاس‌های دمایی سلسیوس و فارنهایت.

- نکته ۴: مقیاس نسبتی نوعی مقیاس اندازه‌گیری است که در آن فاصله معینی در امتداد مقیاس به معنای مفهومی یکسان بوده و مهم نیست که شما کجای مقیاس باشید و عدد "0" روی مقیاس بیانگر غیاب شیء اندازه‌گیری‌شونده است، برای مثال عدد "4" روی چنین مقیاسی به معنای دو برابر مقدار چیزی است که اندازه‌گیری می‌شود، یعنی "2".

- (c) تفسیری: در اینجا اندازه‌گیری وجود ندارد. آیت PT، یک نتیجه اندازه‌گیری، مجموعه‌ای از داده‌ها یا دیگر اطلاعات مرتبط با جنبه تفسیری شایستگی مشارکت‌کننده‌ها می‌باشد.

۲. (a) تک: اقلام PT یک بار تأمین می‌شوند.
- (b) پیوسته: اقلام PT به شکل منظم تأمین می‌شوند.
۳. (a) ترتیبی: آیتم PT به طور ترتیبی از یک مشارکت‌کننده به نفر بعدی به گردش درمی‌آید. در این وضعیت، آیتم PT قبل از اینکه به مشارکت‌کننده بعدی پاس داده شود، ممکن است جهت تعیین اینکه آیا تغییری در آیتم PT اتفاق افتاده یا نه به برگزارکننده PT ارجاع داده شود. همچنین این امکان برای مشارکت‌کنندگان وجود دارد که جهت اندازه‌گیری آیتم PT یکسان در یک مکان مشترک دور هم جمع شوند.
- (b) همزمان: در متداول‌ترین شکل PTها، زیرنمونه‌های انتخاب شده تصادفی از یک ماده توده همگن به طور همزمان برای اندازه‌گیری همزمان بین مشارکت‌کنندگان توزیع می‌شود. پس از دریافت نتایج، برگزارکننده PT، بر اساس تکنیک‌های آماری عملکرد هر مشارکت‌کننده منفرد و یا گروه مشارکت‌کنندگان را به شکل کل ارزیابی خواهد کرد.
۴. (a) پیش - اندازه‌گیری [7]: در این نوع طرح PT، "آیتم PT" می‌تواند آیتمی (برای مثال یک اسباب‌بازی) باشد که مشارکت‌کننده‌ها می‌بایست تصمیم بگیرند کدام اندازه‌گیری‌ها روی آن اجرا شده و یا به کدام مجموعه داده‌ها یا دیگر اطلاعات (برای مثال یک مطالعه موردی) نیاز می‌باشد.
- (b) اندازه‌گیری: تمرکز، به ویژه بر روی فرایند اندازه‌گیری است.
- (c) پساً - اندازه‌گیری: در این طرح PT، "آیتم PT" می‌تواند مجموعه‌ای از داده‌ها باشد که از مشارکت‌کننده درخواست می‌شود تا نظر یا تفسیر خود را در باره آن بیان کنند.
- یک کاربرد خاص PT که اغلب "PT کور" نامیده می‌شود، جایی است که آیتم PT از دیگر آیتم‌های نرمال مشتری یا نمونه‌های دریافت شده از سوی مشارکت‌کننده‌ها غیرقابل تشخیص می‌باشد. تمام انواع طرح‌های PT ذکر شده در بالا می‌توانند به شکل PT کور سازماندهی شوند.
- ۵. انتخاب طرح‌های PT**
- انتخاب طرح PT برای کسب اطمینان از اینکه مشارکت‌کننده بیشترین نفع را از مشارکت ببرد، بسیار حساس است. بنابراین پروسه انتخاب PT مناسب مهم است. از سوی دیگر ضرورت دارد که مشارکت‌کننده شایستگی و صلاحیت برگزارکننده PT را ارزیابی کند. برگزارکننده PT که بر اساس ISO/IEC 17043 عمل می‌کند، می‌تواند به عنوان برگزارکننده شایسته در نظر گرفته شود.
- مشارکت در طرح PT، ابزارهای عینی ارزیابی را برای آزمایشگاه فراهم کرده و قابلیت اطمینان نتایج تولید شده را نیز نشان می‌دهد. لذا مشارکت در طرح PT، با فراهم نمودن معیار خارجی اضافی از قابلیت اطمینان اندازه‌گیری‌های آزمایشگاه می‌تواند مکمل روش‌های اجرایی کنترل کیفیت داخلی آزمایشگاه نیز به حساب آید. به همین خاطر، لازم است تمام آزمایشگاه‌ها یک استراتژی مشارکت PT با هدف شرکت در طرح‌های PT مرتبط و یا در تعداد دفعات متناسب با شرایط خود ایجاد کنند.
- در انتخاب یک طرح PT مناسب در یک زمینه از صلاحیت فنی، آزمایشگاه می‌بایست به سؤالات زیر پاسخ دهد:
- (۱) من به چه سطحی از PT و چه تعداد دفعات مشارکت نیاز دارم؟
 - (۲) آیا طرح‌های PT کامل در زمینه‌های مختلف شایستگی فنی وجود دارند؟
 - (۳) آیا طرح PT مرتبط است؟
 - (۴) آیا برگزارکننده PT دارای صلاحیت می‌باشد؟ یعنی آیا برگزارکننده PT بر اساس ISO/IEC 17043 عمل می‌کند؟
 - (۵) آیا طرح PT انتخاب شده مستقل از هر گونه منفعت‌های بازاریابی یا تولید تجهیزات، واکنشگرها یا مواد کالبره‌کننده در زمینه عملیاتی می‌باشد؟

۱.۵ استراتژی مشارکت در PT

آزمایشگاه‌ها قبل از انتخاب طرح PT، می‌بایست سطح و تعداد دفعات مشارکتشان را برآورد کرده و استراتژی مشارکت PT ایجاد کنند. این ارزیابی می‌بایست با لحاظ کردن زمینه‌های گوناگون شایستگی فنی آزمایشگاه انجام شود. سپس آزمایشگاه قادر خواهد بود تا مناسب‌ترین طرح PT را انتخاب کند.

در مواردی که شرکت در PT‌ها از سوی مراکز نظارتی اجباری باشند، آزمایشگاه شانس برای انتخاب مناسب‌ترین طرح PT و تعداد دفعات نخواهد داشت. با این حال، می‌تواند آن را در استراتژی مشارکت خود در طرح PT برای پوشش‌دهی بعضی از زمینه‌هایی که در آنها مشارکت در PT مفید است، لحاظ کند.

زمینه‌های شایستگی فنی (یا حوزه‌های فرعی) یک آزمایشگاه می‌تواند از طریق یک تکنیک اندازه‌گیری، یک ویژگی و یک فرآورده تعریف شود [6]. بعضی از زمینه‌ها ممکن است دارای بیش از یک تکنیک اندازه‌گیری، ویژگی یا محصول باشند، البته به شرطی که معادل‌سازی و مقایسه‌پذیری بتواند نشان داده شود. وقتی آزمایشگاهی در حال ایجاد استراتژی مشارکت در آزمون مهارت می‌باشد، می‌بایست جنبه‌های زیر را لحاظ کند:

(a) آزمایشگاه پس از آنالیز دقیق دیگر اقدام‌های QA انجام شده (به خصوص مواردی که می‌توان آنها را نشان داده و یا به طور کلی بیان کرد، همینطور پیگیری توسعه بایاس یک مقدار بیان شده) می‌بایست سطح و تعداد دفعات مشارکتش را تعریف کند. لذا میزان مشارکت می‌بایست بر اساس دیگر اقدام‌های QA لحاظ شود، هر چند که نمی‌بایست تنها به موارد زیر محدود شوند:

- استفاده منظم از مواد مرجع (گواهی‌دار)
 - مقایسه آنالیز با تکنیک‌های مستقل
 - مشارکت در توسعه متد/ صحت‌گذاری و/ یا مطالعات تعیین ویژگی ماده مرجع
 - استفاده از اقدامات کنترل کیفیت داخلی
 - دیگر مقایسه‌های درون/ بین‌آزمایشگاهی، برای مثال آنالیز نمونه‌های کور درون آزمایشگاه
- (b) سطح خطر معرفی شده از سوی آزمایشگاه، یعنی بخشی که آزمایشگاه در آن درگیر است و یا روش‌شناسی که از آن استفاده می‌کند. این موضوع می‌تواند برای مثال با لحاظ کردن موارد زیر تعیین شود:
- تعداد اندازه‌گیری‌های اجرا شده
 - میزان جایگزینی (جایابی) پرسنل فنی
 - تجربه و دانش پرسنل فنی
 - منبع قابلیت ردیابی (برای مثال دسترس‌پذیری به مواد مرجع، استانداردهای ملی و...)
 - پایداری/ ناپایداری معلوم تکنیک اندازه‌گیری
 - اهمیت و معنادار بودن و استفاده نهایی از داده‌های اندازه‌گیری (برای مثال، علوم جنایی زمینه‌ای را معرفی می‌کند که مستلزم سطح بالای تضمین می‌باشد)

(c) انواع مختلف PT که می‌تواند از سوی آزمایشگاه‌ها مورد استفاده قرار گرفته و از سوی نهادهای تأیید صلاحیت، مراکز نظارتی یا مشتری‌ها نیز پذیرفته شوند، عبارتند از:

- PT سازماندهی شده از سوی سازمان‌های مستقل نظیر نهادهای تأیید صلاحیت یا سازمان‌هایی نظیر ILAC (سازمان تأیید صلاحیت آزمایشگاهی بین‌المللی)، EA (اتحادیه اروپایی برای تأیید صلاحیت)، APLAC (اتحادیه تأیید صلاحیت آزمایشگاهی آسیا و اقیانوسیه) و IRMM (انستیتوی مواد مرجع و اندازه‌گیری‌ها)
- ILC سازماندهی شده با تعداد کافی از آزمایشگاه‌ها به عنوان یک تمرین پیوسته یا تک
- ارائه یک نمونه داخلی یا هدف به دیگر آزمایشگاه‌ها یا تعداد بیشتری آزمایشگاه‌های خارجی به منظور مقایسه اطلاعات

(d) می‌بایست اذعان کرد بخش‌هایی وجود دارند که در آنها مشارکت در PT به خاطر ویژگی‌های فنی اندازه‌گیری، فقدان طرح‌های PT، تعداد کم آزمایشگاه‌های موجود در این بخش و غیره مشکل است. PT ممکن است در بعضی از زمینه‌ها و تنها برای بخش‌هایی از آزمون/ کالیبراسیون انجام شود (برای مثال، آزمون‌های EMC روی اشیای ساده برای تعداد محدودی از کمیت‌های اندازه‌گیری‌شونده) که از لحاظ اقتصادی امکان‌پذیر و عملی می‌باشد. در اینگونه موارد مناسبت دیگر اقدام‌های QA/QC نیز بسیار مهم است.

(e) هر گونه الزام قانونی برای تعداد دفعات نوع مشارکت

در تعیین زمینه شایستگی فنی (حوزه فرعی)، لحاظ کردن یک رویکرد گام‌به‌گام که از خواص تا محصولات به تکنیک اندازه‌گیری می‌پردازد، مفید می‌باشد. زیرا به احتمال زیاد ممکن است چند محصول و یا خواص مرتبط با یک تکنیک اندازه‌گیری در یک حوزه فرعی معین وجود داشته باشند. حالت عکس آن، یعنی وجود چند تکنیک اندازه‌گیری در یک حوزه فرعی معین برای یک محصول یا خواص مرتبط، کمتر دیده می‌شود.

(a) با توجه به تکنیک اندازه‌گیری: احتمال گنجاندن تکنیک‌های اندازه‌گیری مختلف در حوزه فرعی یکسان ممکن است، هر چند که متداول نمی‌باشد.

(b) با توجه به ویژگی اندازه‌گیری‌شونده، تعیین‌شونده یا شناسایی‌شونده: ممکن است گنجاندن بیش از یک ویژگی (پارامتر) در حوزه فرعی یکسان امکان‌پذیر باشد.

(c) با توجه به محصولات اندازه‌گیری‌شونده: ممکن است گنجاندن محصولات مختلف در حوزه فرعی یکسان امکان‌پذیر باشد، به شرطی که ماتریکس‌ها، اشیای یا مواد لحاظ شده و از طبیعت معادل نیز برخوردار باشند.

آزمایشگاه‌ها می‌بایست قادر باشند تا هر جا که لازم است، مباحث فنی را که به تصمیم آزمایشگاه‌ها مبنی بر سطح و تعداد مشارکت آنها در PT منجر می‌شود، مستندسازی و توجیه کنند.

۲.۵ دسترس‌پذیری طرح‌های PT

اطلاعات مربوط به برگزارکنندگان PT و یا دسترس‌پذیری طرح‌های PT را می‌توان از روش‌های مختلف به دست آورد:

(a) پایگاه داده EPTIS (WWW.eptis.bam.de) فهرست صدها طرح PT اجرا شده در سرتاسر جهان را ارائه می‌کند. تمرکز بر زمینه آزمایشی می‌باشد.

(b) نهادهای تأیید صلاحیت می‌توانند جزئیات صلاحیت برگزارکنندگان PT و دامنه کاربردی مرتبطشان را فراهم کنند.

(c) آزمایشگاه‌های همکار که قبلاً در PTها شرکت کرده یا در باره طرح‌های PT مرتبط مطلع هستند.

(d) برگزارکنندگان PT در کشور خود مشارکت‌کننده که احتمالاً اطلاعات مربوط به طرح‌های PT دیگر برگزارکنندگان را خلاصه کرده‌اند.

(e) جستجو در اینترنت با استفاده از کلمات کلیدی مرتبط نیز می‌تواند اطلاعات مفیدی را فراهم کند.

۳.۵ چگونگی تصمیم‌گیری در باره اینکه آیا طرح PT انتخاب شده برازنده هدف می‌باشد یا خیر؟

اگر طرح‌های PT مشابه در دسترس باشند و لازم باشد انتخابی در این زمینه صورت گیرد، آزمایشگاه می‌بایست طرح‌های PT مختلفی که سطوح مختلف برازش هدف را برآورده می‌کنند نیز لحاظ کند، چرا که معمولاً به ندرت طرح PTی که کاملاً برازنده یک آزمایشگاه باشد وجود دارد. لذا، در عمل می‌بایست طرح PTی که بهترین برازش هدف را برآورده می‌کند انتخاب کرد.

معیارهایی که برازش هدف طرح PT را تعیین می‌کنند، می‌بایست شامل موارد زیر باشند، هر چند که نمی‌بایست فقط محدود به سؤال‌های زیر شوند:

- (a) آیا تعداد نمونه‌ها، ماتریکس‌ها و یا سطوح پیشنهادی طرح PT، مشابه موارد مشاهده شده در آزمایشگاه می‌باشند؟
- (b) آیا طراحی آماری پیشنهاد شده؟ و آیا طرح متدهای اندازه‌گیری مختلف را در نظر می‌گیرد؟
- نکته ۱: طراحی آماری، فرایند برنامه‌ریزی، جمع‌آوری، آنالیز و گزارش داده‌های طرح PT را شامل می‌شود.
- نکته ۲: بسیاری از برگزارکنندگان PT، مثال‌هایی از گزارش‌های PT و یا کپی‌های پروتوکل طرح را در وبسایت‌هایشان دارند که در آنها امکان مشاهده نحوه ارزیابی عملکرد (امتیازدهی) استفاده شده از سوی برگزارکننده وجود دارد.
- (c) آیا تعداد و خاستگاه مشارکت‌کننده‌ها برای طرح PT مناسب است؟
- (d) آیا تعداد دورها کافی است؟
- (e) آیا برگزارکنندگان PT با الزامات ISO/IEC 17043 مطابقت دارند؟
- نکته: تأیید صلاحیت برگزارکننده PT از سوی یک نهاد تأیید صلاحیت ملی، اطمینان بیشتری از شایستگی برگزارکننده‌ها فراهم می‌کند.
- (f) طرح‌های PT همچنین دارای نقش آموزشی مهمی می‌باشند. اگر برگزارکننده PT داده‌ها را بر اساس متدهای آنالیزی مختلف تیمار کند و نتایج عدم قطعیت گزارش شده از سوی مشارکت‌کنندگان را هم تجزیه و تحلیل کند، در این صورت اطلاعات ارزشمندی برای ارزیابی آتی متدهای آنالیزی فراهم خواهد شد.
- توجه به این نکته مهم است که این مسئولیت آزمایشگاه است تا در مورد بررسی معیارهای مربوطه، انجام مقایسه و قضاوت در مورد ارتباط و متناسب بودن طرح PT تصمیم‌گیری کند. فرم ساده پیوست A می‌تواند برای اجرای نظام‌مند این فرایند انتخاب مفید باشد.
- چنانچه طرح PT و نیازهای آزمایشگاه به طور چشمگیری متفاوت از یکدیگر باشند، آنگاه آزمایشگاه می‌بایست ادامه مشارکت را به طور جدی بررسی کند. تعداد زیادی از برگزارکنندگان PT تنها اجازه مشارکت در یک دور را می‌دهند. لذا یک دلیل خوب برای مشارکت در طرح PT وجود ارتباط منطقی آن و نیازهای آزمایشگاه است. بعضی اوقات امکان خرید اقلام PT از دوره‌های قبلی یک طرح PT همراه با گزارش PT امکان‌پذیر می‌باشد. این نکته نیز یک گزینه خوب برای قضاوت در مورد ارتباط یک طرح PT خاص می‌باشد.
- از آنجا که بعضی از برگزارکنندگان PT مشارکت انعطاف‌پذیر نیز ارائه می‌دهند، چنانچه آزمایشگاه بتواند تعداد دفعات مشارکت خود را انتخاب کند، می‌بایست پایداری و بحرانی بودن متد، و ظرفیت پذیرش نمونه‌های اندازه‌گیری روتین را نیز در آزمایشگاه لحاظ نماید.

۶. استفاده از PT از سوی آزمایشگاه‌ها

استفاده اصلی PT برای آزمایشگاه، ارزیابی عملکرد اجرای اندازه‌گیری‌ها یا کالیبراسیون‌های خاص می‌باشد. نتایج و اطلاعات به دست آمده از مشارکت در طرح‌های PT، تأیید خواهند کرد که آیا عملکرد آزمایشگاه رضایت‌بخش است یا خیر و دیگر آنکه نتایج آن می‌تواند به شکل یک شاخص برای وجود مشکلات بالقوه مطرح شده و بر اساس آن تصمیمات لازم اتخاذ شوند.

با این حال، کاربرد نتایج PT خیلی گسترده‌تر از تعیین صلاحیت آزمایشگاه است. آزمایشگاه‌ها، همانطور که در بخش ۱.۴ ذکر

شد، به شکل‌های زیادی می‌توانند از مشارکت در طرح‌های PT سود ببرند [8] که به بعضی از آنها در ذیل اشاره می‌شود:

۱. شناسایی مشکلات/اندازه‌گیری (به عنوان مدیریت خطر و ابزار بهبود عملکرد)

اگر نتیجه یک آزمایشگاه در طرح PT به عملکرد ناراضیت‌بخش اشاره داشته باشد، می‌بایست فرایند بررسی منابع بالقوه خطا شروع شود. بدون مشارکت در طرح PT، چنین منابع خطایی می‌توانستند بدون تشخیص باقی مانده و آزمایشگاه هم قادر نبود تا اقدامات اصلاحی مناسب را انجام دهد. این وضع به نوبه خود باعث می‌شد تا آزمایشگاه به طور پیوسته نتایج ضعیفی را برای مشتری‌ها یا دیگر سهامداران تولید کند.

سرانجام این که، چنین خطاهایی می‌توانستند به خدشه‌دار شدن شهرت آزمایشگاه یا از دست رفتن اعتبار قانونی آن منجر شده و یا ممکن بود عکس‌العمل مشتری‌ها یا دیگر سهامداران نظیر مجموعه‌های نظارتی را نیز در پی داشته باشد. با توجه به این موارد، استفاده از PT می‌تواند به عنوان یک مقوله مدیریت خطر و ابزار بهبود کیفیت در نظر گرفته شود.

۲. مقایسه متدها یا روش‌های اجرایی

برای بعضی آزمایشگاه‌ها، مشارکت ممکن است با استفاده از یک اندازه‌گیری اجرا شده جدید یا نامنظم برای امتحان عملکرد آنها در نظر گرفته شود. در دیگر موارد، مشارکت ممکن است فرصتی را برای مقایسه نتایج به دست آمده از آزمایشگاه، ضمن استفاده از متدهای مختلف (یا سطوح غلظت مختلف و غیره) با نتایج متد نرمال استفاده شده از سوی همان آزمایشگاه فراهم کند. در بعضی موارد خود طرح PT ممکن است خلاصه‌ها و یا مقایسه‌های تمام متدهای آزمایشگاه‌ها یا کیت‌های تجاری، برای مثال روش اجرایی هوشیاری EU مرتبط به ابزارهای تشخیص پزشکی درون لوله آزمایش [9] را فراهم کند. یک چنین داده‌هایی برای فعالیت‌های غیرمعمول یا جدید، بسیار ارزشمند بوده و به انتخاب روش‌شناسی مناسب از سوی آزمایشگاه در آینده کمک کرده و نشان می‌دهند که قبل از پذیرش متدهای جدید آیا به بررسی اضافی نیاز می‌باشد یا خیر.

۳. مقایسه قابلیت‌های اپراتور

وقتی اقلام PT کافی برای بیش از یک اپراتور در یک آزمایشگاه در دسترس باشند، آزمایشگاه قادر خواهد بود تا نتایج بین اپراتورهایش را مقایسه کند. این کار نه تنها کمک می‌کند تا آزمایشگاه عملکرد اپراتورهایش را مقایسه کند، بلکه ورودی‌هایی نیز برای تخمین‌های آزمایشگاهی عدم قطعیت اندازه‌گیری برای اندازه‌گیری‌های مرتبط فراهم می‌کند.

بدین ترتیب آزمایشگاه می‌تواند تا تکرارپذیری بین - اپراتور به دست آمده از سوی آزمایشگاه را با داده‌های منتشر شده (یا داده‌هایی که به شکل دیگر در دسترس باشند) برای اندازه‌گیری‌های مربوطه ارزیابی کند.

طرح PT ممکن است در بعضی موارد به شکلی باشد که نتایج از سوی بیش از یک اپراتور گزارش شوند.

۴. مقایسه سیستم‌های آنالیزی

نتایج PT می‌توانند یک ارزیابی خارجی عینی از عملکرد نسبی سیستم‌های تجزیه‌ای (در مکان‌های یکسان یا مختلف) استفاده شده درون یک سازمان آزمایشگاهی فراهم کنند.

۵. بهبود عملکرد

وقتی آزمایشگاه از نتایج خود در طرح PT راضی نباشد، در این صورت برای مدیریت آزمایشگاه فرصتی فراهم خواهد شد تا بهبود اندازه‌گیری‌های آتی را در دستور کار خود قرار دهد. این کار ممکن است برای مثال شامل آموزش اضافی اپراتور، پذیرش متدهای جدید یا اصلاح شده، افزایش کنترل کیفیت داخلی داده‌ها، اصلاح تجهیزات، کالیبراسیون یا جایگزینی و غیره باشد.

۶. آموزش پرسنل

بسیاری از طرح‌های PT فراهم نمودن اطلاعات روش‌شناسی، تفسیر داده‌ها، تعیین‌های عدم قطعیت و غیره که از نتایج کلی طرح PT و یا از سوی کارشناسان درگیر در ارزیابی نتایج مطرح می‌شوند را به عنوان یکی از اهداف خود لحاظ می‌کنند. بعضی از طرح‌های PT دارای نقش آموزشی جامع برای مشارکت‌کننده‌ها و اپراتورها می‌باشند.

۷. تبادل اطلاعات با برگزارکننده PT

آزمایشگاه‌ها پس از صدور گزارش PT، معمولاً می‌توانند جهت کسب اطلاعات اضافی در مورد نتایج یا توصیه در مورد علل بالقوه نتایج نارضايت بخش با برگزارکننده PT تماس داشته باشند. بعضی از برگزارکنندگان PT «نشست‌های مشارکت کنندگان» را برگزار می‌کنند که اطلاعات بسیار مفیدی را برای آزمایشگاه‌ها فراهم می‌کند.

۸. ایجاد اطمینان در پرسنل، مدیریت و کاربران خارجی خدمات آزمایشگاهی

عملکرد موفقیت‌آمیز در طرح PT می‌تواند در تک‌تک پرسنل و مدیران مستقیم آنها نوعی اطمینان اضافی ایجاد کند. دیگر مدیریت‌ها، شامل مدیریت‌های بدون تخصص فنی مرتبط نیز می‌توانند به عملکرد موفقیت‌آمیز پرسنل آزمایشگاه‌هایشان در اندازه‌گیری‌های با اهمیت بحرانی برای فعالیت‌ها و مسئولیت‌های اجرایی سازمان اعتماد کامل داشته باشند. کاربران خارجی خدمات آزمایشگاهی شامل مشتری‌ها و طرف‌های درگیر در نتایج اندازه‌گیری نیز وقتی مطلع باشند که آزمایشگاه حاضر است تا عملکرد اندازه‌گیری‌اش را به طور منظم از طریق طرح‌های PT ارزیابی کند، به آن آزمایشگاه اعتماد بیشتری خواهند داشت.

عملکرد موفقیت‌آمیز یک آزمایشگاه در طرح PT (یا تصحیح مؤثر مشکلات اندازه‌گیری آن پس از یک عملکرد همراه با عدم موفقیت) برای نهادهای تأیید صلاحیت و نظارتی این اطمینان را به وجود می‌آورد که داده‌های این آزمایشگاه مورد تأیید بوده و از رسمیت قانونی برخوردار می‌باشند. مزیت آشکار تداوم موفقیت‌ها برای آزمایشگاه‌ها حفظ جایگاه آنها به عنوان سازمان‌های دارای شایستگی و صلاحیت می‌باشد.

با این حال، مزایای داخلی آن برای آزمایشگاه‌ها، پرسنل و مدیریت نیز بسیار ارزشمند است. البته چنانچه نگاه آنها به PT به عنوان یک ابزار مهم برای ثبات اطمینان در نتایج آزمایشگاهی و بهبود آن باشد. در این راستا حضور در PT آن هم صرفاً به منظور اهداف تأیید صلاحیتی اهمیت چندانی نخواهد داشت.

۹. عدم قطعیت اندازه‌گیری

نتایج مشارکت آزمایشگاه در PT می‌تواند با احتیاط برای بازبینی عدم قطعیت اندازه‌گیری استفاده شود، چرا که عدم قطعیت تخمینی می‌بایست با پراکندگی نتایج به دست آمده از سوی آن آزمایشگاه در طی چند دور مشارکت در PT سازگار باشد.

«رویکرد PT» می‌تواند در موارد خاص برای برآورد عدم قطعیت نیز استفاده شود. برای مثال، چنانچه تمامی مشارکت کنندگان از متد یکسان در طرح PT استفاده کنند، انحراف استاندارد بدست آمده، معادل با تخمین تجدیدپذیری است که می‌تواند به عنوان انحراف استاندارد تجدیدپذیری به دست آمده در یک مطالعه مشترک نیز استفاده شود [10]، [11]. در پیوست H، یک مثال در مورد چگونگی تصدیق عدم قطعیت اندازه‌گیری آزمایشگاه از طریق مشارکت در طرح PT وجود دارد.

۱۰. استفاده از اقلام PT به عنوان کنترل‌های کیفیت داخلی

در بعضی از طرح‌های PT، که در آن مواد پایدار و کافی برای مشارکت کنندگان فراهم می‌شود، ماده استفاده نشده می‌تواند برای پایش کنترل کیفیت داخلی اندازه‌گیری‌ها و به صورت یک ماده مرجع مفید استفاده شود.

هر جا که مناسب باشد می‌توان از مقادیر مرجع نسبت داده شده به آیت PT (یا مقادیر توافقی به دست آمده هنگام طرح PT) به عنوان مقادیر مرجع داخلی برای کنترل کیفیت اندازه‌گیری، آموزش اپراتور و غیره استفاده کرد.

۱۱. تصدیق عملکرد متد

بعضی از طرح‌های PT، بسته به طراحی، در تعیین دقت (تکرارپذیری و تجدیدپذیری) یا درستی مقایسه‌ای متدهای استفاده شده در طرح PT مفید می‌باشند. در بیشتر موارد، تعیین دقت و درستی متدها هدف اولیه طرح PT به شمار نمی‌آید. برای انجام این کار، اغلب اطلاعات بیشتری لازم است که می‌توان آنها را از برگزارکننده PT به دست آورد.

۷. چطور برگزارکننده PT عملکرد آزمایشگاه را برآورد می‌کند؟

۱.۷ مقدمه

نتایج طرح‌های PT به شکل‌های زیادی تولید شوند که گستره‌ی وسیعی از انواع داده‌ها و توزیع‌های آماری اصلی را در بر می‌گیرند. بنابراین، هدف این بخش معرفی جنبه‌های اصلی طرح آماری استفاده شده از سوی برگزارکنندگان PT می‌باشد تا بدین شکل آزمایشگاه‌ها درک بهتری از ارزیابی‌های عملکرد داشته باشند. لذا این بخش در انتخاب طرح مناسب و تفسیر نتایج به آزمایشگاه کمک می‌کند. با این حال، به خاطر گستره‌ی تکنیک‌های مختلف استفاده شده، بررسی تمام جنبه‌های آماری در این سند امکان‌پذیر نمی‌باشد. مهم است که طرح آماری استفاده شده از سوی برگزارکننده PT برای نوع و هدف طرح PT سازماندهی شده مناسب باشد. به علاوه، طرح استفاده شده از سوی برگزارکننده PT می‌بایست به طور کامل برای مشارکت‌کننده‌ها توضیح داده شود. تکنیک‌های آماری ترجیحی در ISO 13528 توصیف شده‌اند [12]. گرچه دیگر رویکردهای معتبر نیز می‌توانند استفاده شوند، با این حال فرض‌های زیربنایی رویکرد آماری استفاده شده در طرح‌های PT اغلب بر اساس توزیع نرمال داده‌ها پایه‌گذاری شده‌اند. هر چند که معمولاً برای مجموعه‌ای از نتایج مشارکت‌کنندگان و توزیع اساساً نرمال نتایج، امکان مشاهده دنباله‌های سنگین بر اثر آلودگی و تعدادی داده انحرافی (پرت) وجود دارد. در گذشته رویکرد اصلی استفاده شده از سوی برگزارکنندگان PT (که هنوز هم در بعضی طرح‌های PT استفاده می‌شود) استفاده از آزمون‌های آماری برای شناسایی حضور داده‌های انحرافی در مجموعه داده‌ها بود. با این حال، رویکرد متداول امروزی که از سوی برگزارکنندگان PT استفاده می‌شود، همانطور که در ISO 13528 پیشنهاد شده [12]، استفاده از آماره‌های استوار است [13,14]. آماره استوار این مزیت را دارد که سهم داده‌های انحرافی برای محاسبه پارامترهای آماری نظیر میانگین و انحراف استاندارد را کاهش می‌دهد. تعداد زیادی از رویکردهای آماری استوار وجود دارند که بعضی از آنها در ISO 13528 [12] توصیف شده‌اند.

۲.۷ عناصر اصلی در برآورد نتایج PT

یکی از عناصر اصلی در تمام PTها، برآورد عملکرد هر مشارکت‌کننده است که برای انجام آن، برگزارکننده PT می‌بایست برای برآورد عملکرد اساساً دو مقدار را ایجاد کند که عبارتند از:

۱. مقدار نسبت داده شده (یا مقدار اسمی)

۲. انحراف استاندارد برای ارزیابی مهارت

به علاوه از برگزارکننده PT انتظار می‌رود تا تخمینی از عدم قطعیت اندازه‌گیری مقدار اسمی ارائه کرده و قابلیت ردیابی اندازه‌شناسی آن را نیز بیان کند، مفهومی که در ISO/IEC 17043 نیز گنجانده شده است [4]. عملی بودن، وابستگی و الزام این تخمین نیز می‌بایست در طراحی آماری طرح PT تعیین شود.

از متدهای مختلفی برای ایجاد این مقادیر استفاده می‌شود [12,15]. هیچ پروتوکل استاندارد صریح و روشنی وجود ندارد که دقیقاً جزئیات طرح آماری استفاده شده را توصیف کند، با این حال، این طرح می‌بایست با طرح‌های توصیفی مستندات مرجع توافق داشته باشد. طرح آماری می‌بایست به شکل مستند از سوی برگزارکننده PT فراهم شود، این طرح معمولاً در پروتوکل طرح PT و یا در گزارش PT ارائه شده و هنگام انتخاب یک طرح PT می‌بایست لحاظ شود.

۱.۲.۷ مقدار اسمی

همانطور که در ISO 13528 [12] توصیف شده، به طور کلی پنج متد برای رسیدن به مقدار اسمی که در واقع همان تخمین کاری مقدار واقعی است، در دسترس می‌باشد:

۱. فرمولاسیون

۲. مقادیر مرجع گواهی‌دار

۳. مقادیر مرجع

۴. مقادیر توافقی از سوی آزمایشگاه‌های تخصصی

۵. مقدار توافقی از سوی مشارکت‌کنندگان

توضیح مقدار اسمی در پیوست D آمده است.

۲.۲.۷ انحراف استاندارد برای ارزیابی مهارت

همانطور که در ISO 13528 [12] توضیح داده شده، به طور کلی پنج رویکرد برای تعیین انحراف استاندارد برای ارزیابی مهارت، یعنی گستره‌ی قابل قبول نتایج مشارکت‌کنندگان وجود دارد:

۱. مقدار تجویز شده

۲. برداشت ذهنی

۳. مدل عمومی

۴. نتایج آزمایش دقت

۵. داده‌های به دست آمده در یک دور از طرح PT

توصیف انحراف استاندارد برای ارزیابی مهارت در پیوست D ارائه شده است.

در حال حاضر روش معمول برای ایجاد یک مقدار اسمی و انحراف استاندارد ارزیابی مهارت، استفاده از نتایج مشارکت‌کنندگان در PT است. با این حال، قویاً در پروتوکل یکسان‌سازی برای PT در شیمی تجزیه که از سوی آیوپاک منتشر شده [16]، پیشنهاد می‌شود که متدهای امتیازدهی می‌بایست بر اساس معیار برازش هدف که از قبل از سوی برگزارکننده PT در کاربرد خاص و بر اساس شرایط ویژه تعیین لحاظ شده‌اند، پایه‌گذاری شوند. لذا هر جا که امکان داشته باشد برگزارکننده PT به جای استفاده از مقداری که ممکن است از دوری به دور بعد، بسته به پراکندگی نتایج ارائه شده از سوی مشارکت‌کنندگان تغییر کند، می‌بایست انحراف استاندارد ارزیابی هدف را بر اساس مقدار برازش هدف پایه‌گذاری کند. استفاده از مقدار برازش هدف، پایش نمره‌های عملکردی را طی دوره‌های بعدی طرح PT آسان‌تر خواهد کرد.

۲.۲.۷ ارزیابی عملکرد

در برآورد عملکرد (یا امتیاز) ارائه شده از سوی برگزارکننده PT، اعتبار نتایج تجزیه‌ای خام تولید شده از سوی مشارکت‌کنندگان مشخص می‌شود. هدف ارزیابی عملکرد نرمالیزه شده این است که تمام نتایج PT با یکدیگر مقایسه شوند، بدین ترتیب مشارکت‌کننده به راحتی معنای ارزیابی را درک خواهد کرد.

استفاده از عدم قطعیت اندازه‌گیری در برآورد عملکرد به خاطر افزایش درک مفهوم عدم قطعیت در حال افزایش است. دو نوع عدم قطعیت اندازه‌گیری را می‌توان لحاظ کرد:

۱. عدم قطعیت اندازه‌گیری مقدار اسمی

۲. عدم قطعیت اندازه‌گیری نتیجه مشارکت‌کننده

با در نظر گرفتن اهداف مختلف طرح PT، امکان تعریف یک متد ارزیابی عمومی تک وجود ندارد. بنابراین، تعداد زیادی از طرح‌های آماری برای ارزیابی عملکرد در دسترس هستند. فهرست متداول‌ترین آنها در ادامه و در پیوست E ارائه شده است. دیگر طرح‌های آماری که در این سند پوشش‌دهی نشده‌اند را می‌توان در ISO 13528 [12] مطالعه کرد.

(a) «نمره‌ی Z» (که معمولاً استفاده شده و عدم قطعیت اندازه‌گیری در آن لحاظ نمی‌شود)

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

در اینجا:

x : نتیجه گزارش شده از سوی مشارکت کننده

X : مقدار اسمی

$\hat{\sigma}$: انحراف استاندارد برای ارزیابی مهارت

(b) «نمره ی Z' » (عدم قطعیت استاندارد مقدار اسمی لحاظ می‌شود)

$$Z' = (x - X) / \sqrt{\hat{\sigma}^2 + u_x^2}$$

در اینجا:

x : نتیجه گزارش شده از سوی مشارکت کننده

X : مقدار اسمی

$\hat{\sigma}$: انحراف استاندارد برای ارزیابی مهارت

u_x : عدم قطعیت استاندارد مقدار اسمی X

(c) «نمره ی Z » (عدم قطعیت استاندارد مقدار اسمی و نتیجه مشارکت کنندگان لحاظ می‌شود)

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

در اینجا:

u_x : تخمین خود مشارکت کننده از عدم قطعیت استاندارد نتیجه x مربوط به خودش

u_X : عدم قطعیت استاندارد مقدار اسمی X

(d) «عدد E_n » (عدم قطعیت بسط یافته مقدار اسمی و نتیجه مشارکت کنندگان لحاظ می‌شود)

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

در اینجا:

U_x : عدم قطعیت بسط یافته نتیجه x مشارکت کننده

U_{ref} : عدم قطعیت بسط یافته مقدار اسمی X تعیین شده در آزمایشگاه مرجع

قضاوت‌های زیر معمولاً برای نمره‌های Z و Z' و Z در نظر گرفته می‌شوند:

(a) نمره ی $|Z| \leq 2$ ، به عملکرد «رضایت بخش» اشاره داشته و پیامی را ایجاد نمی‌کند.

(b) نمره ی $2 < |Z| < 3$ ، به عملکرد «سؤال برانگیز» اشاره داشته و یک پیام هشدار ایجاد می‌کند.

(c) نمره ی $|Z| \geq 3$ ، به عملکرد «تارضایت بخش» اشاره داشته و یک پیام اقدام ایجاد می‌کند.

قضایات‌های زیر نیز معمولاً برای اعداد E_n در نظر گرفته می‌شوند:

(a) نمره‌ی $|E_n| \leq 1$ ، به عملکرد «رضایت‌بخش» اشاره داشته و به پیامی منجر نمی‌شود.

(b) نمره‌ی $|E_n| > 1$ ، به عملکرد «نارضایت‌بخش» اشاره داشته و به پیام اقدام منجر می‌شود.

اساس ارزیابی می‌بایست در دوره‌های مختلف یک طرح PT یکسان و منسجم باشد، بدین ترتیب نمره‌های به دست آمده در دوره‌های بعدی با یکدیگر قابل مقایسه خواهند بود. تنها بدین ترتیب است که یک مشارکت‌کننده متوجه روند طولانی‌مدت عملکرد خود خواهد شد.

۴.۲.۷ اثر عدم قطعیت مقدار اسمی

عدم قطعیت استاندارد مقدار اسمی به مدت استفاده شده برای به دست آوردن آن بستگی دارد، و چنانچه از اندازه‌گیری‌های مختلف در چند آزمایشگاه به دست آمده باشد، به تعداد آزمایشگاه‌ها نیز بستگی خواهد داشت.

متدهای محاسبه عدم قطعیت استاندارد مقدار اسمی در ISO 13528 [12] بررسی شده است.

اگر عدم قطعیت استاندارد (u_x) مقدار اسمی در مقایسه با انحراف استاندارد ارزیابی مهارت خیلی بزرگ باشد، آنگاه این خطر وجود دارد که بعضی از آزمایشگاه‌ها به خاطر نادرستی در تعیین مقدار اسمی و نه به خاطر علل درون آزمایشگاه‌ها، عملکرد نارضایت‌بخش یا سؤال‌برانگیز دریافت کنند.

به همین خاطر، می‌بایست عدم قطعیت استاندارد مقدار اسمی ایجاد شده و برای آزمایشگاه‌های مشارکت‌کننده در طرح PT گزارش شود.

اگر $u_x \leq 0.3\hat{\sigma}$ [16] باشد، آنگاه عدم قطعیت استاندارد مقدار اسمی قابل چشم‌پوشی بوده و نمی‌بایست در تفسیر نتایج آزمون مهارت لحاظ شود.

اگر معیار بالا برآورده نشود، آنگاه برگزارکننده PT می‌بایست یکی از کارهای زیر را انجام دهد:

(a) استفاده از متد متفاوت برای تعیین مقدار اسمی، طوری که عدم قطعیت آن، معیار بالا را برآورده سازد.

(b) استفاده از عدم قطعیت مقدار اسمی در تفسیر نتایج آزمون مهارت (نمره Z ، زتا و عدد E_n را در بالا ببینید)

(c) اطلاع‌رسانی به مشارکت‌کنندگان آزمون مهارت که عدم قطعیت مقدار اسمی قابل چشم‌پوشی نیست.

۵.۲.۷ طرح‌های PT کیفی

برای این طرح‌های PT، ارزیابی آماری پذیرفته شده مرسوم و قابل قبولی وجود ندارد، مگر اینکه خود برگزارکننده PT نمرات عملکرد را با مقایسه نتایج آزمایشگاه‌ها نسبت به مقدار اسمی و تبدیل نتایج کیفی به داده‌های قابل اندازه‌گیری بر اساس معیارهای از پیش تعیین شده، ایجاد کند.

اگر نمره‌های عملکرد ایجاد نشده باشند، نتایج عمدتاً به شکل نتایج «بله/خیر» یا نتایج «تشخیص داده شد/تشخیص داده نشد» بیان خواهند شد. در حال حاضر برای این نوع نتایج، هنوز رویکرد مرسوم و شناخته شده‌ای وجود ندارد.

۶.۲.۷ داده‌های انحرافی

یک داده انحرافی، مشاهده‌ای است که از نظر عددی از بقیه داده‌ها فاصله دارد. داده‌های انحرافی می‌توانند به شکل تصادفی در هر توزیع مشاهده شوند، و اغلب بیانگر خطای اندازه‌گیری و یا جمعیتی با توزیع دُم بلند است. در حالت خطای اندازه‌گیری، شخص می‌تواند با استفاده از آماره‌هایی که نسبت به داده‌های انحرافی استوار هستند، آنها را کنار بگذارد، در حالی که در توزیع دُم بلند، داده‌های انحرافی بیانگر این نکته هستند که توزیع، پراکندگی بالایی داشته و شخص در استفاده از ابزارها یا شواهدی

که توزیع نرمال را فرض می‌کنند، می‌بایست خیلی محتاط باشد. یک علت معمول مشاهده داده‌های انحرافی، مخلوط دو توزیع است که ممکن است دو زیرجمعیت متمایز از هم باشند، از سوی دیگر مشاهده این وضع ممکن است بیانگر «آزمون صحیح» در برابر «خطای اندازه‌گیری» باشد، که این حالت نیز با یک مدل مخلوط، مدل‌سازی می‌شود.

برگزارکننده PT می‌بایست روش تیمار داده‌های انحرافی در آنالیز آماری را توضیح دهد.

۸. تفسیر آزمایشگاهی نتایج

مشارکت در طرح PT به خودی خود ارزش چندانی ندارد، مگر اینکه آزمایشگاه بتواند از ارزیابی عملکرد و اطلاعات عمومی ارائه شده در گزارش طرح PT استفاده کند. آزمایشگاه نه تنها می‌بایست ارزیابی عملکرد به دست آمده را بپذیرد، بلکه خود نیز می‌بایست آن را ارزیابی و تفسیر کند، و از ارائه هر گونه تفسیرهای غلط و یا تفسیرهای نابجا نیز اجتناب نماید.

می‌بایست برای طرح‌های پیوسته، بعد از هر دور PT، ارزیابی عملکرد آزمایشگاهی و ارزیابی عملکرد زمانی انجام شود.

۱۰.۸ ارزیابی عملکرد از سوی آزمایشگاه

تفسیر عملکرد PT، به تمام سطوح مدیریت آزمایشگاه از اپراتور گرفته تا مدیریت ارشد مربوط می‌شود. پرسنل مسئول اندازه‌گیری می‌بایست با عملیات طرح PT آشنا بوده و معمولاً کار خود را با ارزیابی اولیه شروع کنند.

اگر قرار است بررسی‌هایی انجام شوند، می‌بایست در روش اجرایی عدم انطباق سیستم مدیریت کیفیت آزمایشگاه تیمار شوند. ممکن است مدیریت ارشد همیشه با عملکرد PT آشنا نباشد، در این صورت به شدت توصیه می‌شود تا حداقل با مفاهیم پایه PT آشنایی داشته باشند. از آنجا که آزمایشگاه می‌بایست از متدهای صحه‌گذاری شده همراه با کنترل‌های کیفیت داخلی استفاده کند، لذا هر گونه عملکرد ضعیف می‌بایست جدی گرفته شود، چرا که ممکن است مشکلی در رابطه با صحه‌گذاری و یا کنترل‌های کیفیت داخلی وجود داشته باشد.

در باره تفسیر نتایج PT نکاتی کلیدی وجود دارند که یادآوری آنها قبل از ورود به جزئیات این موضوع می‌تواند مفید باشد. همانطور که قبلاً بیان شد، PT صرفاً به «قبول» یا «رد» یک اندازه‌گیری نمی‌پردازد و درباره‌ی یادگیری به کمک نتایج نیز نکاتی را به ما می‌آموزد. یک عملکرد رضایت‌بخش در یک دور برای یک آزمایشگاه که در آن تمام آزمایشگاه‌ها عملکرد رضایت‌بخش دارند، لزوماً به معنای سطح بالای شایستگی و مهارت نیست، در این حالت، انحراف استاندارد ارزیابی مهارت می‌تواند خیلی بزرگ باشد. به همین ترتیب، عملکرد نارضایت‌بخش در یک دور لزوماً به این معنا نیست که آزمایشگاه فاقد صلاحیت است، بلکه می‌بایست این نتیجه مطالعه و درس‌هایی از آن گرفته شود تا دوباره شاهد عملکرد نارضایت‌بخش نباشیم. با این حال، عملکرد ضعیف پیوسته به وجود مشکلات عمده در آزمایشگاه‌ها و پروسه اندازه‌گیری اشاره داشته و اگر این حالت اتفاق بیفتد آزمایشگاه در مورد اینکه آیا می‌بایست این روش را برای آن اندازه‌گیری خاص تا زمان حل مشکلات ادامه دهد یا نه، توجه جدی داشته باشد.

۱۰.۸ بازبینی نتایج یک دور PT

نتایج هر دور PT، می‌بایست بدون توجه به عملکرد به دست آمده ارزیابی شوند، چرا که یک نتیجه رضایت‌بخش لزوماً به معنای عملکرد خوب نیست.

تمام اطلاعات موجود در گزارش PT و نه فقط نمره عملکرد می‌بایست ارزیابی شوند. برای مثال، عملکرد نارضایت‌بخش در یک دور از PT، که در آن اکثریت مشارکت‌کنندگان تا سطح رضایت‌بخش عمل کرده‌اند، می‌بایست با عملکرد نارضایت‌بخش که در آن تعداد زیادی از مشارکت‌کنندگان عملکرد نارضایت‌بخش داشته‌اند، کنار هم قرار گرفته و مقایسه شوند. هر دو وضعیت می‌بایست دقیقاً بازبینی شوند، چرا که هر دو به وجود مشکلات در فرایند اندازه‌گیری اشاره دارند. به عنوان بخشی از فرایند بازبینی، پرسنل آزمایشگاه همیشه می‌بایست کنترل کنند که نتایج گزارش PT، همان نتایجی هستند که از سوی آزمایشگاه

ارائه شده‌اند، به خصوص زمانی که سیستم نمره‌دهی عملکرد استفاده شده در طرح کاملاً روشن بوده و برازنده هدف باشد. اگر لازم باشد جهت اجتناب از هر گونه تفسیر غلط عملکرد، می‌بایست با برگزارکننده PT تماس گرفته شود.

اگر توجیه‌پذیر باشد، آزمایشگاه می‌تواند جهت ارزیابی مهارت، با استفاده از انحراف استاندارد برازنده‌تر برای هدف، دومرتبه نمره‌ی عملکرد خود را محاسبه کند (پیوست F.2). اگر پس از بررسی دقیق، آزمایشگاه نتیجه‌گیری کند که نتیجه در واقع نارضایت‌بخش است، آنگاه اقدامات اصلاحی می‌بایست انجام شوند (بخش ۲.۸ را ببینید).

نتایج آزمایشگاهی حاصل از مشارکت در PT می‌تواند برای کنترل اعتبار عدم قطعیت اندازه‌گیری آزمایشگاه نیز استفاده شود.

۲.۱.۸ پایش عملکرد PT با گذشت زمان

پس از ارزیابی دقیق نتایج یک دور، جهت شناسایی مشکلات احتمالی مرتبط با بی‌دقتی، خطای سیستماتیک یا خطای انسانی با گذشت زمان، می‌بایست پایش عملکرد PT انجام شود.

جهت پایش عملکرد PT تهیه یک نمودار گرافیکی نمرات عملکرد از دوری به دور بعد بسیار مفید است. این نمودار اغلب از سوی برگزارکننده PT در گزارش PT ارائه می‌شود و گاهی خود مشارکت‌کننده هم می‌تواند آن را رسم کند. این رویکرد باعث می‌شود تا نتایج غیرقابل انتظار یا غیرمعمول بیشتر برجسته شده و به شناسایی روندها نیز کمک می‌کند. معمولاً از طریق روش‌های کنترل کیفیت داخلی آزمایشگاه (IQC) روندهای مرتبط مورد انتظار برای مثال، کالیبراسیون یا نگهداری دستگاهی نادرست، یا استفاده از واکنشگرهای نامناسب تشخیص داده می‌شوند و پایش عملکرد PT در طی زمان نیز می‌تواند به عنوان یک سیستم مکمل برای IQC به کار گرفته شود. جهت تصمیم‌گیری درباره اینکه آیا عملکرد با گذشت زمان در حال بهبود یا کاهش می‌باشد، داده‌های حاصل از دورهای بعدی می‌بایست قابل مقایسه باشند.

با این حال، مجموعه داده‌های حاصل از اندازه‌گیری یکسان از دو دور PT مختلف ممکن است دارای انحراف استاندارد متفاوتی برای ارزیابی مهارت باشند و بسته به گروه آزمایشگاه‌های مشارکت‌کننده همینطور تأثیر دیگر متغیرهای حاصل از نمونه‌های پیچیده یا کمپلکس ممکن است به نمرات عملکردی منجر شوند که به شکل متفاوت محاسبه شده‌اند. از سوی دیگر مشارکت‌کنندگان ممکن است نمرات Z خودشان (یا دیگر نمرات عملکردی) را با استفاده از انحراف استاندارد انتخابی برای ارزیابی مهارت محاسبه کنند، البته اگر مقدار استفاده شده در دورهای متوالی به طور چشمگیری متفاوت از یکدیگر باشند، یک انحراف استاندارد به دست آمده از مقالات علمی (برای مثال، به دست آمده از یک متد استاندارد شبیه ISO، EN، ASTM، DIN) می‌تواند برای ارزیابی مهارت استفاده شود. اگر این مقادیر از مقالات علمی در دسترس نباشند، معیارهای خود آزمایشگاه، بسته به هدف مشارکت در PT یا اهمیت اندازه‌گیری (هر مقدار واقعی می‌تواند استفاده شود، برای مثال ۱۰٪ از مقدار اسمی) می‌توانند انتخاب شوند.

توجه داشته باشید که لازم نیست انحراف استاندارد انتخاب شده برای ارزیابی مهارت، حتماً ثابت باشد، چرا که ممکن است وابسته به غلظت نیز باشد. اگر آزمایشگاهی تصمیم بگیرد که نمرات عملکردی خودش را دوباره محاسبه کند، می‌بایست این قضیه را توجیه کرده و انتخاب خود را به شکل مستند نشان دهد.

مثال‌هایی درباره چگونگی پایش عملکرد آزمایشگاه تک طی زمان در پیوست G ارائه شده است.

۲.۸ بررسی نتایج PT سؤال‌برانگیز یا نارضایت‌بخش

۱.۲.۸ ضرورت بررسی

ممکن است آزمایشگاه‌ها بعضی مواقع دارای نتایج PT سؤال‌برانگیز یا نارضایت‌بخش باشند. وقتی این حالت اتفاق بیفتد، آزمایشگاه می‌بایست به طور شفاف آنها را شناسایی و مستندسازی کند.

میزان بررسی به فاکتورهای زیادی بستگی دارد که از میان آنها می‌توان به بحرانی بودن متد، تعداد دفعات نتایج نارضایت‌بخش

و دلیل وجود بایاس اشاره کرد. در هر حالت آزمایشگاه می‌بایست ارزیابی نتایج را مستند کند، حتی اگر قصد انجام هیچ اقدام خاصی را نداشته باشد.

به عنوان یک اصل بنیادی، هر نمره عملکرد نارضایت‌بخش می‌بایست بررسی شده و مستندات بررسی نیز می‌بایست آشکارا به وجود یک مشکل اشاره کنند. وقتی که مشارکت در طرح PT پیوسته، چند دور در سال انجام شود، معیارهای زیر می‌توانند برای بررسی نتایج سؤال برانگیز از سوی آزمایشگاه انتخاب شوند:

(a) ۲ نمره عملکرد سؤال برانگیز متوالی برای پارامتر یکسان

(b) ۹ نمره عملکردی متوالی برای پارامتر یکسان، که دارای علامت بایاس یکسان در برابر مقدار اسمی هستند

با این حال، ذکر این نکته مهم است که این مسئولیت آزمایشگاه است که با توجه به دفعات مشارکت، برازش هدف طرح و بحرانی بودن اندازه‌گیری و غیره معیارهای خود را برای بررسی تنظیم کند.

نکته کلیدی آن است که عملکرد نارضایت‌بخش و روندهای آن می‌بایست بررسی شوند.

۲.۲.۸ بررسی علت اصلی

وقتی به بررسی کامل نیاز باشد، جهت به حداکثر رساندن احتمال تعیین علت اصلی مشکل، رویکرد گام‌به‌گام ترجیح داده می‌شود. مثالی از یک فرم تأییدکننده این رویکرد در پیوست B ارائه شده است.

روش بررسی گام‌به‌گام شامل مراحل زیر است و پرسنل اجراکننده آنالیزها را نیز شامل می‌شود:

(a) تجزیه و تحلیل مسئله بر پایه داده‌های خام، عملکرد کلی دور، نتیجه مطالعات بین‌آزمایشگاهی متوالی و داده‌های کنترل کیفیت داخلی

(b) برنامه‌ریزی برای اقدام یا اقدامات اصلاحی

(c) اجرا و ثبت اقدام یا اقدامات اصلاحی

(d) بررسی کارآمد بودن اقدام یا اقدامات اصلاحی

۳.۲.۸ علل عملکرد ضعیف

«متأسفانه» دلایل بروز عملکرد ضعیف زیادند و به صرف زمان و بررسی پیچیده نیاز می‌باشد. با این حال، از آنجا که بررسی‌ها می‌بایست منجر به بهبود عملکرد آزمایشگاهی شوند، لذا یافتن علل عملکرد ضعیف بسیار مهم است. جهت تسهیل فرایند بررسی‌ها، مرور علل اصلی عملکرد ضعیف در ذهن می‌تواند مفید باشد، بدین ترتیب بررسی‌ها با تمرکز بیشتری انجام خواهند شد. علل اصلی که در ادامه می‌آیند، از جمله عللی هستند که در یک همه‌پرسی اینترنتی به دست آمده‌اند و [17] به ترتیب اهمیت عبارتند از:

(a) تهیه نمونه (برای مثال توزین، خشکاندن، استخراج، هضم، پاکسازی، رقیق‌سازی و...)

(b) نقص دستگاهی یا مشکل سرویس‌دهی

(c) خطای انسانی (برای مثال آموزش غلط، نسخه‌برداری)

(d) کالیبراسیون

(e) انتخاب متد اندازه‌گیری

(f) خطای محاسبه

(g) مشکل گزارش‌دهی (برای مثال فرمت، واحدها، تشخیص، تفسیر)

(h) مشکل آیتم PT

(i) جابجایی و نگهداری نمونه

(j) نمونه‌برداری اولیه

(k) دیگر مشکلات مربوط به طبقه‌بندی

(l) ردیابی نمونه (برای مثال برچسب‌زنی، زنجیره امانت‌دهی و امضا)

(m) مشکل بر گزارکننده طرح PT

جهت شناسایی علت اصلی عملکرد ضعیف، تمرکز بر علل احتمالی اصلی نظیر:

(a) خطای تحریری

(b) خطای فنی (برای مثال متد، تجهیزات، آموزش، کنترل‌های کیفیت داخلی)

(c) مشکل مربوط به طرح PT (نظیر برنامه ناکارآمد، ارزیابی نادرست)

مهم است.

حتی پس از یک بررسی همه‌جانبه باز هم ممکن است منشأ عملکرد ضعیف همچنان ناشناخته باقی بماند. یک عملکرد ضعیف تک می‌تواند به خطای تصادفی نسبت داده شود و اگر همچنان عملکرد ضعیف تکراری مشاهده شود، آنگاه فرایند تجزیه می‌بایست مورد سؤال واقع شود.

۱.۳.۲.۸ خطای تحریری

گرچه خطاهای تحریری مستقیماً به صلاحیت فنی آزمایشگاه مرتبط نمی‌باشند، اما نشان می‌دهند که آزمایشگاه ممکن است هنگام گزارش نتایج به مشتری‌ها دارای یک مشکل بالقوه باشد.

خطاهای تحریری شامل موارد زیر هستند:

(a) خطاهای نسخه‌برداری

(b) برچسب‌زنی غلط

(c) واحدهای نادرست

(d) خطای محل قرارگیری اعشار

اولین گام مهم در بررسی شناسایی، احتمال وقوع خطای تحریری است. اگر خطاهای تحریری علت همیشگی نتایج نارضایت‌بخش باشند، آنگاه می‌بایست بررسی بر روی جنبه‌های کیفیتی سیستم مدیریت انجام شود.

۲.۳.۲.۸ مشکل فنی

به خاطر پیچیدگی آنالیز، ممکن است مشکلاتی در هر سطح از روش اجرایی آنالیزی واقع شود لذا هر یک از مراحل فرایند آنالیزی که در ادامه می‌آیند می‌بایست هنگام بررسی بازبینی شوند:

(a) نگهداری / پیش تیمار آیتم PT

(b) داده‌های کنترل کیفیت داخلی / متد

(c) تجهیزات / واکنشگرها / کالیبراسیون

(d) شرایط محیطی

(e) پردازش داده‌ها

اگر بررسی فرایند آنالیزی در شناسایی علت اصلی به آزمایشگاه کمک نکند، ممکن است بازبینی صحه‌گذاری متد مفید واقع شود.

۳.۳.۲.۸ مشکل مربوط به طرح PT

عملکرد ضعیف ممکن است ناشی از این واقعیت باشد که طرح PT انتخابی نادرست بوده و یا مشکلی برای آیتم‌های PT رخ داده باشد. نکات زیر می‌بایست بررسی شوند:

(a) تفاوت ماتریکس بین آیتم PT و نمونه‌های روتین

(b) تخریب بالقوه آیتم PT

(c) سطوح غلظت پارامتر خارج از دامنه کاربردی متد

(d) فقدان پایداری یا همگنی نمونه‌ها

(e) ارائه دستورالعمل‌های غلط به مشارکت‌کنندگان

(f) مشکلات نگهداری آیتم PT

(g) گروه آزمایشگاه‌های همکار نادرست

(h) مقدار اسمی نادرست

(i) انحراف استاندارد غلط برای ارزیابی مهارت

(j) ارائه داده‌های غلط از سوی برگزارکننده PT

آزمایشگاه می‌بایست یافته‌هایش را با برگزارکننده PT در میان گذاشته و یا در صورتی که طرح PT انتخابی مناسب باشد می‌تواند آن را ارزیابی کند.

پیوست A: فرم انتخاب طرح PT مرتبط

پارامتر:

آیتم PT:

متد:

برگزارکننده PT:

ارزیابی طرح PT:

ملاحظات	بله خیر	بله	معیارهای انتخاب:
			آیا آن یک پارامتر بحرانی برای آزمایشگاه است؟
			آیا پارامترهای پیشنهادی، معادل یا برابر با پارامترهای آزمایشی روتین می‌باشند؟
			آیا آیتم PT، معادل یا برابر با آیتم آزمایشی روتین است؟
			آیا گستره پارامتر، مناسب آزمایشگاه است؟
			آیا برگزارکننده PT، نتایج به دست آمده را با لحاظ کردن متدهای مختلف استفاده شده تیمار می‌کند؟
			آیا تعداد مشارکت‌کنندگان یا اندازه گروه همکاران مناسب است؟
			آیا تعداد دفعات دورها کافی است؟
			آیا گزارش عدم قطعیت نتیجه امکان‌پذیر است؟
			اگر پاسخ مثبت است، آیا برگزارکننده PT عدم قطعیت‌های گزارش شده از سوی مشارکت‌کنندگان را در آنالیز آماری لحاظ کرده است؟
			آیا برگزارکننده PT، عدم قطعیت مقدار اسمی را گزارش می‌کند؟
			آیا قابلیت ردیابی اندازه‌شناسی مقدار اسمی ارائه شده است؟
			آیا برگزارکننده PT اطلاعاتی را درباره طرح آماری استفاده شده ارائه می‌کند؟
			آیا ارزیابی عملکرد آزمایشگاه‌ها بر اساس معیارهای نمره‌دهی (برای مثال نمره‌ی Z) است؟
			آیا برگزارکننده PT هنگام کسب نتایج ضعیف مساعدت می‌کند؟
			آیا گزارش‌ها شامل آنالیز کافی نتایج و اطلاعات لازم جهت اجرای اقدامات اصلاحی برای آزمایشگاه‌ها می‌باشند؟
			آیا برگزارکننده PT جهت اجرای اقدامات اصلاحی «نمونه‌های اضافی/ تکراری» برای آزمایشگاه‌ها فراهم می‌کند؟
			آیا برگزارکننده PT دارای صلاحیت بوده و آیا توسط شخص ثالث به طور رسمی شناخته شده است؟
			آیا گزارش‌ها به زبانی ویرایش شده‌اند که قابل فهم توسط آزمایشگاه‌ها باشند؟

نتیجه:

آیا این طرح PT مربوط به آزمایشگاه می‌شود؟ بله ☐ خیر ☐

ملاحظات:

تأیید شده از سوی:

تاریخ:

پیوست B: فرم مستندسازی بررسی‌های PT

بررسی انجام شده از سوی: تاریخ:

پارامتر:
آیتم PT:
متد:
طرح PT:
شماره دور:
نتیجه یا نتایج آزمایشگاه:
نتیجه/گستره قابل قبول:
برآورد عملکرد (نمره):
پارامتر بحرانی: بله ☐ خیر ☐
تا چه اندازه طرح PT با آنالیز روتین قابل مقایسه می‌باشد (برای مثال ماتریکس، پارامترها، سطح غلظت و غیره)؟

آیا نتایج دوره‌های قبلی طرح PT به یک رَوند سؤال‌برانگیز یا نارضایت‌بخش اشاره دارند؟
اگر پاسخ مثبت است، تجزیه و تحلیل این رَوند می‌بایست مطرح شود.

ارزیابی اولیه:

آیا آیتم PT در شرایط قابل قبول دریافت شد؟	بله ☐ خیر ☐
اگر جواب منفی است، آیا این شرایط می‌توانند علت نتیجه ضعیف را توضیح دهند؟	
آیا آیتم PT معادل یک نمونه روتین بود؟	بله ☐ خیر ☐
اگر جواب منفی است، آیا این وضع می‌تواند دلیل نتیجه ضعیف را توضیح دهد؟	
آیا آیتم PT به عنوان یک نمونه روتین آزمایش شد؟	بله ☐ خیر ☐
اگر جواب منفی است، آیا این وضع می‌تواند علت نتیجه ضعیف باشد؟	
آیا ارزیابی نتایج بر اساس متد گروه‌بندی شده است؟	بله ☐ خیر ☐
اگر جواب مثبت است، آیا این وضع می‌تواند نتیجه ضعیف را توضیح دهد؟	
بر اساس پیشنهادات ارائه شده در بالا، آیا میزان ارتباط طرح PT می‌بایست بازنگری شود؟	بله ☐ خیر ☐
آیا آیتم PT اولیه پس از دریافت برگه ارزیابی PT دوباره اندازه‌گیری شد؟	بله ☐ خیر ☐
اگر جواب مثبت است، آیا نتیجه قابل مقایسه است؟	
آیا یک آیتم تکراری درخواست شده و دوباره اندازه‌گیری شد؟	بله ☐ خیر ☐
اگر جواب مثبت است، آیا نتیجه قابل مقایسه است؟	

بررسی تحریری:

خطاهای تحریری نوعی برای مثال می‌توانند شامل خطاهای نسخه‌برداری، خطای ثبت داده‌ها، عدم اطلاع‌رسانی برگزارکننده PT درباره تغییر متد و واحدهای نادرست استفاده شده باشند.

آیا نتیجه ضعیف به خاطر خطای تحریری بود؟	☐ بله ☐ خیر
و آیا اقدام اصلاحی انجام شد؟	
نتیجه تصحیح شده:	
آیا این نتیجه هنوز هم سؤال برانگیز یا نارضایت‌بخش است؟	☐ بله ☐ خیر
اگر جواب مثبت است، بررسی می‌بایست ادامه یابد.	

بررسی فنی:

جنبه‌های زیر می‌بایست لحاظ شوند:

روش اجرایی تجزیه:
کنترل کیفیت داخلی:
نگهداری/تهیه آیت‌م PT:
تجهیزات:
شرایط محیطی:
چه اثراتی ممکن است بر روی نتایج روتین گذشته و یا آینده داشته باشند؟
نتیجه:

اقدام یا اقدامات اصلاحی انجام شده:

تأیید شده از سوی:
مدیر فنی: تاریخ:
مدیر کیفیت: تاریخ:

پیوست C: تفسیر داده‌های PT از سوی کاربران

C.1 مقدمه

لازم است آزمایشگاه‌ها صلاحیت خود را به طرف‌های ذی‌نفع نظیر نهادهای تأیید صلاحیت، نهادهای نظارتی و مشتری‌ها نشان دهند. نتایج PT و دیگر فعالیت‌های کنترل کیفیت، یکی از ابزارهای نشان دادن صلاحیت و شایستگی آزمایشگاه‌ها هستند. از آنجا که PT معمولاً یک ارزیابی طرف سوم است، لذا رفته رفته طرف‌های ذی‌نفع پیشنهاد و یا تقاضای مشارکت آزمایشگاه‌ها در طرح‌های PT مناسب را دارند، بدین ترتیب آنها می‌توانند به یک ارزیابی مستقل عملکرد آزمایشگاه‌ها دسترسی داشته باشند. مسئولیت آزمایشگاه این است که هنگام ارائه نتایج PT برای طرف‌های ذی‌نفع، تمام اطلاعات اضافی مناسب (مثل نمره عملکرد محاسبه شده و بررسی‌های مجدد) و غیره را نیز فراهم کند، بدین ترتیب از نحوه‌ی عملکرد آزمایشگاه اطمینان حاصل خواهد شد.

C.2 نهادهای تأیید صلاحیت

نهادهای تأیید صلاحیت و ارزیاب‌های فنی استخدام شده از سوی نهادهای تأیید صلاحیت، معمولاً درک خوبی از نقش PT داشته و در تفسیر نتایج طرح PT به دست آمده از آزمایشگاه‌های دارای تأیید صلاحیت و یا آزمایشگاه‌هایی که در جستجوی کسب تأیید صلاحیت هستند، از مهارت کافی برخوردار می‌باشند. به طور کلی ارزیاب‌های فنی با طرح‌های PT که آزمایشگاه در آن شرکت می‌کند آشنا هستند. پروتکل‌های طرح و دیگر مستندات مطالعه شده و چنانچه لازم باشد، با برگزارکننده PT تماس گرفته تا در مورد مسائل مهم بحث و موضوع کاملاً شفاف شود. سطح عملکرد طرح PT برای هر آزمایشگاه در برابر معیارهای ایجاد شده از سوی برگزارکننده PT تعیین می‌شود. در بعضی موارد، آنچه که معادل با عملکرد نارضایت‌بخش در طرح PT است، ممکن است در دامنه کاربردی تأیید صلاحیت آزمایشگاه قابل قبول یا برازنده هدف باشد، همینطور برعکس؛ آنچه که معادل با عملکرد رضایت‌بخش در طرح PT است ممکن است در دامنه کاربردی تأیید صلاحیت آزمایشگاه قابل قبول و یا برازنده هدف نباشد.

C.3 نهادهای نظارتی

نهادهای نظارتی می‌بایست از کیفیت رضایت‌بخش اندازه‌گیری‌های انجام شده در آزمایشگاه‌هایی که روش آن در قوانین یا رهنمودها ارائه شده، اطمینان حاصل کنند. بنابراین نهادهای نظارتی علاوه بر رویکردهای دیگر نظیر آنالیزهای داور انجام شده یا ارائه نمونه‌های کنترلی برای آنالیز می‌توانند از عملکرد طرح PT به عنوان یکی از راه‌های ارزیابی کیفیت استفاده کنند.

هر جا که نهاد نظارتی نیز در توسعه طرح PT درگیر باشد، طرح، جنبه‌هایی را که با خود آن نهاد دارای ارتباط مستقیم می‌باشند را در هم ترکیب کرده و در نتیجه هدف آن به سادگی درک خواهد شد. برای وضعیت‌هایی که در آن نهاد نظارتی از طرح مستقل برای اهداف خودشان استفاده می‌کنند، پیشنهاد می‌شود که با برگزارکننده PT درباره دامنه کاربردی و پارامترهای عملیاتی طرح وارد بحث و مذاکره شوند. این امر باعث خواهد شد تا نتایج به دست آمده از هر آزمایشگاه تنها در زمینه کاری آن آزمایشگاه به کار گرفته شود. جهت قضاوت درباره‌ی عملکرد آزمایشگاه در ارتباط با تولرانس‌های مجاز در مقررات می‌بایست فرایندهای آماری استفاده شده از سوی برگزارکننده PT برای محاسبه عملکرد آزمایشگاه به خوبی درک شوند. در چنین وضعیت‌هایی برای آنکه تفسیر نادرستی از داده‌های عملکرد طرح PT رخ ندهد، ممکن است به توصیه و مشاوره برگزارکننده PT نیاز باشد.

C.4 مشتریان آزمایشگاه‌های مشارکت‌کننده

مشتری یک آزمایشگاه مشارکت‌کننده در طرح PT می‌تواند از عملکرد طرح PT به عنوان یک ابزار برای پایش کیفیت آن آزمایشگاه استفاده کند. مشتری می‌بایست درک خوبی از چگونگی روند اجرای طرح PT و چگونگی محاسبه عملکرد از سوی برگزارکننده PT در طرح PT داشته باشد.

گرچه بعضی از سیستم‌های تعیین عملکرد در طرح PT عمومی هستند، نظیر استفاده از نمره Z ، اما در عمل سیستم‌های مختلف زیادی وجود دارند. به علاوه، مشتری‌ها می‌بایست درباره روش محاسبه نمرات Z و دیگر شاخص‌های عملکرد آگاه باشند، چرا که این روش‌ها در طرح‌های PT گاهی متفاوت از هم هستند. مشتری‌ها رفته‌رفته معیارهای عملکرد طرح PT را در مستندات برون‌سپاری خود پیوست کرده و از اطلاعات به دست آمده از پیمانکاران احتمالی در مورد عملکرد طرح PT برای تصمیم‌گیری درباره اینکه بهتر است قرارداد با کدام آزمایشگاه منعقد شود، استفاده می‌کنند. مشتری‌ها هنگام استفاده از عملکرد طرح PT به عنوان معیاری در واگذاری کار، می‌بایست اطمینان حاصل کنند که آیا «استاندارد عملکردی» تنظیمی واقع‌بینانه و قابل دسترس است یا خیر. برای مثال، درخواست اینکه آزمایشگاه‌ها می‌بایست برای تمام آنالیت‌ها در تمام دوره‌های طرح PT به نتایج رضایت‌بخش برسند، غیرواقع‌بینانه است. برگزارکننده‌های PT معمولاً اطلاعات مناسب در رابطه با عملکرد کلی طرح PT را در گزارش PT ارائه می‌کنند، بدین ترتیب یک ملاک ارزیابی می‌تواند تنظیم شود. مشتری‌ها همچنین می‌بایست اطمینان داشته باشند که تعیین‌های مورد نظر آنها به طور آشکار بیان شده باشند، چرا که طرحی ممکن است دارای دامنه کاربردی وسیع‌تری باشد، از سوی دیگر، عملکرد آزمایشگاه‌ها در تعیین‌هایی که مستقیماً ارتباطی به خواسته مورد نظر آنها ندارد، نمی‌بایست مهم تلقی شود.

مشتری‌ها می‌بایست داده‌های مرتبط با عملکرد طرح PT مربوط به آزمایشگاه پیمانکار را در زمینه درست خود بررسی کنند، چرا که آزمایشگاه‌ها گاهی می‌توانند به شکلی داده‌ها را به مشتری عرضه کنند که نشان‌دهنده تصویر مثبت غیرواقعی از آزمایشگاه باشد.

جهت به دست آوردن تصویر درست از عملکرد واقعی آزمایشگاه پیشنهاد می‌شود مشتری‌ها موارد زیر را در نظر بگیرند:

(a) کسب اطلاعات لازم درباره دامنه کاربردی و عملیات طرح PT (برای مثال، پروتوکل طرح PT) از آزمایشگاه یا برگزارکننده PT

(b) نگاهی به عملکرد آزمایشگاه در طول زمان، چرا که یک دور مشارکت در طرح PT تنها نمای کلی از عملکرد آزمایشگاه را نشان می‌دهد.

(c) بازبینی عملکرد کل تمام مشارکت‌کنندگان، جهت قضاوت درباره‌ی نحوه‌ی عملکرد آزمایشگاه

(d) تقاضای کپی‌های گزارش‌های طرح PT (هر جا که محرمانه بودن مسئله‌ساز نباشد) برای تأیید داده‌های عملکرد طرح PT، برگزارکننده PT ممکن است این داده‌ها را فراهم کند، گرچه معمولاً توافق مشارکت‌کنندگان نیز لازم است.

یک نتیجه نارضایت‌بخش در هر دور از طرح PT آزمایشگاه را ضعیف نمی‌کند، همین‌طور رسیدن ۱۰۰٪ به نتایج رضایت‌بخش در هر دور از طرح PT نیز الزاماً به معنای خوب بودن آزمایشگاه نیست.

روشی که طی آن آزمایشگاه به یک نتیجه نارضایت‌بخش عکس‌العمل نشان می‌دهد، معمولاً اطلاعات بیشتری را درباره آن آزمایشگاه به ما می‌دهد تا صرفاً وقوع یک نتیجه نارضایت‌بخش در آن آزمایشگاه.

پیوست D: جنبه‌های آماری PT

یکی از عناصر اصلی در تمام PTها ارزیابی عملکرد هر مشارکت‌کننده است. بدین منظور برگزارکننده PT می‌بایست دو مقدار را ایجاد کند که برای ارزیابی عملکرد استفاده می‌شوند:

۱. مقدار اسمی که اغلب مقدار توافقی نیز نامیده می‌شود.

۲. انحراف استاندارد ارزیابی مهارت

همانطور که در بخش ۲.۷ بدان اشاره شد، برگزارکننده PT می‌بایست عدم قطعیت اندازه‌گیری مقدار اسمی را نیز تخمین بزند.

D.1 مقدار اسمی و عدم قطعیت استاندارد مقدار اسمی

همانطور که در ISO 13528 توضیح داده شده، پنج متد اصلی برای رسیدن به مقدار اسمی و عدم قطعیت استاندارد مربوطه وجود دارد:

۱. فرمولاسیون: افزودن مقدار معلوم یا غلظت معلومی از آنالیت به یک ماده پایه‌ی فاقد آنالیت. این متد در بسیاری از حالت‌ها، به خصوص برای حالت‌هایی که مقدار کل آنالیت (و نه غلظتی از آن) در معرض اندازه‌گیری است، به شکل رضایت‌بخش عمل می‌کند. البته ممکن است این متد دقیقاً پیچیدگی روش‌های اجرایی تهیه نمونه نرمال [که ممکن است شامل استخراج و تشکیل گونه‌ی خاص باشند] و مشکلات بازیابی را شامل نشود.

عدم قطعیت استاندارد با ترکیب عدم قطعیت‌ها و استفاده از رویکرد توصیف شده در راهنمای بیان عدم قطعیت در اندازه‌گیری (GUM) تخمین زده می‌شود [18].

۲. مقادیر مرجع گواهی‌دار: وقتی آیتم PT یک ماده مرجع گواهی‌دار (CRM) باشد، مقدار مرجع گواهی‌دار آن به عنوان مقدار اسمی استفاده می‌شود. در این حالت مزیت ردیابی مقدار اسمی وجود دارد، اما این رویکرد گران است و اغلب CRMهای مناسب نیز در دسترس نیستند.

عدم قطعیت استاندارد، از اطلاعات عدم قطعیت ارائه شده در گواهی CRM به دست می‌آید.

۳. مقادیر مرجع: مجموعه‌ای از اقلام PT تهیه شده از سوی یک آزمایشگاه منتخب، با استفاده از متد اولیه یا همراه با یک ماده مرجع گواهی‌دار (CRM) اندازه‌گیری می‌شوند. مقدار اسمی مستقیماً از متد اولیه استفاده شده یا از کالیبراسیون در برابر مقدار مرجع گواهی‌دار CRM به دست می‌آید. در این روش نیز از متد اولیه و CRM استفاده می‌شود، لذا مقدار قابل ردیابی نیز ایجاد خواهد شد. اما روش به نتایج حاصل از آزمایشگاه تک متکی بوده و متدهای اولیه یا CRMهای مناسب نیز ممکن است همیشه در دسترس نباشند.

عدم قطعیت استاندارد از نتایج آزمایش آزمایشگاه منتخب و عدم قطعیت‌های مقادیر مرجع گواهی‌دار CRM به دست می‌آید.

۴. مقادیر توافقی حاصل از آزمایشگاه‌های تخصصی: تعیین یک مقدار توافقی به دست آمده از نتایج گروهی از آزمایشگاه‌های داور یا تخصصی که در به کارگیری متدهای آنالیزی با تجربه می‌باشند. این روش احتمالاً دقیق‌ترین رویکرد به دست آوردن مقادیر واقعی برای اقلام PT است، هر چند که اجرای آن ممکن است گران باشد. مشکل دیگر آنکه، اغلب یافتن گروهی از آزمایشگاه‌های مرجع یا تخصصی که در تخصص‌شان هیچ شک و تردیدی وجود نداشته باشد و از سوی تمام مشارکت‌کنندگان PT نیز پذیرفته شوند، سخت و یا گاهی غیرممکن است. این مشکل برای PTهای بزرگ و بین‌المللی و با مشارکت تعداد زیاد کشورها بیشتر صادق است. برای تعداد زیاد آنالیزها، مقدار واقعی در اصل بر اساس متد استفاده شده تعریف می‌شود. در این حالت‌ها، آزمایشگاه داور یا آزمایشگاه تخصصی همگی می‌بایست از متد یکسان استفاده کرده و دستورات آن را جزیه‌جز اجرا کنند. ممکن است یک بایاس مجهول در نتایج گروهی از آزمایشگاهی متخصص وجود داشته باشد. قبل از اجرای PT، آزمایشگاه‌های کارشناس و متدهای به کار رفته می‌بایست بیان شوند.

اگر هر آزمایشگاه تخصصی یک تخمین عدم قطعیت استاندارد را گزارش کند، آنگاه عدم قطعیت استاندارد از طریق رابطه‌ی زیر تخمین زده می‌شود:

$$u_x = \frac{1.25}{P} \times \sqrt{\sum_{i=1}^P u_i^2}$$

در اینجا:

u_i : عدم قطعیت استاندارد حاصل از آزمایشگاه تخصصی

P : تعداد آزمایشگاه‌های تخصصی

چنانچه عدم قطعیت استاندارد تخمین زده نشود، می‌بایست از روش مقدار توافقی مشارکت‌کنندگان استفاده شود.

۵. مقدار توافقی مشارکت‌کنندگان: استفاده از مقدار توافقی به دست آمده در هر دور PT و بر اساس نتایج به دست آمده از سوی مشارکت‌کنندگان. مقدار توافقی معمولاً با استفاده از تکنیک‌های آماری استوار تخمین زده می‌شود. رویکرد توافقی به طور آشکار ساده‌ترین رویکرد بوده و در بعضی حالت‌ها برای مثال، هنگام استفاده از نمونه‌های با ماتریکس طبیعی، اغلب تنها روش تخمین مقدار واقعی است.

عدم قطعیت استاندارد از رابطه‌ی زیر تخمین زده می‌شود:

$$u_x = 1.25 \times S^* \sqrt{P}$$

در اینجا:

S^* : انحراف استاندارد استوار نتایج مشارکت‌کنندگان

P : تعداد مشارکت‌کنندگان

محدودیت‌های این رویکرد عبارتند از:

- (a) ممکن است توافق واقعی بین مشارکت‌کنندگان وجود نداشته باشد.
- (b) توافق ممکن است بر اثر استفاده عمومی از روش‌شناسی ناقص همراه با بایاس بوده و این بایاس در عدم قطعیت استاندارد مقدار اسمی محاسبه شده منعکس نشود.

D.2 انحراف استاندارد برای ارزیابی مهارت

همانطور که در ISO 13528 [12] بیان شده، پنج رویکرد برای تعیین انحراف استاندارد ارزیابی مهارت، یعنی گستره قابل قبول نتایج مشارکت‌کنندگان وجود دارد:

۱. مقدار تجویز شده: انحراف استاندارد ارزیابی مهارت ممکن است در مقدار مورد نیاز برای کار خاص تفسیر داده‌ها تنظیم شده و یا از الزامات ارائه شده در اسناد قانونی به دست آید.

با این رویکرد، انحراف استاندارد ارزیابی مهارت معادل با بیان «برازش هدف» برای متد اندازه‌گیری خواهد شد.

۲. برداشت ذهنی: انحراف استاندارد ارزیابی مهارت را می‌توان در مقداری تنظیم کرد که با سطح عملکردی که برگزارکننده PT خواهان آن است مطابق باشد.

با این رویکرد، انحراف استاندارد ارزیابی مهارت معادل با بیان «برازش هدف» برای متد اندازه‌گیری خواهد شد.

۳. مدل عمومی: مقدار انحراف استاندارد ارزیابی مهارت ممکن است از یک مدل عمومی برای تجدیدپذیری متد اندازه‌گیری به دست آید.

عیب این رویکرد آن است که تجدیدپذیری واقعی یک متد اندازه‌گیری خاص ممکن است به طور چشمگیری از مقدار ارائه شده در مدل تفاوت داشته باشد، چرا که استفاده از یک مدل عمومی به طور ضمنی به این نکته اشاره دارد که تجدیدپذیری فقط به سطح غلظت پارامتر بستگی دارد و به خود پارامتر، روش اندازه‌گیری یا اندازه نمونه بستگی ندارد.

۴. نتایج آزمایش دقت: وقتی که متد اندازه‌گیری استفاده شده در طرح PT استاندارد باشد و اطلاعات تکرارپذیری و تجدیدپذیری متد در دسترس باشند، انحراف استاندارد ارزیابی مهارت را می‌توان با استفاده از این اطلاعات محاسبه کرد.

۵. از داده‌های به دست آمده در خود طرح PT: با این رویکرد، انحراف استاندارد ارزیابی مهارت در یک دور، از نتایج گزارش شده از سوی مشارکت‌کنندگان در همان دور به دست می‌آید و باید انحراف استاندارد استوار نتایج گزارش شده از سوی تمام مشارکت‌کنندگان باشد.

عیب این رویکرد آن است که مقدار مورد نظر ممکن است از یک دور تا دور بعدی تفاوت چشمگیری داشته باشد. در این حالت استفاده از مقادیر نمره Z برای بررسی روندهایی که در چند دور به طور دائمی مشاهده می‌شوند مشکل خواهد شد.

پیوست E: برآورد عملکرد PT

در ادامه متداول‌ترین آماره‌های (قبلاً در بخش ۷.۲.۳ نیز به آنها اشاره شد) برآورد عملکرد ارائه می‌شوند. دیگر طرح‌های آماری که در این سند به آنها اشاره نشده را می‌توانید در ISO 13528 مشاهده کنید [12].

E.1 «نمره‌ی Z»

یکی از عناصر اصلی و مشترک در تمام PTها، استفاده از شاخص عملکرد برای تعیین کمی عملکرد تجزیه‌ای هر مشارکت‌کننده [12,16] است.

نمره Z اغلب به عنوان یک شاخص عملکرد توصیه می‌شود. نمره Z معیاری از انحراف نتیجه از مقدار اسمی است و به شکل زیر محاسبه می‌شود:

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

در اینجا:

X: نتیجه گزارش شده از سوی مشارکت‌کننده

X: مقدار اسمی

$\hat{\sigma}$: انحراف استاندارد ارزیابی مهارت

فرض اصلی در استفاده از نمره‌ی Z این است که نمره‌های منفرد دارای توزیع نرمال یا گاوسی با میانگین صفر و انحراف استاندارد یک می‌باشند. بر این اساس، نتایج آنالیز را می‌توان به شکل «خوش‌رفتار» توصیف کرد. یک طبقه‌بندی متداول بر اساس نمره‌های Z می‌تواند انجام شود.

E.2 «نمره‌ی Z'»

$$Z' = (x-X) / \sqrt{\hat{\sigma}^2 + u_x^2}$$

در اینجا:

X: نتیجه گزارش شده از سوی مشارکت‌کننده

X: مقدار اسمی

$\hat{\sigma}$: انحراف استاندارد ارزیابی مهارت

u_x : عدم قطعیت استاندارد مقدار اسمی X

هنگام تصمیم‌گیری در رابطه با اینکه کدام یک از نمره‌های Z یا Z' می‌بایست استفاده شوند، برگزارکننده PT باید جنبه‌های زیر را در نظر بگیرد.

(a) اگر $u_x \leq 0.3\hat{\sigma}$ [16] باشد، آنگاه عدم قطعیت استاندارد مقدار اسمی قابل چشم‌پوشی بوده و احتمالاً استفاده از نمره‌های Z' هیچ نفعی نخواهد داشت.

(b) اگر $u_x \geq 0.3\hat{\sigma}$ باشد، آنگاه عدم قطعیت استاندارد مقدار اسمی قابل چشم‌پوشی نبوده و استفاده از نمره‌های Z' پیشنهاد می‌شود.

(c) پیامدهای اینکه نتایج آزمایشگاهی به پیام اقدام یا هشدار منجر شوند چقدر جدی است؟ آیا نتایج برای سلب صلاحیت آزمایشگاه‌ها در اجرای متد اندازه‌گیری برای گروهی از کاربران استفاده می‌شوند؟

E.3 «نمره ی زتا»

رفته‌رفته آزمایشگاه‌ها تشویق می‌شوند تا عدم قطعیت نتایج خودشان را تخمین بزنند، بدین ترتیب برگزارکنندگان PT نیز می‌توانند یک چنین اطلاعاتی را در نمره‌دهی عملکرد PT ضمیمه کنند.

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

در اینجا:

u_x : تخمین آزمایشگاه از عدم قطعیت استاندارد نتیجه x

u_X : عدم قطعیت استاندارد مقدار اسمی X

معمولاً معیارهای عملکرد زیر برای نمره‌های Z و Z' و زتا شناخته شده‌اند:

(a) اگر $|Z| \leq 2$ ، نمره به عملکرد «رضایت‌بخش» اشاره داشته و به پیامی منجر نمی‌شود.

(b) اگر $2 < |Z| < 3$ ، نمره به عملکرد «سؤال‌برانگیز» اشاره داشته و به پیام هشدار منجر می‌شود.

(c) اگر $|Z| \geq 3$ ، نمره به عملکرد «نارضایت‌بخش» اشاره داشته و به پیام اقدام منجر می‌شود.

E.4 «عدد E_n »

دیگر سیستم نمره‌دهی مناسب، استفاده از «عدد E_n » است که عدم قطعیت بسط‌یافته را لحاظ می‌کند.

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

در اینجا:

U_x : عدم قطعیت بسط‌یافته نتیجه x از سوی مشارکت‌کننده

U_{ref} : عدم قطعیت بسط‌یافته مقدار اسمی X که در آزمایشگاه مرجع تعیین می‌شود

وقتی که مقدار اسمی از سوی آزمایشگاه مرجع ایجاد شده باشد، از عدد E_n استفاده می‌شود. با این حال، لازم است مشارکت‌کنندگان درک خوبی از عدم قطعیت‌شان داشته و تمامی مشارکت‌کنندگان به شکل یک‌دست و منسجم آن را گزارش کنند.

معمولاً قضاوت زیر برای اعداد E_n انجام می‌شود:

(a) اگر $|E_n| \leq 1$ ، نمره به عملکرد «رضایت‌بخش» اشاره داشته و به پیامی منجر نمی‌شود.

(b) اگر $|E_n| > 1$ ، نمره به عملکرد «نارضایت‌بخش» اشاره داشته و به پیام اقدام منجر می‌شود.

در اینجا، به جای استفاده از رقم ۲ که در نمره‌های Z استفاده می‌شود، از مقدار بحرانی ۱ استفاده می‌شود، چرا که عدد E_n بر اساس عدم قطعیت‌های بسط‌یافته و نه بر اساس عدم قطعیت‌های استاندارد، محاسبه می‌شود.

E.5 نمره‌دهی برای زمانی که معیار برازش هدف مرسوم وجود ندارد – نمره‌های Z اصلاح شده:

گاهاً مشارکت‌کنندگان طرح PT متوجه می‌شوند که انحراف استاندارد ارزیابی مهارت استفاده شده از سوی برگزارکننده PT برای بعضی و یا تمام کاربردهایی که آنها درگیر آن هستند، مناسب نمی‌باشد. در نتیجه نمره Z حاصل گمراه‌کننده خواهد بود. در چنین مواردی، مشارکت‌کننده می‌تواند به شکل منطقی و بر پایه انحراف استاندارد مناسب ارزیابی مهارت که برای هدف شخصی مناسب است، یک سیستم نمره‌دهی جداگانه ایجاد کند.

یعنی [16]:

$$Z = \frac{(X - X_{av})}{\sigma_f}$$

در اینجا:

X_{av} : مقدار اسمی

σ_f : مقدار مناسب انحراف استاندارد برای ارزیابی مهارت

هر چند، این کار نیاز به دخالت برگزارکننده PT ندارد.

پیوست F: مثال‌هایی از رویکردهای مختلف در برآورد عملکرد

F.1 نتایج مقایسه‌ای مقدار اسمی و دقت بین آزمایشگاهی متدهای آماری کلاسیک و متدهای آماری استوار

در اجرای این مقایسه برای رسیدن به تخمین مقدار واقعی و انحراف استاندارد ارزیابی مهارت از پروتوکل آماری استفاده شد. بر این اساس، به منظور برآورد نتایج آنالیزی دریافت شده از مشارکت‌کنندگان (جدول F.1) چند رویکرد آماری در نظر گرفته شد. یک متد قراردادی بر پایه محاسبات ISO 5725 و رویکرد دوم بر اساس کاربرد آماره‌های استوار به کمک الگوریتم A توصیف شده در ISO 13528 [12] انتخاب شدند.

با توجه به مقدار اسمی، اختلاف‌های قابل ملاحظه‌ای (کمتر از 0.2%) بین متدهای آماری به کار رفته، مشاهده نشد. لذا معدل استوار تمام نتایج مشارکت‌کنندگان، بدون شناسایی داده انحرافی به عنوان بهترین تخمین مقدار واقعی غلظت آنالیت مدنظر در این طرح انتخاب شد (معدل استوار = 586.7%).

با این حال، همانند تخمین دقت بین‌آزمایشگاهی، دو پروتوکل ذکر شده در بالا با یکدیگر مقایسه شده و تفاوت‌های عمده‌ای در انحراف استاندارد ارزیابی مهارت مشاهده شد. تیمار آماری نتایج نشان داد که درصد انحراف استاندارد (RSD%) به دست آمده از متدهای کلاسیک تقریباً دو برابر بزرگتر از انحراف استاندارد تخمین استوار نتایج با ملاحظه یکسان درباره داده‌های انحرافی خواهد شد.

جدول F.1: نتایج مقایسه‌ای مقدار اسمی و دقت بین آزمایشگاهی بین متدهای آماری کلاسیک و استوار

	ISO 5725	ISO 13528
مقدار اسمی	586.5 ‰	586.7 ‰
انحراف استاندارد ارزیابی مهارت	0.56 ‰	0.31 ‰
RSD	0.10	0.05

بعضی از روش‌های آماری، میانه و دامنه‌ی بین‌چارکی نرمالیزه شده IQR را انتخاب می‌کنند که به ترتیب معیار مرکز و پراکندگی داده‌ها هستند و به شکل مشابه برای میانگین و انحراف استاندارد به کار گرفته می‌شوند. میانه و IQR نرمالیزه شده به عنوان آماره‌های استوار، کمتر تحت تأثیر داده‌های انحرافی در مجموعه داده‌ها قرار می‌گیرند.

میانه، مقدار وسط گروه است، یعنی نصف نتایج، بالاتر از آن و نصف دیگر، پایین‌تر از آن قرار می‌گیرد. اگر N (تعداد کل نتایج) عدد فرد باشد، میانه، مقدار مرکزی تک است. اگر N، عدد زوج باشد، میانه، میانگین دو مقدار مرکزی است.

IQR نرمالیزه شده معیار تغییرپذیری نتایج است و برابر با IQR ضرب در فاکتور (0.7413) است، که آن را به مقدار قابل مقایسه با انحراف استاندارد تبدیل می‌کند.

F.2 بحثی پیرامون ارزیابی عملکرد آزمایشگاه‌ها

هدف این بخش توصیف دامنه وسیع احتمالاتی است که برگزارکننده طرح PT ممکن است برای ارزیابی داده‌های ارائه شده از سوی مشارکت‌کننده‌ها لحاظ کند. لذا، برای پوشش اصول آماری چندین حالت با رویکرد متفاوت ارزیابی شدند [19]، [20]، [21].

به طور کلی، جهت برآورد ارزیابی عملکرد هر آزمایشگاه با تفسیر مقادیر نمره Z مقادیر لحاظ شده در بیان نمره Z (مقدار اسمی و انحراف استاندارد ارزیابی مهارت) در هر یک از چهار رویکرد آماری زیر محاسبه شده‌اند:

۱. ایزو 5725 [1]: میانگین کل و انحراف استاندارد تجدیدپذیری با تشخیص داده انحرافی

۲. متد میانه و NIQR: میانه کل داده‌ها و دامنه بین‌چارکی نرمالیزه شده [20]

۳. ایزو 13528: معدل استوار و انحراف استاندارد استوار محاسبه شده بر اساس الگوریتم‌های A و S ، بدون تشخیص داده انحرافی

۴. معیار برازش هدف: معدل استوار و انحراف استاندارد هدف برای مقدار ارزیابی مهارت بر اساس $RSD\%$ تثبیت شده از دوره‌های مناسب PT قبلی در این سطح غلظت [14]

پس از محاسبه و بر اساس یکی از چهار پروتوکل بالا (جدول F.2) مشاهده شد که بدون توجه به متد آماری به کار رفته، چهارده مشارکت‌کننده مقادیر نمره Z قابل قبولی ($|Z| \leq 2$) داشتند. علت مشاهده این وضع در این واقعیت نهفته است که این آزمایشگاه‌ها هستند که به خاطر تکنیک‌های آنالیزی مشترک متعادل‌ترین نتایج با تعداد انحراف‌های کمتر در پراکندگی داده‌ها را ایجاد می‌کنند، به خاطر همین پروتوکل آماری بر عملکرد آنها تأثیری نداشته است.

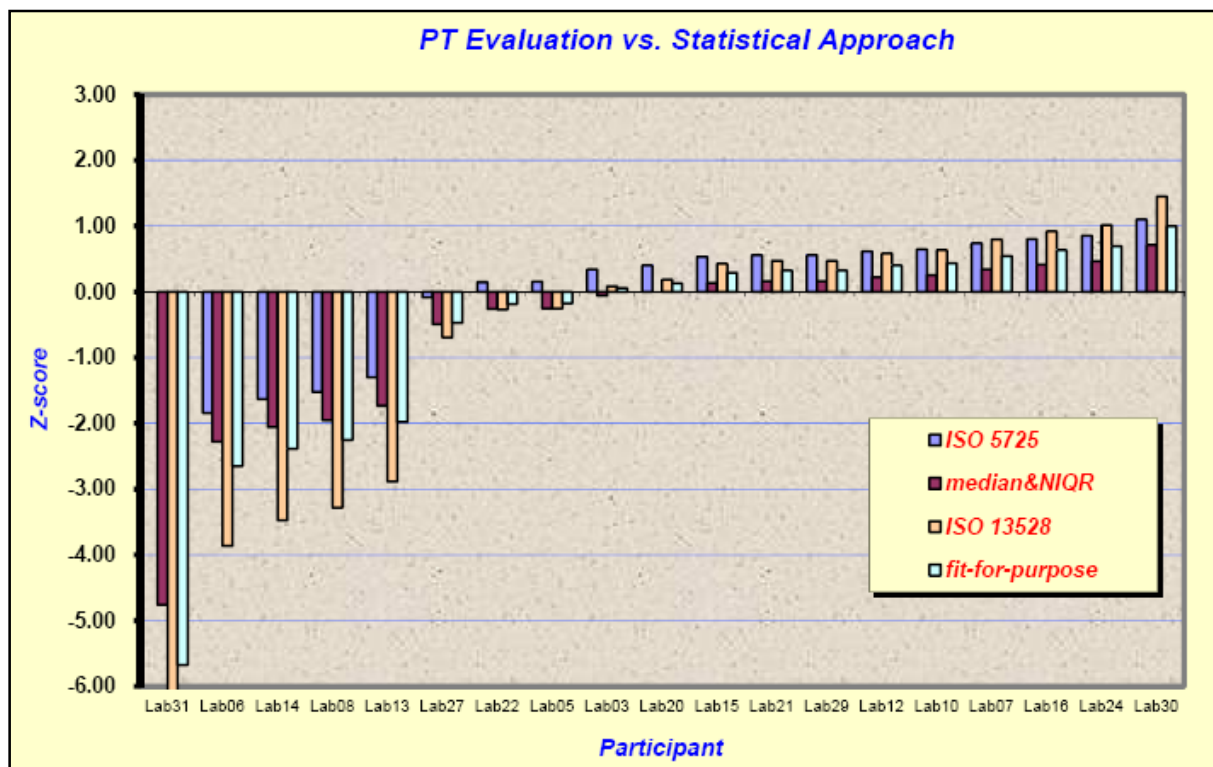
به علاوه، بر اساس توزیع مقادیر نمره Z ، در پنج آزمایشگاه یک روند مشخص در پراکندگی آشکار می‌شود. بنابراین، بر اساس ISO 5725 [1]، مقادیر نمره Z با معیارهای پذیرش مطابقت دارند، به خصوص وقتی که داده‌های حاصل از یک آزمایشگاه به عنوان داده‌های انحرافی رد شده باشند.

جدول F.2: گزارش خلاصه نتایج نمره Z کل به دست آمده از آزمایشگاه‌های مشارکت‌کننده با پیروی از پروتوکول‌های آماری مختلف

Participant	ISO 5725	Median & NIQR	ISO 13528	Fit-for-purpose
Lab.31	outlier	-4.76	-8.27	-5.68
Lab.06	-1.84	-2.28	-3.86	-2.65
Lab.14	-1.63	-2.06	-3.48	-2.39
Lab.08	-1.52	-1.95	-3.28	-2.25
Lab.13	-1.30	-1.73	-2.89	-1.98
Lab.27	-0.09	-0.50	-0.69	-0.48
Lab.22	0.15	-0.26	-0.27	-0.19
Lab.05	0.16	-0.25	-0.26	-0.18
Lab.03	0.34	-0.06	0.08	0.06
Lab.20	0.40	0.00	0.19	0.13
Lab.15	0.53	0.13	0.43	0.29
Lab.21	0.56	0.16	0.47	0.32
Lab.29	0.56	0.16	0.47	0.32
Lab.12	0.62	0.22	0.58	0.40
Lab.10	0.65	0.25	0.64	0.44
Lab.07	0.74	0.34	0.80	0.55
Lab.16	0.81	0.41	0.92	0.63
Lab.24	0.86	0.46	1.01	0.70
Lab.30	1.10	0.71	1.45	1.00

به عبارت دیگر، برای اجتناب از تأثیر نتایج اکستریم، کاربرد متدهای آماری استوار [22]، [23]، به خاطر حذف نشدن داده‌های انحرافی، به طور چشمگیری به مقادیر نمره Z بزرگتری منجر می‌شوند. بر این اساس، چنانچه محاسبه با استفاده از متد میانه و NIQR انجام شود، مقادیر نمره Z کمی کوچکتر از مقادیر نمرات Z متناظری خواهند شد که بر اساس متد استوار در ISO 13528 تخمین زده شده‌اند [12]. در این حالت، یک آزمایشگاه پیام هشدار دریافت می‌کند در حالی که چهار آزمایشگاه مقادیر نمره Zی نشان می‌دهند که به پیام اقدام منجر می‌شوند، بنابراین نتایج آزمایشگاهی که قبلاً در رویکرد پارامتری به عنوان داده انحرافی لحاظ شده، نیاز به بررسی ویژه دارد.

سرانجام آنکه، مشاهده می‌شود هنگام به کارگیری معیار برازش هدف بر اساس الزام کاربر که در آن نسبت عملکردی از سوی خود برگزارکننده PT تعیین می‌شود و به خاطر فرمول پروتوکل آماری، هیچ داده انحرافی رد نخواهد شد. لذا مقادیر نمره Z به نتایج منجر می‌شوند که برای پانزده آزمایشگاه نتیجه رضایت‌بخش و برای سه مشارکت‌کننده پیام هشدار و یک آزمایشگاه پیام اقدام در پی داشته که مستلزم بررسی بیشتر است.



شکل F.1: خلاصه گرافیکی نتایج نمره Z کل به دست آمده از سوی آزمایشگاه‌های مشارکت‌کننده با پیروی از پروتوکل‌های آماری مختلف

F.2.1 ملاحظات نهایی

با توجه به این واقعیت که داده‌های مشارکت‌کنندگان PT معمولاً دارای چولگی سنگین هستند و حضور داده‌های انحرافی نیز بسیار متداول است، لذا آماره‌های کلاسیک به تخمین بیش از حد انحراف استاندارد ارزیابی مهارت منجر می‌شوند. بدین ترتیب تعداد زیادی از نمره‌های Z به شکل رضایت‌بخش لحاظ خواهند شد. لذا استفاده از متدهای آماری استوار توصیه می‌شود، چرا که این متدها نسبت به داده‌های انحرافی حساسیت کمتری دارند. این حالت به طور آشکار در آماره‌های مقایسه‌ای ارائه شده در مجموعه داده‌های مثال بالا مشاهده می‌شود.

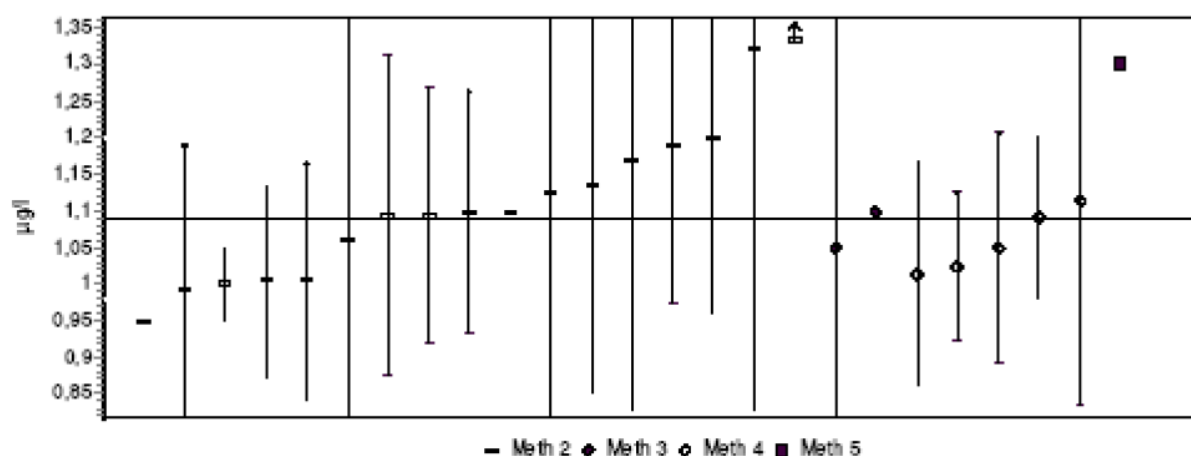
پروتوکل‌های استوار بحث شده در این سند جدای از حالت‌هایی که در آن فرض می‌شود تمام مشارکت‌کنندگان، عملکرد آنالیزی یکسانی ندارند، به خصوص برای داده‌های با توزیع نرمال و حداکثر ۱۰٪ داده‌های انحرافی که تک‌مندی و نسبتاً متقارن می‌باشند، قابل کاربرد است. با این حال، مشاهده می‌شود که متد میانه و NIQR برای داده‌های نامتقارن استوارتر است، در حالی که در حالت‌های چندمندی یا توزیع داده‌های همراه با چولگی، کاربرد مدل‌های مخلوط و توابع دانسیته هسته‌ای می‌بایست در نظر گرفته شود. در این مثال، به کارگیری متد آماری کلاسیک و محکم برای داده‌های PT آشکارا نشان می‌دهد که مقادیر میانگین تقریباً مشابه هستند، در حالی که تفاوت‌های معنی‌داری در مقدار انحراف استاندارد مشاهده می‌شود و در بعضی حالت‌ها این مقدار برای برازش هدف PT خیلی بزرگ است. به علاوه، رسیدن به تخمین مناسب از پارامتر انحراف استاندارد کل که از طریق آن بررسی دقت متد آنالیزی امکان‌پذیر شده، همین‌طور فراهم کردن یک ارزیابی عملکرد سازگار با الزامات بین‌مقایسه‌ای نیز مهم است.

سرانجام اینکه، کاربرد معیار برازش هدف می‌بایست الزام کاربر را توصیف کرده و از دوری به دور دیگر نیز منسجم باشد، بدین ترتیب نمره‌های عملکرد در دوره‌های متوالی قابل مقایسه خواهند بود. ویژگی انحراف استاندارد ارزیابی مهارت بر اساس انحراف استاندارد نسبی، به جای توصیف ساده نتایج داده‌ها، بخش زیادی از هدف کیفیتی را برآورده کرده، و منعکس‌کننده برازش هدف می‌باشد.

F.3 مقدار اسمی وابسته به متد

برگزارکننده PT می‌بایست قبل از تصمیم نهایی درباره مقدار اسمی اختلاف‌های ممکن بین نتایج به دست آمده از متدهای آنالیزی مختلف را لحاظ کند. برای مثال در آنالیز غلظت مقادیر پایین فلز در آب، چنانچه از چند متد مختلف استفاده شده باشد، نتایج، تغییر قابل ملاحظه‌ای را نشان می‌دهند. متد ICP-MS حساس‌ترین متد برای اندازه‌گیری غلظت‌های پایین فلز است و مقدار میانگین نتایج به دست آمده از این متد ممکن است در تعیین مقدار اسمی مناسب‌تر باشد. روش‌های پیش‌تیمار متفاوت می‌توانند بر روی نتایج تأثیر داشته باشند. اگر نتایج به دست آمده با استفاده از اسیدهای مختلف در پیش‌تیمار، برای مثال اسیدهای هضم متفاوت در آنالیز فلزات در خاک یا لجن، به طور جداگانه تیمار شوند، در این صورت مقادیر میانگین متفاوتی ایجاد خواهد شد.

مثال F.1: تعیین کادمیم (مقدار میانگین = $1.09 \mu\text{g/L}$ ، تخمین زده شده با محاسبه) با استفاده از متدهای تجزیه‌ای مختلف (متد ۲ = GAAS، متد ۳ = ICP-OES، متد ۴ = ICP-MS، متد ۵ = دیگر متدها)



مثال F.2: مقادیر میانگین نتایج به دست آمده با استفاده از اسیدهای متفاوت برای تیمار نمونه لجن در یک PT برای آنالیز فلزها

فلز	پیش تیمار	میانگین (mg/kg)	SD	N
Cr	LN1	23.36	4.36	15
	LO1	34.91	10.69	7
Al	LN1	8918	1200	13
	LO1	16451	3778	8

LN1– digestion with HNO_3 or with $\text{HNO}_3 + \text{H}_2\text{O}_2$

LO1 – digestion $\text{HNO}_3 + \text{HCl}$ (تیزاب سلطانی)

پیوست G: مثال ارزیابی عملکرد طولانی مدت

پایداری عملکرد اندازه‌گیری‌های آزمایشگاه در طی زمان مهم است. این مثال [24] که بر پایه آنالیز حاصل از متد ایستایی خون می‌باشد روشی را برای ارزیابی عملکرد آنالیزی طولانی مدت آنالیز کمی نشان می‌دهد، البته به شرطی که داده‌های کافی وجود داشته و غلظت‌های متفاوت نیز در طرح PT استفاده شوند.

نتایج تست آزمایشگاهی منفرد، با مقدار توافقی، که به شکل میانگین مقدار تمام نتایج تست‌ها یا مقدار میانگین یک گروه متد همکار تعیین شده‌اند، مقایسه می‌شود. در این مدل، مقدار توافقی می‌تواند به سادگی با یک مقدار اسمی یا هدف جایگزین شود.

دو متغیر برای ارزیابی عملکرد آزمایشگاه استفاده می‌شوند:

۱. CV تجزیه‌ای طولانی مدت (LCVa)

۲. بایاس کل طولانی مدت

عملکرد تجزیه‌ای طولانی مدت بر پایه رگرسیون خطی (بر پایه حداقل مربعات) و استفاده از مقدار توافقی (X) به عنوان متغیر مستقل و مقدار آزمایشگاهی (Y) به عنوان متغیر وابسته بنا نهاده می‌شود.

G.1 ضریب تغییر تجزیه‌ای طولانی مدت (LCVa)

ضریب تغییر تجزیه‌ای طولانی مدت (LCVa) بر پایه تغییرپذیری خط رگرسیون (Sy/x) و مقدار میانگین تمام مقادیر توافقی ($\bar{X}$) بنا نهاده می‌شود.

برای مقایسه LCVa بین آزمایشگاه‌ها، مقدار آن می‌بایست پس از تنظیم بایاس (b) محاسبه شود. لذا، LCVa، با استفاده از فرمول زیر محاسبه می‌شود:

$$LCVa = \frac{(S_{y/x}/b)}{\bar{X}} \times 100\%$$

G.2 بایاس کل طولانی مدت

بایاس کل طولانی مدت (B) می‌تواند از فرمول زیر محاسبه شود:

$$B = \sqrt{\frac{n-1}{n} \cdot (b-1)^2 \cdot S_x^2 + (\bar{Y} - \bar{X})^2}$$

بایاس کل طولانی مدت شامل دو مؤلفه است: بایاس متناسب (PB) و بایاس ثابت (CB)

بایاس متناسب بر اثر انحراف شیب ایجاد شده و به تغییرپذیری مقدار توافقی بستگی دارد.

$$PB = \sqrt{\frac{n-1}{n} \cdot (b-1)^2 \cdot S_x^2}$$

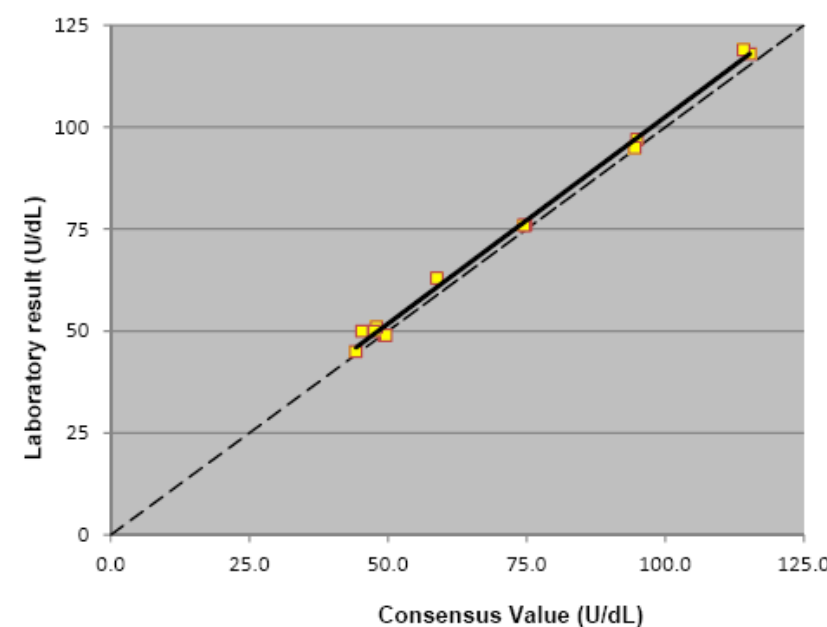
بایاس ثابت، منعکس‌کننده انحراف از مقدار توافقی است.

$$CB = \sqrt{(\bar{Y} - \bar{X})^2}$$

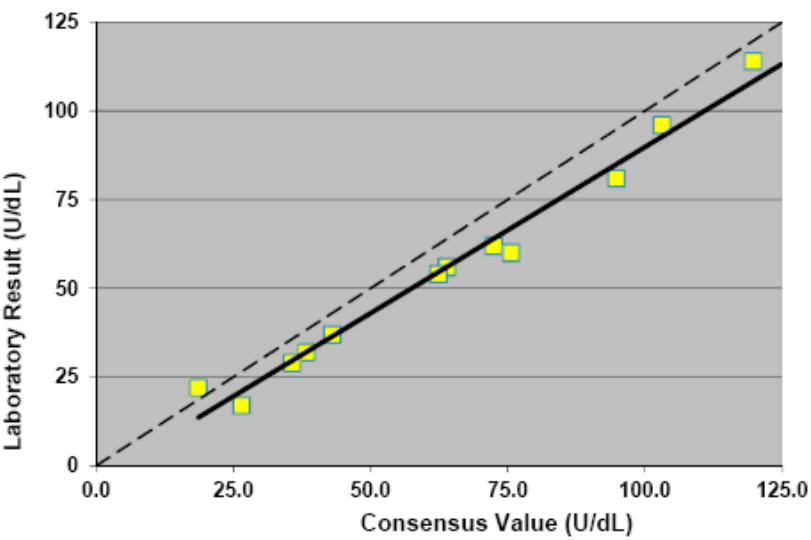
نسبت بایاس متناسب و ثابت نشان می‌دهد که آیا بایاس عمدتاً بر اثر خطاهای کالیبراسیون (بایاس متناسب) یا بر اثر دیگر عوامل نظیر اثرات ماتریکس (بایاس ثابت) ایجاد شده است یا خیر.

ارزیابی طولانی‌مدت دو پارامتر در زیر ارائه شده است.

جدول G.1: ارزیابی طولانی‌مدت آنتی‌ترومبین

Round	Lab Result (U/dL)	Cons. Value (U/dL)	Antithrombin			
2005-1		49.8				
2005-2	50.0	45.3				
2005-3	118.0	115.3				
2005-4	49.0	49.3				
2006-1	76.0	74.8				
2006-2	119.0	114.1				
2006-3	49.0	49.6				
2006-4	76.0	74.5				
2007-1	97.0	94.9				
2007-2	51.0	47.9				
2007-3	45.0	44.2				
2007-4	63.0	58.8				
2008-1	50.0	47.5				
2008-2	95.0	94.5				
Mean	72.2	68.6				
SD		26.2				
Number	13					
LCVa		2.6%	Intercept	1.08	Variability	1.84
Bias		5.2%	Slope	1.01	Corr.coeff	0.9958

جدول G.2: ارزیابی طولانی‌مدت لخته پروتئین C

Round	Lab Result (U/dL)	Cons. Value (U/dL)	Protein C Clot			
2004-1		125.0				
2004-2	62.0	72.4				
2004-3	22.0	18.6				
2004-4	114.0	119.7				
2005-1	32.0	38.3				
2005-2	60.0	75.6				
2005-3	96.0	103.1				
2005-4	29.0	35.7				
2006-1	81.0	94.8				
2006-2	17.0	26.5				
2006-3	56.0	63.9				
2006-4	37.0	43.0				
2007-1	54.0	62.4				
Mean	55.0	67.6				
SD		35.0				
Number	12					
LCVa		7.0%	Intercept	-3.86	Variability	4.46
Bias		18.9%	Slope	0.94	Corr.coeff	0.9800

پیوست H: مثال استفاده از عدم قطعیت اندازه‌گیری

در سال ۲۰۰۸، مؤسسه تأیید صلاحیت (IFEP) آزمون مهارت «تست سختی راکول» را سازماندهی کرد. مشارکت در این آزمون برای آزمایشگاه‌های تأیید صلاحیت و آزمایشگاه‌های فاقد صلاحیت آزاد بود. این آزمون مهارت بر اساس استاندارد آزمون 1-6508 EN ISO [25] پایه‌گذاری و بر اساس راهنمای 1-43 ISO/IEC [26] و ISO 13528 [12] طراحی شد.

مشارکت‌کنندگان بلوک‌های مرجع گواهی‌دار (دارای گواهی از MPA NRW دورتموند) در سه سطح مختلف سختی (25 HRC, 45 HRC, 60 HRC) دریافت کردند. علاوه بر این بلوک‌ها یک قطعه ماده آزمایشی استاندارد فولاد سردکاری شده با سختی تقریبی 50 HRC برای مشارکت‌کنندگان ارسال شد.

روش اجرایی تست به شکل زیر تعریف شد:

تکلیف A:

اجرای پنج اندازه‌گیری HRC بر اساس 1-6508 EN ISO [25] بر روی هر یک از سه بلوک مرجع گواهی‌دار.

تکلیف B:

ارزیابی عدم قطعیت اندازه‌گیری: تهیه یک قطعه ماده آزمایشی استاندارد و اجرای پنج اندازه‌گیری سختی بر اساس 1-6508 EN ISO [25].

مشارکت‌کنندگان:

۷۶ آزمایشگاه از ۲۸ کشور در این آزمون مهارت شرکت کردند. ۵۷ مشارکت‌کننده دارای تأیید صلاحیت بر اساس [2] ISO/IEC 17025 بودند.

H.1 ارزیابی آزمون مهارت

ارزیابی نتایج بر اساس ISO 13528 [12] و راهنمای 1-43 ISO/IEC [26] پایه‌گذاری شد. تکلیف A، بر اساس معیار «خطای مجاز ماشین تست» و «تکرارپذیری مجاز ماشین تست» بر پایه 1-6508 EN ISO ارزیابی می‌شود، جدول ۵ [25]. نتایج تکلیف B، تنها برای محاسبه عدم قطعیت اندازه‌گیری استفاده شدند. خطای ماشین تست، E بر اساس معادله (۱) محاسبه می‌شود:

$$(۱) \quad E = \bar{H} - \bar{X}_{CRM}$$

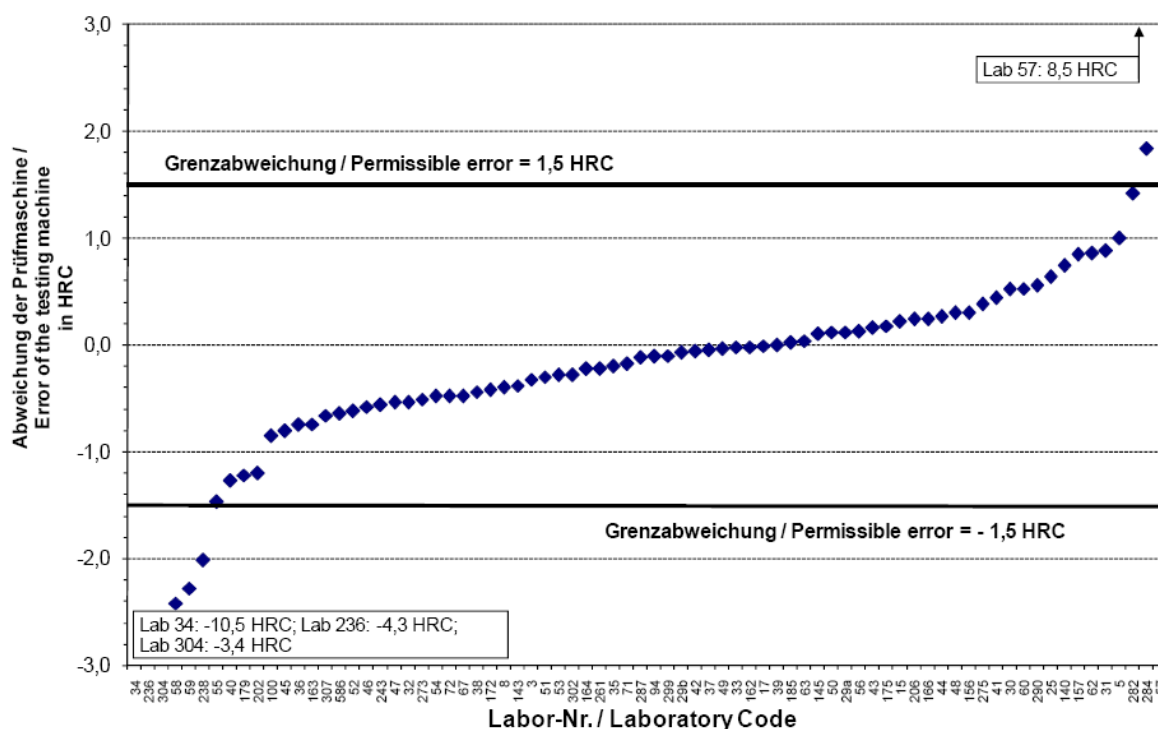
$\bar{H}$: مقدار میانگین (ریاضی) برای پنج اندازه‌گیری روی یک بلوک با سختی معین،

و $\bar{X}_{CRM}$: مقدار مرجع گواهی‌دار مربوط به سختی هر بلوک است

نتایج این تمرین را می‌توان در شکل H.1 مشاهده کرد.

خطای مجاز ماشین تست GA (1a) در 2-6508 EN ISO بیان می‌شود، جدول ۵ [25]:

$$(۱a) \quad -1.5 \text{ HRC} \leq E \leq 1.5 \text{ HRC}$$



شکل H.1: تست A. مثال سطح 45 HRC: خطای ماشین تست

H.2 تعیین عدم قطعیت اندازه‌گیری

ارزیابی عدم قطعیت اندازه‌گیری می‌بایست دقیقاً بر اساس EN ISO 6508-1 [25]، و رویکرد متد ۱ (به شکل مخفف M1) انجام شود که بر پایه کد عملیاتی UNCERT، شماره ۱۴، [28] پایه‌گذاری می‌شود. علاوه بر آزمایش ماده مرجع گواهی‌دار، این متد نیازمند پنج اندازه‌گیری سختی بر روی یک ماده استاندارد است.

از تمام مشارکت‌کنندگان تقاضا شد تا عدم قطعیت اندازه‌گیری بسط‌یافته‌شان را برای اندازه‌گیری انجام شده روی ماده استاندارد (تکلیف B) گزارش کنند. IFEP برای تمام مشارکت‌کنندگان، عدم قطعیت اندازه‌گیری منفرد را بر اساس معادلات (۴) و (۵) بر پایه [29] و [30] محاسبه کرد.

$$(4) \quad U = 2 * \sqrt{u_E^2 + u_{CRM}^2 + u_H^2 + u_{\bar{x}}^2 + u_{ms}^2}$$

$$(5) \quad \tilde{U} = \frac{U}{\bar{X}_{CRM}} \times 100$$

در اینجا:

U : عدم قطعیت اندازه‌گیری بسط‌یافته

$\tilde{U}$: عدم قطعیت اندازه‌گیری بسط‌یافته نسبی

u_E : عدم قطعیت اندازه‌گیری استاندارد ماشین تست

u_{CRM} : عدم قطعیت اندازه‌گیری استاندارد بلوک مرجع گواهی‌دار

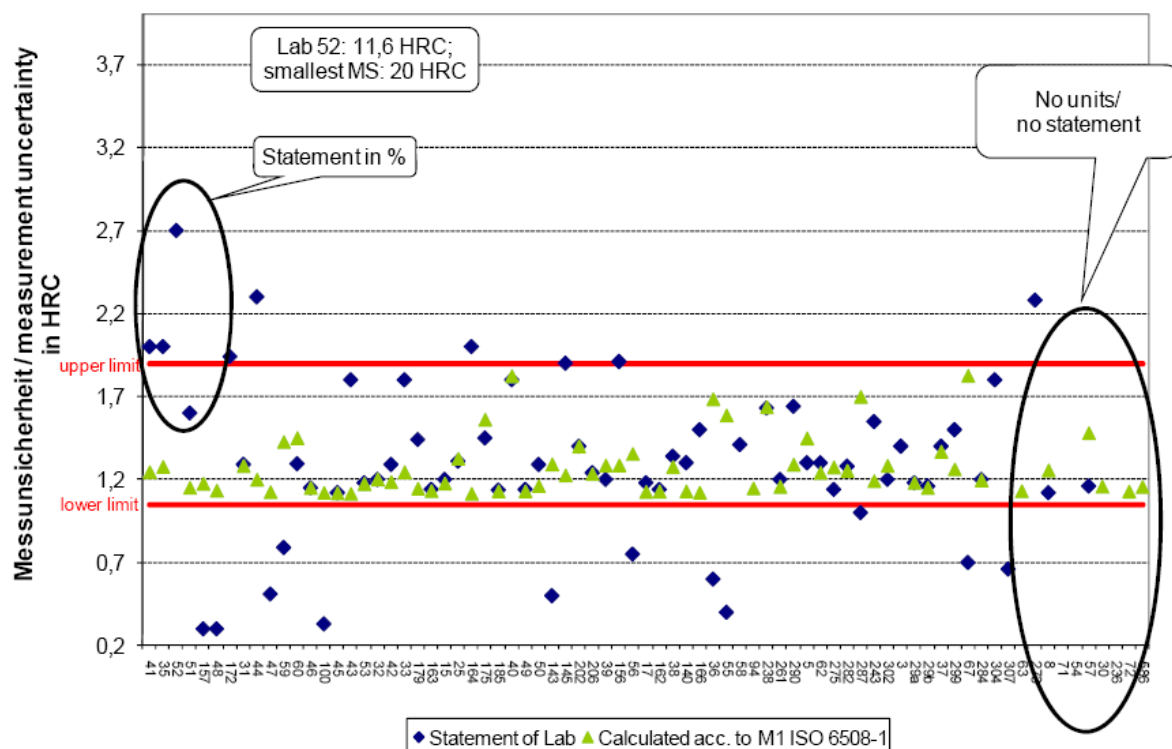
u_H : عدم قطعیت اندازه‌گیری استاندارد ماشین تست آزمایشگاهی که سختی بلوک مرجع گواهی‌دار را اندازه می‌گیرد

$u_{\bar{x}}$: عدم قطعیت اندازه‌گیری استاندارد ناشی از تست ماده

u_{ms} : عدم قطعیت اندازه‌گیری استاندارد بر اساس تفکیک‌پذیری ماشین تست

$\bar{X}_{CRM}$: مقدار مرجع گواهی‌دار بلوک مرجع گواهی‌دار

حداقل سطح عدم قطعیت اندازه‌گیری بسط‌یافته نسبی $\bar{U}$ ، با ترکیب فاکتورهای ثابت u_E ، u_{ms} و u_{CRM} ارائه می‌شود، که مقدار آن 2.1% است. نتایج مقادیر عدم قطعیت اندازه‌گیری گزارش شده از سوی مشارکت‌کنندگان را می‌توان در شکل ۲ مشاهده کرد. به علاوه، برای به تصویر کشیدن تفاوت‌ها، محاسبات انجام شده با IFEP نیز نشان داده شده است. برای مشارکت‌کنندگانی که تمام اطلاعات درخواستی را فراهم نکرده بودند عدم قطعیت اندازه‌گیری محاسبه نشد.



شکل H.2: مقایسه عدم قطعیت اندازه‌گیری بیان شده از سوی مشارکت‌کنندگان و محاسبه شده از سوی IFEP بر اساس متد M1، ISO 6508-1

H.3 نتایج و بحث

بر پایه روش اجرایی متد آزمایشی، مقادیر عدم قطعیت اندازه‌گیری گزارش شده از سوی آزمایشگاه‌های مشارکت‌کننده در PT، همبستگی خوبی بین نتایج آزمایشگاه‌ها و نتایج محاسبه شده از سوی سازمان‌دهنده را نشان می‌دهند. حدود ۲۰٪ از مشارکت‌کنندگان مقادیری را گزارش دادند که خیلی کوچک بود و بر اساس فاکتورهای تثبیت شده‌ای که نمی‌توانند پایین‌تر از حداقل تعریف شوند، منطقی نبودند. این مشارکت‌کنندگان می‌بایست متد تخمین عدم قطعیت اندازه‌گیری‌شان را تصدیق کرده و از نتایج این آزمون مهارت درس‌های لازم را بگیرند. مشارکت‌کنندگانی هم که تمام مقادیر لازم برای محاسبه را بیان نکرده یا عدم قطعیت اندازه‌گیری را بیش از حد مورد انتظار تخمین زده بودند، ابزارهایی را برای تصدیق متدهایشان در اختیار داشتند. برای این متد آزمایشی خاص، محاسبه عدم قطعیت اندازه‌گیری مربوطه بر اساس استاندارد آزمون پایه‌گذاری می‌شود. برگزارکننده این آزمون مهارت از طریق فرمول‌های ارائه شده می‌توانست علاوه بر نتیجه آزمایش، گستره قابل قبول مقادیر عدم قطعیت اندازه‌گیری برای مشارکت‌کنندگان را نیز نشان دهد. اثر آموزشی یک چنین آزمون مهارتی ارزشمندتر از PT است که تنها آماره‌های عملکردی را گزارش می‌کند. در آینده، مشارکت‌کنندگان در PT می‌بایست برگزارکننده را بر اساس فرصت‌هایی نظیر فرصت آموزشی اضافه شده به طرح PT انتخاب کنند.

کتاب‌شناسی

- [1] ISO 5725-1994, Accuracy (trueness and precision) of measurement methods and results
- [2] ISO/IEC 17025:2005, General requirements for the competence of testing and calibration laboratories
- [3] ISO 15189:2007, Medical laboratories – Particular requirements for quality and competence
- [4] ISO/IEC 17043:2010, Conformity assessment — General requirements for proficiency testing
- [5] ISO/IEC Guide 99-12:2007/JCGM 200:2008, International vocabulary of metrology — Basic and general concepts and associated terms (VIM, 3rd edition)
- [6] EA-4/18, Guidance on the level and frequency of proficiency testing participation
- [7] EURACHEM - Information leaflet on Pre- and post-analytical proficiency testing, First English edition, 2009-05-14
- [8] ILAC Brochure: 2008, Benefits for laboratories participating in proficiency testing programs
- [9] DIRECTIVE 98/79/EC of the European parliament and of the council of 27 October 1998 on in vitro diagnostic medical devices, Official Journal of the European Communities.
- [10] Eurolab Technical Report 1/2007 – Measurement uncertainty revisited, Alternative approaches to uncertainty evaluation March 2007
- [11] ISO 21748:2010, Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation
- [12] ISO 13528:2005, Statistical methods for use in proficiency testing by interlaboratory comparisons
- [13] Analytical Methods Committee - Robust Statistic Part I & II. Analyst 114, 1693-1702, (1989)
- [14] Thompson, M. and Ellison, S.L.R., Fitness for purpose – the integrating theme of the revised Harmonised Protocol for Proficiency Testing in Analytical Chemistry Laboratories. Accred. Qual. Assur. 11, 373-378, (2006)
- [15] Tholen, D.W., Statistical treatment of proficiency testing data, Accred. Qual. Assur., 3 (1998), 362-366
- [16] The International Harmonized Protocol for the Proficiency Testing of Analytical Chemistry Laboratories. Pure Appl. Chem., 78, 1; 145-196, (2006)
- [17] VAM Bulletin - Ellison, S.L.R., LGC, , Issue 33 - Autumn 2005, 21-22
- [18] BIPM, IEC, IFCC, ISO, IUPAC, OIML: Guide to the Expression of Uncertainty in Measurement. ISO, Geneva, Switzerland, First Edition (1993)
- [19] Uhlig, S. and Lischer, P - Statistically-based performance characteristics in laboratory performance studies, Analyst, 123 , 167-172 (1998)
- [20] Rosario P., Martínez JL, Silván JM . - Comparison of different statistical methods for evaluation of proficiency test data, Accred. Qual. Assur. 13:493:499 (2008)
- [21] Harms, AV. - A new approach for proficiency test exercise data evaluation, Accred. Qual. Assur. 14:253:261 (2009).
- [22] Rousseeuw P.J. and Leroy A.M. - Robust regression and outlier detection. Wiley, New York (1987)
- [23] Analytical Methods Committee - Robust Statistic Part I & II. Analyst 114, 1693-1702 (1989)
- [24] Meijer, P., DE Maat, M.P.M, Kluft, C., Haverkate, F., and Hans C. van Houwelingen, H.C. - Long-Term Analytical Performance of Hemostasis Field Methods as Assessed by Evaluation of the Results of an External Quality Assessment Program for Antithrombin, Clinical Chemistry 48:7 1011–1015 (2002)

- [25] EN ISO 6508-1:2006, Metallic materials – Rockwell hardness test. Part 1: Test method (scales A, B, C, D, E, F, G, H, K, N, T), Beuth Verlag, Berlin, March 2006
- [26] ISO/IEC Guide 43-1: 1997, Selection and use of proficiency testing scheme by laboratory accreditation bodies. (This standard was withdrawn on 29.01.2010 and revised by ISO/IEC 17043: 2010.)
- [27] EN ISO 6508-2:2006, Metallic material – Rockwell hardness test. Part 2: Verification and calibration of testing machines (scales A, B, C, D, E, F, G, H, K, N, T), Beuth Verlag, Berlin, March 2006
- [28] UNCERT COP 14:2000, Gabauer, W., Manual of Codes of Practice for the Determination of Uncertainties in Mechanical Tests on Metallic Materials, The Estimation of Uncertainties in Hardness Measurements, Project, No. SMT4-CT97-2165.
- [29] Weißmüller, C., Frenz, H. - Modelle zur Ermittlung der Messunsicherheit in der Härtpfung; Statistische Auswertung eines Ringversuchs mit 90 Teilnehmern, In: Tagungsband Werkstoffprüfung, Neu-Ulm 25.-26.11.2004, MAT INFO, Frankfurt 2004.
- [30] Weißmüller, C., Frenz, H. - Berechnung der Messunsicherheit für mechanisch- technologische Prüfverfahren, Seminarunterlagen, unveröffentlicht, Recklinghausen, 2005.

Selection, Use and Interpretation of Proficiency Testing (PT) Schemes

Second Edition - 2011