

技能試験の概要とその統計的方法

外部精度管理である技能試験への継続的な参加は、食品、環境、製品安全、臨床検査の分野で広がりつつある。技能試験結果から、最大限の情報を得て、技能の向上に活かしていくためには、技能試験の透明で質の高い運用と、信頼性のある技術的な検討が欠かせない。本稿では、① 技能試験がどう設計され、② その報告値がどう統計処理され、③ その結果が技能の確認と向上にどう利用されるかについて解説する。

城野 克広，津越 敬寿

1 はじめに

食品の安全性がどう確保されているか、という話題からこの解説を始めよう。食のトレーサビリティなどと言われて久しいが、普通の食卓ではなんとなく買った食材を、なんとなく食べている方も多いだろう。

実は、飲食に起因する衛生上の危害の発生を防止し、国民の健康の保護を図ることを目的とする食品衛生法により、リスクのある食品に対しては規格を定めることができ、国立あるいは都道府県や政令指定都市設置の衛生研究所、又はその他の登録検査機関で規格に基づいた検査を実施することが定められている。この検査機関として登録されるには、設備が十分に整っているとか、適切に要員が配置されているとかいう要件を満たさなくてはならない。このような観点から検査品質を確保することは、もちろん重要である。

一方、衛生研究所及び登録検査機関には、平成9年から「外部精度管理調査」なるものへの継続的な参加が求められている。一例として、平成17年に財団法人食品薬品安全センター（以下、食薬センター）が提供した外部精度管理調査¹⁾²⁾の一項目である「残留農薬検査」の試験を紹介しよう。この試験では、ほうれんそうペーストに、農薬（クロルピリホス）の標準液を加え、よく混合したものを取り分け、参加を表明した180以上の衛生研究所、登録検査機関に一斉に配付した。参加試験所（この場合は衛生研究所や登録検査機関）は約1ヶ月のうちに農薬の含有量の測定を実施し、測定値を食薬センターに知らせる。このとき、参加試験所には農薬がどれだけ加えられたかは知らされない。試験方法については、それぞれの参加試験所で普段通りの分析が実施されるよう通常の標準作業手順書（SOP）に従うことが指示される。測定値の報告を受けた食薬センターは、その結果を取りまとめ、他の参加試験所を特定できないよ

う統計処理した結果を参加試験所に通知する。

簡単に言えば、他の多くの参加試験所と同じような値を報告した参加試験所のパフォーマンスは満足なものと言えるし、それとは異なる値を報告した参加試験所については、試験手順や試験環境を是正するきっかけとなる。外部精度管理調査に継続的に参加し、悪いパフォーマンスとなった場合には是正を行うことは、参加試験所が十分な技能を持っていることの一つの根拠になる。このような管理の下、消費者は自分で育てたわけでも、自分で分析したわけでもない野菜の農薬の残存について、気にすることもなく野菜を口にすることができるのである。

図1に示すように、精度管理には内部精度管理と外部精度管理がある。外部精度管理とは技能試験への継続的な参加のことであり、食薬センターが提供する外部精度管理調査はその例である。技能試験とは、ISO/IEC 17043の一致規格として2011年に発行されたJIS Q 17043「適合性評価-技能試験に対する一般要求事項」³⁾によれば、「試験所間比較による、事前に決めた基準に照らしての参加試験所のパフォーマンスの評価」のことである。なお、試験所間比較は「事前に定めた条件に従って、二つ以上の試験所が、同一品目又は類似品目で行う、測定又は試験の企画、実施及び評価」を意味して

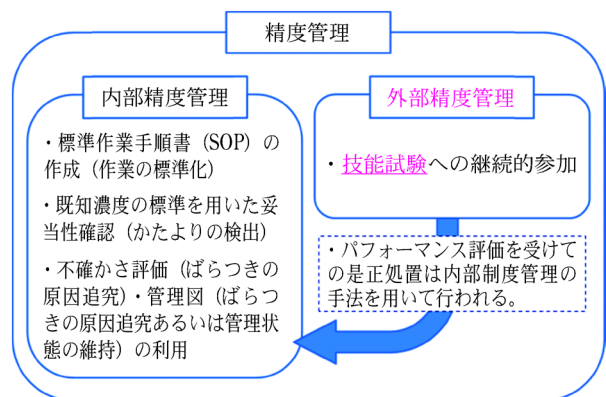


図1 精度管理の概念図

いる。近年、技能試験の重要性が高まってきている。その理由は、上のような試験所認定・登録制度が様々な分野で普及しつつあることが大きい。

特に国際規格に基づく試験所の認定・登録制度が最近話題である。独立行政法人製品評価技術基盤機構認定センター（以下、IAJapan）では四つの認定・登録プログラムを設け、平成25年3月現在で計691の試験所及び校正機関（校正とは測定器の指示値の評価）が認定・登録を受けている。この認定・登録は、試験所のマネジメントシステムの国際規格ISO/IEC 17025の同一規格JIS Q 17025「試験所及び校正機関の能力に関する一般要求事項」⁴⁾への適合性評価をもって行われているものが多い。JIS Q 17025では、技能試験への参加が強く推奨されており、IAJapanの認定・登録機関のほとんどが技能試験に参加している。民間の認定機関の代表格である公益財団法人日本適合性認定協会（以下、JAB）では平成25年3月現在で、ISO/IEC 17025に基づいて277の試験所及び校正機関を認定しているなど計379機関の認定を実施している。JABでも認定を受ける機関のほぼすべてが技能試験に参加している。

計量法に基づく登録制度である計量証明事業も技能試験に大きくかかわる。計量証明事業者として登録する際には、外部精度管理の実績は求められない。しかし、公益社団法人日本分析化学会（以下、分析化学会）と一般社団法人日本環境測定分析協会（以下、日環協）は、長らく大気・水質汚染のモニタリングなどの環境分析分野の技能試験を提供し、試験の精確さの向上に努めてきた。現在では、地方自治体の環境計量証明事業において、日環協や分析化学会が提供する技能試験への参加実績が入札の条件になっていることもある。一例として、滋賀県での水質汚濁対策事業にかかる入札条件を図2に示す。

このような状況にあり、技能試験により試験所の技能を評価する重要性が高まるとともに、『技能試験の品質』が問われるようになってきている。JIS Q 17043は技能試験提供者のマネジメントシステム規格である。この規格では、技能試験提供者が適合すべき要求事項が示されている。また、技能試験におけるパフォーマンスの評価のためには、統計学的な知見が必要である。これに対応し、JIS Z 8405「試験所間比較による技能試験のための統計的方法」⁵⁾（ISO 13528の同一規格）があり、JIS Q 17043を補足する役割を果たしている。

本稿は、技能試験がどう設計され（2章）、その報告値がどう統計処理され（3章）、その結果が技能の確認と向上にどう利用されるか（4章）についての解説である。これにより、技能試験をより身近に、より広く、より正しく利用していただくきっかけとなることを願うものである。

2. 入札参加条件

(1) 滋賀県内および滋賀県の隣接府県（京都府、福井県、岐阜県、三重県）において、環境計量証明事業（濃度(水・土壌)）の登録がある事業者とする。ただし、(1)河川環境基準監視調査については、滋賀県内で登録のある事業者とする。

(2) (1)の登録の環境計量証明事業所において、3. に示すとおり(社)日本環境測定分析協会が実施する「ISO/IEC 17043に基づく技能試験」(以下、「技能試験」という。)に参加していること。

(3) 入札に付する事業ごとの入札参加条件等については、上記に示す条件に加えて、事業者ごとの入札の公告に示すことから、十分に確認すること。

3. 技能試験への参加について

3. 1. 必須とする技能試験項目について

	河川環境基準 監視調査	地下水概況 調査	地下水定期 モニタリング調査	工場・事業場 排水等監視調査
水中の生活環境 項目試験	○			○
水中の富栄養化 成分分析	○			○
水中の陰イオン 分析	○	○	○	○
水中の金属分析	○	○	○	○
水中の残留農薬 分析	○	○	○	○
水中の揮発性成 成分分析	○	○	○	○

図2 滋賀県での水質汚濁対策事業にかかる入札参加条件の例 (<http://www.pref.shiga.lg.jp/d/biwako/files/files/h26nyusatsusankazyoken.pdf>, 2013年9月17日最終確認)

2 技能試験の概要（企画・設計）

技能試験では、測定値（分野によって計測値や分析値ともいうが本稿では測定値ということにする）の相互比較を行い、その一致度を評価することが行われる。真に正しい値と比較することができれば、測定値が正しいか誤っているかとの判定も可能であろうが、残念ながらその真に正しい値を知ることはできない。よって、より正しいと皆が納得できる値と比較することになる。この皆が納得できる値を合意値と呼ぶ。技能試験では、付与値（誤解を恐れずに言う「正解」となる値）に合意値を用いる。そして、合意値には、例えば標準物質や認証標準物質（候補を含む）を参加試験所に配付する場合にはそれらに付された値（それぞれ特性値や認証値と呼ばれる）を用いることができるが、参加試験所が報告した測定値から求めることも多い。詳細は3・2節に記す。

技能試験提供者は、このような配付試料（JIS Q 17043では“技能試験品目”として定義されている）の準備・調製、参加試験所への配付、測定値の報告の受領、その報告値の統計解析、参加試験所のパフォーマンス評価結果の通知（報告書）を行う。この際に用いるマネジメントシステムに対し、JIS Q 17043の要求事項への適合が技能試験の品質を担保するために求められる。

2・1 JIS Q 17043の概要

JIS Q 17043は5章から構成されており、1～3章は他の規格同様「適用範囲」、「引用規格」、「用語及び定義」である。4章「技術的要求事項」がこの規格の肝である。5章「管理上の要求事項」も非常に重要なのだが、JIS Q 9001やJIS Q 17025といったマネジメント規格にはほぼ共通する内容が記されているので、別の機会に確

表1 JIS Q 17043 4章の概略

項タイトル	各項の内容
4.1 一般	技能試験提供者に求められる能力
4.2 要員	適切な要員の準備、特定の要員へ権限を与えることなど
4.3 機器、設備及び環境	技能試験スキームの運用に適切な設備を用意することなど
4.4 技能試験スキームの設計	技能試験スキームの目標、目的および基本設計を示した計画（参加条件、参加者が測定を行う事柄の情報、談合やかいざんを防ぐ手立て、統計分析、結果を公表する範囲 etc.）など
4.5 方法又は手順の選択	参加者は日常の手順で試験をすることが期待されていること、技能試験提供者が測定方法を指定してもよいことなど
4.6 技能試験スキームの運用	参加者への事前通知、指示書の内容（日常のサンプルと同様に取り扱う必要性、試験又は校正に影響する要素の詳細 etc.）、品目の包装及びラベリングなど
4.7 データ分析及び技能試験スキーム結果の評価	パフォーマンスの評価は外部に委託してはならないこと、過去の試験や類似試験との比較や方法・手順間のばらつきなどの事項に関して専門的所見を提示することなど
4.8 報告書	最終報告書の承認を外部に委託してはならないこと、報告書に含まれる事項（技能試験提供者の名称・連絡先、外部委託した活動の名称、参加者のパフォーマンスに関するコメント、統計分析の手順 etc.）など
4.9 参加者との連絡	参加者に提供しなくてはならない情報、パフォーマンスの評価についての異議申立ての手順など
4.10 機密保持	機密保持について

この表中で「参加者」としているものは、本文中の「参加試験所」に該当する。JIS Q 17043 中では、表中と同様に「参加試験所」ではなく「参加者」の用語が使用されているので、注意されたい。

認していただくほうがよいだろう。4章の概要を表1に示す。技能試験の設計においては、様々な点に注意しなくてはならない。例えば、参加試験所のパフォーマンスの評価や最終報告書の承認は外部委託してはならない。外部委託した活動については報告書に記載することが求められる。また、参加試験所が普段の手順で測定をすることを前提に技能試験を設計する必要がある。報告書には参加試験所のパフォーマンスに関する技術的コメントを含め、参加試験所の改善活動に資するものにしなければならない。

2.2 技能試験の留意点—用語を通じて

JIS Q 17043 の特に混乱しやすい点は、実は用語を理解することにより明らかにできることが多い。本節では、用語を洗い直すことを通じて、技能試験において留意すべき点について述べる。

まず「技能試験」との呼び名であるが、もちろん proficiency testing の日本語訳である。ここで、「試験」

という単語の持つイメージがしばしば問題となる。そのイメージとはいわゆる入学試験や資格試験のような「合否判定を行うもの」である。その結果、技能試験でのパフォーマンス評価結果で一喜一憂するような状況が見られる場合がある。ところが、技能試験は精度管理の一環として行われるものであり、試験所のマネジメントシステムが（測定の技術的能力も含めて）機能しているかをチェックし、PDCA (plan, do, check, act) サイクルの check を担うものである。従って、「満足」の評価を得たから合格とするものではなく、技能試験への参加を通してマネジメントシステムの点検を行うのが本来である。

次に技能試験の「技能」である。これも広く見られる理解が「技術的能力」としての取り扱いである。proficiency の訳であるが、proficiency を辞書で引くと「熟達、堪能」とあり、これは technical competence に限定するものではなく、試験所のマネジメントシステムの習熟度を意味すると解される。したがって、分析そのもので正しいと目される測定値を得たとしても、報告時点で（例えば転記ミスや単位の取り違いによって）桁違いの報告値となれば、技術的能力は高くても、技能試験で言うところの技能ではパフォーマンス評価結果として不満足となる。逆に測定値が外れていてもなんらかの原因で付与値に近い報告値となれば満足となり得る。

これらのように、精度管理の一環として行われる技能試験について、その誤解は多く散見される。例えば「精度管理のための技能試験」と言われると、「馬から落馬」のような印象である。

一つの例として、分析化学会が提供する「トレーサビリティと不確かさ理解のための分析技能試験」⁶⁾に触れておく。上記のことを考えると、マネジメントシステムのチェックではなく、参加試験所にトレーサビリティと不確かさの理解を与えることを趣旨とするこの“技能試験”は、本来的には技能試験と呼べる範囲にはないと言える。もちろん、試験所間比較には違いないわけで、例えば、試験所間比較を用いた講習と考えるべきであろう。JIS Q 17043 は技能試験以外を目的とする試験所間比較の運用活動にも適用できることを、その序文に記述しており、JIS Q 17043 を講習のために使用しても差し支えない。

このように技能試験と試験所間比較は同一ではない。JIS Q 17043 の序文に記述されているが、試験所間比較を用い、試験所のパフォーマンスを確定するのが技能試験である。目的を異にする試験所間比較、例えば、ある分析法の妥当性や特性を評価するために行われる試験所間比較は、参加試験所の能力を前提に参加試験所を選定するもので、参加試験所のパフォーマンスを評価するものではなく、これはしばしば「共同実験」と呼ばれる。JIS Q 17043 の旧規格である JIS Q 0043 にて inter-laboratory testing schemes を共同実験スキームと訳し

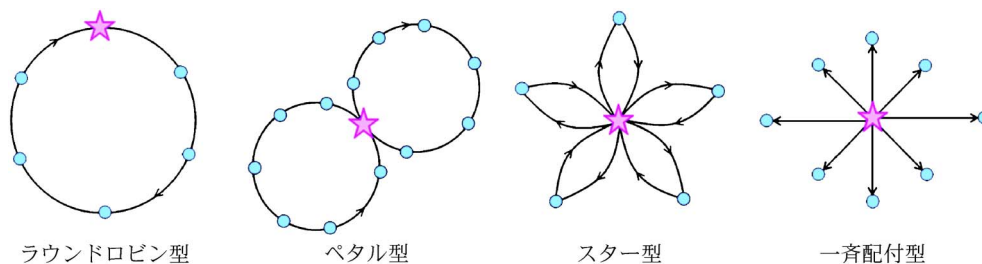


図3 試験所間比較における試料配付の形態 (☆が技能試験提供者, ○が技能試験参加試験所, 矢印が試料回付(配付)の方向)

ていたために、試験所間比較＝共同実験と誤解されている例も見受けられるが、共同実験 (collaborative trials) は試験所間比較を用いる一形態である。技能試験と言った場合には、参加試験所のパフォーマンスの評価までを含み、そこまで責任を持って運用することを意味することになる。技能試験の提供者はその点に注意する必要があるだろう。

試験所間比較にはいくつかの形態がある。図3に示すように、技能試験品目の配付方法で分けられる。例えば同じ品目を参加試験所から参加試験所に回付していくものをラウンドロビン型 (あるいはラウンドロビンテスト) と呼ぶ。品目の回付中に、例えば経時変化のないことの確認等のために技能試験提供者にいったん戻し、さらに次の参加試験所に回付していく場合、あるいは複数の同等品目を複数ルートで回付する場合にはペタル型と呼ぶ。また、品目を参加試験所に配付し、測定後に回収して品目の確認を行う形態をスター型と呼ぶ。ここで、化学系の技能試験やその他の試験所間比較についても「ラウンドロビン」と呼ばれることが多い。しかし、どの試験所間比較を見ても、同一品目を参加試験所間に回付するというものではない。小分けした試料を一斉に参加試験所に配付する方式が採られるが、測定後に回収するものでもない (測定にて試料は消費されるため回収不能である)。この一斉配付型をラウンドロビンテストと呼ぶのは明らかに誤用であり、terminology (術語学) の整備が求められる大きな一例と言える。

2.3 JIS Q 17043 による適合性評価の現状と展望

試験所認定において、IAJapan や JAB などの認定機関とは別の機関が提供した技能試験の結果を認定・登録の材料として利用する場合と、他に適切な技能試験がなく認定機関が提供する場合とがある。本節では、認定機関とは別の機関が提供する技能試験の品質がどのように確かめられているかについて説明する。

JAB では平成 25 年 8 月から JIS Q 17043 による技能試験提供者認定制度を開始したが、JAB の利用する技能試験提供者のすべてが JIS Q 17043 の認定取得を目指す状況にはない (実際のところは、JIS Q 17043 の取得とは別に、試験所から提出された技能試験項目の

みを対象に、JAB が提供する技能試験と同等の品質があるかの確認を行っているとのことである)。

IAJapan でも、平成 24 年度以降に JIS Q 17043 をベースにした評価を行っており (認定ではない)、Web サイトなどで「認定センターが JIS Q 17043 (ISO/IEC 17043) に適合した技能試験として、結果を利用するものです。」という表現により技能試験の品質が確認されたことを明らかにしている⁷⁾。

一方、分析化学会の技能試験もそうであるが、「ISO/IEC 17043 に基づく技能試験」と自ら冠しているのみである場合は、自己適合の意味であろう。今後は、技能試験結果が JIS Q 17025 などの認定に使用されることを前提とする場合には、(本来なら認定取得が望ましいが、) 上のような簡易な評価なども含め、第三者による品質確認は公正さの観点から必須のものになると考えられる。

3 技能試験における統計的方法

3.1 パフォーマンス統計

performance statistics は JIS Q 17043 では「パフォーマンス統計」と訳され、JIS Z 8405 では (JIS Q 17043 制定後の改定がなされていないため、) 「成績を表す統計量」と訳されている。本稿では「パフォーマンス統計」とする。これは技能試験の結果を評価する指標である。いくつかの提案があるが、実際によく使われているのは以下の二つである。

$$z = \frac{x_i - X}{\hat{\sigma}} \dots\dots\dots(1)$$

$$E_n = \frac{x_i - X}{\sqrt{U_{\text{lab}}^2 + U_{\text{ref}}^2}} \dots\dots\dots(2)$$

z は z (ゼット) スコア、 E_n は E_n (イーエヌ) 数と呼ばれる。 X は「付与値」と呼ばれ、この値に近いほどよい結果とされる、いわば『正解のようなもの』である。式(1)中の $\hat{\sigma}$ は「技能評価の標準偏差」と呼ばれ、その試験で許容される試験所のばらつきの標準偏差に相当する。式(2)中の U_{lab} 、 U_{ref} は各試験所が報告した x_i と付与値 X の拡張不確かさである。パフォーマンスの評価は以下のように与えられる。

1) zスコア

- $|z| \leq 2$: “満足” なパフォーマンスを示し、シグナルを出さない。
- $2 < |z| < 3$: “疑わしい” パフォーマンスを示し、警告シグナルを出す。
- $|z| \geq 3$: “不満足” なパフォーマンスを示し、処置シグナルを出す。

2) E_n 数

- $|E_n| \leq 1$: “満足” なパフォーマンスを示し、シグナルを出さない。
- $|E_n| > 1$: “不満足” なパフォーマンスを示し、処置シグナルを出す。

これらのシグナルの基準はzスコア及び E_n 数が標準正規分布に従うことを前提とし、 $|z|=2$ 及び3はそれぞれ5%及び0.3%の、 $E_n=1$ は5%の有意水準を基に定められている。伝統的に校正分野で E_n 数がよく用いられる。試験の分野で不確かさをういつつも、従来のzスコアと同じスケールで解釈ができるように、 ξ (ゼータ) スコアなる指標も提案されている³⁾⁵⁾。

図4aに平成20年に日環協が提供した技能試験「水中の重金属分析」⁸⁾での鉄の含有量について、zスコアによるパフォーマンス評価を示す。この試験では $|z| > 3$ となったのは、全体の6.3%であり、有意水準の0.3%よりかなり大きい。この原因は、簡単に言えば、技能試験結果の分布が正規分布ではないことである。これについては、どのように「付与値」と「技能評価の標準偏差」を算出するかと関連させて、後ほど説明を加える。

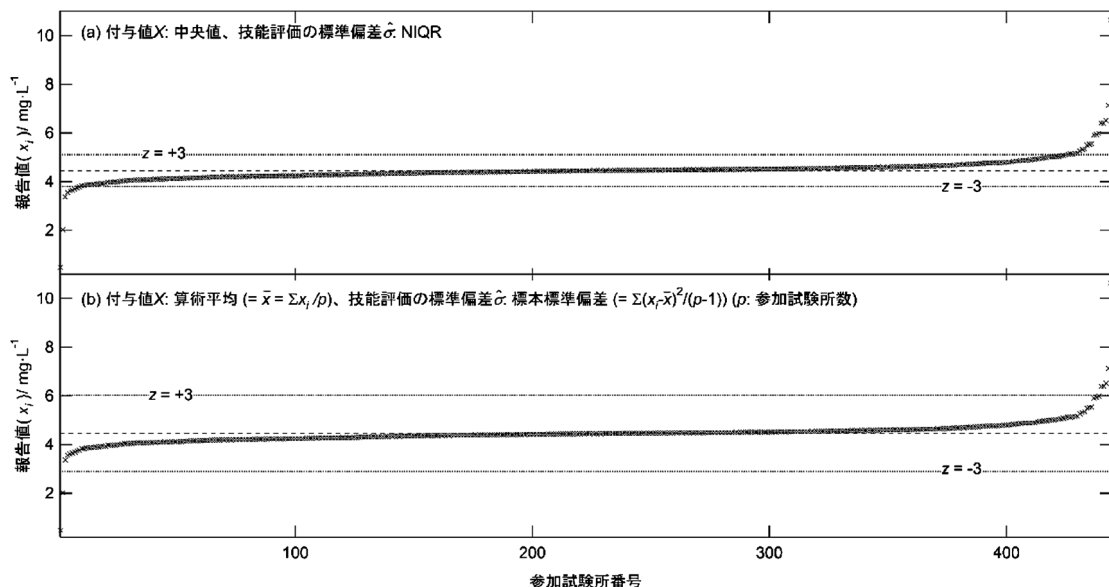
3.2 付与値

付与値の決め方の候補にはいくつかのものがある。配付試料調整時の仕込み量から濃度が分かる場合など一定以下の不確かさで値が分かる試料を配付する場合、その値を付与値としてもよいが、実際にそのような値を付与値としている例は少ない。 E_n 数で評価する場合、熟練試験所が技能試験に参加し、その機関の報告値を付与値とすることが多い。zスコアで評価する場合は、技能試験の報告値の結果を受け、その代表的な値を付与値とする方法がよく用いられる。このとき、単純な算術平均は、外れた値の影響を受けやすいので付与値には不向きであることに注意しなくてはならない。技能試験の目的を考えれば、当然外れ値の存在は想定される。外れ値の影響を受けにくいことをロバストといい、JIS Z 8405ではロバストな付与値の決定方法として、以下のアルゴリズムAなる方法を提案している。p個の試験所が値 x_1, x_2, \dots, x_p を報告しているとき、

- i) x^* を x_i の中央値として、 $s^*=1.483\Delta$ (Δ は $|x_i - x^*|$ の中央値)を計算する。
- ii) $\delta=1.5s^*$ として、 $x_i < x^* - \delta$ のとき、 $x_i^* = x^* - \delta$, $x_i > x^* + \delta$ のとき $x_i^* = x^* + \delta$, それ以外では $x_i^* = x_i$ とする。
- iii) 以下の計算により新しい x^* と s^* を求める。

$$x^* = \sum_i^p x_i^* / p \dots\dots\dots (3)$$

$$s^* = 1.134 \sqrt{\sum_i^p (x_i - x^*)^2 / (p-1)} \dots\dots\dots (4)$$



(a) 付与値に中央値、技能評価の標準偏差に NIQR を用いて解析したもの（詳細は本文 3.2 節および 3.3 節を参照）、(b) 同じく算術平均、標本標準偏差を用いて解析したもの。それぞれ、 $|z| > 3$ となった試験所の割合は (a) 6.3% (全 444 中 28 試験所) と (b) 1.5% (同じく 7 試験所)。

図4 平成20年度日環協「水中の重金属分析」(日環-42)⁸⁾の鉄の技能試験結果

iv) 更新後にこの x^* , s^* の値が更新前と比べて数字の第3有効数字が変化する場合, $x_i = x_i^*$ として ii) に戻る。もし変化しなければ収束したとみなす。

なお、中央値とは、それより大きい値と小さい値の個数が同じになる値のことである。JIS Z 8405 では、このアルゴリズムで求められる x^* を「ロバストな平均」、 s^* を「ロバストな標準偏差」とし、ロバストな平均を付与値とすることが提案されている。ロバストな平均は文献9に示された推定量の特殊なケースである。JIS Z 8405 によれば、健全かつ統計的な基礎があり、計算方法を文書によって規定しうる場合は、他の方法を使用しても構わない。例えば、中央値はロバストな統計量として知られており、実際に付与値としてよく使われる。

ロバストな方法は、この方法を適用する前に外れ値の除去をする必要がないのが強みである。しかし、国際純正・応用化学連合（以下、IUPAC）のプロトコル¹⁰⁾では、これらの方法を適用する前に、中央値の50%以下及び150%以上の値をあらかじめ除去する手順を推奨している。

これらの手法を用いて実際に付与値を計算した結果を表2に示す。正の方向と負の方向に外れたものが同程度あり、影響が相殺されているため、どれも同じような値となる。もちろん、正負の一方に大きくかたよる場合には外れ値の影響が出るので、ロバストな手段により解析することはやはり大切である。

3.3 技能評価の標準偏差

パフォーマンス評価に使用される技能評価の標準偏差は、その試験で許容されるばらつきの標準偏差と言える。例えば、濃度の代表値を質量分率 c で表したときに

表2 平成20年度日環協「水中の重金属分析」(日環-42)⁸⁾の鉄の技能試験結果を用い、様々な方法で算出された付与値及び技能評価の標準偏差

	付与値		技能評価の標準偏差	
	値	標準偏差	値	標準偏差
外れ値に強い(ロバストな)手法	ロバストな平均値	4.43	ロバストな標準偏差	0.23
	中央値	4.45	NIQR (正規四分位範囲)	0.22
	IUPAC 推奨の外れ値除去後のロバストな平均値	4.43	IUPAC 推奨の外れ値除去後のロバストな標準偏差	0.23
	IUPAC 推奨の外れ値除去後の中央値	4.45	IUPAC 推奨の外れ値除去後のNIQR	0.22
ロバストでない手法	算術平均	4.47	標本標準偏差	0.52

実際の試験では中央値とNIQRの組が使われた。IUPAC推奨の外れ値除去とは、中央値の50%以下と150%以上を除く操作のこと。ロバストな平均とロバストな標準偏差のように同じ行の方法が対になって使われやすい。算術平均と標本標準偏差は技能試験で用いられることは少ない。

$\hat{\sigma} = 0.02c^{0.8495}$ で与えられる Horwitz 曲線⁵⁾は再現精度の一般的モデルであり、その精度は許容されるとして、技能評価の標準偏差として使用されることがある。

許容されるばらつきが『技能試験の報告値の大部分が従う分布のばらつき』であると考えるときには、技能試験の結果から、技能評価の標準偏差を算出することになる。アルゴリズムAのロバストな標準偏差はその一例となる。ロバストな平均を付与値とした場合、ロバストな標準偏差を採用することが多い。

付与値として中央値を用いる場合、正規四分位範囲NIQR (normalized interquartile range) がよく用いられる。NIQRは四分位範囲IQRを0.7143倍したもので、IQRとは Q_1 より小さい報告値の個数と、 Q_3 より大きい報告値の個数がそれぞれ報告値の総数の4分の1になる Q_1 と Q_3 により、 $IQR = Q_3 - Q_1$ で与えられる。正規分布において25%点と75%点との間の大きさが標準偏差の0.7143倍であることから、標準偏差の推定値としてNIQRが導かれる。NIQRは上位及び下位25%の値の効果が直接的でなく、外れ値の影響を受けにくい。

紹介した様々な方法で算出した技能評価の標準偏差を表2に示す。ロバストと呼ばれる手法では、いずれも同じような結果になる一方、標本標準偏差は他のものの2倍以上大きい。これは外れ値の影響である。図4bに図4aと同じデータを算術平均と標本標準偏差を用いてパフォーマンス評価した例を示す。 $|z| > 3$ となる試験は1.5%となり、ロバストな中央値とNIQRを用いた結果の6.3%と比べて緩い判定基準になっている。NIQRによる評価では、50%を超える報告値が一つの正規分布に従うことを前提にしている。標本標準偏差による評価では、すべての値が一つの正規分布に従うことを前提にしている。この違いが大きな差を生むのである。

3.4 均質性・安定性の評価

配付試料の均質性・安定性は、その技能試験でパフォーマンスの評価ができるかという観点から重要である。均質性について、JIS Z 8405では以下の確認手順を与えている。瓶詰の液体が配付試料である場合、 $g \geq 10$ なる g 本の瓶をランダムに選ぶ。1本につきそれぞれ2回、計 $2g$ 回ランダムな順序で濃度を測定する。その結果から、瓶間の標準偏差を求める。それが技能評価の標準偏差の0.3倍よりも小さければ、均質性は問題ないとする。

安定性の試験は、均質性の試験と同じ試験所が測定の実施期間と同等の期間をおいてから行う。均質性試験の $2g$ 回の平均値と、安定性のための測定の平均値の差の絶対値が技能評価の標準偏差の0.3倍よりも小さければ、安定性は問題ないとする。

均質性・安定性の試験においては、実際に0.3倍以下になるかは試験を終えないと分らず、その基準を超え

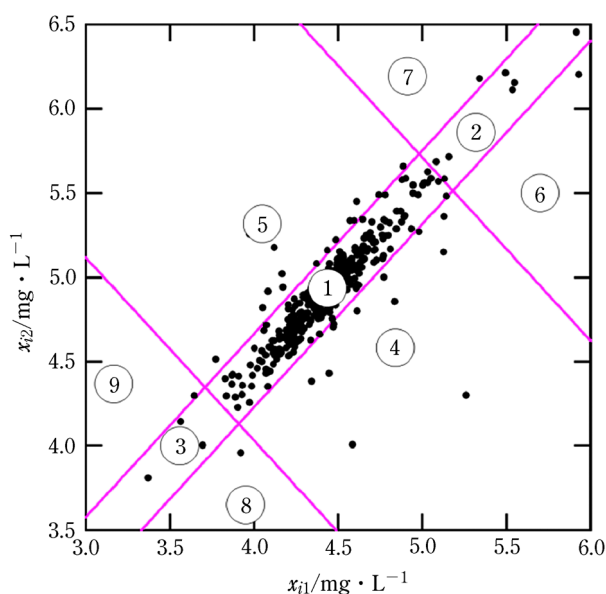
の場合について JIS Q 17043, JIS Z 8405 に記載はあるが、十分な解決策を提示してはいない。新しい技能試験を立ち上げる際はもちろん、継続的に試験を提供していく場合にも、この点に十分な配慮が必要である。

4 技能試験結果の利用

4.1 参加試験所の視点での技能試験の利用

パフォーマンス評価の結果が不満足だった場合、試験手順などを含むマネジメントシステムの見直しが必要になる。筆者ら¹¹⁾は、複数の試験試料が配付される試験において、ユーデンプロットを使用し、是正を効率的に実施する方法を提案した。ユーデンプロットとは、試料 1 と試料 2 があつたとき、試験所 i の二つ報告値 x_{i1} , x_{i2} について (x_{i1}, x_{i2}) (又はそれらの z スコア (z_{i1}, z_{i2})) をプロットしたものである。図 5 に日環協が平成 20 年に提供した技能試験「水中の重金属分析」⁸⁾における鉄の含有量の試験結果のユーデンプロットを示す。ユーデンプロットにより、再現性(試験条件の一部が変化したときの試験結果のばらつきの小ささ)や併行性(試験条件の全部が同一の場合の試験結果のばらつきの小ささ)として説明できる試験所内効果と、異なる試験者や試験手順・試験環境などの影響による試験所間の効果とを分けて検討することができる。

ユーデンプロット中に描かれた実線は、試験所間の z スコアと、試験所内の z スコアとも呼べるものを算出し、描いた評価枠である。日環協ではこれとほぼ同じ評価枠が作成され、参加試験所に提供されている。評価枠



この技能試験では、2種の試料(試料1および試料2)が回付された。グラフの横軸および縦軸は、試験所 i の試料1の報告値 x_{i1} および試料2の報告値 x_{i2} 。試験所内の効果に比べ、試験所間の効果が強いほど、この散布図の直線的な傾向が強くなる。

図5 平成20年度日環協「水中の重金属分析」(日環-42)⁸⁾の鉄の含有量のユーデンプロットと評価枠

の作るそれぞれの区分に該当する値を報告した試験結果について、パフォーマンス評価は表3のとおり与えられる。残念ながら、最も興味のある試験所間効果を、完全に試験所内効果と分離することはできない。このために、試験所間の z スコアの絶対値が3を超える場合でも、再現性・併行性が小さいことを確認するための測定が必要になる。

試験試料が単一であるときにも、不満足なパフォーマンス評価を得た場合には、再測定などにより原因を探ることが必要である。この場合も含め、技能試験の結果を受けて試験手順などを見直すためのフローチャートを図6に示す。重要なことは、技能試験の結果を活かし、試験所の技能の改善につなげることである。

4.2 第三者の視点での技能試験の利用

参加試験所以外の技能試験の結果を利用する者には認定機関と直接の顧客が考えられる。ここでは第三者である認定機関、特に IAJapan と JAB がどのように技能試験の結果を利用しているかを解説する。直接の顧客が技能試験結果を利用する場合にも参考になるだろう。

4.2.1 技能試験参加計画について

認定における技能試験について、国際試験所認定協力機構(以下、ILAC)は技能試験方針文書 ILAC P9¹²⁾を定めている。この方針文書中では、「認定機関は技能試験関連活動の監査及び認定プロセスにおける使用についての方針を持っていないとてはならない」とされている。IAJapan では ILAC P9 に対応して方針文書 URP24¹³⁾を定めている。その中で「技能試験参加計画」について言及がある。

IAJapan では技能試験参加計画の作成を、プログラムにより必須あるいは望ましい要求事項としている。技能試験参加計画のポイントとして、① 能動的であるこ

表3 評価枠による区分と試験所間及び試験所内の効果についてのパフォーマンス評価の対応

図2中の区分	試験所間の z スコア: z_u	試験所内の z スコア: z_v	評価
①	$-3 < z_u < 3$	$-3 < z_v < 3$	試験所間のかたより及び試験所内のばらつきは不満足でない。
②	$z_u \geq 3$	$-3 < z_v < 3$	試験所間のかたより及び/又は試験所内のばらつきが不満足である。
③	$z_u \leq -3$	$-3 < z_v < 3$	
④	$-3 < z_u < 3$	$z_v \geq 3$	試験所内のばらつきが不満足である。
⑤	$-3 < z_u < 3$	$z_v \leq -3$	
⑥	$z_u \geq 3$	$z_v \geq 3$	試験所内のばらつきが不満足である上、試験所間のかたよりも不満足な疑いがある。
⑦	$z_u \geq 3$	$z_v \leq -3$	
⑧	$z_u \leq -3$	$z_v \geq 3$	
⑨	$z_u \leq -3$	$z_v \leq -3$	

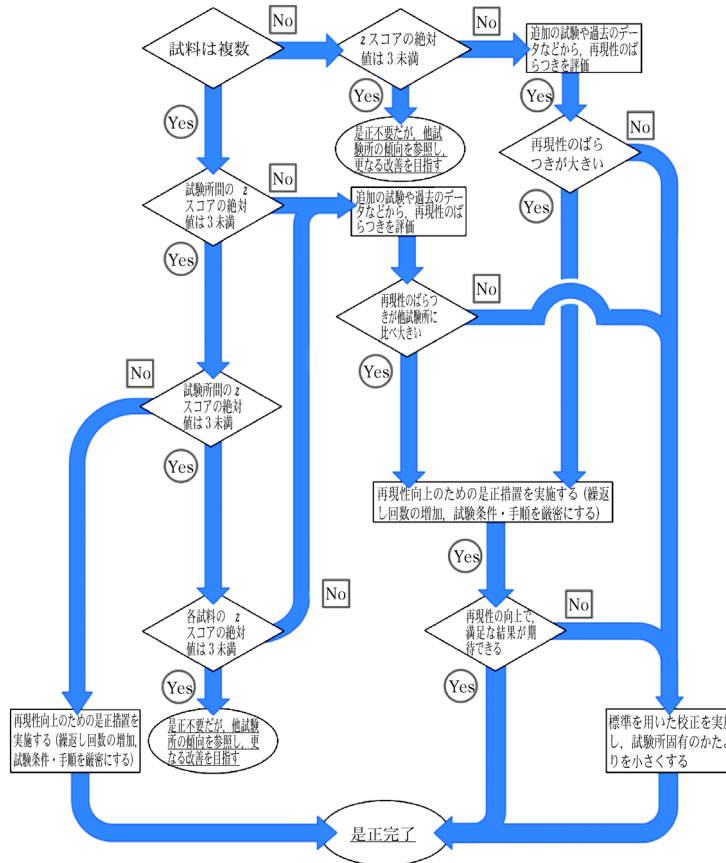


図 6 技能試験結果を用いた是正処置のフローチャートの一例

と、② 計画性があること、③ 分野が特定できること、④ 頻度が定量的であることを挙げている。「能動的である」とは、例えば「IAJapan の要請に応じて技能試験に参加する」という表現では認められないということである。技能試験への参加の頻度に関しては、試験分野の相場観に合わせて定められる。ただし、その間隔が 4 年を超えない継続的な活動として位置づけられていることが原則である。JAB でも様式は問わないが、技能試験参加計画の作成が求められている。

4.2.2 外部機関の技能試験結果の利用について

2.3 節に記したように、IAJapan や JAB で、JIS Q 17043 の認定を取得していない他機関が提供した技能試験の結果を試験所認定・登録の材料とする場合、その技能試験の品質確認は JIS Q 17043 の要求事項のエッセンスへの適合確認によっても実施される。このようにして提供された技能試験結果について、両機関とも認定・登録をしている試験所の評価が疑わしい又は不満足の場合には、試験所から認定機関に直ちに通知するように定めがある。このため、外部機関の技能試験を利用している場合でも、両機関に認定・登録された試験所のパフォーマンス不良の原因が追究されず、次の更新時の審査まで放置されることはなく、直ちに必要な是正処置が行われる。

4.2.3 技能試験の結果と是正処置について

JAB ではおおむね以下の手順でパフォーマンス評価の結果が是正処置に生かされているとのことである。

- ① $|z| \leq 2$ の場合、試験結果の考察の報告を求めない。
- ② $2 < |z| < 3$ の場合、即時に試験結果の考察を求めることは基本的にないが、新規の認定又は更新の審査の際には、その原因についての見解を確認することがあり、場合によっては是正を求める。
- ③ $|z| \geq 3$ の場合には、(i) 試験所に試験結果の原因追求の実施を求め、見解書を提出させる。(ii) JAB がパフォーマンス評価と見解書の記載を合わせ、「不満足項目が認定対象外であるか」、「既に是正が完了しているか」、「材料起因等のため不満足でない」と判断されるか」を基準に合否の判定を実施する。(iii) 原因の追求が不十分で、尚も試験所の技能に疑問が残る場合には是正処置を求める。

E_n 数を用いる場合には、 $|z| \geq 3$ を $|E_n| \geq 1$ と置き換えてよい。また、JAB が提供する試験では、4.1 節で説明したような試験所間と試験所内の z スコアに着目することもある。このような是正処置を実施することにより、JAB で認定・登録を受けた試験所は、パフォーマンスの不良が放置されることがない仕組みになっている。およそその手順は IAJapan でも同じである。

試験所の技能試験結果を試験所の顧客が利用する場合に、第三者の認定のような管理するのは難しい。しか

し、外部精度管理実績を確認することは、品質確認の一つの手段となりうる。この際、①参加の頻度、②直近のパフォーマンスが不満足でないか、③不満足の場合に適切な是正が行われているかの3点について確認するのが基本だろう。

5 技能試験のこれから

1章に技能試験の重要性が増す背景には、試験所認定・登録の仕組みの重要性が増していることが述べた。この動きは国際的なものである。例えば、JIS Q 17025に基づいて、試験所を認定し、その試験所が出した試験の結果は、世界中どこでも有効に使うことができる仕組みが現在すでに運用されている。製品安全の国際規格を整備したとしても、それが機能的に運用されるには、その規格に基づいた試験の品質が確保されていなくてはならない。このときに技能試験の果たす役割は大きい。試験所認定の審査において「お国柄やお人柄が出ているのでは？」との印象に対して、客観性の高いツールとなるからである。食品、環境、製品安全、臨床検査のような分野では既にそうであるが、今後さらに技能試験は存在感を發揮していくだろう。

その際、技能試験結果の利用者（多くの場合は試験所の顧客と考えられる）には、「技能試験への参加を通して、試験所のマネジメントシステムの点検が行われている」という認識を、より強める必要がある。例えば、ある測定の地方公共団体による入札公告で、参加資格に「技能試験の全試験項目の結果が（中略）いずれも「満足」の評価であること」としていたものがあった。筆者（津越）が担当者と面会する機会がたまたまあり、その際にこの条件の目的を聞いたところ、やはり「技能試験で満足の結果を得た試験所が高い技術的能力を持つと考えられる」という回答であった。しかし、技能試験をえて熟練技術者に担当させ、通常業務より測定回数を大幅に増やした測定値を報告するようなことが仮にあったとしたら、依頼業務の精度が達成されているかはわからないということも考えられる。本稿で触れている技能試験の趣旨を説明したところ、参加条件は変更され、精度管理として技能試験に参加することのみを求めることとなった。もちろん、技能試験に参加する（check）のみではなく、改善（act）されなければならないことは言うまでもない。

重要性が増すほどに、技能試験提供者のみならず技能試験利用者にも上のような言わば計量計測リテラシーが望まれる。もちろん、技能試験提供者には透明で質の高い技能試験の運用が求められるし、参加試験所は技能試験の結果で一喜一憂せず、継続的な参加と改善を心がけねばならない。

謝辞 本稿の執筆にあたり、JAB 保坂守男様、小島勇夫様、NITE 石毛浩美様、大高広明様、食薬センター 渡辺卓穂様、鈴木達也様、日環協 福田俊一様に技能試験の運営、技能試験の認定における利用などについて、ご意見いただいた。日環協には既に公開されている技能試験データの提供をいただいた。ここに感謝の意を表したい。

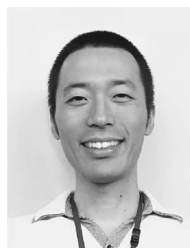
文 献

- 1) 食品薬品安全センター、食品薬品安全センター 外部精度管理、<http://www.fdsc.or.jp/hatanoken/gaibuseidokanri.html> (2013年9月18日最終確認)。
- 2) 渡辺卓穂、勝村利恵子、高坂典子、福光 徹、鈴木達也、大島赴夫：食品衛生研究，**60**, 17 (2010)。
- 3) 日本規格協会，JIS Q 17043，適合性評価—技能試験に対する一般要求事項 (2011)。(ISO/IEC 17043の同一規格)
- 4) 日本規格協会，JIS Q 17025，試験所及び校正機関の能力に関する一般要求事項 (2005)。(ISO/IEC 17025の同一規格)
- 5) 日本規格協会，JIS Z 8405，試験所間比較による技能試験のための統計的方法 (2008)。(ISO 13528の同一規格)
- 6) 日本分析化学会，(公社)日本分析化学会が行っている技能試験，<http://www.jsac.or.jp/ginousiken/st06.html> (2013年9月18日最終確認)。
- 7) 製品評価技術基盤機構 認定センター，認定センターで実施する技能試験，<http://www.iajapan.nite.go.jp/iajapan/gaiyou/ginoushiken.html> (2013年9月18日最終確認)。
- 8) 日本環境測定分析協会，ISO/IEC Guide 43-1に基づく技能試験報告書—日環 42-08WM (日環-42) (2008)。
- 9) Analytical Methods Committee: *Analyst*, **114**, 1693 (1989)。
- 10) M. Thompson, S. L. R. Ellison, R. Wood: *Pure Appl. Chem.*, **78**, 145 (2006)。
- 11) K. Shirono, K. Iwase, H. Okazaki, M. Yamazawa, K. Shikakume, N. Fukumoto, M. Murakami, M. Yanagisawa, T. Tsugoshi: *Accredit. Qual. Assur.*, **18**, 161 (2013)。
- 12) ILAC, ILAC-P9: 11/2010, Policy for Participation in Proficiency Testing Activities (2010)。
- 13) 独立行政法人 製品評価技術基盤機構 認定センター，URP24 IAJapan 技能試験に関する方針 (2011)。

城野克広 (Katsuhiko SHIRONO)

(株)産業技術総合研究所計測標準研究部門計量標準システム科計量標準基盤研究室 (〒305-8565 茨城県つくば市東1-1-1 中央第5)。東京大学大学院博士課程修了。博士 (環境学)。《現在の研究テーマ》測定における統計的方法の活用。《趣味》読書、英会話。

E-mail: k.shirono@aist.go.jp



津越敬寿 (Takahisa TSUGOSHI)

(株)産業技術総合研究所計測標準研究部門 (〒305-8563 茨城県つくば市梅園1-1-1 中央第3)。東京理科大学大学院理学研究科博士後期課程化学専攻修了。理学博士。《現在の研究テーマ》発生気体分析の高度化とその応用、技能試験運営のマネジメントシステム。《趣味》飲み歩き、子供に遊んでもらうこと。

