UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

FILIPE DE MEDEIROS ALBANO

DESENVOLVIMENTO DE MELHORIAS NO PROCESSO DE PROVISÃO DE ENSAIOS DE PROFICIÊNCIA POR COMPARAÇÃO INTERLABORATORIAL

TESE DE DOUTORADO

Porto Alegre

2016

FILIPE DE MEDEIROS ALBANO

DESENVOLVIMENTO DE MELHORIAS NO PROCESSO DE PROVISÃO DE ENSAIOS DE PROFICIÊNCIA POR COMPARAÇÃO INTERLABORATORIAL

Tese submetida ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul, como requisito parcial à obtenção do título de Doutor em Engenharia na área de concentração em Sistemas de Qualidade.

Orientadora: Prof[a]. Carla Schwengber ten Caten, Dr[a].

Porto Alegre

2016

FILIPE DE MEDEIROS ALBANO

DESENVOLVIMENTO DE MELHORIAS NO PROCESSO DE PROVISÃO DE ENSAIOS DE PROFICIÊNCIA POR COMPARAÇÃO INTERLABORATORIAL

Esta tese foi julgada adequada para a obtenção do título de Doutor em Engenharia e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul.

_____

Prof[a]. Carla Schwengber ten Caten, Dr[a].

Orientador

_____

Prof. José Duarte Ribeiro, Dr.

Coordenador PPGEP-UFRGS

Banca Examinadora:

Prof. Michel Jose Anzanello, Dr. (Departamento de Engenharia de Produção e Transportes da Universidade Federal do Rio Grande do Sul – DEPROT/UFRGS)

Prof[a]. Morgana Pizzolato, Dr[a]. (Faculdade de Engenharia da Universidade Federal de Santa Maria - UFSM)

Prof. Katia Fernanda Streit, Dr[a]. (INMETRO - Superintendência Regional do Rio Grande do Sul – INMETRO/SURRS)

Dedico este trabalho para minha esposa e colega de doutorado, que encarou este desafio ao meu lado, apoiando, compreendendo e auxiliando em todos os momentos necessários. Este caminho não seria o mesmo sem este apoio e amor.

Agradeço imensamente aos meus pais, que sempre valorizaram a educação e me mostraram que o investimento em coisas não materiais, como o conhecimento, nos faz evoluir.

Um agradecimento especial para universidade pelo ensino de alta qualidade; para minha orientadora pela confiança; para a banca, pelas valiosas contribuições e aos meus colegas de trabalho da Rede Metrológica RS pelo apoio nesta jornada!

LISTA DE SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
APLAC	*Asia Pacific Laboratory Accreditation Cooperation*
CGCRE	Coordenação Geral de Acreditação Do Inmetro
CNAS	*China National Accreditation Service for Conformity Assessment*
COFRAC	*Association Chargée De L'accréditation des Laboratoires*
EA	*European Accreditation*
EP	Ensaio de Proficiência
IAAC	*Inter American Accreditation Cooperation*
IEC	*International Electrotechnical Commission*
ILAC	*International Laboratory* Accreditation *Cooperation*
INMETRO	Instituto Nacional de Metrologia, Qualidade e Tecnologia
ISO	*International Organization for Standardization*
PT	*Proficiency Testing*
RMRS	Rede Metrológica do Rio Grande do Sul
UKAS	*United Kingdom Accreditation Service*

RESUMO

Esta tese apresenta uma contribuição para a melhoria do processo de desenvolvimento e execução de Ensaios de Proficiência (EP) por comparação interlaboratorial. Este tema está diretamente relacionado ao desenvolvimento de qualidade nos processos analíticos de laboratórios, de forma a ampliar a confiabilidade de resultados emitidos por estas entidades. Desta forma, os objetivos desta tese são: (i) Identificar e compreender os principais aspectos conceituais relacionados com o desenvolvimento de EP, levantando as principais lacunas teóricas acerca deste tema; (ii) Analisar a relação existente entre a realização de EP, validação de métodos e a estimativa da incerteza de medição de ensaios e calibrações realizadas em laboratórios; (iii) Investigar a relação entre o desenvolvimento de EP e conhecimentos utilizados na gestão dos projetos das comparações interlaboratoriais; (iv) Propor e aplicar uma sistemática para identificar parâmetros representativos para execução de testes de homogeneidade conduzidos em EP; (v) Analisar a influência do tipo de distribuição de probabilidade na avaliação de desempenho de laboratórios quando se trabalha com valor de consenso baseado nos dados dos participantes das rodadas de EP. Como resultados, foi possível: a identificação da relação de causa e efeito entre EP, validação de métodos e estimativa da incerteza de medição; a identificação de áreas de melhoria na gestão de projetos de EP, destacando-se a gestão de riscos e custos; o desenvolvimento de uma sistemática para selecionar variáveis em testes de homogeneidade e estabilidade de EP, sendo apoiada pela técnica de análise de componentes principais; e, por fim, a identificação do impacto do coeficiente de curtose e tipo de distribuição de probabilidade em EP que utilizam avaliação de desempenho através de valor de consenso entre os laboratórios participantes da comparação, bem como comprovação da equivalência de dois métodos robustos de análise de desempenho em EP.

Palavras-chave: Ensaios de Proficiência, Comparação Interlaboratorial, Laboratórios, Metrologia.

ABSTRACT

This thesis presents a contribution to improve the development and implementation of Proficiency Testing (PT) for laboratory comparison. This is directly related to quality development in laboratories, in order to increase the reliability of results issued by these entities. The objectives of this thesis are: (i) Identify and understand the main conceptual aspects related to the development of PT, finding the main theoretical gaps on this subject; (ii) analyze the relationship between PT, method validation and estimation of the uncertainty of measurement performed in laboratories; (iii) To investigate the relationship between the development of PT and knowledge used in the project management of interlaboratory comparisons; (iv) To propose and implement a method for identifying representative variables for performing homogeneity and stability tests conducted in PT; (v) analyze the influence of type of probability distribution in laboratories performance evaluation when consensus value based on data from participants are used. As results, it was possible: to identify the cause and effect between PT, methods validation and estimation of uncertainty measurement; to identify areas for improvement in the PT projects, mainly costs and risk management; to develop a method for selecting variables in homogeneity and stability tests using principal component analyses technique and; to identify the impact of kurtosis coefficient and type of the probability distribution in PT that use performance evaluation by consensus value between the participating laboratories, as well as verification of the equivalence of two robust methods used to assess laboratories performance in PT.

Keywords: *Proficiency Testing, Interlaboratory Comparison, Laboratories, Metrology.*

SUMÁRIO

CAPÍTULO 1

1.1 INTRODUÇÃO

As exigências das indústrias, do governo e do setor de serviços em relação à qualidade de ensaios e calibração têm ampliado ao longo dos últimos anos. A partir desta realidade, os laboratórios de diferentes áreas da metrologia passaram a se adequar a critérios de qualidade mais rigorosos, com embasamento em normas internacionais.

Atualmente, a Coordenadoria Geral de Acreditação (CGCRE) do Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO) é o órgão responsável pela acreditação de laboratórios no país, tendo um acordo de reconhecimento com o *International Laboratory Acreditation Cooperation* (ILAC). A acreditação é concedida através de uma avaliação da conformidade de terceira parte, por avaliadores qualificados (ISO, 2005a). Para a avaliação da conformidade de laboratórios, a norma de referência utilizada é a ABNT NBR ISO/IEC 17025:2005, que trata sobre sistemas de gestão da qualidade na área de ensaios e calibração (ISO, 2005b).

As Redes Metrológicas estaduais também são órgãos que realizam avaliação de laboratórios, porém esta atividade é chamada de Reconhecimento de Competência Técnica, sendo a CGCRE o organismo de acreditação oficial do país. A diferença entre estes processos é que a acreditação possui caráter internacional, devido ao seu reconhecimento pelo ILAC (PIZZOLATO, 2006). De acordo com CGCRE (2011) e RMRS (2014), laboratórios que buscam acreditação ou reconhecimento devem participar de Ensaios de Proficiência (EP).

Em países como França, Inglaterra, Estados Unidos, China, Canadá, entre outros, a evolução da exigência na área de laboratórios é uma realidade. De acordo com NIST (2005), organismos de acreditação como *United Kingdom Accreditation Service* (UKAS), A*merican Association for Laboratory Accreditation* (A2LA), *Comité français d'accréditation* (COFRAC), *Standards Council of Canada* (SCC), *China National Accreditation Service for Conformity Assessment* (CNAS), possuem processos estruturados para fornecer confiança nas avaliações de conformidade realizadas em laboratórios. Estas instituições estão em um processo de crescimento, devido à demanda de solicitações de avaliação da conformidade e às exigências de mercado (do governo e do setor produtivo), embasadas na garantia da qualidade

de ensaios e calibrações. Os organismos internacionais exigem participação compulsória em EP para laboratórios que almejam a acreditação.

Os EP são ferramentas importantes e que fornecem subsídio para confiabilidade de ensaios e calibrações, sendo programas de comparação de resultados entre um grupo de laboratórios, objetivando avaliar a competência técnica para desempenhar um método de ensaio ou calibração (ILAC, 2011). Depois de participar de um EP o laboratório tem uma espécie de diagnóstico a respeito de suas medições (ensaio ou calibração) e verifica se é proficiente nos resultados fornecidos aos seus clientes (WONG, 2011).

Os EP são realizados através de uma sistemática que tem como objetivo apoiar os laboratórios de ensaios e calibração na garantia dos serviços prestados, fornecendo apoio ao Sistema de Gestão da Qualidade (SGQ) da empresa (HOWERTON *et al*., 2010). Por meio das comparações interlaboratoriais é possível avaliar o desempenho de laboratórios para ensaios ou medições específicas, identificar problemas analíticos, estabelecer uma comparabilidade de métodos de ensaio ou calibração, prover confiança adicional aos clientes do laboratório, capacitar os participantes com base em resultados das comparações interlaboratoriais, validar a incerteza declarada e atribuir valores para materiais de referência (ABNT, 2011).

As empresas que promovem rodadas de comparação entre laboratórios são chamadas de provedores de EP. Recomenda-se que estas organizações sigam a norma ABNT NBR ISO/IEC 17043, que foi elaborada para fornecer uma base consistente a todas as partes interessadas para determinar a competência de organizações provedoras de EP. Atualmente a CGCRE realiza, também, a acreditação deste tipo de organização (CGCRE, 2011).

Em cada EP deve ser definido o desenvolvimento e operação do programa, definindo os possíveis participantes, amostras, frequência do programa, entre outras informações (RMRS, 2011). Segundo ABNT (2011), os programas de comparação podem variar de acordo com as necessidades do setor em que eles são utilizados, a natureza dos itens de EP, os métodos em uso e o número de participantes. A natureza do ensaio ou medição efetuada no EP define o método de comparação de desempenho, que pode ser quantitativo, qualitativo ou interpretativo.

Um dos fatores críticos na provisão de EP é a garantia de que as amostras distribuídas são homogêneas e estáveis (THOMPSON, 2004). Já para equipamentos, utilizados em

programas de calibração, é essencial garantir que os mesmos possuam uma boa repetitividade e sejam estáveis, de forma que não apresentem deriva significativa dos seus valores (GUST, 2007). Outro fator relevante é o tratamento estatístico que o provedor do EP aplica nos dados dos laboratórios participantes, pois o método de análise de dados utilizado pode influenciar diretamente o resultado obtido na rodada de comparação (BREZIS, 2011).

O presente trabalho contempla o tema EP por comparações interlaboratoriais, de forma que a pesquisa foi conduzida com objetivo de contribuir cientificamente com o desenvolvimento desta atividade. Na próxima seção estão apresentados os objetivos desta tese.

1.2 **OBJETIVOS**

A tese está focada em um objetivo geral de pesquisa, o qual orientou o desenvolvimento do trabalho realizado. O mesmo está derivado em objetivos específicos.

1.2.1 **Objetivo Geral**

O objetivo geral da tese foi analisar e propor melhoria nos processos de provisão de EP por comparação interlaboratorial.

1.2.2 **Objetivos Específicos**

a) Identificar e compreender os principais aspectos conceituais relacionados com o desenvolvimento de EP, levantando as principais lacunas teóricas acerca deste tema;

b) Analisar a relação existente entre a realização de EP, validação de métodos e a estimativa da incerteza de medição de ensaios e calibrações realizadas em laboratórios;

c) Investigar a relação entre o desenvolvimento de EP e conhecimentos utilizados na gestão dos projetos das comparações interlaboratoriais;

d) Propor e aplicar uma sistemática para identificar parâmetros representativos para execução de testes de homogeneidade e estabilidade conduzidos em EP;

e) Analisar a influência do tipo de distribuição de probabilidade na avaliação de desempenho de laboratórios quando se trabalha com valor de consenso baseado nos dados dos participantes das rodadas de EP.

1.3 **JUSTIFICATIVA**

Atualmente, resultados ensaios ou calibrações executadas por laboratórios servem como auxílio para tomada de decisões nas indústrias, áreas da saúde, meio ambiente, entre outras. Por meio destes resultados aceitam-se ou rejeitam-se matérias-primas, diferencia-se o desempenho de fornecedores, processos produtivos são modificados, atua-se sobre a saúde das pessoas e dos animais (MOURA, 2014). Desta maneira, percebe-se a importância da confiabilidade dos resultados analíticos.

Com base nesta lógica, o mercado europeu, começou a exigir qualificação dos laboratórios que realizam estes testes, como uma garantia de o produto estar conforme requisitos especificados (THOLEN, 2011). Além da questão de exportação, diversos órgãos do governo, tais como Ministério da Agricultura Pecuária e Abastecimento, Órgãos Ambientais, Agências Reguladoras, entre outros, passaram a exigir em Instruções Normativas e Regulamentos que laboratórios tivessem acreditação e/ou reconhecimento de competência técnica (BRASIL, 2013; BRASIL, 2012). Esta ação também direciona os laboratórios para um processo de qualificação compulsório, sendo que para obter uma acreditação ou reconhecimento, o laboratório deve participar de EP.

A participação em EP é um pré-requisito para solicitação de uma avaliação de laboratório com base na norma ABNT NBR ISO/IEC 17025 pela CGCRE ou pela Rede Metrológica RS (CGCRE, 2011; RMRS, 2014). Com este tipo de exigência de órgãos que avaliam laboratórios, percebe-se que os EP são necessários na rotina de entidades que buscam sua qualificação e um reconhecimento de terceira parte (THOLEN, 2011).

Na avaliação de um laboratório, uma das principais dificuldades encontradas está ligada ao processo de validação de métodos e garantia da qualidade (ROSA *et al*., 2011). A garantia da qualidade pode ser executada de forma interna ou externa. Os EP fazem parte dos

controles externos de qualidade (KISETS, 2006). A validação de métodos também pode ser apoiada por EP, bem como a estimativa da incerteza de medição (BRIDWELL *et al.* , 2010). De acordo com Tholen (2011), outro aspecto positivo no processo de avaliação de laboratórios é o ingresso na participação em controles externos de qualidade, que são fundamentais para confiabilidade de resultados.

Autores como Pizzolato, Caten e Jornada (2008), destacam que um resultado satisfatório em EP (na área de calibração e de ensaios) é tão relevante quanto à acreditação ou reconhecimento de um laboratório, pois atesta a sua competência técnica. Para Tawfik e Fatah (2010), a conquista de uma acreditação é mais simples do que sua manutenção ao longo do tempo. Entretanto, em termos práticos percebe-se que laboratórios tem apresentado maior dificuldade nos últimos anos na obtenção da acreditação. Um dos fatores é a demanda por participação de EP antes de solicitar formalmente o processo de avaliação da conformidade.

Diferentes pesquisadores destacam a importância da participação em EP. De acordo com estudos realizados por Howerton *et al*. (2010), foram analisados resultados de EP realizados nos Estados Unidos em laboratórios clínicos de 1994 até 2006 e percebeu-se uma diferença significativa na melhoria dos laboratórios participantes, onde observou-se a redução de resultados insatisfatórios. Neste estudo foram avaliados 36.000 resultados de participantes de EP na área de análises clínicas. Estes EP são aprovados pela *Clinical Laboratory Improvement Amendments* (CLIA), que é uma agência regulatória da área de saúde Norte Americana.

Cabe destacar que, segundo Moura (2014), existe uma demanda crescente por EP no Brasil. Atualmente esta demanda não é suprida, pois existem poucos provedores capacitados para atender todas as áreas que existem nos laboratórios acreditados.

Como justificativa da importância do tema, também destaca-se que EP são pouco explorados em termos de publicações científicas, por ter como característica o envolvimento de diversas área de conhecimento. Ainda, especificamente no Brasil, os EP estão ganhando importância por estarem sendo demandados com mais frequência nos últimos anos.

Os argumentos citados anteriormente constituem a justificativa para o desenvolvimento desta pesquisa. A seção a seguir apresenta o delineamento da elaboração deste trabalho.

1.4 **DELINEAMENTO DO ESTUDO**

Uma vez definidos os objetivos da pesquisa, foi estabelecido o delineamento pelo qual os objetivos foram alcançados, bem como o método de trabalho e de pesquisa adotados.

1.4.1 **Descrição dos estudos**

O método de trabalho da tese está apresentado na Figura 1, a qual descreve o mapa conceitual sobre os estudos desenvolvidos. O mapa apresenta a lógica do desenvolvimento dos cinco artigos que estruturaram o desenvolvimento da pesquisa.

O Artigo 1 apresenta uma Revisão Sistemática acerca do tema EP, descrevendo os principais conceitos e práticas identificados na literatura consultada. Este artigo fornece o referencial teórico da pesquisa, na qual constam 151 referências, sendo oriundas de artigos internacionais, normas e publicações de institutos de metrologia ou organismos de avaliação da conformidade. Este trabalho proporcionou o aprofundamento teórico do tema e identificação de lacunas teóricas para desenvolvimento da pesquisa e dos artigos subsequentes.

O Artigo 2 descreve os resultados de uma pesquisa qualitativa com especialistas, o qual explora a relação entre EP, validação de métodos e incerteza de medição. Os especialistas selecionados para o desenvolvimento do trabalho foram avaliadores de laboratórios, profissionais que se envolvem com o processo de aprovação de laboratórios no processo de reconhecimento da Rede Metrológica RS e de acreditação na CGCRE e especialistas internacionais de outros organismos de acreditação. A pesquisa utilizou o método *Web Delphi*. Os resultados foram compilados em um diagrama de relações, o qual permitiu identificar as relações de causa e efeito entre os fatores analisados, proporcionando um maior entendimento entre os itens estudados.

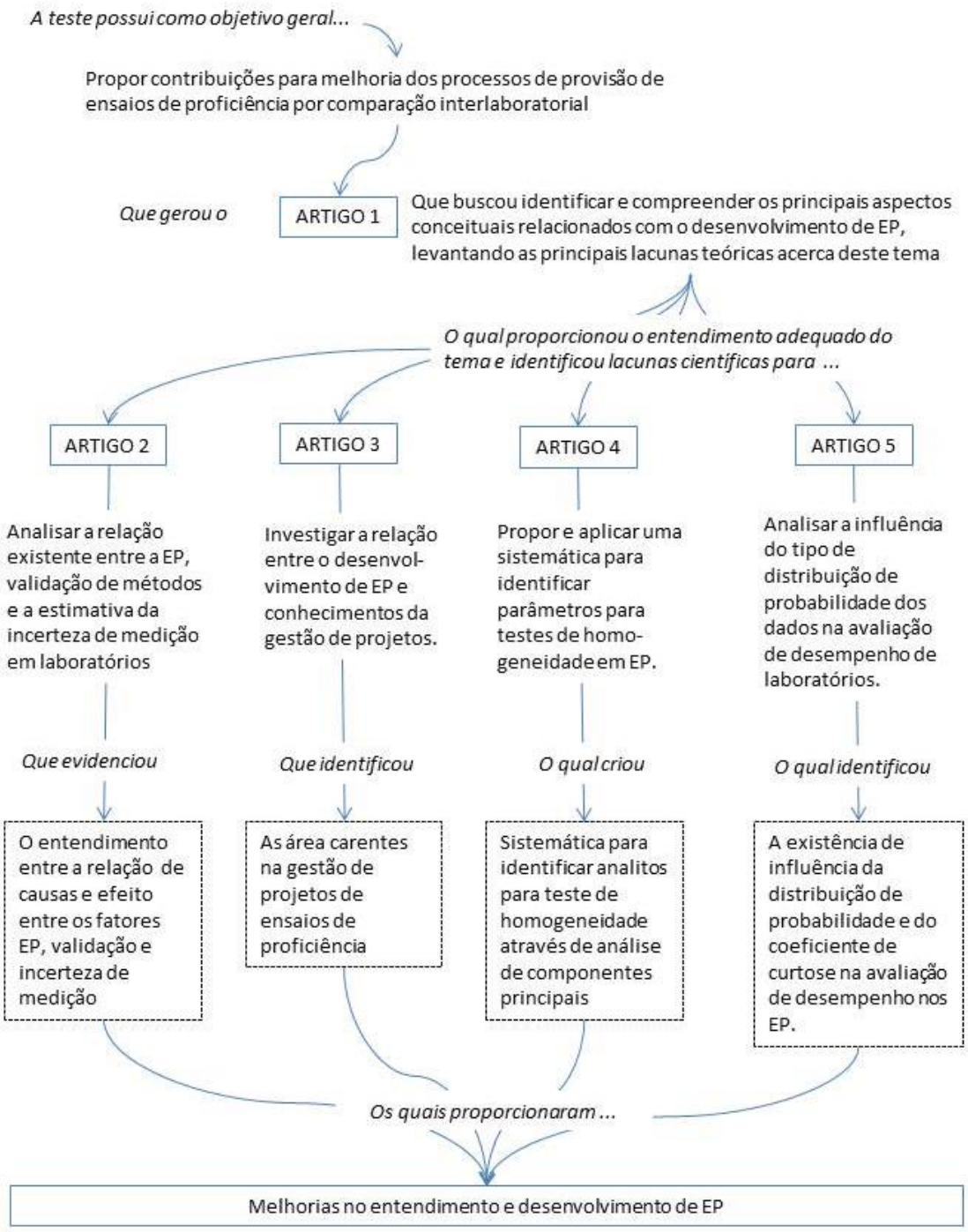

Figura 1 – Mapa conceitual da Tese

O Artigo 3 apresenta uma pesquisa realizada com os provedores de EP brasileiros acreditados pela CGCRE. A pesquisa foi realizada através da aplicação de questionários para identificar os conceitos da área de gestão de projetos utilizados no desenvolvimento de EP. Os

resultados indicaram áreas carentes em relação à gestão dos provedores de EP e possibilidades de melhorias.

O Artigo 4 apresenta resultados da aplicação de uma sistemática para seleção de parâmetros para realização de testes de homogeneidade e estabilidade em EP. O estudo experimental foi conduzido com dados do EP de bebidas, carvão e meio ambiente da Rede Metrológica RS. Neste estudo foi criado um indicador para seleção das variáveis que seriam utilizadas na realização dos testes de homogeneidade e estabilidade. O indicador combina resultados da Análise de Componentes Principais, sendo validados com a opinião de especialistas na área dos EP. Os resultados indicam que a proposta da sistemática é adequada para este fim e que foi capaz de selecionar variáveis representativas para realização do teste de homogeneidade e estabilidade.

O Artigo 5 descreve um estudo experimental para avaliação do tipo de método estatístico utilizado na avaliação de desempenho de laboratórios e do coeficiente de curtose e tipo de distribuição de probabilidade dos dados em rodadas de EP. O estudo foi conduzido com dados do programa de análises ambientais da Rede Metrológica RS, considerando 27 analitos diferentes. Ainda, para complementar o banco de dados, foi utilizada a simulação de Monte Carlo. O estudo constatou influência do tipo de método de avaliação de desempenho utilizado e o coeficiente de curtose dos dados analisados. Ambos os fatores apresentaram diferenças significativas no % de resultados satisfatórios obtidos pelos laboratórios nos programas de EP. Ainda, percebeu-se forte correlação entre as distribuições bimodais e o coeficiente de curtose negativo, demonstrando ser este um fator que pode gerar impacto significativo no tratamento de dados por provedores de EP (quando se trabalha com valor de consenso com base nos dados dos participantes).

O método de trabalho da pesquisa contribuiu para identificação de melhorias no desenvolvimento de EP. As mesmas são contribuições de cunho gerencial e também técnico.

1.4.2 **Método de Pesquisa**

A Figura 2 apresenta um resumo sobre o método de pesquisa da tese de doutorado. A lógica de estruturação da pesquisa está dividida por artigo desenvolvido.

Artigo	Questão de Pesquisa	Natureza da Pesquisa	Abordagem da Pesquisa	Objetivos da Pesquisa	Procedimentos Técnicos da Pesquisa
1	Quais são os principais conceitos relacionados com o desenvolvimento de EP? (i);	Básica	Qualitativa	Exploratória	Pesquisa Bibliográfica (revisão sistemática)
2	Qual a relação existente entre o desenvolvimento e provisão de EP, validação de métodos e estimativa da incerteza de medição?	Básica	Qualitativa	Descritiva	Estudo de Caso
3	Como são desenvolvidos os projetos de EP em provedores brasileiros?	Aplicada	Quantitativa	Exploratória	Estudo de Caso
4	Quais critérios o provedor deve ter ao selecionar analitos para os testes de homogeneidade e estabilidade das amostras?	Aplicada	Quantitativa	Explicativa	Experimental
5	Qual influência a distribuição de probabilidade dos dados pode ter na avaliação de desempenho de laboratórios em EP quando se utiliza valor de consenso?	Aplicada	Quantitativa	Explicativa	Experimental

Figura 2 – Classificação da pesquisa desenvolvida

O Artigo 1 apresentou uma pesquisa Básica, que, segundo Gil (1991), objetiva gerar conhecimentos novos e relevantes para o avanço científico, sem necessariamente possuir uma aplicação prática prevista. A abordagem da pesquisa foi Qualitativa, pois não demandou o uso de métodos e técnicas estatísticas e o pesquisador analisou os dados indutivamente (GIL, 1991). Em relação ao objetivo, foi classificada como Exploratória, uma vez que envolveu levantamento bibliográfico e análise de casos que estimulem a compreensão do tema (LAKATOS; MARCONI, 1993). Os procedimentos foram enquadrados como pesquisa Bibliográfica, pois foi construída a partir de material já publicado, constituído principalmente artigos de periódicos e material teórico disponibilizado em sites de organismos relacionados com o tema do trabalho (GIL, 1991).

O Artigo 2 apresentou uma pesquisa Básica e Qualitativa. Em relação aos seus objetivos, foi classificada como Descritiva, já que visa descrever as características de determinados estratos da população e estabelecer relações entre as variáveis pesquisadas (LAKATOS; MARCONI, 1993). Como procedimento técnico foi adotado o Estudo de Caso,

pois envolveu a análise profunda de poucos objetos, de maneira que se permita o seu amplo e detalhado conhecimento (LAKATOS; MARCONI, 1993).

O Artigo 3 trata-se de uma pesquisa Aplicada, pois objetiva gerar conhecimentos para aplicação prática, sendo dirigidos à solução de problemas específicos (GIL, 1991). A abordagem da pesquisa foi Quantitativa, a qual considera que tudo pode ser quantificável, o que significa traduzir em números opiniões e informações para classificá-las e analisá-las, demandando uso de recursos e de técnicas estatísticas (LAKATOS; MARCONI, 1993). Em relação ao seu objetivo, a pesquisa foi identificada como Exploratória, utilizando como procedimento técnico o Estudo de Caso.

O Artigo 4 apresentou uma pesquisa Aplicada e Quantitativa. Como objetivo de pesquisa, classificou-se a pesquisa como Explicativa, que visa identificar os fatores que determinam ou contribuem para a ocorrência dos fenômenos (GIL, 1991). O procedimento técnico utilizado foi a pesquisa Experimental, onde se determina um objeto de estudo, selecionam-se as variáveis que seriam capazes de influenciá-lo, definem-se as formas de controle e de observação dos efeitos que a variável produz no objeto (GIL, 1991). O último artigo obteve a mesma classificação em relação ao método de pesquisa que o Artigo 4.

1.5 **DELIMITAÇÕES DO TRABALHO**

A pesquisa foi desenvolvida e estruturada com estudos que devem ser delimitados devido aos fatores de influência dos resultados. O primeiro artigo, intitulado "*Proficiency tests for laboratories: a systematic review (*Ensaios de Proficiência para Laboratórios: uma revisão sistemática*)*" foi desenvolvido considerando publicações no período de 2005 a 2012. Ainda, este trabalho não considerou publicações em *grey literature* (teses e dissertações). Também não foram considerados relatórios de EP publicados por provedores nacionais e internacionais.

O segundo artigo, que possui como título "*Analysis of the relationships between Proficiency Testing, Validation of Methods and Estimation of Measurement Uncertainty: a qualitative study with experts (*Análise da contribuição dos ensaios de proficiência para confiabilidade de resultados emitidos por laboratórios: um estudo qualitativo com especialistas)" abordou uma pesquisa com profissionais, onde participaram colaboradores da

Rede Metrológica e INMETRO. No total 8 especialistas brasileiros e 4 especialistas internacionais de diferentes áreas contribuíram com o trabalho, onde foi desenvolvido um diagrama de relações para explicar o objeto estudado. O diagrama pode ser limitado à percepção dos participantes nas entrevistas realizadas. O Artigo 3, intitulado "*Management analysis of proficiency testing projects developed by Brazilian accredited providers (*Análise da gestão dos projetos de ensaios de proficiência para laboratórios de provedores acreditados no Brasil*)*" possui como limitação o instrumento de pesquisa utilizado, o qual foi desenvolvido para o propósito da pesquisa, com questões específicas. Ainda, os conceitos sobre gestão de projetos trabalhados no artigo relacionam-se com as publicações do PmBOK e da norma ISO10.006, não abordando demais referências relacionadas com este tema.

O Artigo 4, que possui como título "*Principal component analysis for selection of variables in homogeneity and stability tests applied to proficiency testing (*Análise de componentes principais para seleção de variáveis em testes de homogeneidade e estabilidade aplicados em comparações interlaboratoriais)", teve como estudo de caso três programas de EP da Rede Metrológica RS (na área de bebidas, meio ambiente e carvão). Os dados oriundos dos EP apresentavam *missing values*, os quais foram substituídos por um algoritmo matemático. Na prática é complexo obter-se bancos de dados de EP com todos os valores preenchidos, visto que nem sempre todos os laboratórios participantes da intercomparação realizam todos os ensaios propostos. O método desenvolvido também contou com a participação de especialistas para validar o resultado do indicador proposto para seleção de variáveis no teste de homogeneidade e estabilidade.

O último artigo, intitulado "*The influence of the probability distribution, kurtosis coefficient and performance assessment method on statistical analysis of data from proficiency testing using consensus value* (Análise do impacto da distribuição de probabilidade, coeficiente de curtose e método de avaliação de desempenho em comparações interlaboratoriais que utilizam valores de consenso)" possui como limitação o banco de dados utilizado, onde parte dos dados foram obtidos por simulação através do método de Monte Carlo e não são oriundos de comparações reais. Os dados reais utilizados também se referem a um programa de análises ambientais, não abrangendo outras áreas da metrologia.

1.6 ESTRUTURA DA TESE DE DOUTORADO

Esta tese está estruturada em três capítulos. No primeiro, é apresentada a lógica geral da pesquisa proposta, seu contexto e sua importância como contribuições para lacunas teóricas e práticas sobre o tema. Além disso, são apresentados os objetivos, a justificativa, o delineamento do estudo, a descrição dos artigos propostos e as delimitações da pesquisa.

O segundo capítulo apresenta os artigos propostos e resultados obtidos, sendo que o formato dos mesmos está alinhado com os periódicos definidos como foco para a pesquisa da Tese. Os artigos foram desenvolvidos em inglês. O último capítulo apresenta uma síntese dos resultados da pesquisa, as conclusões e sugestões para trabalhos futuros.

CAPÍTULO 2 – ARTIGOS

ARTIGO 1 - Proficiency tests for laboratories: a systematic review.

ARTIGO 2 – Analysis of the relationships between Proficiency Testing, Validation of Methods and Estimation of Measurement Uncertainty: a qualitative study with experts

ARTIGO 3 - Management analysis of proficiency testing projects developed by Brazilian accredited providers.

ARTIGO 4 - Principal component analysis for selection of variables in homogeneity and stability tests applied to proficiency testing

ARTIGO 5 - The influence of the probability distribution, kurtosis coefficient and performance assessment method on statistical analysis of data from proficiency testing using consensus value

ARTIGO 1 - Proficiency tests for laboratories: a systematic review

Artigo publicado no Accreditation and Quality Assurance Journal for Quality, Comparability and Reliability in Chemical Measurement. ISSN 0949-1775. Volume 19. Number 4. Accred Qual Assur (2014) 19:245-257. DOI 10.1007/s00769-014-1061-8

Qualis Cappes: B1

Proficiency Tests for laboratories: a systematic review

Abstract: Laboratories analytical results must be reliable. Proficiency Tests (PT) main objective is to provide independent demonstrations of laboratory competence. The demand for these activities is increasing on the world scenario, as well as its importance. The main objective of this study is to identify the important publications of the PT theme from 2005 to 2012 based on a systematic review procedure. The method proposed reached a total of 113 papers published in indexed journals and over 34 additional references, including standards, guidelines and recommendations of international or regional accreditation body cooperation, international standards organization and international metrological institutes. All selected references were clustered based on its approach and then the main practices were presented. The approaches identified were related to performance assessment, calculation method for performance assessment, use of PT for validation and/or estimation of measurement uncertainty and management of PT. Results indicate some opportunities to develop researches, such as: project management related to PT, the importance of analyses of data probability distribution function when consensus value is used, criteria to select the parameter to homogeneity and stability tests and to explore the link between PT, method validation and measurement uncertainty, among others.

Keywords: Proficiency test, systematic review, interlaboratory comparisons, laboratories.

Introduction

The requirements of the industries, the government, and the service sector in relation to the quality of testing and calibration have expanded over the past few years. From this fact, the laboratories of different areas have come to fulfill with accreditation criteria's.

Within the context of the qualification of laboratories, there are the requirements of ISO / IEC 17025 and ISO 15189, in which it is explicit that the laboratory should monitor the validity of tests and calibrations performed through a procedure of quality control. Such monitoring can be accomplished through participation in Proficiency Tests (PT) [1].

The PT are important programs that support the reliability of tests and calibrations. They are programs that compare results among a group of laboratories, with the goal of evaluating the technical competence for performing a method of testing or calibration [2]. After participating in a PT, the laboratory has evidence regarding its measurements, checking its proficiency.

In Brazil, the participation in PT is a prerequisite for requesting accreditation by National Institute of Metrology, Quality, and Technology (INMETRO) or Rede Metrológica RS. PT are needed in the routine entities seeking their qualification and recognition of third parties [3]. Laboratories have difficulty validating methods and evaluating measurement uncertainty, which are some of the most points in meeting international normative criteria [4]. These activities can be supported by PT.

The PT are conducted through a system that aims to support the testing and calibration laboratories, ensuring the services offered and providing important information to the company quality management [5]. Through PT it is possible to evaluate the performance of laboratories for specific tests or measurements, identifying analytical problems, establish comparability of methods for testing or calibration, provide additional assurance to laboratory customers, enabling participants based on results of interlaboratory comparisons, validate the declared uncertainty and assign values to reference materials [6].

The comparison programs may vary according to the needs of the industry in which they are used, sample characteristics, methods in use and the number of participants. The nature of the test or measurement taken in PT defines the method of comparison of performance, which can be quantitative, qualitative, or interpretive [6].

Within this context, we highlight the following research question: What are the main practices and knowledge developed and implemented in the performance assessment of laboratories in PT?

The purpose of this study is to identify and analyze the knowledge and leading practices developed and implemented in PT. Our specific goal is to identify the key trends in this area and the theoretical gaps in the development of PT. This article is structured in four sections: introduction, description of the protocol of the systematic review, analysis of results and conclusions.

Protocol of the Systematic Review

This is a research of applied nature and has goals of exploratory character and is dependent on knowledge of the primary sources consulted. The approach of this study is considered qualitative. The proposed logic to perform the systematic review is described in Figure 1, which shows the method used to perform the search, critical appraisal, and synthesis of the information selected.

The proposed method is based on the concepts presented by Akonbeng [7]. The systematic review was chosen for this study based on the statements made by the author cited above, who highlights the fact that this kind of work enables incorporating a larger number of contributions of relevant results, rather than just limiting the completion of some authors, allowing generalizability of the results.

The first stage of the protocol consisted on the elaboration of the research question which underlies the research proposal. The next step was to identify in what language the search would be performed, which was defined as only English. The survey was conducted in six scientific databases, where the initial focus of the search was papers published in indexed journals.

The keywords of the search were defined using Boolean logic, with applications such as OR, AND, and *. The survey period was also limited between 2005 and 2012 (until June). The first result was 10,563 papers. Subsequently areas that dealt only with specific matters, without addressing the research question and the theme of PT, were excluded. At this point, the number decreased to 2,354 papers. As inclusion criteria for journals to prioritize the search, it was stipulated that those with the area of PT in its scope and papers that were directly related to the research topic would be considered, amounting to 125 studies selected. The analysis and selection of articles were conducted through a critical reading of their abstracts and 12 papers were discarded because they were not directly related to the research question. Thus, 113 papers were selected.

The last step was a secondary search in the references of selected papers, identifying key standards, guides and recommendations from entities related to Metrology and Quality areas, in which 34 more references were added. Other details about the protocol of the systematic review are described in Figure 1.

As Figure 1 demonstrates, the implementation of the protocol generated a total of 113 articles and 34 additional references. The next section, which discusses the results of the study, presents a critical analysis of selected documents as well as the major theoretical and practical contributions identified, in order to develop considerations on the subject exposed.

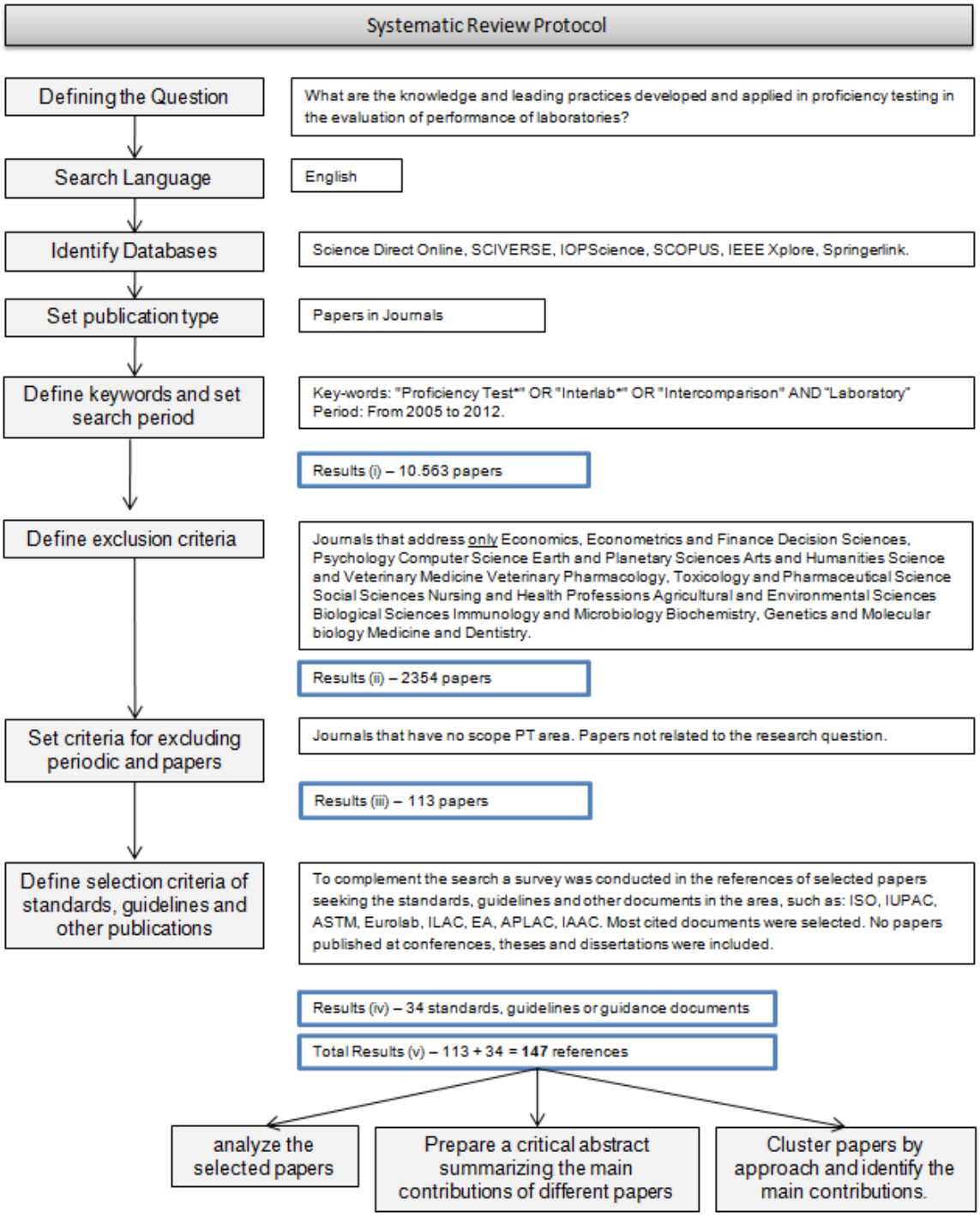

Figure 1 – Protocol of the systematic review

Analysis of the selected documents

The analysis of the references researched was divided into two distinct parts: papers published in journals and standards and other documents in the field of PT. Following are the key concepts and practices identified.

Papers published in journals

We selected 113 papers published in journals that presented the topic PT and were directly related to the research. The main journals selected were: Accreditation and Quality Assurance (80), Flow Measurement and Instrumentation (2), IEEE Transactions on Instrumentation and Measurement (9), Measurement Transaction (11), and Metrologia/BIPM (11). The selected studies were critically analyzed and classified into four sub-areas, according to their approach and the application of the research conducted, namely: performance assessment in PT; calculation method for performance assessment in PT; use of PT for validation and/or estimation of measurement uncertainty, and management and improvements obtained in PT. There are some papers that can be related to more than one approach, but we analyzed them and classified in the approach that has more correlation to the article. This classification is presented in Table 1.

Table 1 – Approaches and Selected Papers

Approach	**Paper**
Performance Assessment in PT	[9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [21] [23] [29] [40] [49] [60] [63] [95] [96] [97] [104] [105] [106] [108] [109] [110] [111] [112] [113] [114] [115]
Calculation method for performance assessment in PT	[19] [22] [24] [26] [27] [28] [30] [31] [32] [33] [34] [35] [36] [37] [39] [41] [42] [44] [45] [46] [50] [52] [54] [55] [56] [57] [58] [59] [61] [65] [66] [68] [69] [71] [73] [76] [78] [79] [81] [83] [85] [87] [89] [90] [91] [93] [98] [100] [103] [107] [116] [118]
Use of PT for validation and/or estimation of measurement uncertainty	[8] [20] [25] [38] [43] [53] [62] [64] [67] [70] [75] [77] [80]
Management of PT and Improvements obtained in PT	[47], [48], [51], [72], [74], [82], [84], [86], [88], [92], [94], [99], [101], [102], [117], [119], [120]

Papers about performance assessment

Several publications analyzed are connected to the use of PT for performance assessment of laboratories, where they are used to confirm modifications or improvements made in measuring methods, and may also be used to assess different measurement systems [9, 11, 13, 14, 23, 40, 49]. This PT is usually made with reference laboratories involved,

which can come from National Metrology Institutes (NMI) [10]. Comparisons are also frequently made between NMI, called key-comparison, which are relevant to ensure the measurements made by NMIs are equivalent [11]. The purpose of key comparison is to support equivalence of measurements of NMIs. Comparisons with long rounds can use NMI reference laboratories and also pivot laboratories, which make intermediate measurements and are considered to be sub-references, and can participate in the stability study of the artifact [13].

Another common practices are bilateral comparisons, generally made between two laboratories, where the one that has the best measurement capability, that is the lowest uncertainty, is designated as the reference [15, 16, 17, 18, 21]. Bilateral comparisons can be made with or without the presence of an NMI.

PT performance assessment also allows predictions concerning the analytical performance of laboratories in one country or a large organization. Research indicates that, in the field of microbiology, it was possible to assess the performance of Belgian and Canadian laboratories in a project for technical improvement of laboratories [60]. A similar approach was presented in comparisons made in other countries such as Croatia, Finland, France, Germany, Hungary, Russia, Slovenia, Spain and Switzerland, where one can have an overview of the participants and can assess the quality of results issued broadly, identifying regional deficiencies [66].

The performance evaluation of laboratories can also be accomplished through the use of Certified Reference Materials (CRM), with property values already known. In this case, as the number of participating laboratories increases, the cost of PT increases, since these materials are expensive [63]. Another possibility is to use a consensus value or historical value of other PT. This approach is also discussed in the next section.

Bilateral comparisons are more frequent in the area of calibration or physical measurement systems. In the testing area more specifically in the chemical and biological areas, the most frequent type of PT is the simultaneous samples, where there are rounds of comparison with several laboratories (usually more than 20 involved).

Papers about calculation method for PT performance assessment

Most of the papers analyzed discusses the statistical methods used to evaluate the performance of laboratories in PT. Surveys indicate that there is a reasonable harmonization

in the use of indexes such as z-score and Error Normalized, but the procedure used to set the assigned value and the standard deviation or uncertainty of reference are not harmonized [24, 41, 43, 46, 61, 68, 71, 73, 90, 98].

A result is considered satisfactory when the z-score is less or equal than |2|, questionable when it is between |2| e |3|, and unsatisfactory when it is equal to or larger than | 3 |. Already the standard error (En) should be equal to or smaller than |1|, so the results were satisfactory. The estimated reference standard deviation and measurement uncertainty need to be reliable. When they are not correctly estimated, the performance evaluation can be considered inconsistent[46].

There are different approaches to obtain the assigned values in PT. The safest way is to obtain the value of a known sample, such as a CRM, or a reliable reference laboratory such as an NMI, for example. Accredited laboratories could also be considered to be a reference, but for this, they should, in addition to accreditation, provide a suitable measurement capability (a reduced uncertainty) [41]. In the latter case a prior demonstration of proficiency would also be advisable.

One of the common approaches in terms of calculation methods for PT performance assessment is the use of consensus value, calculated by classical or robust statistics. The reliability of the determination of the consensus value is relevant, since the mean, median, or mode calculated will be designated as the reference value for PT. The estimated standard deviation also plays a key role in the evaluation of performance, so it must be assessed by the PT provider with caution [59, 61].

A study of PT providers from different European countries and the United States (in the health area with hemoglobin and leukocytes analyses) indicated that the method of calculation used for performance assessment does not have a standard [66]. The exclusion of outliers was performed by providers who participated in the survey, but using different procedures (Russia considered outliers values above 2s, Finland, Spain, France, Hungary, and Slovenia considered values above 3s and Germany considered values greater than + - 40% of the median, for instance). In this same survey, the designation of the assigned value was performed in different ways: Germany and Slovenia worked with reference labs, Croatia, France, Russia, Spain, and Finland worked with the mean, Belgium and Switzerland worked with the median, and Hungary used the mean and value of a specialist laboratory as the

assigned value. This demonstrates the lack of standardization of the different providers. The criterion used for satisfactory results was also variable. Half of the countries surveyed used a criterion of% variation amount over the designated target value, which ranged from 3-25%. These values are usually stipulated by the legislation of those countries. The other providers work with the criterion based on the deviation of PT, and the range of satisfactory ranged from 1-2s [66]. Besides the differences observed, the decision about the PT Scheme design in performance evaluation is not a cultural issue, not even a regional issue. The decision is based on the PT Scheme provider, except in the case that designs are set by regulation.

Several studies indicate that the probability distribution of the PT data, when working with consensus value, should also be considered [30, 32, 50, 116]. Ideally it should follow a Gaussian distribution, that is, symmetrical. If the associated probability distribution is not normal, the assessment by consensus value may be impaired (in the case of bimodal or asymmetric distributions, for instance) [30].

Another important issue is when the number of laboratories is reduced in a PT (less than 30, for example), because one should be more careful in performance assessment, since the reliability regarding the estimated reference standard deviation tends to decrease significantly [57]. Another case that deserves special attention is when the amounts of the analyses of interest are very low, because in this case the use of the standard deviation of consensus may not be the best alternative. Studies indicate that the proposal of Horwitz or the determination of deviations based on historical data of rounds, considering the mass fraction of the element analyzed, prove to be the most appropriate alternatives for the designation of reference standard deviation [68, 73, 89].

Researchers have also conducted simulations to verify the suitability of the use of consensus values of the PT through the Monte Carlo method [41, 79]. In these studies it is clear that the concentration of the analyses, the method bias, the tendency of the laboratory and its repeatability can affect the consensus value. Even so, the approach of using the consensus value was considered adequate (considering the different simulated scenarios). It is worth highlight that studies comparing the use of consensus value with the use of CRM as a reference value were also conducted.

It was observed that performance assessment by En is more frequent in the calibration area. This index is the ratio between the difference of a value measured by a

laboratory and a reference laboratory and the root of the quadratic sum of the expanded uncertainties (of the laboratory being assessed and of the reference laboratory). Usually this value must be less or equal than | 1 | to be satisfactory, but it is also possible that the evaluation criteria is less than | 2 | when working with a standard uncertainty [28, 39]. However, it is not possible to assess the performance of En only mathematically. This index is valid only if the uncertainty of the reference value is less than or equal to the uncertainty of the laboratory being assessed. Studies show that even labs with Enless than 1 still have inadequate results compared to the others [59]. Other publications also comment on the necessary caution when comparing results with high uncertainty, which benefits the laboratories with a high random error. Moreover, it is commented on the problem of using only the z-score between laboratories in PT, which only evaluates the accuracy of the laboratory, but does not account for its repeatability. Therefore, it is always necessary to consider uncertainty or its components in a consistent performance evaluation [65].

Other related approaches with methods for PT performance assessment can also be highlighted, such as: applications of a new statistical method ordinal analysis of variance (ORDANOVA) for interlaboratory comparisons with measurement or semi-quantitative (ordinal) and qualitative (binary) test results [78]; development of methods for quantitative analysis of PT [118], taking as an example the case of the Organization for the Prohibition of Chemical Weapons working with PT to verify that laboratories are able to identify prohibited chemical substances and hazardous samples [76]; using an average weighted by the uncertainty of the laboratories to create a consensus value and weighted averages with different criteria [26, 37]; application of ANOVA and ISO 5725 for performance assessment of PT participants [42]; use of PT participants' results to perform assessment of homogeneity and stability of the data rounds of comparison [44], among others.

Papers about method validation and estimation of measurement uncertainty

Papers classified within the approach of this section are linked to the use of PT in the validation of a method and to support the estimation of its uncertainty. It is possible to estimate measurement uncertainty through PT [20] using alternative approaches so that the comparison data can be combined with data from internal quality control of a laboratory, thus obtaining a combination of different sources of variability focusing on a reasonable estimate of the uncertainty of a trial. Different authors also comment about the use of PT in the

validation of methods that have been modified from their original proposal and, after a comparison with other laboratories, may consider that the changes were consistent and appropriate [25]. This two uses of PT Schemes are not pointed in ISO/IEC 17043.

Still on the validation methods, it can be stated that PT results could be used as an alternative to meet certain requirements such as analytical precision and uncertainty [43]. Furthermore, the samples of PT could be used in internal quality control. This additional use of PT can help laboratories to reduce the financial impact of its quality assurance procedure [43].

The adequacy of performance assessment performed in a PT is linked to uncertainty of the assigned value. Within this context, it is possible to work with a "target uncertainty." The importance of implementing the "target measurement uncertainty" was indicated in different areas (testing and calibration). For a proper comparison, it was recommended that uncertainty target was at least three times less than the uncertainty of the participating laboratory [53]. This way, the laboratory can identify whether or not its uncertainty is appropriate [64].

Since the publication of the Guide to the Expression of Uncertainty in Measurement (GUM), many projects have been carried out to develop an alternative practice when it is technically or economically difficult to obtain a suitable mathematical model of the measurement [62]. Many laboratories are also reluctant to apply the law of propagation of uncertainty with its apparent mathematical complexity. These alternative practices can use the experimental data available from laboratories, such as repeatability, reproducibility, control charts, PT, among others. The only point to be noted in this approach is the fact that the standard uncertainty used based on the PT may be higher, because this proposal takes into account all the variability introduced by the different analytical methods. A more promising method for estimating uncertainty would be to use a combination of PT data and internal validation data of the method or quality control [62].

There is a mathematical model that was tested to estimate uncertainty of a laboratory, relating it to the standard deviation of the measurement and with the concentration of an analysis. This model was evaluated through a meta-analysis considering different PT, where its wide application was evident. The proposed mathematical function may be represented by the square root of the quadratic sum of α and $C.\beta$, where C is the analyzed concentration and

the other components are related to the standard deviation of the method [67]. The parameter α is connected to the detection limit of the method and intermediate β, to the relative accuracy of the method.

With these two parameters a curve can be developed, where on the x-axis there is the mass fraction of the element that is being analyzed and on the y-axis, the standard deviation related to the concentration. Thus, it is possible to obtain the constants alpha and beta of the mathematical model mentioned before and to obtain the standard deviation for reproducibility of the measurement system for any concentration value. This can be done with different analytical parameters. Obviously, a good estimate of model data depends on different concentrations of PT and preferably with a large number of participants. As the reproducibility standard deviation is the major component of the estimate of uncertainty, the same could be obtained by multiplying this amount by the factor k of 2, to obtain the expanded uncertainty [67]. Other research on the same topic claim that this approach is useful and if applied appropriately makes available equations related to the performance of different analytical methods, besides the fact that the measurement uncertainty can be estimated for different concentrations [80]. It is worth highlight that these equations can be used to obtain an indication of the average quality of analytical results in a specific field and can be used by regulatory bodies to formulate legislation requirements according to the quality of existing measurement in the area [80].

Finally, other researchers indicate that the two most important concepts in metrology are certainly traceability of standards used and its measurement uncertainty, and its concepts are related to PT Schemes [75]. In areas such as chemistry and biology, many problems remain to be resolved to support international agreements related to these concepts. Therefore, NMIs laboratories in these areas have developed strategies so that conclusions in PT are feasible and increasingly frequent [75], due to its importance and connection with traceability and uncertainty.

Papers about management and improvement of PT

The PT is developed by providers, who must also have proven their qualifications through an assessment of an accreditation body. These assessments are relatively recent, beginning through pilot programs, mainly in Europe, in 2005 [48, 88]. In Brazil, this activity

became an official accreditation only in 2011, after the implementation of a pilot project by INMETRO.

International research conducted with 160 different providers from 32 countries show a strong tendency for accreditation of PT [47]. According to these surveys, it was found that this type of evaluation is based on various combinations of normative documents, which may illustrate a lack of harmonization of accreditation bodies. Furthermore, it was shown that some customers have an appeal to their suppliers to seek accreditation. However, among the providers consulted, less than half expect an improvement in their quality through accreditation and more than half expect a significant increase in their costs [47].

Another interesting approach is the possibility of organizing interlaboratory collaborative studies with a purpose of assessing the performance of the analytical test method and not only from laboratories [72]. Within this context, researchers recommend care in the management and conduct of a trial for purposes of performance assessment methods, as well as their statistical analysis. Issues such as the choice of participating laboratories and the designation of the assignment values are important. Therefore, it is clear that it is possible to establish a standard method for analysis through rounds of interlaboratory collaborative studies, with greater assurance that the developed method provides reproducibility in different operating conditions [72].

Requirements applicable to PT are similar to those considered in the production of reference materials [78]. The samples of PT should have a degree of homogeneity and stability for the purpose of identifying differences between the laboratories. Based on this logic, the process used to prepare the samples held by the provider must be appropriate and shall ensure the quality of the items that will be sent to laboratories in the comparison rounds [78]. Tests for homogeneity and stability are essential in this context.

Normally PT are performed in rounds that occur during one year. Studies in the field of occupational medicine indicate that 28% of PT run with 4 rounds per year [72]. Similar results were observed in hematology and microbiology, with a median of 3 rounds per year. The median of biochemistry was 6 rounds per year, where 33% of the PT have intervals of one month. The number of samples per round varied between 1 (31%) to over 20 (0.5%), where most providers offer between 1 and 3 samples per round (83%) [72].

The implementation of the PT has a wide area. Initially they were most in demand in the area of calibration, being performed mainly by reference laboratories. The medical area also started with PT compulsory participation, due its importance. According to accreditation bodies, today the demand for PT in different areas is greater than its supply and availability. The expansion of PT is increasingly perceived in the field of chemical, biological, geological, agricultural tests and even in the veterinary area [51]. Nowadays, most of the PT done in the world is in medical areas.

Different international regulatory agencies also consider the PT as an appropriate way to ensure the reliability of laboratory results and, on several occasions, make participation in these activities compulsory [86]. Yet, research indicates that laboratories participating in PT over time tend to improve their results, as well as the providers improve the management and reliability of their programs [99, 101].

Providers also had to adapt and start work focused on better management of its activities, seeking compliance with standards such as ISO/IEC 17043 [117, 119, 120]. This standard addresses technical and managerial issues that should be followed by PT providers, however it is still not compulsory use in many countries. Meeting this standard, in an isolated way, when not assessed by a third part like an accreditation body, does not guarantee proper operation of the PT developed, since an adequate managerial capacity installed in companies and an appropriate technical knowledge on the subject are necessary.

Other selected references

The second stage of the systematic review focused on the pursuit of standards and guideline of renowned entities in the PT area. We selected the most cited references in the articles that were considered in the previous step. Over 34 references were identified, from International Organization for Standardization (41.2%), American Society for Testing and Materials (14.7%), Asia Pacific Laboratory Accreditation Cooperation (14.7%), International Union of Pure and Applied Chemistry (5.9%), European Co-operation for Accreditation (5.9%), European Federation of National Associations of Measurement, Testing and Analytical Laboratories (5.9%), NORDTEST (2.9%), Bureau International des Poids et Mesures (2.9%), International Laboratory Accreditation Cooperation (2.9%) and InterAmerican Accreditation Cooperation (2.9%).

The selected references were classified into three approaches. The division performed is shown in Table 2. After the classification a summary of the approach of these documents according to their classification is shown.

Table 2 – Approaches and other selected publications

Approach	**Standards, Guides and Guideline Documents**
Definitions, Management, operation and use of PT Programs	[1], [2], [6], [121], [124], [126], [128], [129], [130], [134], [135], [136], [137], [144], [145], [146], [147], [148], [149], [150], [151],
Statistical Methods for PT performance assessment	[122], [123], [125], [127], [131], [132], [133], [138], [139]
Use of PT to estimate measurement uncertainty	[127], [140], [141], [142], [143]

Definitions, Management, operation and use of PT Programs

Standards that address definitions of PT are mostly published by ISO. Some norms are for guidance [126, 128, 135, 134, 136, 137], addressing specific PT in technical areas such as tissues, microbiology, petroleum products, among others. There are, in this group, standards that are used to accredit laboratories [1, 129], that address PT in the field of quality assurance of testing or calibration.

Other standards are also used in the accreditation of reference material producers and providers of PT [124, 6], the latter of which establishes the technical and management requirements that must be followed to conduct a PT appropriately. Reference material producers and PT providers are different types of organizations, and they should not be confounded. Among the surveyed standards, ISO / IEC 17043 is the most complete, and is used globally by different providers in different areas [6].

In this category there are also standards [130] and other documents published by organizations that establish major policies for the accreditation process for laboratories and providers PT [144, 145, 146]. These documents establish the minimum frequency of participation in PT, which should be the policies of the bodies to assess inadequate results obtained in PT and how these factors may influence an accreditation process.

Statistical methods for PT performance assessment

Several standards and guideline documents have different approaches to PT performance assessment [122, 123, 125, 127, 131, 132, 133, 138, 139]. Most documents converge in the use of the same indicators for performance rating, the most common being the z-score (and its variations as Z'-score, zeta-score, etc.) and Error Normalized. However, the method of calculation or estimation of reference values shows much divergence and relative lack of standardization. The standards usually present examples of the application of its procedures to set the assigned values, but they are general. It is common to need a "fit for purpose" in each specific PT Scheme developed.

Most documents propose the evaluation of repeatability, reproducibility and accuracy of the results of the participating laboratories in comparisons, but in a general way. Still, regarding the tests of homogeneity and stability of the items that are compared (samples or artifacts) we emphasize that the references do not provide details regarding how many analyses / parameters should be selected to consider testing representative and consistent. The documents cited in this section do not address in detail the influence that the probability distribution of the data may have on the results of PT.

Use of PT to estimate measurement uncertainty

Documents for estimation of the measurement uncertainty were also frequently referenced in the articles selected in this systematic review. Neither is focused only on PT, since they address methods for estimation of uncertainty in testing or calibration [140, 141, 142, 143].

Furthermore, some documents suggest alternative approaches to calculating uncertainty, considering the results of PT [141, 142]. These approaches should be selected carefully, as the result of uncertainty can be strongly influenced by the performance of the participants of the comparison. Still, these alternative approaches are recommended when there is little information on the sources of variation of the method or when getting values associated with measurement accuracy is complex.

Regarding to uncertainty measurement of the assigned values, we think that is a point to be improved. The standard ISO13528 gives a very simple approach to establish the uncertainty of the assigned value when the provider uses consensus value. In this case, the

uncertainty can vary drastically according to the number of laboratories that are in the round [127].

Identification of gaps to be exploited

We can see the importance of the topic and the increasing demand for participation in PT, whether it is required by the government, accreditation and conformity assessment bodies. Due to the numerous areas of laboratories, providers are not yet prepared to meet all existing demands. Still, there is a perceived need for the structuring of these organizations in terms of obtaining adequate standards in the area and agile management to meet the market demands. Several PT are developed in different countries and different areas, but approaches that assist management of providers with a view of projects were not found in the sources researched. The main area of the reference standard, ISO / IEC 17043 also does not address the issue of development and management of PT with the project vision. It is likely that this is an issue to be explored. This standard does not consider areas such as risk, costs, strategy and time management, for example, which are typical from project management knowledge, and could be useful in PT Schemes.

Although publications related to the topic often address the link between method validation, measurement uncertainty, and PT, it is clear that there is not a document that presents a logical interface between these themes. This ends up creating doubts and does not always clear up what the actual intended use of PT is.

Another important issue, discussed by different researchers, is the impact that the probability distribution of the data can have on performance assessment. These issues, in most cases, are not considered by the providers and may have a high impact on the statistical treatment of data, especially when working with consensus value (with references generated with data from the participants of the PT). Still, the standards of the area do not report details on this fact, nor procedures for assessing probability distributions obtained in PT.

It is noteworthy that a factor cited in different studies was the homogeneity and stability of the samples prepared in PT and the need for ensuring this point to increase confidence in the round of comparison. However, the standards and publications do not make clear what the criteria are for selection of parameters for these tests should be, as well as how many parameters would be representative for an adequate test of homogeneity and stability of

the samples. This fact deserves attention, since a false sense of homogeneity or stability may compromise the trust of a PT.

Finally, the ISO Standards related to PT are, sometimes, general and not specific, because there are an enormous variety of measurement fields, national regulations, and "fit for purpose" needs - one lab's needs for accuracy and precision is not always the same as another's.

Conclusions

This study presented a systematic review that covered the period from 2005 to 2012 (June) considering publications related to the theme PT. A total of 147 references were selected, including articles, standards, and guideline documents.

Thus, it is considered that the objective was achieved, since we analyzed the expertise and main practices related to the theme PT in the research sources listed above. These shortcomings were raised, being related to the management of PT projects; analysis of the link between validation, PT, and measurement uncertainty; preliminary evaluation of the probability distribution of the data from PT; selection of variables for testing homogeneity and stability. The shortcomings are not limited to these topics, though this analysis is based on the perception of the main factors analyzed. In future researches or reviews about this theme is advisable to include published PT reports offered by international cooperation's (for example IMEP and APLAC) and private schemes that are offered internationally.

References

[1]. ISO/IEC 17025:2005 General requirements for the competence of testing and calibration laboratories.
[2]. APLAC (2008). PT-02: Testing - Interlaboratory Comparisons. Issue No. 6.
[3]. Tholen D, (2011) Metrology in service of society: the role of proficiency testing. AccredQualAssur 16: 603-605 Doi: 10.1007/s00769-011-0836-4
[4]. Rosa D, Pizzolato M, Albano FM, Sartori T (2011) Analysis of non conformyties of testing and calibration laboratory. In: Metrologia 2011- Metrology as Factor of Quality, Innovation and Competitiveness. Natal - RN. Book of abstracts.
[5]. Howerton, D; Krolak, J; Manasterski, A; Handsfield, J. (2010) Proficiency testing performance in US laboratories. Arch Pathol Lab Med. 134:751–758
[6]. ISO/IEC 17043:2010. Conformity assessment—general requirements for proficiency testing.
[7]. Akobeng, A. K. (2005) Understanding systematic reviews and meta-analysis. Arch Dis Child 2005;90:845-848 Doi: 10.1136/adc.2004.058230

[8]. Choi HM, Park K, Youn KO, Choi YM (2010) Uncertainty evaluation procedure and intercomparison of bell provers as a calibration system for gas flow meters. Flow MeasInst 21:448-496 , Doi: 10.1016/j.flowmeasinst.2010.07.002

[9]. Standiford DM, Lee M (2012) Inter-laboratory comparison results for Coriolis mass flowmeter calibration facilities. Flow Meas Inst, Doi: 10.1016/j.flowmeasinst.2011.12.004

[10]. Hamilton CA, Kupferman SL, Salazar MT, Deaver D, Wood BM (2005) Interlaboratory comparison at 10 V DC. IEEE Transactions on Instrumentation and Measurement Doi: 10.1109/TIM.2004.838134

[11]. So E, Angelo D, Tsuchiyama T, Tadokoro T, Waltrip BC, Nelson TL (2005) Intercomparison of calibration systems for ac shunts up to audio frequencies. IEEE Transactions on Instrumentation and Measurement doi: 10.1109/TIM.2004.843418

[12]. Bieler M, Seitz S, Spitzer M, Hein G, Pierz K, Siegner U et al (2007) Rise-time calibration of 50-ghz sampling oscilloscopes: intercomparison between PTB and NPL. IEEE Transactions on Instrumentation and Measurement Doi: 10.1109/TIM.2007.890797

[13]. Tang Y, Hamilton CA, Deaver D, Parks H, Wood BM (2007) The seventh intercomparison of Josephson voltage standards in North America. IEEE Transactions on Instrumentation and Measurement Doi: 10.1109/TIM.2007.890797

[14]. Jung JK, So E, Park YT, Kim M (2009) KRISS-NRC intercomparisons of calibration systems for instrument transformers with many different ratios at power frequency. IEEE Transactions on Instrumentation and Measurement Doi: 10.1109/TIM.2008.2006966

[15]. Callegaro L, D'Elia V, Bohá˘cek J (2009) Four-terminal-pair inductance comparison between INRIM and CTU. IEEE Transactions on Instrumentation and Measurement Doi: 10.1109/TIM.2008.928402

[16]. Tang Y, Wood BM, Hamilton CA (2009) A two-way josephson voltage standard comparison between NIST and NRC. IEEE Transactions on Instrumentation and Measurement Doi: 10.1109/TIM.2008.2006960

[17]. Landim RP, Tang Y, Afonso E, Ferreira V (2011) Josephson voltage standard comparison between NIST and INMETRO as a link to BIPM. IEEE Transactions on Instrumentation and Measurement Doi: 10.1109/TIM.2010.2099370

[18]. Kang J, Kang N, Gentle DG, MacReynolds K, Francis MH (2011) Intercomparison of standard gain horn antennas at w-band. IEEE Transactions on Instrumentation and Measurement Doi: 10.1109/TIM.2010.2103413

[19]. Pavese F (2006) A metrologist viewpoint on some statistical issues concerning the comparison of non-repeated measurement data, namely MRA key comparisons. Measurement 39:821–828, Doi: 10.1016/j.measurement.2006.05.007

[20]. Désenfant M, Priel M (2006) Road map for measurement uncertainty evaluation. Measurement 39:841–848, Doi: 10.1016/j.measurement.2006.04.008

[21]. Chung JW, Borys M, Firlus M, Lee WG, Schwartz R (2007) Bilateral comparison of buoyancy artefacts between PTB and KRISS. Measurement 40:774–778, Doi: 10.1016/j.measurement.2006.08.003

[22]. Ferrero C, Casaril M (2009) Proficiency testing programs to improve traceability in chemical analysis. Measurement 42:1502–1509, Doi:10.1016/j.measurement.2009.07.005

[23]. Galliana F, Capra PP, Gasparotto E (2009) Report of the Italian inter-laboratories comparison of high dc resistance on the calibration of a 10 MX and a 1 GX resistors. Measurement 42:1532–1540, Doi:10.1016/j.measurement.2009.07.014

[24]. Ferrero C (2009) The equivalence in accreditation: robust or standard statistical analysis of interlaboratory comparisons. Measurement 42:1541–1545, Doi: 10.1016/j.measurement.2009.07.011

[25]. Astrua M, Pisani M (2009) Validation of a novel technique for the measurement of the refractive index of a prism by means of interlaboratory comparison. Measurement 42:1546–1549, Doi: 10.1016/j.measurement.2009.08.012

[26]. Brunetti L, Oberto L, Sellone M, Terzi P (2009) Establishing reference value in high frequency power comparisons. Measurement 42:1318–1323, Doi: 10.1016/j.measurement.2008.08.008

[27]. Bergoglio M, Malengo A, Mari D (2011) Analysis of interlaboratory comparisons affected by correlations of the reference standards and drift of the travelling standards. Measurement 44:1461–1467, Doi:10.1016/j.measurement.2011.05.012

[28]. Cascetta F, Comazzi M, Musto M, Rotondo G (2012) An experimental intercomparison of gas meter calibrations. Measurement Transaction, doi: 10.1016/j.measurement.2012.04.020

[29]. Galliana F, Capra PP, Gasparotto E (2012) Inter-laboratories comparison at 100 GX and 1 TX level to evaluate the traceability transfer from INRIM in the field of high dc resistance. Measurement 45:615–621, Doi:10.1016/j.measurement.2011.10.028

[30]. Willink R (2006) Forming a comparison reference value, from different distributions of belief. Metrologia, doi: 10.1088/0026-1394/43/1/002
[31]. Hornikova A, Zhang NF (2006) The relation between the En values including covariance and the 'exclusive' En statistic. Metrologia, doi: 10.1088/0026-1394/43/1/N01
[32]. Lira I (2007) Combining inconsistent data from interlaboratory comparisons. Metrologia, doi: 10.1088/0026-1394/44/5/019
[33]. Kacker RN, Forbes A, Kessel R, Sommer K (2008) Bayesian posterior predictive p-value of statistical consistency in interlaboratory evaluations. Metrologia 44, doi:10.1088/0026-1394/45/5/004
[34]. Kacker RN, Forbes A, Kessel R, Sommer K (2008) Classical and Bayesian interpretation of the Birge test of consistency and its generalized version for correlated results from interlaboratory evaluations. Metrologia 45: 257–264, doi:10.1088/0026-1394/45/3/001
[35]. Khlevnoy B, Sapritsky V, Rougie B, Gibson C, Yoon H, Gaertner A et al (2009) CCPR-S1 Supplementary comparison for spectral radiance in the range of 220nm to 2500nm. Metrologia 46: 174-180, doi: 10.1088/0026-1394/46/4/S08
[36]. Zhang W,Zhang NF, Liu H (2009) A generalized method for the multiple artefacts problem in interlaboratory comparisons with linear trends. Metrologia 46:345–350, doi: 10.1088/0026-1394/46/3/024
[37]. Rukhin AL (2009) Weighted means statistics in interlaboratory studies. Metrologia 46:323–331, doi: 10.1088/0026-1394/46/3/021
[38]. Arp U, Klein R, Li Z, Paustian W, Richter M, Shaw PS et al (2011) Synchrotron radiation-based bilateral intercomparison of ultraviolet source calibrations. Metrologia 48:261–267, doi: 10.1088/0026-1394/48/5/004
[39]. Fellmuth B, Wolber L, Head DI, Hermier Y, Hill KD, Nakano T, Pavese F et al (2012) Investigation of low-temperature fixed points by an international star intercomparison of sealed triple-point cells. Metrologia 49:257–265, doi: 10.1088/0026-1394/49/3/257
[40]. Kim M, Pratt JR, Brand U, Jones CW (2012) Report on the first international comparison of small force facilities: a pilot study at the micronewton level. Metrologia 49:70–81, doi: 10.1088/0026-1394/49/1/011
[41]. Wong SK (2006) A comparison of performance statistics for proficiency testing programmes. AccredQualAssur 12:59–66, doi: 10.1007/s00769-006-0233-6
[42]. Svegl F, Strupi JS, Svegl IG (2006) Proficiency testing of chloride content in different types of Portland cement. AccredQualAssur 11: 414–421, doi: 10.1007/s00769-006-0162-4
[43]. Detaille R, Maetz P (2006) Practical uses of proficiency testing as valuable tools for validation and performance assessment in environmental analysis. AccredQualAssur 11: 408–413, doi: 10.1007/s00769-006-0156-2
[44]. Tholen D, Chappel R, Francis B (2006) Use of participant EQA results to assess sample homogeneity and stability for qualitative analytes. AccredQualAssur 11: 400–407, doi: 10.1007/s00769-006-0167-z
[45]. Thompson M, Ellison SRL (2006) Fitness for purpose – the integrating theme of the revised harmonised protocol for proficiency testing in analytical chemistry laboratories. AccredQualAssur 11: 373–378, doi: 10.1007/s00769-006-0137-5
[46]. Boley N, van der Veen AMH, Robouch P, Golze M, van de Kreeke J et al (2006) Comparability of PT schemes – what did we learn from COEPT?. AccredQualAssur 11: 391–399, doi: 10.1007/s00769-006-0106-z
[47]. Schmidt A, Örnemark U, Golze M, Henriksen GM (2006) Surveys on the accreditation of providers of proficiency testing and external quality assessment schemes. AccredQualAssur 11:379–384, doi: 10.1007/s00769-006-0110-3
[48]. Monnery K (2006) Experiences from the assessment of proficiency testing providers. AccredQualAssur 11:385–387, doi: 10.1007/s00769-006-0138-4
[49]. de Leer EWB (2006) The metrological infrastructure for gas analysis. AccredQualAssur 11:175–181, doi: 10.1007/s00769-006-0123-y
[50]. Kuselman I (2006) Nonparametric assessment of comparability of analytical results obtained in proficiency testing based on a metrological approach. AccredQualAssur 10:659–663, doi: 10.1007/s00769-006-0092-1
[51]. Colling A, Jeggo M, Louvandini H, Lelenta M, Robinson M (2007) The establishment of quality systems in veterinary diagnostic testing laboratories in developing countries: experiences with the FAO/IAEA External Quality Assurance Programme. AccredQualAssur 13:33–45, doi: 10.1007/s00769-007-0329-7
[52]. Rienitz O, Schiel D, Güttler B, Koch M, Borchers U (2007) A convenient and economic approach to achieve SI-traceable reference values to be used in drinking-water interlaboratory comparisons. AccredQualAssur 12:615–622, doi: 10.1007/s00769-007-0320-3

[53]. Serapinas P (2007) Approaching target uncertainty in proficiency testing schemes: experience in the field of water measurement. AccredQualAssur 12:569–574, doi: 10.1007/s00769-007-0310-5
[54]. Kuselman I, Belli M, Ellison SLR, Fajgelj A, Sansone U, Wegscheider W (2007) Comparability and compatibility of proficiency testing results in schemes with a limited number of participants. AccredQualAssur 12:563–567, doi: 10.1007/s00769-007-0309-y
[55]. Polkowska-Motrenko H, Rossbach M (2007) Preparation and characterization of a proficiency test material ‘‘Polish mushrooms’’ supported by the International Atomic Energy Agency (IAEA), Vienna. AccredQualAssur 12:343–350, doi: 10.1007/s00769-007-0293-2
[56]. Sucha´nek M, Robouch P, Aregbe Y, Herrero A, Taylor P (2007) AccredQualAssur 12:399–403, doi: 10.1007/s00769-007-0297-y
[57]. Belli M, Ellison SLR, Fajgelj A, Kuselman I, Sansone U, Wegscheider W (2007) Implementation of proficiency testing schemes for a limited number of participants. AccredQualAssur 12:391–398, doi: 10.1007/s00769-006-0247-0
[58]. Thompson M, Mathieson K, Owen L, Damant AP, Wood R (2008) The relationship between accreditation status and performance in a proficiency test. AccredQualAssur 14:73–78, doi: 10.1007/s00769-008-0457-8
[59]. Heydorn K (2008) The application of data from proficiency testing to laboratory accreditation according to ISO 17025. AccredQualAssur 13:703–709, doi: 10.1007/s00769-008-0435-1
[60]. Vernelen K, Noble MA, Libeer J (2008) External quality assessment in microbiology: comparison of results from Belgian and Canadian laboratories with regard to their ability to identify Streptococcus pyogenes. AccredQualAssur 13:501–504, doi: 10.1007/s00769-008-0421-7
[61]. Rosario P, Martínez JL, Silván JM (2008) Comparison of different statistical methods for evaluation of proficiency test data. AccredQualAssur 13:493–499, doi: 10.1007/s00769-008-0413-7
[62]. Fisicaro P, Amarouche S, Lalere B, Labarraque G, Priel M (2008) Approaches to uncertainty evaluation based on proficiency testing schemes in chemical measurements. AccredQualAssur 13:361–366, doi: 10.1007/s00769-008-0402-x
[63]. Fisicaro P, Champion R, Labarraque G, Lalere B, Mace T, Sutour C (2008) AccredQualAssur 13:367–372, doi: 10.1007/s00769-008-0394-6
[64]. Chajduk E. Polkowska-Motrenko H, Dybczynski RS (2008) A definitive RNAA method for determination of selenium in biological samples: uncertainty evaluation and assessment of degree of accuracy. AccredQualAssur 13:443–451, doi: 10.1007/s00769-008-0377-7
[65]. Heydorn K (2008) Proficiency testing in the light of a new rationale in metrology. AccredQualAssur 13:217–222, doi: 10.1007/s00769-008-0368-8
[66]. Van Blerk M, Albarède S, Deom A, Gutiérrez G, Heller S, Nazor A et al (2008) Comparison of evaluation procedures used by European external quality assessment scheme organizers for haemoglobin concentration and leukocyte concentration. AccredQualAssur 13:145–148, doi: 10.1007/s00769-008-0364-z
[67]. Thompson M, Mathieson K, Damant AP, Wood R (2008) A general model for interlaboratory precision accounts for statistics from proficiency testing in food analysis. AccredQualAssur 13:223–230, doi: 10.1007/s00769-008-0356-z
[68]. Ellison SLR, Mathieson K (2008) Performance of uncertainty evaluation strategies in a food proficiency scheme. AccredQualAssur 13:231–238, doi: 10.1007/s00769-007-0353-7
[69]. Koch M, Baumeister F (2008) Traceable reference values for routine drinking water proficiency testing: first experiences. AccredQualAssur 13:77–82, doi: 10.1007/s00769-007-0349-3
[70]. Analytical Method Committee, The Royal Society of Chemistry (2010) The role of proficiency testing in method validation. AccredQualAssur 15:73–79, Doi: 10.1007/s00769-009-0560-5
[71]. Ellison SLR (2009) Performance of MM-estimators on multi-modal data shows potential for improvements in consensus value estimation. AccredQualAssur 14:411–419, Doi: 10.1007/s00769-009-0571-2
[72]. Thomas A (2009) External Quality Assessment in laboratory medicine: is there a rationale to determine frequency of surveys?. AccredQualAssur 14:439–444, Doi: 10.1007/s00769-009-0563-2
[73]. Arnaud J, Jones RL, LeBlanc A, Lee M, Mazarrasa O, Parsons P et al (2009) Criteria to define the standard deviation for proficiency assessment for the determination of essential trace elements in serum: comparison of Z-scores based on the Horwitz function or on biological variability. AccredQualAssur 14:427–430, Doi: 10.1007/s00769-009-0561-4
[74]. Belli M (2009) PT/EQA standards and guidelines: some remarks on quality and reliability of test items. AccredQualAssur 14:421–426, Doi: 10.1007/s00769-009-0554-3

[75]. Priel M, Amarouche S, Fisicaro P (2009) Metrological traceability is not always a straight line. AccredQualAssur 14:593–599, Doi: 10.1007/s00769-009-0540-9
[76]. Dubey V, Velikeloth S, Sliwakowski M, Mallard G (2009) Official proficiency tests of the organisation for the prohibition of chemical weapons: current status and future directions. AccredQualAssur 14:431–437, Doi: 10.1007/s00769-009-0520-0
[77]. Priel M (2009) From GUM to alternative methods for measurement uncertainty evaluation. AccredQualAssur 14:235–241, Doi: 10.1007/s00769-009-0518-7
[78]. Bashkansky E, Gadrich T, Kuselman I (2012) Interlaboratory comparison of test results of an ordinal or nominal binary property: analysis of variation. AccredQualAssur 17:239–243, Doi: 10.1007/s00769-011-0856-0
[79]. Wong S (2011) Performance evaluation for proficiency testing with a limited number of participants. AccredQualAssur 16:539–544, Doi: 10.1007/s00769-011-0816-8
[80]. Koch M, Magnusson B (2012) Use of characteristic functions derived from proficiency testing data to evaluate measurement uncertainties. AccredQualAssur, Doi: 10.1007/s00769-012-0880-8
[81]. Koch M, Baumeister F (2012) On the use of consensus means as assigned values. AccredQualAssur, Doi: 10.1007/s00769-012-0884-4
[82]. Ellison SLR, Hardcastle WA (2012) Causes of error in analytical chemistry: results of a web-based survey of proficiency testing participants. AccredQualAssur, Doi: 10.1007/s00769-012-0894-2
[83]. Côté I, Robouch P, Robouch B, Bisson D, Gamache P, LeBlanc A et al (2012) Determination of the standard deviation for proficiency assessment from past participant's performances. AccredQualAssur, Doi: 10.1007/s00769-012-0906-2
[84]. Jalukse L, Vabson V, Leito I (2006) In situ interlaboratory comparisons for dissolved oxygen concentration and Ph. AccredQualAssur 10: 562–564, Doi: 10.1007/s00769-005-0058-8
[85]. Thompson M (2006) Using mixture models for bump-hunting in the results of proficiency tests. AccredQualAssur 10: 501–505, Doi: 10.1007/s00769-005-0053-0
[86]. Stanton NV, Fritsch T, Geraghty C, Verostek MF, Weiner B, Parsons PJ (2006) The role of proficiency testing in the detection and resolution of calibration bias in the LeadCare R blood lead analyzer; limitations of peer-group assessment. AccredQualAssur (2006) 11: 590–592, Doi: 10.1007/s00769-006-0189-6
[87]. Gluschke M (2006) QUASIMEME results of laboratories involved in the German Marine Monitoring Programme for the North and Baltic Seas. AccredQualAssur 11: 470–473, Doi: 10.1007/s00769-006-0155-3
[88]. Van Berkel E, Boley N, Ellison S, Libeer J, Mäkinen I, Sandberg S (2006) Proficiency testing in analytical chemistry,microbiology and laboratory medicine – working group discussions on current status, problems and future directions. AccredQualAssur 11: 446–450, Doi: 10.1007/s00769-006-0111-2
[89]. Visser RG (2009) Reliability of proficiency test results for metals and phthalates in plastics. AccredQualAssur 14:29–34, Doi: 10.1007/s00769-008-0465-8
[90]. Mäkinen I (2009) Use of uncertainty estimates as reported by participants in proficiency testing for the evaluation of their results: pros and cons. AccredQualAssur 14:35–41, Doi: 10.1007/s00769-008-0462-y
[91]. Mitani Y, Lara-Manzano JV, Rodriguez-Lopez A (2008) Proficiency testing scheme for the harmonization and comparability of analytical measurements. AccredQualAssur 13:421–426, Doi: 10.1007/s00769-008-0398-2
[92]. Gardner M (2008) Long-term proficiency testing for the UK Acid Waters Monitoring Network. AccredQualAssur 13:255–260, Doi: 10.1007/s00769-008-0367-9
[93]. Tao L, Xiang H, Xu H, Sun J, Zhang M, Wang K et al (2010) Application of the FAPAS protocol to a proficiency testing on Pb and Cd extracted from glazed ceramic. AccredQualAssur 15:155–162, Doi: 10.1007/s00769-009-0609-5
[94]. Drnovsˇek J, Pusˇnik I, Bojkovski J, Begesˇ G (2010) Regional cooperation in proficiency testing: strategy planning and practical experiences in Western Balkan countries. AccredQualAssur 15:125–131, Doi: 10.1007/s00769-009-0569-9
[95]. Grimaldi M, Mancin M, Trevisan R, Mioni R (2009) Proficiency testing in food microbiology ''Aqua'': enumeration of Staphylococcus aureus. AccredQualAssur 14:503–505, Doi: 10.1007/s00769-009-0572-1
[96]. De Souza GB, Nogueira ARA, Del-Santo VR, Picchi CMC, Guimarães ES, BarioniJr W (2009) Proficiency testing of animal nutrition laboratories. AccredQualAssur 14:455–460, Doi: 10.1007/s00769-009-0562-3
[97]. Poulsen ME, Christensen HB, Herrmann SS (2009) Proficiency test on incurred and spiked pesticide residues in cereals. AccredQualAssur 14:477–485, Doi: 10.1007/s00769-009-0555-2

[98]. Bayrak ÖT, Okandan E, Üçkardes H (2009) Inter-laboratory comparison scheme for fuel sector, LABKAR in Turkey. AccredQualAssur 14:497–502, Doi: 10.1007/s00769-009-0542-7
[99]. Gaunt W, Whetton M (2009) Regular participation in proficiency testing provides long term improvements in laboratory performance: an assessment of data over time. AccredQualAssur 14:449–454, Doi: 10.1007/s00769-009-0523-x
[100]. Cotman M, Drolc A, Tisˇler T (2009) Interlaboratory studies on wastewater toxicity using Daphnia magna. AccredQualAssur 14:319–327, Doi: 10.1007/s00769-009-0519-6
[101]. Whetton M, Finch H (2009) Analytical performance is improved by regular participation in proficiency testing: an analysis of data from the Aquacheck proficiency testing scheme. AccredQualAssur 14:445–448, Doi: 10.1007/s00769-009-0498-7
[102]. De Graaf DC, Ritter W, Jacobs FJ, Brunain M, Imberechts H, Mintiens K et al (2009) Lessons from the first international proficiency test for the detection of spores from the honey bee pathogen Paenibacillus larvae. AccredQualAssur 14:273–276, Doi: 10.1007/s00769-009-0495-x
[103]. Drolc A, Cotman M (2009) Integration of metrological principles and performance evaluation in a proficiency testing scheme in support of the Council Directive 98/83/EC. AccredQualAssur 14:199–205, Doi: 10.1007/s00769-009-0494-y
[104]. Urquiza MP, Mitani Y, Jaimes HON (2009) Proficiency tests to evaluate commercially available IVD kits for glucose and cholesterol measurements. AccredQualAssur 14:269–271, Doi: 10.1007/s00769-009-0493-z
[105]. Mäkinen I, Hanste U (2009) Proficiency testing for measurement of radon (222Rn) in drinking water. AccredQualAssur 14:473–476, Doi: 10.1007/s00769-009-0488-9
[106]. Antin L, Armishaw P (2010) Aspects of proficiency testing studies of trace elements in environmental samples with a focus on laboratory performance. AccredQualAssur 15:467–471, Doi: 10.1007/s00769-010-0672-y
[107]. Kim Y, Lee J, Song K, Kang Y, Shin S, Jung G et al (2010) A review of proficiency testing scheme in Republic of Korea: in the field of water analysis. AccredQualAssur 15:251–254, Doi: 10.1007/s00769-010-0635-3
[108]. Chung SWC, Tran JCH, Wong WW (2012) Capability of laboratories to determine core nutrients in foods: results of an international proficiency test. AccredQualAssur 17:35–44, Doi: 10.1007/s00769-011-0833-7
[109]. Agbenin JO, Cantarella H (2011) Performance of commercial soil laboratories in a proficiency test program in Brazil. AccredQualAssur 16:553–559, Doi: 10.1007/s00769-011-0814-x
[110]. Chelani AB, Moghe CA, Nimsadkar S, Gandhi K, Bodhe GL, Dhopte SM et al (2011) Evaluation of bias, precision, and systematic errors in proficiency testing of Cl2 and Cu concentration in water. AccredQualAssur 16:379–382, Doi: 10.1007/s00769-011-0782-1
[111]. Pankratov I, Elhanany S, Henig S, Zaritsky S, Ostapenko I, Kuselman I (2010) Development of a proficiency testing scheme for a limited number of participants in the field of natural water analysis. AccredQualAssur 15:459–466, Doi: 10.1007/s00769-010-0673-x
[112]. Aslan B, Gun-Munro J, Flynn GJ (2012) Role of proficiency testing in monitoring of standardization of hemoglobin A1c methods. AccredQualAssur, Doi: 10.1007/s00769-012-0907-1
[113]. Whetton M, Baryla K, Finch H (2012) The effect of the choice of method for determining assigned value on the assessment of performance in melting point analysis. AccredQualAssur, Doi: 10.1007/s00769-012-0908-0
[114]. Mancin M, Grimaldi M, Barco L, Trevisan R, Ruffa M, Mioni R et al (2012) Proficiency testing in food microbiology: experience from implementation of ISO/IEC 17043 and ISO/TS 22117. AccredQualAssur, Doi: 10.1007/s00769-012-0904-4
[115]. Múrtula R, Soria E, Yáñez MA, Catalán V (2012) Proficiency testing schemes for the assessment of Legionella PCR methodologies. AccredQualAssur, Doi: 10.1007/s00769-012-0903-5
[116]. Cordeiro F, Robouch P, Linsinger T, De La Calle B (2012) Determination of brominated flame retardants: a proficiency test. AccredQualAssur, Doi: 10.1007/s00769-012-0900-8
[117]. Mbwambo K, Koch M (2012) Establishing PT schemes in developing countries: examples from Africa. AccredQualAssur, Doi: 10.1007/s00769-012-0893-3
[118]. James VLA (2012) Evaluating participant performance in qualitative PT/EQA schemes. AccredQualAssur, Doi: 10.1007/s00769-012-0895-1
[119]. Lehmann C (2012) Accrediting PT/EQA providers to ISO/IEC 17043. AccredQualAssur, Doi: 10.1007/s00769-012-0892-4
[120]. Gun-Munro J (2012) The challenges and benefits of implementing the requirements of ISO/IEC 17043 by PT/EQA providers. AccredQualAssur, Doi 10.1007/s00769-012-0887-1

[121]. ISO 5725-1 (1994). Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions.
[122]. ISO 5725-2 (1994). Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method
[123]. ISO 5725-4 (1994). Accuracy (trueness and precision) of measurement methods and results — Part 4: Basic methods for the determination of the trueness of a standard measurement method.
[124]. ISO Guide 34 (2009). General requirements for the competence of reference material producers.
[125]. ISO 21748 (2010). Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation.
[126]. ISO/TS 22117(2010). Microbiology of food and animal feeding stuffs - Specific requirements and guidance for proficiency testing by interlaboratory comparison.
[127]. ISO 13528 (2005). Statistical methods for use in proficiency testing by interlaboratory comparisons.
[128]. ISO/TS 20612(2007). Water quality - Interlaboratory comparisons for proficiency testing of analytical chemistry laboratories.
[129]. ISO 15189 (2007). Medical laboratories - Particular requirements for quality and competence.
[130]. ISO/IEC 17011 (2004). Conformity assessment - General requirements for accreditation bodies accrediting conformity assessment bodies.
[131]. ISO Guide 35(2006). Reference materials - General and statistical principles for certification.
[132]. ISO/TR 22971 (2005). Accuracy (trueness and precision) of measurement methods and results - Practical guidance for the use of ISO 5725-2:1994 in designing, implementing and statistically analysinginterlaboratory repeatability and reproducibility results.
[133]. ASTM E2489-11. Standard Practice for Statistical Analysis of One-Sample and Two-Sample Interlaboratory Proficiency Testing Programs.
[134]. ASTM E2027-09 Standard Practice for Conducting Proficiency Tests in the Chemical Analysis of Metals, Ores, and Related Materials.
[135]. ASTM D6674 - 01 Standard Guide for Proficiency Test Program for Fabrics.
[136]. ASTM D6792-07 Standard Practice for Quality System in Petroleum Products and Lubricants Testing Laboratories
[137]. ASTM D7372-07 Standard Guide for Analysis and Interpretation of Proficiency Test Program Results :
[138]. IUPAC (2006). The international harmonized protocol for the proficiency testing of analytical chemistry laboratories, Pure Appl. Chem. 78: 145 – 196.
[139]. IUPAC/CITAC Guide (2010). Selection and use of proficiency testing schemes for a limited number of participants—chemical analytical laboratories (IUPAC Technical Report).
[140]. EUROLAB (2006). Technical Report 1/2006: Guide to the Evaluation of Measurement Uncertainty for Quantitative Results http://www.eurolab.org/documents/1-2006.pdf Accessed 3 Mar 2013.
[141]. EUROLAB (2007). Technical Report: Measurement uncertainty revisited: Alternative approaches to uncertainty evaluation. http://www.eurolab.org/documents/1-2007.pdf Accessed 20 Mar 2013.
[142]. NORDTEST (2003). Technical Report 537: Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories. www.nordicinnovation.net/nordtest.cfm Accessed 22 Mar 2013.
[143]. BIPM, IEC, IFCC, ILAC, IUPAC, ISO, OIML (2008). Guide to the Expression of Uncertainty in Measurement – GUM. JCGM 100, http://www.bipm.org/en/publications/guides/gum.html Accessed 22 Mar 2013.
[144]. EA-3/04 (2001). Use of proficiency testing as a tool for accreditation in testing. http://www.european-accreditation.org/publication/ea-3-04-g Accessed 19 Mar 2013.
[145]. EA-4/18 (2010) Guidance on the level and frequency of proficiency testing participation.http://www.european-accreditation.org/publication/ea-4-18-ta Accessed 18 Jan 2013.
[146]. ILAC (2010). P9: ILAC Policy for Participation in Proficiency Testing Activities https://www.ilac.org/documents/ILAC_P9_11_2010.pdf Accessed 11 Jan 2013.
[147]. IAAC (2013). MD 038/13: Application for the Assessment and Accreditation of Proficiency Testing Providers.http://www.iaac.org.mx/Documents/Controlled/Mandatory/MD%20038%20ACCREDITATION%20OF%20PTP.pdf Accessed 17Mar 2014.
[148]. APLAC (2008). PT 001. Calibration Interlaboratory Comparisons. Issue No. 5.
[149]. APLAC (2010). PT 003. Proficiency Testing Directory. Issue No. 15.
[150]. APLAC (2010). PT 005. Artefacts for Measurement Audits. Issue No. 2.
[151]. APLAC (2010). PT 006. Proficiency Testing Frequency Benchmarks. Issue No. 2.

ARTIGO 2 – Analysis of the relationships between Proficiency Testing, Validation of Methods and Estimation of Measurement Uncertainty: a qualitative study with experts

Artigo aceito para publicação no Accreditation and Quality Assurance Journal for Quality, Comparability and Reliability in Chemical Measurement. ISSN 0949-1775.

Qualis Cappes: B1

Analysis of the relationships between Proficiency Testing, Validation of Methods and Estimation of Measurement Uncertainty: a qualitative study with experts

Abstract: Proficiency Testing (PT) is an important part in the process ensuring quality of results. Validation of methods and estimation of measurement uncertainty are also vital elements in a laboratory qualification process. The main objective of this paper is to identify the relationships between PT, validation of methods and calculation of measurement uncertainty through the drawing up of a Mind Map. The method proposed used the web-based Delphi method to conduct a qualitative research with experts of the area (laboratory assessors and accreditation body members). Experts from five accreditation bodies (Brazil, U.S.A., France, Portugal and China) took part in the research. The study led to the drawing up of a Mind Map showing the connections of cause and effect between the factors under study. It emerged that PT effectively contributes to the reliability of results, since it is directly related to the validation of methods and to measurement uncertainty, so as to increase laboratory's competitiveness.

Keywords: Proficiency Testing, validation of methods, measurement uncertainty, qualitative study, web-based Delphi.

Introduction

Proficiency Testing (PT) is an important tool, helping to ensure reliability of tests and calibrations. PT is a programme for comparisons of results within a group of laboratories, aiming at assessing technical skills to carry out a test or calibration method [1].

Laboratories wishing to receive accreditation for their tests or calibrations must take part in PT, validate their analytical methods and perform the estimation of measurement uncertainty. PT participation is a prerequisite to apply for accreditation assessment. With this kind of requirement set by accreditation bodies, it is possible to observe that PT is a necessary part in the routine of laboratories [2]. It is worth noting that laboratories have difficulties related to matters such as validation of their analytical methods and calculation of measurement uncertainty, these being some of the most delicate aspects to meet international regulations criteria [3].

Laboratory qualification does not only limit to private service providers, not even to a specific metrological area [4]. There are advantages for a company to implement the ISO/IEC 17025 standard, such as improvement in the laboratory's image and in the reliability of results, detection of non-conformities in the management system of the company through internal quality checks and PT participation [5]. On the other hand, costs related to this process cannot be neglected and caution needs to be taken not to increase bureaucracy in the company management. Another positive aspect of the laboratory assessment process is the

participation in external quality checks, which are fundamental for enhancing the reliability of results, such as PT [2].

Laboratory method validation can be interpreted as the confirmation, through objective evidence, that specific measurement requirements are met for an intended use [6]. The main source of error in a laboratory measurement process may arise from sampling, environmental conditions, equipment, handling of testing items or calibration, analysts or from the method itself [7]. Therefore, it is important to validate every stage of the test or calibration in the laboratory environment. Furthermore, this validation can be performed externally and/or internally. External method validation can be supported by PT participation [3].

Literature reveals several studies related to PT, validation of methods and uncertainty calculation, but these issues are treated mostly independently [1, 3, 8, 9; 10, 11, 12; 13, 14]. In this context, the laboratory seeking qualification ends up repeating its efforts, since it calculates uncertainty itself, carries out validation of its analytical method and takes part separately in PT. This fact is relevant since it involves a high workload and its related costs.

In this scenario, the following research questions are presented: which are the relationships between proficiency testing, validation of methods and measurement uncertainty and how do these activities contribute to a reliable measurement result?

The objective of this work is to identify the relationships that exist between proficiency testing, validation of methods and calculation of measurement uncertainty. Furthermore, the specific goal is to investigate how these activities contribute to a reliable measurement result, by showing data in a Mind Map. The study is conducted taking into consideration the point of view of experts in this area.

Methodology

The work herein presented is an applied research with a qualitative approach. As far as its objectives are concerned, this is an exploratory study using as a technical procedure a case study applied through research conducted with laboratory assessment experts. The method proposed is based on five stages, described in Table 1. Data collection was performed through the web-based Delphi method, a structured communication technique, which relies on a panel

of experts. The experts answer questionnaires in two or more rounds related to a research topic or problem and the interaction are made using web tools (e-mails, chats, web conference) [15].

Table 1 – Description of the stages of the proposed method

Stages	Description
Selection of Experts	Selecting experts in laboratory assessments with at least 5 years of experience and working in the testing or calibration area, or in the assessment process management of accreditation bodies. The initial stage was developed with 8 Brazilian experts, since the origin country of the research is Brazil. In the end, 4 more international experts took part in the study.
1st Round: Exploratory Analysis	Drafting open questions focused on the study purposes. Questions must be comprehensive, so that participants can discuss the topic. Questions were submitted by email, using the web-based Delphi method. The questionnaire was sent individually to each selected expert, who was asked to reply within a specific deadline.
2nd Round: Content Analysis	After receiving replies a content analysis was performed in order to group replies in tables according to criteria established by the researchers. Tables with replies to the specific questions were then submitted by email to the participants. They were asked to confirm their replies or to complement them, if necessary. In this round, the researcher was able to engage the participants to complement a specific question or to explain doubts that might have arisen in the 1st Round.
3rd Round: Mind Map drawing up	Analysing data consolidated in the 2nd Round and verifying the cause-and-effect relationships between the topics presented and the criteria used in the previous round. Drawing up a Mind Map in order to show the non-linear relationships that may clarify the understanding of the research topic and the relationships between PT, validation of methods and estimation of measurement uncertainty.
4th Round: Confirmation of the Mind Map	Submitting the Mind Map created to the Brazilian experts, explaining its logic. Experts will be asked to evaluate the diagram, confirming or not the proposal. In this stage, participants can also complement the diagram. The latter was then modified according to the experts' replies. Lastly, sending the Mind Map to experts in laboratory assessment and/or proficiency testing from 4 different countries, who are members of Accreditation Bodies signatories to the ILAC recognition arrangement. The objective of this stage was to confirm the Mind Map drafted with professionals who have different perspectives from Brazilian experts. After this stage, the researchers analysed the final diagram and consolidated it. The final interpretation of the cause-and-effect relationships presented in the Mind Map was then carried out.

The method proposed was to collect data by email, since it is an effective means for gathering information. The limitations of synchronous communication were disregarded. The fact of participants not being present during data collection was not prejudicial, since the

method proposed involves four rounds until the final confirmation of opinions, ideas and concepts, thus offering several interaction opportunities.

Results

Selection of experts and questionnaire drafting

Initial rounds (from 1st to 3rd) of the case study were carried out with 8 Brazilian experts in laboratory assessment. Two of them participate in the accreditation process management. The other six experts work as laboratory assessors (three in the calibration area and three in the testing one).

The experts selected have 5 to 16 years (median of 12.5) of experience with laboratory assessments. 87.5 % of the selected professionals have an engineering background, while 12.5 % are chemistry undergraduates. 87.5 % of the group has a postgraduate degree, 62.5 % of them holding a doctoral degree.

The open questionnaire included two open questions related to the research topic focused on the objectives of this work, as follows:

1) Which is the importance of laboratories' participation in Proficiency Testing?
2) Which is the relationship between proficiency testing, validation of methods and measurement uncertainty calculation in your work as an assessor?

Open and broad questions were chosen to engage the experts to talk about the topic and explain their opinions and ideas on it. Furthermore, questions led the participants to discourse about the link or connection between the items under discussion, demonstrating the possible cause-and-effect relationships between the factors studied. It is worth noting that the questionnaire had no line limitations and that its introduction motivated the respondent to elaborate complete answers.

Data collection and analysis

Data collection and analysis were performed in four stages, being directly related to the issues mentioned initially. The analysis of the first question was carried out in two steps. First, the expressions mostly used in the first part of the answers were analysed and then the

topics presented were grouped and classified by similarity according to different criteria. As for the reply to Question 1 (Which is the importance of laboratories' participation in Proficiency Testing?), it was possible to observe that there were differences in the expressions used, even though they were considered similar. According to the expressions used in the replies ("Fundamental", "Utmost importance", "Best and most efficient way to assess laboratory performance", "important"), it was possible to conclude that experts converge on this aspect.

The second part of the answers to Question 1 and answers to Question 2 were analysed and the main contents were identified. Criteria were set up in order to classify them and conduct a general analysis on the perceptions of the experts consulted. Four criteria were identified, as follows: reliability, performance improvement, assessment of results and competitiveness.

After grouping answers according to the criteria, they were submitted again to the experts by email, this representing the second round of the web-based Delphi method. 2 out of the 8 experts complemented their answers while 6 left them unaltered, as presented in the 1st round. The mechanism referring to the confirmation of answers was considered to be relevant since they could have been improved if necessary.

A connection of cause and effect was observed between the contents of the answers provided. In general, the importance of PT is related to the fact that they are necessary to assess results issued by laboratories.

This assessment leads those involved in PT activities to improve their performance as far as conduction of tests or calibrations is concerned. This action, in turn, increases reliability of results, this being the technical ability the laboratory has in order to adequately conduct tests or calibrations. Furthermore, the mentioned condition is required in an accreditation process, thus strengthening the need for PT.

Performance improvement is related to the fact that the laboratory can analyse its results and deepen the knowledge about its analytical method, take corrective and preventive actions if deemed necessary, check for tendencies in its results and act in its process so as to improve its routine. Improvements must be carried out through structured plans that may ensure quality assurance of results issued.

Assessment of results is related to the performance improvement. In PT, laboratories can compare themselves with bodies that perform the same activities and even with their competitors. In this context, it is necessary that PT participants conduct a careful analysis of their participation, so as to technically perfect their performance. There is no point in a company participating in a PT without carrying out an adequate assessment of the results from the comparison rounds, otherwise even corrective or preventive unnecessary actions cannot be taken.

One of the consequences of PT participation is the increase in reliability of the participating laboratory, since it improves its performance through an assessment of PT results and implementation of improvements. This may influence its competitiveness on the market since in specific areas PT participation is compulsory and is assessed by government bodies. An example is the case of requirements in regulations demanding qualified laboratories. These regulations are laws that cannot be neglected.

Upon analysis of the content in the experts' answers, it was possible to observe a direct relationship between PT, validation of methods and measurement uncertainty calculation. Most of the experts consider that the topics are related and that validation of methods can be confirmed in PT, as well as the laboratory measurement uncertainty calculation, since in these activities results are compared with other kinds of companies performing the same test or calibration.

Other approaches have shown the possibility to calculate measurement uncertainty through data obtained from PT rounds, although the International Standard ISO/IEC 17043 states that this is not a valid use of PT. It is important to stress that laboratory uncertainty and its validation also influence its performance in PT; in fact, if these activities are inadequately carried out, it is likely that the laboratory PT result will reflect this situation, in other words, it may lead to an unsatisfactory or questionable result.

Different experts also stated that PT is a relevant tool to provide Quality Assurance of results, since it assesses accuracy and laboratory data trend. It is worth observing that the ISO/IEC 17025 standard addresses these factors in their technical requirements, PT being quoted in the Quality Assurance section, item 5.9 of the mentioned standard. As a consequence, these tools help obtaining analytical reliability.

Mind Map drawing up

After carrying out content analysis, presented above, a Mind Map was drawn up to summarise and give a graphical overview of the non-linear relationships and the logic of relationships between the items studied in this research. Figure 1 presents the map drawn up.

The eight Brazilian experts who took part in the study received the diagram by email for assessment. All of them considered the diagram as adequate and that it represents the relationships between the factors under study. Two experts suggested modifications, such as the inclusion of "Education labs" in the "PT participants" item and the exclusion of one of the relationships between "interested parties" and "PT participants", since there were two arrows in opposite directions.

In the fourth and last round of the method, the Mind Map was also sent to international experts belonging to other accreditation bodies. The Portuguese Institution for Accreditation (IPAC), the American Association for Laboratory Accreditation (A2LA), the French Committee for Accreditation (COFRAC) and the China National Accreditation Service (CNAS) took part in this stage. Each body appointed an expert in laboratory assessment to critically evaluate the Mind Map, which was submitted by email. The CNAS expert was also member of the PT Committee of the Asia Pacific Laboratory Accreditation Cooperation (APLAC). This stage lasted approximately 1 month. Experts made few adjustments, such as adapting nomenclature (concerning the term "accuracy"), including the Standardised Organisations and proposing the inclusion of another direct relationship (between the validation of methods and the estimation of measurement uncertainty boxes). Some experts provided a positive feedback on the work, such as:

- The diagram relates to how PT benefits the related parties. It is the first time for me to see that such a comprehensive map covers the various points scattered in different places and demonstrate the inter relationships between them. It is very good. (CNAS expert);
- The Mind Map developed correctly addresses the general relationships and interfaces between PTs, Method validations, Uncertainty estimations and the interested parties. (COFRAC expert);
- This is wonderful! Nice job capturing the various roles, results and perspectives (A2LA expert).

According to Figure 1, it is possible to realise that PT may generate data for the estimation of measurement uncertainty and for validation of methods and both of them can be verified by the laboratory when it takes part in PT. Validation and PT also generate data and information on the accuracy of methods that may support the estimation of uncertainty and the assessment of PT results, which is considered as the critical analysis of the laboratory performance.

This analysis, when adequately performed, leads the laboratory to take corrective and preventive actions in its routine, bringing about a performance improvement. This action is the basis for analytical reliability. Currently, laboratory reliability is one of the relevant factors for it to ensure competitiveness over its competitors and on the market, adapting to government and accreditation bodies' requirements.

Interested parties include clients, government, standardisation organisations and bodies that perform conformity assessments, represented in violet in Figure 1. These bodies have direct influence on the PT provider, since they assess the provider according to the standards' requirements, such as ISO/IEC 17043. Furthermore, these bodies can also be PT providers themselves.

Accreditation bodies also demand for reliability of results from laboratories that are accredited and assessed according to the ISO/IEC 17025 standard. It is important to underline that the accreditation body can assess laboratories from the most diverse areas (research, services, industry, among others). Moreover, PT participation is compulsory for laboratories that are accredited or going through the accreditation process.

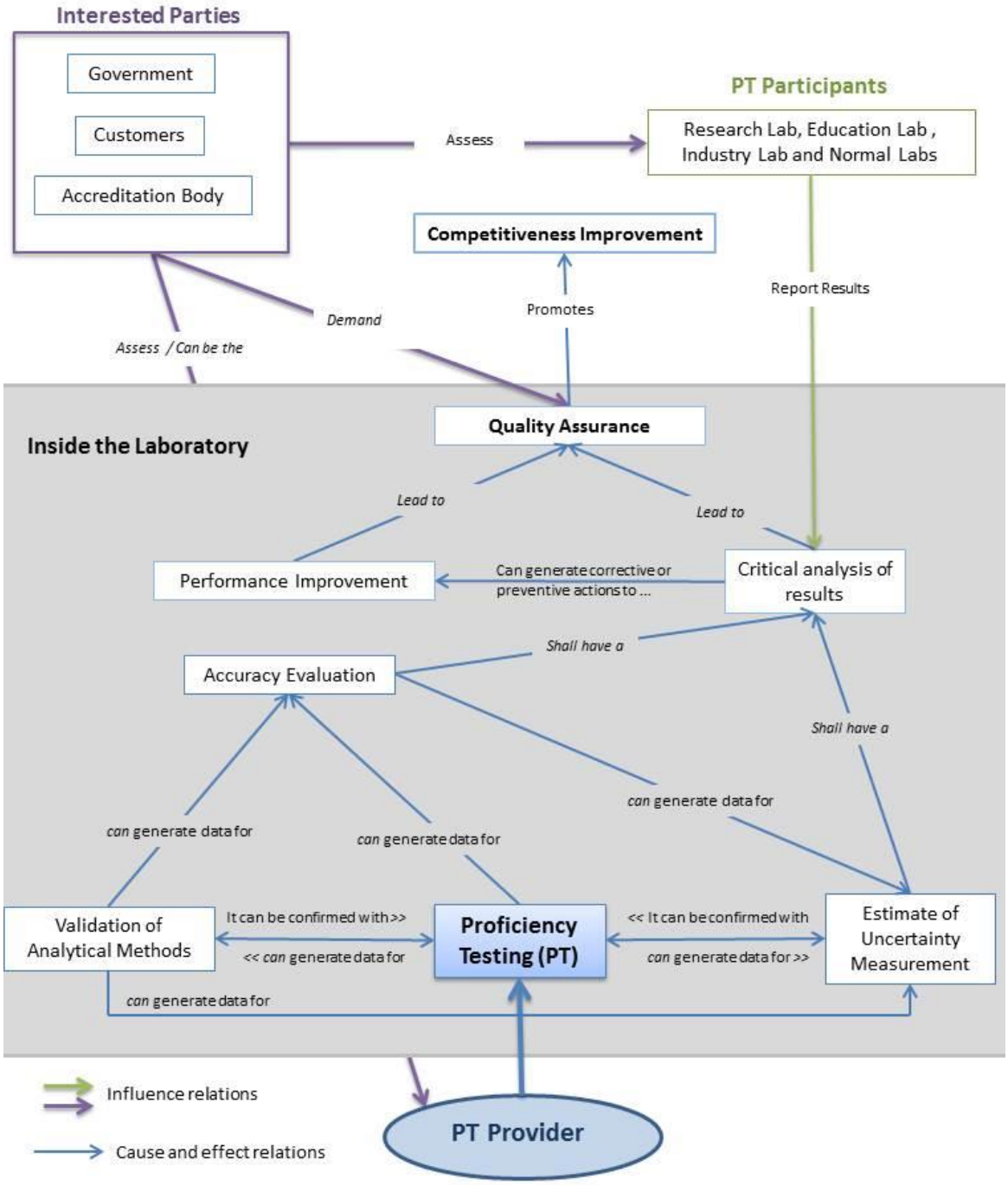

Figure 1 – Mind Map

PT participants have a direct relationship in the stage of "Critical Analysis of Results" since, in reports published by PT providers, the performance and tests or calibrations from all involved in the comparison process appear. Normally, laboratories are not identifiable and their confidentiality is guaranteed by passwords. PT participants have the possibility to

compare results with their competitors, partner laboratories, universities, among others, to check the adequacy of their measurement system and to compare results from different measurement methods.

In view of the above, the relationships presented in the Mind Map of Figure 1 are justified, leading to a better understanding of PT contributions to the reliability of a result issued by a laboratory. Experts also pointed out that the Figure showed connections that many times are not clear to laboratories and not even to assessors, and that they were well represented by the diagram.

Web-based Delphi method has proven to be suitable for conducting this study, since it allowed researchers to simultaneously collect information from all experts through email contact, apart from facilitating registration of data in the questionnaires. The rounds performed were important because they confirmed the perceptions and allowed participants to complement their answers, when necessary.

Moreover, Brazilian experts participated from the beginning to the end in the research and showed their interest in the final result. One of the experts made comments on the feedback provided during the study, which was positive and compensated the efforts deployed to answer the questionnaire. On the other hand, one of the weak points of this system was the fact that not always participants' answers were immediate and many times the researcher had to put pressure on them to receive questionnaires within the established deadline.

Conclusions

The goal of this work was to identify the relationships existing between PT, validation of methods and measurement uncertainty calculation. This goal was achieved because, through the method applied, the relationships between the mentioned factors, based on the experts' opinions, were verified. The specific goal aimed at understanding how PT, validation and measurement uncertainty contribute to a reliable measurement result. This contribution was evident in the Mind Map: through a cause-and-effect connection between the factors studied it was possible to observe that analytical reliability was obtained, and laboratory competitiveness could also be promoted.

Suggestions for future studies include the confirmation of the Mind Map developed with other experts from different accreditation bodies signatories of the ILAC arrangement that were not consulted for this research.

References

[1]. ISO/IEC 17043 (2010) Conformity assessment—general requirements for proficiency testing. International Organization for Standardization, Geneva.

[2]. Tholen D (2011) Metrology in service of society: the role of proficiency testing. Accred Qual Assur 16:603–605. doi:10.1007/s00769-011-0836-4

[3]. Albano, F, Caten, C (2014). Proficiency tests for laboratories: a systematic review. Accred Qual Assur 19:245–257. doi 10.1007/s00769-014-1061-8

[4]. Grochau, I. Ferreira, Caten, C (2012). A process approach to ISO/IEC 17025 in the implementation of a quality management system in testing laboratories. Accred Qual Assur. 17: 519-527. doi 10.1007/s00769-012-0905-3

[5]. Tawfik, H. Fatah, A (2011). ISO/IEC 17025 Accreditation: Between the Desired Gains and the Reality. Quality Assurance Journal. 13: 21–27 doi: 10.1002/qaj.465

[6]. ISO/IEC 17025 (2005) General requirements for the competence of testing and calibration laboratories. International Organization for Standardization, Geneva.

[7]. Brezis, M (2011). Warning: false sense of certainty from the illusion of accuracy and precision in measurements. Accred Qual Assur 16: 599–602 doi 10.1007/s00769-011-0821-y

[8]. United Kingdom Accreditation Service (UKAS). M3003: The Expression of Uncertainty and Confidence in Measurement. UKAS, 2. ed. www.ukas.com/Library/downloads/publications/M3003.pdf. Accessed 22 Mar 2014

[9]. Coordenação Geral De Acreditação (CGCRE). Orientação para a acreditação de laboratórios, produtores de materiais de referência e provedores de ensaio de proficiência. file:///C:/Users/falbano/Downloads/DOQ-Cgcre-1_10%20(2).pdf Accessed 22 Mar 2015

[10]. Jardim, W. Sodré, F (2009). Desempenho analítico de laboratórios prestadores de serviço na determinação de metais em águas. Quimica Nova. 32-4: 1083-1088 On-line version ISSN 1678-7064.

[11]. Eurachem/CITAC Guide (2007). Use of uncertainty information in compliance assessment. www.eurachem.org

[12]. European Federation of National Associations of Measurement, Testing and Analytical Laboratories (EUROLAB). Measurement uncertainty revisited: Alternative approaches to uncertainty Evaluation. Technical Report No. 1. March, 2007.

[13]. Hauswaldt, A; Rienitz, O; Jährling, R; Fischer, N; Schiel, D; Labarraque, G; Magnusson, B (2012). Uncertainty of standard addition experiments: a novel approach to include the uncertainty associated with the standard in the model equation. Accred Qual Assur 17:129–138 doi 10.1007/s00769-011-0827-5

[14]. Willink R (2006) Forming a comparison reference value, from different distributions of belief. Metrologia. doi:10.1088/0026-1394/43/1/002

[15]. Okoli, C; Pawlowski, S (2004). The Delphi method as a research tool: an example, design considerations and applications. Information & Management 42: 15–29. doi:10.1016/j.im.2003.11.002

ARTIGO 3 - Management analysis of proficiency testing projects developed by Brazilian accredited providers

A) Artigo publicado no 8th WORKSHOP PROFICIENCY TESTING IN ANALYTICAL CHEMISTRY, MICROBIOLOGY AND LABORATORY MEDICINE - Berlin, Germany, 6-9 October 2014.

B) Artigo publicado na Revista Qualidade Banas, v. Ano XV, p. 76-91, 2015.

Management analysis of proficiency testing projects developed by Brazilian accredited providers

Abstract: Proficiency Testing (PT) schemes are developed by providers, who seek their recognition through accreditation to the ISO/IEC 17043 standard. Other guides and standards, such as PMBoK and ISO10.006, could also be used to manage aspects of the development of PT projects. In this context, the main goal of this study is to analyze the management of projects developed by PT accredited providers in Brazil. The research considered all accredited PT providers in the country (12 organizations), from testing and calibration areas. The results indicate that the providers declared knowledge level of the ISO/IEC 17043 is higher and significantly different than that of PMBoK and ISO10.006. PT providers use some important knowledge areas related to project management, but not all of them. It was identified that risk and cost management are the main areas for improvement. It was found that the PT projects become processes after they have been successfully developed by the providers. Also, most of the Brazilian accredited PT providers (83%) consider themselves as being mature in PT schemes development.

Keywords: Proficiency testing, project management, provider.

Introduction

Laboratories that are ISO/IEC 17025 accredited must seek a continuous improvement process and the reliability of their results [1]. According to Tawfik and Fatah, obtaining accreditation is simpler than keeping it over time [2]. In order to be qualified enough to obtain and maintain this third-party acknowledgment, laboratories must enroll in a Proficiency Testing (PT), to identify if they are technically competent in the tests and calibrations performed.

One of PT objectives can be to support test and calibration laboratories at assuring their reliability [3]. Through PT it is also possible to assess the performance of laboratories concerning tests or specific measurements, to identify analytical problems, to compare testing or calibration methods, to qualify participants based on results of interlaboratory comparisons and to validate the declared uncertainty [14].

In every PT, is relevant to define the program development, operation and goals, determining possible participants, samples and program frequency [8]. PT schemes may vary according to the needs of the sector in which they are used, PT items, methods being used and number of participants. The schemes can be classified in five different types: sequential, simultaneous, interpretative, sample review and split samples [4].

In Brazil, market demands for PT are higher than the current supply [6]. It is relevant to mention the great diversity of existing tests and calibrations areas, which makes the development of several types of PT possible. From 2010 on, CGCRE/INMETRO, the Brazilian accreditation body, developed an accreditation program for PT providers. This

accreditation is carried out based on the ISO/IEC 17043 standard, which refers to technical and managerial issues connected to the conformity assessment of PT providers.

According to Brazilian researches, there is a tendency of growth in the quantity of laboratories with satisfactory results in the PT of microbiological analyses of Rede Metrológica RS, a Brazilian PT provider [7]. In PT reports published by the same provider, it is also evident that laboratories have a tendency of improving their analytic performance over time in different areas (environmental water analyses, physico-chemical tests in wines and cachaça, coal analyses, among others) [5].

Due to the increasing demand for PT schemes in the country, providers face the need to develop projects to structure the supply of new services for the market, in addition to seek their qualification through accreditation [6]. Unfortunately, the standard related to PT schemes accreditation (ISO/IEC 17043) does not consider all areas that are typical from project management [1].

In addition, PT projects are activities which require some detailing, due to their complexity – a factor that justifies special attention to develop a project of a new PT scheme [14]. There are some benefits in Project Management practices, such as a good control (the project's progress can be measured through comparison with defined performance goals and standards) and greater flexibility (since it can employ or add experts for limited periods) [9].

A project is a temporary effort employed to create a service, a product or a certain result [10]. For this reason, it has a deadline and a result that is different from a routine. The management of a project deals with the execution of five managerial processes: initiation, planning, execution, monitoring and control, and closure [10].

In order to implement projects, the Project Management Institute (PMI) proposes, along the steps previously mentioned, the integrated action of nine knowledge areas: scope management, time, cost, quality, human resources, communication, risks, acquisitions, and integration. These areas are usually managed by the project manager, who is the responsible for carrying out tasks connected to the project development [10].

Project management practices have been developed in different areas. There are researches that relate project management with the ISO9001 standard. It was seen that companies which got certification based on this standard have better performance in their projects when compared to non-certified companies. In these companies it was noticed the importance of the

communication processes for obtaining an appropriate project [11]. Other authors also address good practices of project management with the communication among stakeholders [12].

Another world-renowned reference is ISO 10006 standard, which deals with quality management systems focusing on guidelines for quality management in projects. This standard is applicable to projects of varying complexity, small or large ones, of short or long-term duration, in different environments and regardless the type of product or process involved [13]. This standard only gives orientation about project development and it is not used in accreditation or certification processes.

There are the following research questions in the PT and Project Management subjects: How is the management of PT projects developed by Brazilian accredited providers? Do the providers use recommendations of project management standards to develop their activities?

The general objective of this paper is to analyze the management of PT projects of Brazilian accredited providers. The specific objectives are: to assess the declared knowledge of providers concerning the ISO 10006, ISO/IEC 17043 standards and the PMBoK Guide of PMI (i); to assess the common areas among the references mentioned above (ii); and to assess the opportunities for improvement of the management of PT developed in Brazil (iii).

Method

This research was an exploratory work, where a case study was applied through a survey with Brazilian PT providers. This work has a quantitative approach. The method proposed is based on 4 steps, presented in Table 1.

Table 1 – Method phases

Phases		Description
1	Identification of accredited providers	Identification of Brazilian PT providers that are accredited by CGCRE/INMETRO. The research was conducted in the website http://www.inmetro.gov.br/credenciamento/acre_prod_ep.asp
2	Questionnaire preparation and submission	Preparation of the questionnaire to be sent to PT providers with questions related to the research's objective. Items included in the identification of the provider profile: number of employees, experience time, company setting (public or private) and operating area. Questions that were made to the providers: a) How does your PT provider declare your knowledge about ISO/IEC 17043, ISO 10006 and PMbOK Guide? Responses were registered on a Likert scale from 1 to 5, where 5 is high stated knowledge.

		b) Which of these knowledge areas do you consider that are included in the ISO/IEC 17043? The answers could be one or more of the 9 areas of the PMBOK Guide. c) Which of these knowledge areas do you use in projects related to PT schemes developed by your company? The answers were register as described in question "b". d) Do you consider the PT developed by your company as being projects? There were 3 options for answer the question: "Yes", "No" and "Partly". When "Partly" was chosen as a response, the provider needed to justify the choice. e) Do you think your business is mature in PT projects development? Why? The answers were done in a free text format. The questionnaire was submitted via e-mail in .doc format to technical or quality manager of the providers identified in the previous step.
3	Result analysis	For data analysis we used descriptive statistical methods in Microsoft Excel software. Statistical analysis (Kruskal-Wallis test) were performed using the Statistical Package for Social Sciences (SPSS). Comparison between the requirements of ISO/IEC 17043, ISO 10006 and PMbOK Guide were also done in a cross check table.
4	Conclusions	General discussion of the results, identifying possible opportunities of improvement for the management of PT projects. Analysis of relationships existent between the researched items, discussing the characteristics of the management carried out by Brazilian PT providers.

Results

Profile analysis of Brazilian accredited PT providers

The research was carried out in 2013, where 12 companies were consulted. It represented 100% of the Brazilian accredited providers. All of them answered all the questions of the questionnaire.

The first part of the questionnaire presented questions about the PT provider's background, aiming to identify its experience, number of employees, academic background, operating area, and company type (public, private or non-profit private). It was noted that the Brazilian providers has a median experience of 10 years (from 2.5 to 35). The number of employees also was variable (from 2 to 100, with median of 4.5). Major of providers (58.3%) are from private nonprofit organizations. 25% of the providers are from public areas and 16.7% are private normal companies.

Most of the companies that provide PT in Brazil and are ISO/IEC 17043 accredited works in the testing area (75%). Only 8.3% works with testing and calibration PT schemes and 16.7% are focused in calibration area.

Regarding academic background, we noticed that this activity is intellectually demanding. This can be hold by the fact that 41.7% of the PT providers have staffs holding a

doctoral or master's degree completed or in progress. 16.7% have at least MBA course and 41.7% have only graduation complete.

Analysis of declared knowledge about PMBoK, ISO 10006 and ISO/IEC 17043

The following step consisted in performing an analysis about the level of declared knowledge of PT providers in relation to the project management guide PMBoK, ISO 10006 and ISO/IEC 17043 standards. Figure 1 presents the average of the responses.

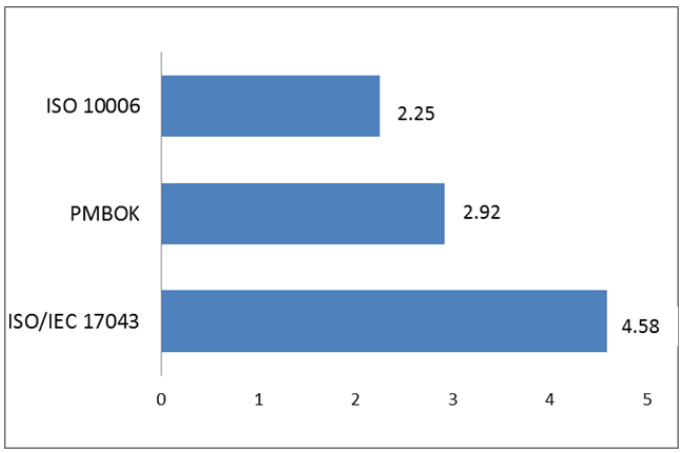

Figure 1 – PT providers declared knowledge of ISO/IEC 17043 and existing guidance for project management, expressed as the average of responses according to the Likert scale (0.0, low; 5.0, high)

It was possible to notice that the declared knowledge in relation to the ISO/IEC 17043 standard was high (4.58 average), which was something that did not occurred in the other cases. Kruska-Wallis test results indicated significant difference, since the *p-value* was 0.0007 (alpha equal to 0.05). Post hoc test was applied (alpha equal to 0.05), in order to check which factor presented significant differences from the others. This evaluation showed that the PT providers declared knowledge about PMBoK and the ISO10006 standard have no significant difference (*p-value* > 0,05). This means that the knowledge concerning this standard and this guide are equally low and significantly different from the declared knowledge about the ISO/IEC 17043 standard. Thus, it was also possible to notice that all PT providers declared to have high knowledge of the ISO/IEC 17043 standard and equally low knowledge of the ISO10006 and the PMBoK.

Analysis of project management areas used in PT schemes in Brazil

The Brazilian accredited PT providers also responded questions regarding the project knowledge areas they considered that the ISO/IEC 17043 standard approaches and which ones they actually use when developing their PT schemes.

The results of this topic are presented in Figure 2. The nine PMBoK knowledge areas were included as a way to check if they are considered in PT development. The area "strategy" was quoted in the research as the alignment of projects with the company's strategic planning (connection with its objectives, vision, values and mission). This topic was included in the research because it is also important in project development, but "strategy" is not one of the nine knowledge areas of PMbOK.

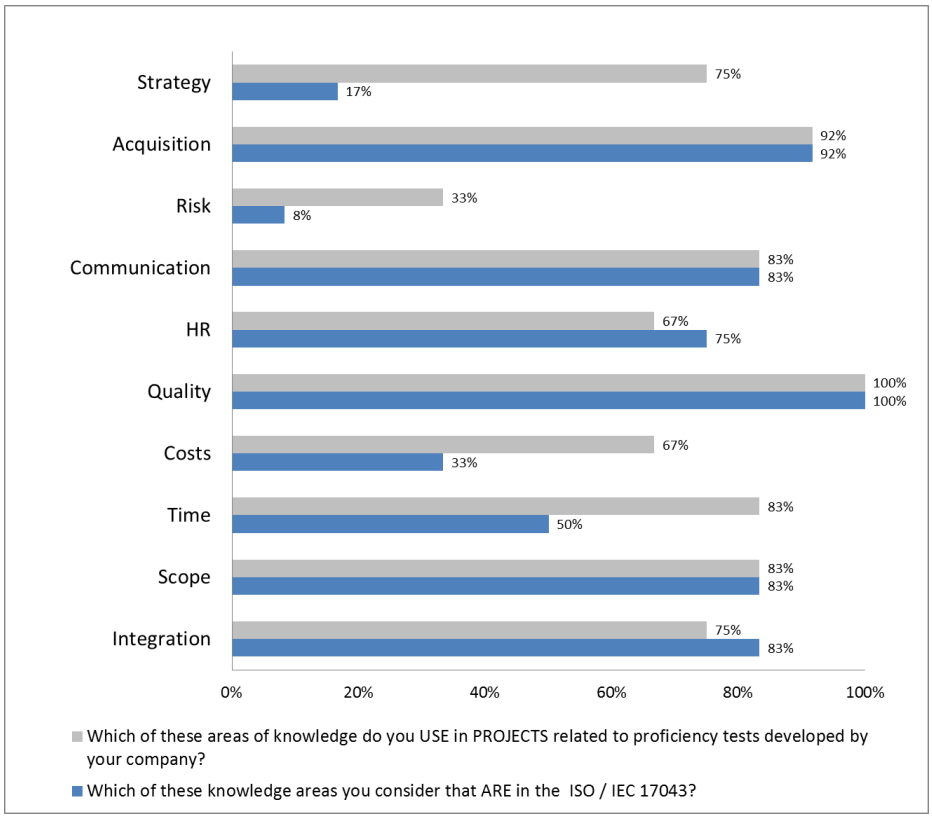

Figure 2 – Project Management areas and the relation with ISO/IEC 17043

It was possible to see that the providers agreed that strategy, costs, risks and time are poorly approached in the standard used for the PT development, since they presented less than 51% in the graph of Figure 2. In order to support the analyses, a matrix was prepared crossing information from PMBoK, ISO 10006 and ISO/IEC17043 (Table 2). In this matrix, it is possible to see that these areas are indeed not addressed by ISO/IEC 17043.

Table 2 – Relation between standards and guide

PMBoK		ISO 10006		ISO/IEC 17043	
Item	**Chap**	**Item**	**Req**	**Item**	**Req**
Project life cycle and organization	2	Characteristics of the business	4.1	-	-
Managing processes	3	Management Commitment	5.1	Management system	5.2
Integration among activities of initiation, planning, execution, control and closure.	4	Processes related to interdependence	7.2	-	-
Project scope	5	Processes related to scope	7.3	Model of proficiency testing programs	4.4
		Critical analysis of management and program evaluation	5.3	Choice of method or procedure	4.5
Time	6	Processes related to deadlines	7.4	-	-
Cost	7	Processes related to costs	7.5	-	-
Quality	8	Quality Management System	4.2	Equipment, facilities and environment	4.3
		Processes related to improvement	8.1	Operation of proficiency testing programs	4.6
		Measurement and Analysis	8.2	Data analysis and result evaluation of PTs	4.7
		Continuous improvement	8.3	Reports	4.8
				Management system	5.2
				Document control	5.3
				Critical analysis of orders, proposals and contracts	5.4
				Control of nonconforming work	5.9
				Improvement	5.10
				Corrective actions	5.11
				Preventive actions	5.12
				Record control	5.13
				Internal audits	5.14
				Critical analyses by management	5.15
Human resources	9	Processes related to resources	6.1	Staff	4.2
		Processes related to personnel	6.2		
Communication	10	Processes related to communication	7.6	Communication with participants	4.9
				Confidentiality	4.10
				Customer service	5.7
				Complaints and appeals	5.8
Risk	11	Processes related to risks	7.7	-	-
Purchase(acquisition)	12	Processes related to purchase	7.8	Subcontracting	5.5
				Purchasing services and supplies	5.6

In Table 2 we noticed that other areas, such as “Integration”, are not addressed in an explicit way in the ISO/IEC 17043 standard either, although most providers consider that it is

approached, because it presents 83% in the graph of Figure 2. “Integration” is a key part of any Quality Management System, so it is evident that those with practice of implementing ISO/IEC 17043 consider this area in PT development.

In the same graph, the project management areas used by the PT providers accredited in Brazil are presented. Even if not approached by ISO/IEC 17043, most providers (75%) consider issues related to strategy when carrying out a PT scheme. Other factors that were considered were cost and time, with results of 67% and 83%, respectively. It was possible to notice that cost was not considered as relevant as other areas, such as quality, communication, acquisition and scope. In Table 2, it was possible to see that these other areas (quality, communication, acquisition and scope) are considered in ISO/IEC 17043. This may be an explanation for this fact.

In Figure 2 it was also possible to see that the most mentioned area was quality (100%), something which was already expected, because this is an area linked to conformity assessment. The least mentioned area was Risk (33%), which may represent a possible area for improvement of PT provider’s management. It is noteworthy that the areas in Table 2 that show the strongest and the weakest connection with the ISO/IEC 17043 standard are quality and risk, respectively. ISO CASCO also confirm the importance of risk management and this topic will be included in ISO 9001:2015, and will therefore be included in the revision of ISO/IEC 17025 and any future revision of ISO/IEC 17043.

The last two issues addressed in the research were those presented in Figure 3. Most PT providers consider that their activities are developed through projects (58%). 25% of the participants mentioned that they consider their activities only partially as a project. The reason for calling it “partial” was coherent, since it was reported that PT schemes are projects when they are developed for the first time. After they are validated, commonly with pilot rounds, they are then frequently developed and become processes, integrating the provider’s routine. It was possible to interpret that PT which are frequently done, without significant changes in their scopes, may be considered a process, not a project.

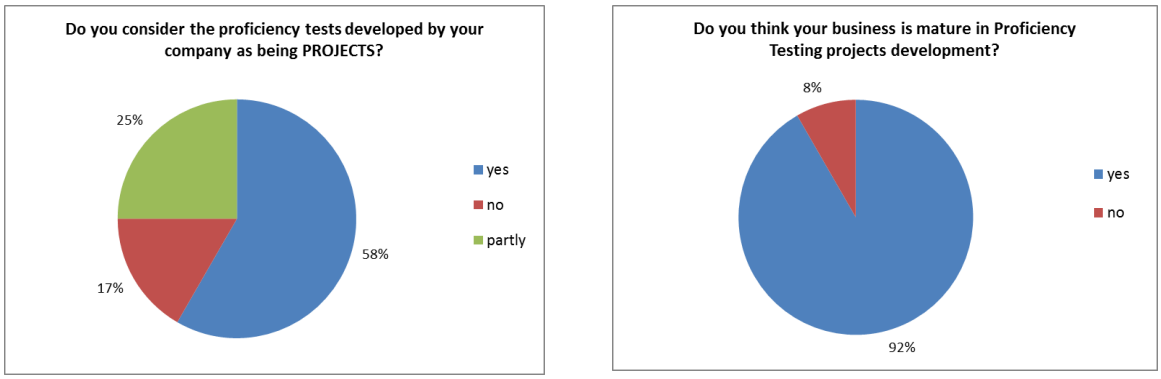

Figure 3 – Project development and PT (a) provider's maturity (b)

Most providers (83%) consider themselves mature at developing PT. They report that this fact is related to a few factors, such as: experience with PT and ISO/IEC 17043 accreditation, besides some providers have less than 3 years of experience.

Conclusions

The improvement of PT providers in Brazil is relevant to support the development of laboratories accreditation process. The general objective of this paper was to analyze the management of PT projects of Brazilian accredited providers, which was accomplished. It was also possible to assess the declared knowledge of providers concerning the ISO 10006 and ISO/IEC 17043 standards, and the PMBoK guide, to analyze relationship areas (common areas) between these references and analyze the declared maturity level of PT projects management.

The level of declared knowledge in relation to standards and project management guides like PMBoK and ISO10006 was considered low when compared to the ISO/IEC 17043 standard, which is the reference used for PT development. Despite this fact, it was possible to see that the providers also use knowledge of project management areas to develop their activities, excluding risk management, which is one of the biggest gaps in PT activities. The cost area also evidenced possible improvements. It is worth mentioning that both of them are not included in the ISO/IEC 17043 standard.

It was also observed that most providers consider PT as projects (58%). Even so, it was found that when the same PT is developed several times, it becomes a routine. In this context, the providers develop a project at the first time they provide a PT scheme. If the PT

achieves its goals and the provider decide to cyclically repeat this activity, this PT could be considered a process.

As a suggestion for future researches, one possible proposal would be develop procedures for managing costs and risks in PT, in order to analyze which practices would be applicable for this type of activity. Also, we suggest carrying out this survey with other accredited providers around the world, aiming to compare the results between Brazil and other nations.

References

1. Albano, F, Caten, C (2014). Proficiency tests for laboratories: a systematic review. Accred Qual Assur 19:245–257. doi 10.1007/s00769-014-1061-8
2. Tawfik, H. Fatah, A (2010). ISO/IEC 17025 Accreditation: Between the Desired Gains and the Reality. Quality Assurance Journal. Vol. 13. 21–27. doi: 10.1002/qaj.465
3. Howerton, D; Krolak, J; Manasterski, A; Handsfield, J (2010). Proficiency Testing Performance in US Laboratories: Results Reported to the Centers for Medicare & Medicaid Services, 1994 Through 2006. Archives of Pathology & Laboratory Medicine: Vol. 134, No. 5, pp. 751-758.
4. ISO/IEC 17043 (2010). Conformity assessment - General requirements for proficiency testing.
5. Associação Rede de Metrologia e Ensaios do Rio Grande do Sul (RMRS). Relatório do Programa de EP em Análises Ambientais. http://www.redemetrologica.com.br/interlaboratoriais/1 Acessed in Mar 2014.
6. Moura, M (2013). Análise do atendimento à demanda por ensaios de proficiência no Brasil por meio da acreditação. 7 Congresso Brasileiro de Metrologia. SBM, Ouro Preto, Brazil, ISBM 978-85-86920-08-0
7. Pizzolato, M. Ribeiro, J (2007). Análise do desempenho dos laboratórios nos ensaios de proficiência de análises microbiológicas de águas naturais realizados pela rede metrológica do rio grande do sul. SBM, Curitiba, Brazil, ISBM 978-85-86920-08-0
8. Tholen, D (2011). Metrology in service of society: the role of proficiency testing. Accreditation and Quality Assurance Journal. 1-9 p. 2011. doi: 10.1007/s00769-011-0836-4
9. Chen, C, Y (2011). Managing projects from a client perspective: The concept of the meetings-flow approach. International Journal of Project Management, Volume: 29 Issue: 6 pp.671-686 . Issn: 0263-7863
10. Project Management Institute (2013). A Guide to the Project Management Body of Knowledge (PMBOK Guide) - Fifth Edition. ISBN 13:9781935589679.
11. Din, S. Abd-Hamid, Z. Bryde, D (2011). ISO 9000 certification and construction project performance: The Malaysian experience. International Journal of Project Management, Volume: 29 Issue: 8 p.1044-1056 Issn:0263-7863
12. Pegoraro, C. Saurin, T. Paula, I, C (2011). O papel da gestão de requisitos em projetos de ambientes construídos: um estudo de caso. Revista Produção Online. Florianópolis, SC, Brasil. v.11, n. 4, p. 965-994. doi http://dx.doi.org/10.14488/1676-1901.v11i4.648
13. ISO 10006 (2003). Quality management systems - Guidelines for quality management in projects.
14. Koch M, Baumeister F (2008) Traceable reference values for routine drinking water proficiency testing: first experiences. Acred Qual Assur 13:77-82. doi 10.1007/2Fs00769-007-0349-3

ARTIGO 4 - Principal component analysis for selection of variables in homogeneity and stability tests applied to proficiency testing[1]

Artigo revisado e resubmetido para publicação no Accreditation and Quality Assurance Journal for Quality, Comparability and Reliability in Chemical Measurement. ISSN 0949-1775.

Qualis Cappes: B1

[1] Artigo possui coautoria do Dr. Michel Anzanello.

Principal component analysis for selection of variables in homogeneity and stability tests applied to proficiency testing

Abstract: Proficiency Tests (PT) based on interlaboratory comparisons are activities aimed at assessing the technical competence of laboratories in carrying out specific measurements. The analyses of homogeneity and stability of prepared samples are an important step in ensuring the reliability of the comparison rounds, since improper selection of the parameter to carry out this evaluation can influence the promoted comparison. This paper proposes a method for selecting the most relevant variables aimed at improving homogeneity and stability tests in PT. For that matter, the approach relies on a variable importance index derived from Principal Components Analysis (PCA) parameters. The proposed method was applied to three different PT schemes (beverage, water and coal) in Brazil. Results indicate that the use of PCA was adequate to help the variable selection of homogeneity and stability tests in PT schemes. The selected subset of variables was corroborated by experts in the PT schemes analyzed.

Keywords: Proficiency tests, interlaboratory comparison, homogeneity, stability, principal component analysis

Introduction

Laboratories seeking accreditation should participate in PT Schemes, which are important instruments that provide confidence for testing and calibration [1]. In the testing area, one of the important items in PT is the preparation of samples, which must ensure that such samples are homogeneous and stable so that the variability between them is negligible across the variability of laboratories [2]. That becomes a relevant fact, since the non-homogeneity or stability of a batch of samples can provide the laboratory with a false impression of adequate or inadequate results in a comparison program.

Homogeneity and stability tests should always be conducted in PT in order to ensure the reliability of the comparison rounds [2]. These tests are often expensive, and it typically becomes hard to choose which chemical or biological variable, among others, should be used in the homogeneity and stability tests [3]. Since it is often economically unfeasible to test all variables, the most critical ones (higher breakdown or more sensitive) are chosen to conduct the tests. Despite its importance, studies attesting the correlation between variables and the real representativeness of the variable chosen by experts to represent the sample are rarely carried out [4].

In light of the aforementioned, we present the following research questions: how to discover if the variable chosen by experts to test homogeneity and stability of the PT samples really represents the variability in sample preparation? Is selecting a single variable sufficient for conducting homogeneity and stability tests?

To examine the homogeneity and stability of samples it is necessary to select at least one variable representing the samples being compared. Any failure in this process can compromise a PT round. A reliable course of action tailored to select variables is the use of Principal Components Analysis (PCA), which allows identifying the variables that better represent the variability of data and increases the efficiency of their interpretation [5].

Mathematically, PCA performs an orthogonal transformation to convert a set of correlated variables in a set of new uncorrelated variables by linearly combining the original variables. This procedure yields the so called principal components (PC), which are typically in smaller number when compared to the original variables [6].

The transformation of data carried out by PCA aims at maximizing the variance between the PCs, and is performed in such a way that the first PC represents the largest possible variance, the second represents the second larger variance and so on [7]. Besides reducing the problem dimensionality, the new variables generated by PCA are orthogonal between themselves, justifying the wide use of that technique in datasets affected by highly correlated variables [5]. In terms of its practical contributions, PCA has been widely applied in several segments with different scopes: (i) exploratory data analysis, preparation of forecasting and variable selection models [8]; (ii) analysis and discussion in terms of the components' scores, also called factor scores and their loadings [9]; (iii) reduction of the number of original variables [10]; (iv) unveiling of internal structure of data aimed at obtaining proper explanation of its variance [8]; and (v) graphical analysis and clustering of observations [7].

This paper proposes a PCA-based method to select relevant variables in homogeneity and stability tests of samples in PT. For that matter, parameters generated by the PCA give rise to a variable importance index aimed at identifying the most important variables to be assessed in PT. Variables presenting the highest indices are qualitatively assessed by process experts, and a decision on their retention is made. As secondary objectives, the following aspects stand out: (i) application of the proposed framework to three different PT schemes; and (ii) discussion on the advantages and disadvantages of applying PCA in PT scenarios.

PT Schemes data

The developed research presents an applied character and follows a quantitative approach. As for the objectives, this is an exploratory work that uses as technical procedure a case study applied through the analysis of real data from PT schemes promoted by Rede Metrológica RS, a Brazilian accredited ISO/IEC 17034 provider. The study was conducted in three different PT schemes (beverages – red wine, water – waste water, and mineral coal), based on chemical analyses. Details on the data of PT are depicted in Table 1.

Table 1 – PT schemes data details

PT scheme	Variables	Laboratories (N)	Variables (N)
Beverage (Red Wine)	1-propanol (mg/100mL), ethyl acetate (mg/100mL), acetaldehyde (mg/100mL), 2-methyl propanol (mg/100mL), ashes (g/L), sugar (g/L), pH (-log[H+]), chlorides (mg/L), malic acid, alcohol (%), density (g/cm^3), total acidity (mEq/L), free SO_2 (mg/L), total SO_2 (mg/L), volatile acidity (mEq/L), adjusted volatile acidity (mEq/L), methanol, 3-Methyl-2-butanol (mg/100mL)	32	18
Waste Water	As, Fe, Ni, Hg, Ca, Cd, Pb, Cr, Na, Ba, Mn, Es (mg/L)	77	12
Mineral Coal	moisture (%), ashes (%), volatile matter (%), fixed carbon (%), total S (%), calorific power (J/g)	20	6

The missing data in such datasets was replaced using an interpolation technique that relies on solving a direct linear system of equations for missing elements; that procedure was carried out in Matlab® 2014. The software used to perform the PCA analyses was SPSS (Statistics for Windows, version 21.0., Armonk, NY).

Results were obtained in a fourteen-day interval and samples were evaluated by different laboratories. The comparison was aimed at assessing the performance of the various participants and checking whether they were technically competent in measuring the proposed

variables. The PT schemes were structured according to ISO/IEC 17043 and the performance was evaluated according to ISO 13528 standard (through the consensus value – Annex C).

Results and discussion

We initially assessed the correlation matrix of the original data of each PT, depicted in Tables 2, 3 and 4. Some variables are highly correlated, especially in mineral coal data (see Table 4).

Table 2 – Beverage variables correlation matrix

Correlation	sugar	Alc.	Dens.	Vol. Acid.	Ajus. Vol. Acid.	Tot. Acid.	pH	free_ SO2	total_ SO2	Meth.	Acet.	Eth. Acet.	Malic Acid	1-prop.	2-met prop	3-met 2-but	ashes	Chlo.
sugar	1.00	-0.28	0.08	0.00	-0.55	0.08	-0.05	-0.43	0.14	-0.66	0.05	0.18	0.33	0.12	0.20	-0.05	0.15	0.04
Alc.		1.00	-0.05	0.25	0.30	0.63	-0.71	0.31	0.44	-0.01	0.24	0.01	-0.22	-0.12	-0.01	0.20	-0.10	0.13
Dens.			1.00	-0.10	-0.09	-0.06	0.04	-0.12	-0.07	-0.18	0.03	-0.10	-0.16	-0.09	-0.18	-0.18	-0.07	-0.12
Vol. Acid.				1.00	0.36	0.53	-0.48	-0.09	0.33	0.07	-0.25	-0.19	0.05	-0.34	-0.24	0.41	0.15	0.25
Ajus.Vol. Acid.					1.00	0.51	-0.51	0.52	0.25	0.49	-0.13	-0.27	-0.23	-0.25	-0.13	0.27	-0.06	0.02
Tot. Acid.						1.00	-0.94	0.18	0.63	0.03	0.05	-0.13	-0.19	-0.27	-0.07	0.30	-0.13	0.09
pH							1.00	-0.22	-0.66	-0.01	-0.09	0.13	0.17	0.24	0.03	-0.30	0.06	-0.12
free_SO2								1.00	0.02	0.35	0.19	0.00	-0.19	0.04	0.12	0.11	-0.18	-0.08
total_SO2									1.00	0.00	-0.24	-0.33	-0.21	-0.38	-0.11	0.43	-0.14	0.01
Meth.										1.00	-0.16	-0.23	-0.11	-0.15	-0.06	0.37	-0.12	0.01
Acet.											1.00	0.75	0.09	0.70	0.40	-0.66	0.11	-0.07
Eth. Acet.												1.00	0.45	0.85	0.64	-0.61	0.35	0.08
MalicAcid													1.00	0.53	0.55	-0.13	0.28	0.29
1-prop.														1.00	0.80	-0.68	0.49	0.22
2-met prop															1.00	-0.21	0.53	0.33
3-met 2-but																1.00	-0.07	0.21
ashes																	1.00	0.68
Chlo.																		1.00

Table 3 – Waste water variables correlation matrix

Correlation	As	Fe	Ni	Hg	Ca	Cd	Pb	Cr	Na	Ba	Mn	Es
As	1.00	0.11	-0.01	-0.01	-0.10	0.14	0.12	-0.03	-0.17	-0.01	0.23	-0.04
Fe		1.00	0.47	0.43	0.02	0.40	0.43	0.42	-0.15	0.32	0.47	0.09
Ni			1.00	0.16	-0.07	0.76	0.77	0.77	-0.45	0.52	0.42	-0.16
Hg				1.00	0.19	0.09	0.11	0.12	0.12	0.10	0.26	0.14
Ca					1.00	-0.17	-0.12	-0.26	0.43	-0.19	-0.22	-0.32
Cd						1.00	0.87	0.68	-0.55	0.48	0.52	-0.13
Pb							1.00	0.74	-0.57	0.61	0.60	-0.19
Cr								1.00	-0.36	0.65	0.48	-0.16
Na									1.00	-0.43	-0.41	-0.08
Ba										1.00	0.45	0.00
Mn											1.00	-0.01
Es												1.00

Table 4 – Mineral coal variables correlation matrix

Correlation	moisture	ashes	volatile mat.	fixed carbon	total S	calorific power
moisture	1.00	0.28	-1.00	-1.00	-1.00	0.99
ashes		1.00	-0.27	-0.29	-0.27	0.32
volatile matter			1.00	1.00	1.00	-0.99
fixed carbon				1.00	1.00	-0.99
total S					1.00	-0.99
calorific power						1.00

PCA was then applied to the correlation matrix of each PT; parameters of interest emerging from PCA include the amount of variance explained by the j^{th} retained PC, Ev_j, and the loading of variable i in each PC j, L_{ij}. For the variable selection procedure we propose an importance index, vs, to highlight variables that explain most variance on data; see Equation (1). The proposed index relies on the following rationale: variables with high loadings contribute substantially on explaining variability inside a specific PC; we use the absolute loading value to avoid cancelling the importance of variables when summing up the loadings from retained PCs. In addition, relevant variables emerging from PCs that explain high amounts of data variance should have their influence increased, so the absolute loadings are weighted by Ev_j. In our propositions, we retained a number of PCs that explains at least 70% of the cumulative variance in the data. Finally, the higher the vs_i, the more important we understand variable i for explaining data variance in the PT; we recommend retaining the

variables yielding the three highest *vs* values. Next, such variables are qualitatively assessed by process experts for a final decision on retaining or not that variable for PT.

$$vs_i = \sum Ev_j \times |L_{ij}| \quad \text{for } i=1,\ldots,I \quad (1)$$

Tables 5, 6 and 7 depict the PCA loadings and estimated *vs* for the three PT schemes (beverage, waste water and mineral coal). Once the recommended variables are selected by the proposed index, experts qualitatively assess whether such variables are suitable for PT purposes.

Table 5 – Principal components of Beverage PT scheme

	Principal Component and loadings				
Variable	**1**	**2**	**3**	**4**	***VS***
Sugar	-0.31	0.04	0.76	-0.33	4.24
Alc.	0.45	0.61	-0.14	-0.28	5.12
Dens.	-0.02	-0.23	0.06	-0.37	1.76
Vol. Acid.	0.50	0.33	0.37	0.19	4.86
Ajus.Vol. Acid.	0.61	0.34	-0.43	0.23	5.63
Tot. Acid.	0.63	0.63	0.14	-0.28	**6.11**
Ph	-0.61	-0.69	-0.13	0.28	**6.16**
free_SO2	0.24	0.29	-0.68	0.07	3.91
total_SO2	0.61	0.33	0.32	-0.21	5.33
Meth.	0.35	-0.02	-0.54	0.55	4.30
Acet.	-0.52	0.49	-0.38	-0.45	6.01
Eth. Acet.	-0.75	0.48	-0.15	-0.15	5.97
MalicAcid	-0.52	0.26	0.29	0.32	4.84
1-prop.	-0.84	0.46	-0.19	0.04	**6.28**
2-met prop	-0.59	0.57	-0.01	0.24	5.35
3-met 2-but	0.70	-0.03	0.28	0.44	5.22
Ashes	-0.40	0.41	0.30	0.49	5.09
Chlo.	-0.10	0.44	0.35	0.58	3.91
Eigenvalue	5.07	3.16	2.40	2.07	
% of Variance	28.16	17.57	13.32	11.48	
Cumulative %	28.16	45.73	59.05	70.53	

In the PT of Beverage analyses (Red wine), the variables with highest *vs* were 1-propanolol, pH and total acidity (in Table 5, *vs* of such variables are highlighted). Those variables were qualitatively assessed by a process expert with more than 20 years of

experience in wine analyses, who also is the technical responsible for one of the official beverage laboratories of Brazilian government. According to his opinion, parameters 1-propanolol and pH properly represent the wine sample: we have one variable, pH, that represents the group of traditional physic-chemical analyses and is highly correlated (|0.94|) with the total acidity, and a second variable, 1-propanolol, which represents the group of chromatographically testing (such variable is moderately correlated ($|r|>0.6$) with ethyl acetate, acetaldehyde, 2-methyl propanol and 3-Methyl-2-butanol). In addition, the expert also commented that these two variables (traditional physic-chemical and chromatographical) have different magnitudes in terms of uncertainty measurement. The selected variables also are sensible in terms of stability, what is relevant for the PT purpose. Due to this fact, is also important to consider variables that can represent these two groups.

Table 6 depicts the results of *vs* for the waste water PT scheme. Lower values of *vs* were assigned to variables As, Hg and Es; such variables were deemed the most problematic ones in this set according to the expert as the testing demands many dilutions and increase the complexity of the variable behavior. The expert, who presents 26 years of experience in water analysis in an accredited ISO / IEC 17025 laboratory, also reported that the Pb parameter (responsible for the highest *vs*) is a suitable parameter for the homogeneity testing and stability in PT, since it properly represents the group of metals. In addition, the expert complemented his assessment saying that the parameters with highest *vs* (Pb, Ni, Cr) present near wavelengths in the metal analysis through the ICP method.

Results of the variable selection procedure for the Mineral Coal PT scheme are presented in Table 7. The expert consulted for the that data is a manager of a fuel laboratory accredited ISO/IEC 17025 with more than 30 years of experience in coal analysis. The highest *vs* values were accounted by total sulfur (Total_S), volatile matter (Volat_mat) and fixed carbon (Fixed_Carbon). The sulfur and volatile matter are the most sensitive variables due to potential oxidation, while fixed carbon is directly related to volatile matter calculation (justifying its high *vs* score). On the other hand, ash is a mineral residue and presents very stable properties, so it is not representative in a PT. Moisture is also not representative because all results are reported on a dry basis. In light of that, we recommend using variables total sulfur and volatile matter in the analysis due to the high *vs* and expert approval. In

addition, the volatile matter is also recommended in the homogeneity and stability tests, since it is cheaper to be obtained.

Table 6 – Principal components of Waste Water PT scheme

Variable	Principal Component and loads				
	1	**2**	**3**	**4**	***vs***
As	0.14	-0.17	0.15	0.90	2.19
Fe	0.57	0.42	0.41	0.07	4.15
Ni	0.85	0.15	-0.18	-0.14	**4.81**
Hg	0.20	0.65	0.55	-0.03	2.80
Ca	-0.27	0.76	-0.30	0.12	3.07
Cd	0.87	-0.02	-0.15	0.06	4.58
Pb	0.92	0.03	-0.18	0.04	**4.86**
Cr	0.84	0.04	-0.15	-0.21	**4.65**
Na	-0.63	0.52	-0.01	-0.06	4.02
Ba	0.72	-0.06	0.00	-0.23	3.88
Mn	0.71	0.03	0.28	0.25	4.19
Es	-0.09	-0.31	0.79	-0.29	2.36
Total (Eigenvalue)	4.91	1.59	1.39	1.11	
% of Variance	40.94	13.29	11.62	9.24	
Cumulative %	40.94	54.23	65.84	75.09	

Table 7 – Principal components of Mineral Coal PT scheme

Variable	Principal Component and loadings		
	1	**2**	***vs***
Moisture	-0.99	-0.07	5.12
Ashes	-0.35	0.93	2.62
Volat_mat	0.99	0.08	**5.13**
Fixed_Carbon	0.99	0.06	**5.13**
Total_S	0.99	0.08	**5.13**
calorific_power	-0.99	-0.03	5.07
Eigenvalue	5.082	0.901	
% of Variance	84.703	15.009	
Cumulative %	84.703	99.712	

The Annex B of the ISO 13528 standard addresses the assessment of homogeneity and stability tests. It takes at least 10 samples analyzed twice (each sample) to evaluate homogeneity and 2 samples analyzed in duplicate to assess stability. The tests shall be performed in repeatability conditions. A PT with many variables involved can require an elevated number of homogeneity and stability tests, which can become unaffordable in terms

of cost and time demanded for analysis. The standards related to PT schemes do not present clear criteria on the systematics or methods for selecting the variables to be used in homogeneity and stability tests in PT.

Selection of representative variables from one sample for homogeneity and stability testing is a significant event in the planning and execution of a PT. Such variables should be able to represent the behavior of the sample during the period of PT implementation. The application of PCA, as presented this paper, was deemed helpful to define the most representative variables on a PT. Variables selected by the framework were confirmed by experts in the tests, corroborating the robustness and usefulness of the proposed importance index.

The disadvantages of this application are associated to the need of computational resources to develop the PCA analysis, and technical and statistical knowledge to properly interpret data. Furthermore, in order to develop this research, it is necessary to rely on real PT databases, whereas it is not possible to carry out the proposed steps without prior rounds. However, it is not always possible to count on databases comprised of all tests; using algorithms to fulfill the missing data may affect the actual representation of variability retained by the PC, requiring additional analysis.

Conclusions

This study aimed at proposing a PCA-based framework to choose parameters in homogeneity and stability testing of samples conducted in PT. For that matter, parameters emerging from the PCA were used to build a variable importance index aimed at identifying the most important variables to be evaluated in PT. Variables yielding the highest indices were then qualitatively assessed by process experts regarding their retention or removal in PT.

When applied to data from three Brazilian PT programs, the proposed framework led to satisfactory results as it was able to consistently select variables to evaluate the homogeneity and stability of the PT samples in beverage, waste water and mineral coal datasets. The variables pointed as relevant by the method were corroborated as important when qualitatively assessed by process experts.

Future developments include the application of the method proposed in other PT schemes in different areas, and the proposition of homogeneity and stability tests based on the results of the participating labs.

References

1. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO/IEC 17043 (2010). Conformity assessment - General requirements for proficiency testing. Geneva.
2. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 13528 (2015) Statistical methods for use in proficiency testing by interlaboratory comparisons. Geneva.
3. THOLEN, D. CHAPPEL, R. FRANCIS, B. In: Accred Qual Assur. 11: 400–407 DOI 10.1007/s00769-006-0167-z. 2006.
4. ALBANO, F, CATEN, C (2014). Proficiency tests for laboratories: a systematic review. Accred Qual Assur 19:245–257. doi 10.1007/s00769-014-1061-8
5. HAIR JR, ANDERSON RE, TATHAM RL & BLACK WC. Multivariate Data Analysis (7th Edition). 2009. ISBN-13: 978-0138132637
6. RENCHER, A. C. Methods of Multivariate Analysis. New York: Wiley, 1995.
7. ROBERTS Y. L. et al. Quantitative comparison of the variability in observed and simulated shortwave reflectance. In: Atmos. Chem. Phys. Discuss. 12, 28305-283041, 2012. Doi:10.5194/acpd-12-28305-2012.
8. ANZANELLO, M. J.; ALBIN, S.; CHAOVALITWONGSE, W. A., Selecting the best variables for classifying production batches into two quality levels. Chemometrics and Intelligent Laboratory Systems. v. 97, 2009. http://dx.doi.org/10.1016/j.chemolab.2009.03.004.
9. GUEDES, H. Aplicação da análise estatística multivariada no estuda da qualidade da água no rio Pomba, MG. In: Revista Brasileira de Engenharia Agrícola e Ambiental v.16, n.5, p.558–563, 2012.
10. REIS, P. ET AL. Mineral composition of Lippia alba (Mill.) N.E. Brown leaves. J. In: Braz. Chem. Soc. vol.21 no.10 São Paulo 2010. http://dx.doi.org/10.1590/S0103-50532010001000015

ARTIGO 5 - The influence of the probability distribution, kurtosis coefficient and performance assessment method on statistical analysis of data from Proficiency Testing using consensus value

Artigo submetido para publicação no Accreditation and Quality Assurance Journal for Quality, Comparability and Reliability in Chemical Measurement. ISSN 0949-1775.

Qualis Cappes: B1

The influence of the probability distribution, kurtosis coefficient and performance assessment method on statistical analysis of data from Proficiency Testing using consensus value

Abstract: The activity of interlaboratory comparison makes part of laboratories' accreditation process and is an important quality tool. Most of the time, providers estimate reference values used in Proficiency Testing (PT) by means of statistical procedures, leading to consensus values. This work aims at analysing the influence of the kurtosis coefficient and of the type of data probability distribution on the performance assessment using consensus value in PT. The analysis also includes the evaluation of different assessment approaches concerning laboratories performance, by applying robust and classical statistics. Data were obtained from a PT programme in the environmental area and by Monte Carlo simulation. Results show that bimodal and platykurtic distributions are inadequate to assess laboratories' performance in comparisons using consensus value. Furthermore, it emerged that the robust methods analysed for performance assessment are equivalent, even in asymmetric distributions.

Keywords: Interlaboratory Comparison, Proficiency Testing, Monte Carlo, Consensus Value, Robust Statistics.

Introduction

Proficiency Testing (PT) providers are responsible for assessing the performance of laboratories participating in comparison rounds of results. This assessment must be carefully conducted, since it has a direct influence on laboratory routine and on possible actions that must be taken when a questionable or unsatisfactory result is obtained.

One of the standards provided by the ISO/IEC 17043 for the statistical analysis of results is the ISO 13528 [1]. PT reference values can be determined according to the participants' consensus values (all participants, or a subgroup of experts) or based on designated values independently determined (only by a laboratory or through certified reference materials) [2]. The performance of each participating laboratory may be assessed through statistical analysis of the results that have been sent, the real value estimate (designated value) being defined through consensus. It is important to note that also the prepared samples are analysed using homogeneity and stability testing [2].

The most widely used index for performance assessment in the testing field is the *z*-score [3]. This value represents the difference between the laboratory's result and the established reference value, dividing it by the variability related to the reference value. The *z*-score result must be lower or equal to |2| in order to obtain a satisfactory performance. Results between |2| and |3| are considered to be questionable and results higher or equal to |3| are unsatisfactory [2].

The consensus value can be calculated through a robust mean of the data from laboratories participating in the comparison, generated through an algorithm with an outlier removal process. This process also generates a robust standard deviation [2]. Another manual, published by an Australian PT provider, suggests the use of the median as consensus value and of the normalised interquartile range as variability measure. The latter is calculated according to the difference between the 3rd and the 1st quartiles of participating laboratories' data, multiplying this result by a constant of 0.7413 [4]. Both approaches carry out performance assessment with *z*-score.

Researchers carried out studies with PT providers investigating statistical methods being used for the performance assessment. Answers were received from 1 company in the U.S. and from 11 European providers: Belgium, Croatia, Finland, France, Germany, Hungary, Russia, Slovenia, Spain (two providers), and Switzerland. Five out of the 11 European providers use standards established based on the experience, biological variation, state of art, or professional consensus [5]. The other providers used variable standards based on the statistical analysis of the results obtained. Except from the German, Hungarian and Slovenian providers, all European ones that were consulted use consensus target values. Therefore, it is possible to realise that there is no alignment of practices for performance assessment, despite most of the providers using *z*-score as a performance measurement, its satisfactory value being lower or equal to |2| [5].

Other researchers conducted a study that analysed the performance of laboratories based on two statistical procedures: robust statistics and repeatability and reproducibility analysis proposed by the ISO5725-2 standard. The analysed data were taken from the CNACL (*China National Accreditation Committee for Laboratories*). Robust procedures used the median and the normalised interquartile range as central tendency and variability parameters, respectively. The analysis by robust statistics (assessed with *z*-score) has proven to be more demanding than the one proposed by ISO5725-2, showing a higher number of unsatisfactory results [6].

Studies were also developed with the intent to analyse the use of robust statistics for the performance assessment in PT, as far as strength and hardness are concerned. The research results show that the use of *z*-score with robust parameters is more rigorous than the normalised error with target measures established by a reference laboratory [7].

Another author also comments on a problematic issue, that is, the probability distribution with assumption of consensus is not necessarily the most realistic one. The author considers the Bayesian approach as acceptable, but in some cases it might not be the most suitable [8]. Other research studies also report that the results of an interlaboratory assessment are statistically consistent if they fit into a normal consistency (Gaussian) model, which postulates that results have the same expected unknown value [9]. According to this logic, it is possible to observe that data distribution may have an influence on the laboratory performance results.

Studies were carried out comparing different performance assessment methods by simulation, such as *z*-score and E_z (similar to *z*-score but it considers the laboratory measurement uncertainty and the reference value uncertainty) with unimodal and multimodal distributions [10]. In this study it was perceived that multimodality can have a great impact on the performance assessment and must be considered by PT providers. This work also suggested that E_z is the most recommended performance index, but in order to use it participants must declare their measurement uncertainty.

Research studies were also developed with simulated data, revealing that statistical protocols for data assessment through consensus value worked well when the distribution was approximately symmetric and unimodal [11]. The same study reports that many PT data analysis protocols assumed a normal distribution, even though this fact was not always confirmed [11].

The general objective of this work is to evaluate the impact that probability distribution, kurtosis coefficient and data performance assessment method may have on PT using consensus value. Secondary objectives include verifying the existence of correlation between data concentration and two factors: percentage of satisfactory results and the variation coefficient regarding the group of laboratories participating in PT.

Method

As far as its objectives are concerned, the research is classified as an explanatory work that uses an experimental study as technical procedure. This work has a quantitative approach, with an applied focus. The proposed method is based on 4 stages, presented in Figure 1.

Stage		Description
1	**Identification of the Database (DB) to be used**	To define the PT programme that will be used in the study, so as to contemplate different types of probability distribution (symmetric, asymmetric and bimodal). If necessary, data simulation by Monte Carlo method will be conducted, generating pseudorandom numbers to complement the DB.
2	**Generating the response variable related to the performance assessment methods used**	To select data from the DB and to calculate the response variable with three performance assessment methods (robust method proposed by ISO13528, robust method proposed by *Proficiency Testing Australia* - PTA - and use of classical statistics with no outlier removal). Each analyte is supposed to generate a result of Laboratory Performance Index (LPI), which includes a value from zero to 1.0, indicating the % of laboratories with satisfactory results (with *z*-score lower or equal to |2.0|). To report the probability distribution, the Distortion Coefficient (DC) and the kurtosis coefficient resulting from the data analysis of each tested parameter.
3	**Analysis of results**	To analyse the correlation between the controllable factors (DC, kurtosis coefficient and performance assessment method). To perform the Variance Analysis considering the pertinent controllable factors so as to identify the influence of these on the response variable. To conduct Post-hoc tests, if ANOVA presents significant differences. Tests were conducted with the *Action* ® statistical software.
4	**Discussion of results**	To carry out a general discussion of results, identifying the influence of controllable factors on the performance analysis process in interlaboratory comparisons by consensus value.

Figure 1 – Description of the stages in the working method

Results

Description of the DB used for the research

The real PT data that made part of the research are taken from a *Rede Metrológica RS* programme in the environmental area, in which the test matrix is water. All analysed parameters are physicochemical characteristics. This PT is accredited according to the ISO/IEC 17043 standard. The analytes taken into consideration for the study were: arsenic, iron, nickel, calcium, cadmium, chemical oxygen demand, ammoniacal nitrogen, nitrates, pH, conductivity, alkalinity, turbidity, chlorides, sulphates, lead, chromium, sodium, barium, manganese, strontium, phosphorus, surfactants, fluoride, biological oxygen demand, zinc, aluminium and hardness.

These 27 parameters resulted in 1,756 analyses conducted in this PT. To complement the DB, 4 more parameters with bimodal distributions were generated, since just one of the real parameters (hardness) showed this kind of behaviour. The simulation was carried out by Monte Carlo method and added 278 more data elements. The final DB totalled 31 analytes and 2,034 data elements.

Estimate of the experiment response variable (LPI)

Each DB parameter was entered into 3 spreadsheets, each one of them containing an algorithm for assessing laboratories performance (by *z*-score calculation), as shown in Figure 2.

Performance Assessment Method (Algorithm)	**Value designated as reference**	**Value designated as variability estimator**
Attachment A of ISO 13528	Robust Mean (after outlier removal process)	Robust Standard Deviation
Proficiency Testing Australia (PTA)	Median	Normalised interquartile range
Classical Statistics	Mean	Standard Deviation

Figure 2 – Algorithms used to generate the LPI

The response variable that was generated was the LPI, related to the % of laboratories' satisfactory results. The lower the LPI, the better the algorithm manages to distinguish between satisfactory, questionable and unsatisfactory results.

Each DB analyte used also provided information related to other controllable factors that were assessed in the experiment, such as: kurtosis coefficient, DC and type of probability distribution. Each analyte generated 3 LPI results (one for each performance assessment method). Therefore, the experiment was conducted with 93 (31x3) LPI data elements. Figure 3 schematically explains the experiment carried out.

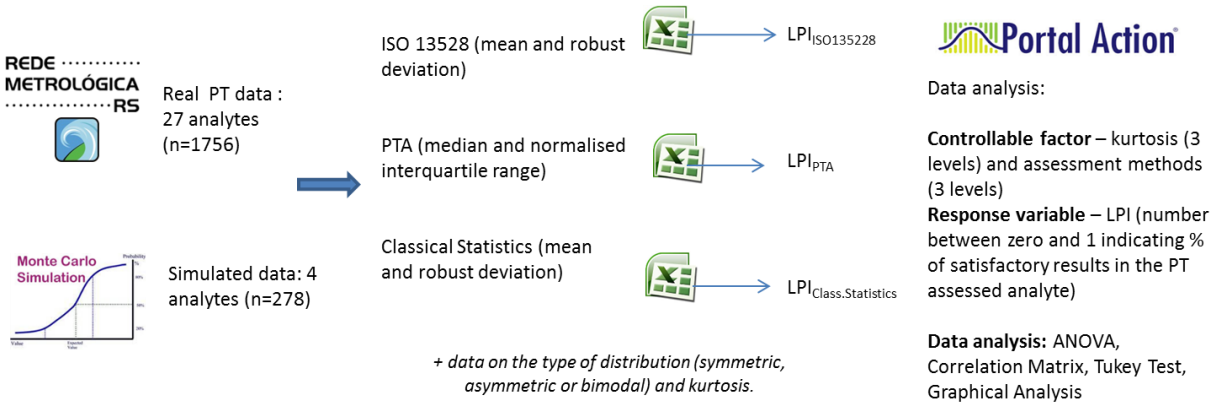

Figure 3 – Illustrative scheme of the experiment

Analysis of results

ANOVA considering kurtosis and the type of performance assessment method

Kurtosis, a dispersion measure characterising the flattening of the probability distribution function curve, was one of the controllable factors. It indicates if the probability distribution is platykurtic (more flattened than normal distribution), mesokurtic (same flattening as normal distribution) or leptokurtic (peaked and more concentrated than normal distribution).

Another factor analysed was the DC of the distribution, which represents the degree of asymmetry of a distribution around its mean, being connected to the type of probability distribution shown. Ideally, symmetric distributions have a DC close to zero. Positive asymmetric distributions have positive coefficients while negative asymmetric distributions present a negative distortion. For classification purposes of the distributions, the following criterion was established: symmetric distributions (DC between -0.55 and 0.55), positive asymmetric (DC≥0.55) and negative asymmetric ones (DC≤0.55).

Upon analysis of correlation between controllable factors, it was possible to realise that kurtosis is strongly correlated (r 0.869) with the modulus of the DC. The higher the kurtosis value, the higher the distribution distortion in modulus (positive or negative asymmetry). By analysing the type of probability distribution (based on Kernel density estimation of data that generated LPIs), it was evident that asymmetric distributions, with high DC (in modulus), show the most leptokurtic behaviour (high kurtosis). Distributions

classified as symmetric present a kurtosis coefficient with intermediate values, yet they were classified as leptokurtic. Instead, bimodal distributions showed a negative kurtosis coefficient, being more platykurtic. Figure 4 provides a graphical overview.

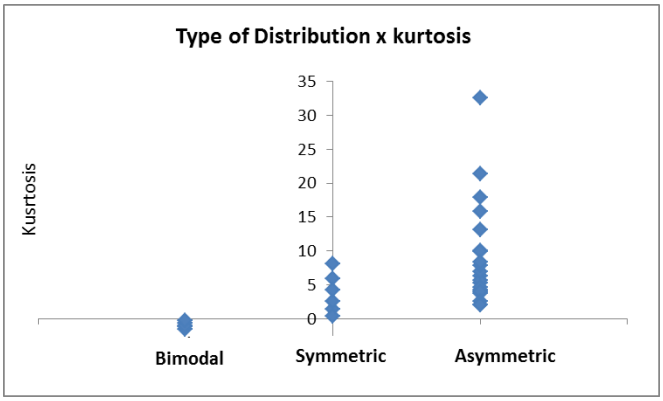

Figure 4 – Type of Distribution and kurtosis

Therefore, it was decided to use kurtosis as controllable factor in the experiment, distinguished at three levels (platykurtic, leptokurtic and highly leptokurtic distribution). Mesokurtic distribution was not used as a level of the controllable factor because in the DB no distributions were obtained with kurtosis coefficients being equal or very close to zero. To classify distribution in relation to kurtosis, distributions with kurtosis from -1.54 to -0.29 were considered platykurtic; distributions with kurtosis between 0.37 and 4.28 were considered leptokurtic, while with kurtosis between 5.24 and 32.52 distributions were considered highly leptokurtic. It is important to stress that kurtosis does not have higher limit. Distributions with high kurtosis may exist, but its lower limit is -2, due to Bernoulli's equation with p of 1/2.

Another controllable factor used was the type of data performance method assessment, with 3 levels (according to the algorithms that were used and are described in Figure 2). After data collection and organisation, ANOVA was conducted. Table 1 shows the results obtained.

Table 1 – ANOVA between kurtosis and performance assessment method

ANOVA Table	*Degrees of Freedom*	*Sum of Squares*	*Mean Square*	*F-ratio*	*p-value*
Kurtosis	2	0.1518	0.0759	35.2330	0.0000
Assessment Method	2	0.1409	0.0704	32.6970	0.0000
Kurtosis x Method	4	0.0388	0.0097	4.5050	0.0024
Residuals	84	0.1810	0.0022		

The kurtosis and assessment method factors showed a significant difference (p<0.05) in the LPI results. The interaction between factors was also considerable. This makes it possible to conclude that PT results are influenced by the performance method used and by distribution kurtosis (which is related to the type of data probability distribution). Figure 5 shows the interaction graph, where it is possible to observe the effect of controllable factors.

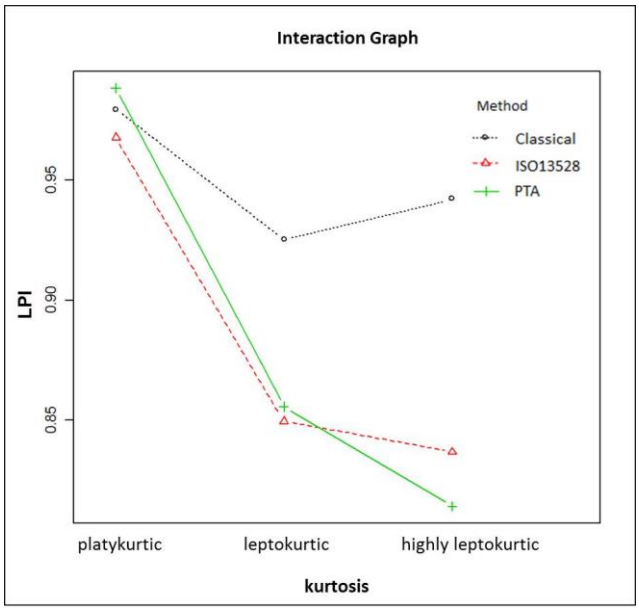

Figure 5 – Interaction between controllable factors

Post-hoc test (Tukey Test)

Tukey test pointed out the significant differences between levels of controllable factors. Table 2 shows the result of the post-hoc test for the main factors and for interaction.

Table 2 – Tukey Test

p-value for the comparison between Distributions **(kurtosis)**	Platykurtic Distribution	Leptokurtic Distribution	Highly Leptokurtic Distribution
Platykurtic Distribution	-	0.000	0.000
Leptokurtic Distribution	-	-	0.958

p-value for the comparison between **Methods**	ISO 13528	PTA	Class. Stat.
ISO 13528	-	0.999	0.000
PTA	-	-	0.000

Differences in the **Interaction of factors**	ISO 13528	PTA	Class. Stat.
Platykurtic Distribution			
Leptokurtic Distribution			
Highly Leptokurtic Distribution			

Key:	
No significant difference – Tukey test (alpha 0.05): p>0.05	

Distribution kurtosis has a significant influence on the performance assessment of laboratories in PT. Platykurtic distributions are typically more flattened and can usually be related to multimodal distributions, mainly those with more negative kurtosis. This fact demands the attention of PT providers, since when they work with consensus value based on participants' data, the value designated as true may be masked by multimodal distributions, characterised by a low kurtosis coefficient. In this case, it would be important to check which factor may have generated this behaviour, with measurement techniques used by PT participants or specific conditions that may provide results with different central tendencies (means) in an interlaboratory comparison round.

In the DB used, all platykurtic distributions showed a bimodal behaviour. In this scenario, LPI was close to 1.0, which clearly demonstrates the incapability to differentiate satisfactory results from questionable and unsatisfactory ones. The above corroborates the fact that multimodal and platykurtic distributions are inadequate in PT, even with procedures based on robust statistics generated through consensus value (proposed by ISO13528 and PTA). It is recommended to confirm the type of distribution with Kernel density estimation, since the kurtosis coefficient and the DC alone are not enough to adequately characterise the type of probability distribution.

It is important to note that symmetric distributions may present a certain degree of flattening and a negative kurtosis (usually not very peaked). If distribution is perfectly symmetric, performance assessment methods would also be equivalent, since it would be an ideal normal distribution, leading to the same LPI of approximately 0.95, independently from the performance assessment method adopted (in which the value established by robust mean, median and mean would be very close). This is due to the fact that 95.45% of the data in a perfectly symmetric distribution are between -2 and 2 standard deviations from the mean, being equal to the *z*-score value of satisfactory results.

According to data in Table 2, performance assessment methods proposed by ISO13258 and by PTA do not show significant differences as far as LPI is concerned ($p>0.05$). Nevertheless, both methods present a considerable difference when compared with the use of classical statistics ($p<0.05$). The provider could choose between the two methods (which do not show any differences) to treat PT data in which consensus value established by laboratories is used, as long as distribution unimodality and its kurtosis are verified. This

happens even in leptokurtic and highly leptokurtic distributions, which typically present a marked asymmetry (positive or negative).

The most frequent distributions in the DB used show leptokurtic or highly leptokurtic behaviour, with reasonable asymmetry (the mean of moduli of DC was 1.5, real DB data being taken into consideration). This may be associated to the fact that real PT data have heavy-tailed distributions, characterised by few laboratories that present results very far from reference value and by many laboratories being close to the established consensus value.

Therefore, robust methods for data analysis (ISO13528 and PTA) have proven to be adequate even in asymmetric and leptokurtic distributions.

Concentration of the physicochemical parameter and LPI

It was also possible to analyse, with data from the experiment, the correlation between the median of the physicochemical parameters under evaluation and the LPI obtained. Figure 6 is a graph showing dispersion between these factors, which presented a weak (r 0.32) and a not significant correlation (p>0.05). This fact makes it possible to infer that the concentration of elements does not necessarily influence a high or low LPI.

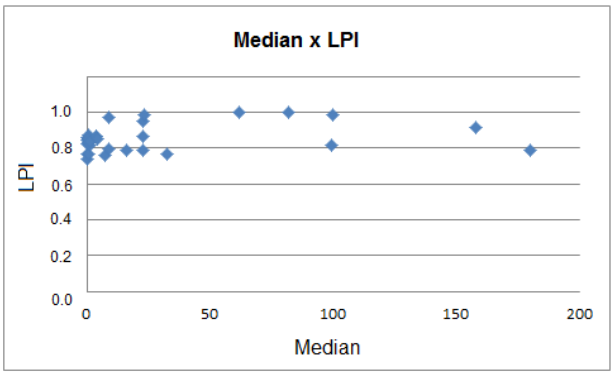

Figure 6– Relation between the Median of physicochemical parameters of the DB and the LPI

In Figure 7, which shows the median of physicochemical parameters and its respective RSD(%), it is possible to observe that as the element concentration decreases, the RSD(%) increases. Factors show a negative (r 0.53) and significant correlation (p 0.002). This was expected, since variability is proportionally high in low concentration values (especially in values lower than 1.0). This fact may indicate that high RSD(%) in low concentration values

are expected, but it does not necessarily have an impact on the LPI obtained by the group of laboratories.

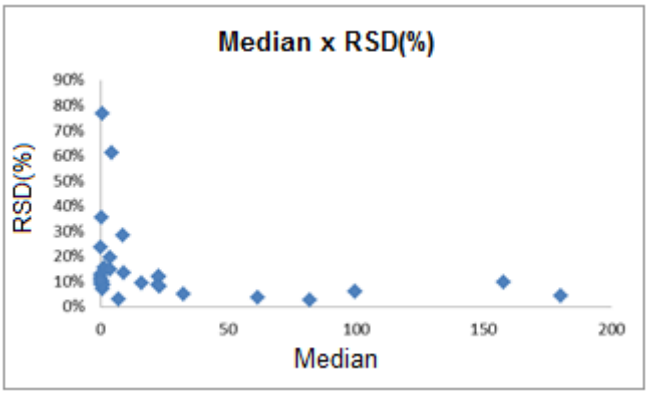

Figure 7– Relation between the Median of physicochemical parameters of the DB and the RSD

Furthermore, it can be inferred that a high RSD(%) in very low concentrations (usually lower than 1.0) may not be a representative indicator of the dispersion of the laboratories group, since this relation is mathematically high, but this does not make it easy or difficult to obtain satisfactory PT results (represented by LPI).

Considerations upon the results obtained

Figure 8 shows the general discussions on the experiment developed and on the results obtained.

Factor	Observations related to the DB evaluated and the results obtained
Kurtosis coefficient and probability distribution	-The kurtosis coefficient shows a strong correlation with the distribution asymmetry. Leptokurtic or highly leptokurtic distributions tend to be asymmetric in PT. -The kurtosis coefficient of the distribution has a significant difference in the LPI of PT working with consensus value. - More platykurtic distributions (with negative kurtosis coefficient) have proven to be more associated to bimodal distributions. In some cases, symmetric and unimodal distributions can also show a negative kurtosis coefficient, but not in a marked way. - Bimodal distributions masked data in PT performance assessments by consensus value. - It is relevant to confirm the type of data probability distribution before conducting performance assessment by consensus value. Therefore, the use of Kernel density estimation is suggested to represent the distribution behaviour. - The most frequent distributions originated from PT data accredited in the environmental area were asymmetric and leptokurtic.
Performance Assessment Method	-Methods that make use of robust statistics to assess performance in PT (Annex A of ISO13528 and PTA) have proven to be equivalent and adequate, even in asymmetric and leptokurtic distributions.

	-Performance assessment methods under evaluation present a significant difference when compared with classical statistics (which does not use outlier removal). This fact tends not to be there in symmetric distributions (where the logic suggests that the three methods should have a LPI of approximately 0.95) and bimodal distributions (in which the three methods tend to show a LPI close to 1.0). In this last case, PT performance assessment by consensus value is not adequate.
Further observations	-The median value, which represented the concentration of physicochemical analytes in the experiment, does not have a significant correlation with the LPI. Therefore, it is possible to obtain equally high or low LPIs in different concentration values. - The median value is significantly correlated with the RSD(%) value of the laboratories group. The lower the median is, the higher the RSD(%) tends to be. This makes it possible to infer that, for very low concentration values, RSD(%) is not an adequate indicator to assess the general performance of the laboratories group, since any small deviation could mathematically have a large impact on the median, leading to a high RSD(%).

Figure 8– Summary of the discussion of results

Conclusions

The experiment presented underlined the importance of evaluating the probability distribution of data and its kurtosis coefficient, in order to assess whether the procedure of statistical treatment by consensus value can be used. In cases of multimodal distributions with negative kurtosis coefficient, the adoption of another procedure for establishing reference values is recommended, such as working with Certified Reference Materials (CRM) or with values established by a small group of specialised laboratories. This research also pointed out that the asymmetry of data used was not a problem for robust algorithms, which have proven to be adequate and equivalent (ISO13528 and PTA's algorithms).

The general objective of this work was to evaluate the impact that the probability distribution, the kurtosis coefficient and the performance assessment method of data can have on PT working with consensus value. It can be concluded that the objective was satisfactorily achieved, since the influence of the mentioned factors was analysed in a DB presenting real PT data from an accredited provider, expanded by means of Monte Carlo simulation.

The specific objectives of verifying the existence of correlation between data concentration with percentage factors of satisfactory results (LPI) and the RSD(%) of the PT participating laboratories group were also achieved. It was possible to observe the weak correlation between data concentration and the LPI and the significantly negative correlation

between data concentration and the RSD(%). This information is relevant for PT and laboratory data analysis providers in comparison rounds using consensus value.

Suggestions for future works include the application of the same procedure presented in this work with data from other PT programmes. Moreover, it is suggested to evaluate the influence of measurement uncertainty of participating laboratories on the establishment of consensus value and which impact this factor may have on data probability distribution.

References

1. ALBANO, F, CATEN, C (2014). Proficiency tests for laboratories: a systematic review. Accred Qual Assur 19:245–257. doi 10.1007/s00769-014-1061-8
2. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO/IEC 17043 (2010). Conformity assessment - General requirements for proficiency testing. International Organization for Standardization, Geneva.
3. BLERK, M. Comparison of evaluation procedures used by European externalquality assessment scheme organizers for haemoglobin concentration and leukocyte concentration. In:Accred Qual Assur (2008) 13:145–148. DOI 10.1007/s00769-008-0364-z.
4. FERRERO, C. CASARIL, M. Proficiency testing programs to improve traceability in chemical analysis. In: Measurement 42 (2009). 1502–1509.
5. FERRERO, C. The equivalence in accreditation: Robust or standard statistical analysis of interlaboratory comparisons. In: Measurement 42 (2009). 1541–1545.
6. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 13528 (2005) Statistical methods for use in proficiency testing by interlaboratory comparisons. International Organization for Standardization, Geneva.
7. KACKER, R. FORBES, A., KESSEL, R. KLAUS-DIETER SOMMER. Bayesian posterior predictive p-value of statistical consistency in interlaboratory evaluations. In: Metrologia 45 (2008). 512–523 DOI:10.1088/0026-1394/45/5/004.
8. PROFICIENCY TESTING AUSTRALIA (2014). Guide to Proficiency Testing. Australia. http://www.pta.asn.au/documents/Guide-to-Proficiency-Testing-Australia.pdf Accessed 26 Jan 2015.
9. ROSARIO, P. MARTINEZ, J SILVAN, J. Comparison of different statistical methods for evaluation of proficiency test data. In: Accred Qual Assur (2008) .13:493–499. DOI 10.1007/s00769-008-0413-7
10. WILLINK, R. Forming a comparison reference value, from different distributions of belief. In: In: Metrologia 43 (2006). 12–20. DOI:10.1088/0026-1394/43/1/002.
11. WONG, S. A comparison of performance statistics for proficiency testing programmes. In: Accred Qual Assur (2007). 12:59–66. DOI 10.1007/s00769-006-0233-6.

CAPÍTULO 3

Este capítulo apresenta uma síntese dos resultados e das limitações encontradas, caracterizando as conclusões do trabalho. Também são apresentadas as sugestões de trabalhos futuros.

3.1 CONCLUSÕES

O objetivo principal da pesquisa foi analisar e propor melhoria nos processos de provisão de EP por comparação interlaboratorial. Inicialmente, para se entender claramente o estado da arte do tema em questão, foi estabelecido um objetivo de cunho teórico: (i) Identificar e compreender os principais aspectos conceituais relacionados com o desenvolvimento de EP, levantando as principais lacunas teóricas acerca deste tema. Este objetivo foi atingido através da elaboração de uma revisão sistemática, a qual abordou 151 publicações no período de 2005 a 2012 sobre EP. Esta revisão permitiu o entendimento detalhado do contexto da pesquisa proporcionou uma base adequada para o desdobramento da tese.

As principais lacunas teóricas identificadas foram: necessidade de melhorias gerenciais nos projetos de EP desenvolvidos; necessidade de esclarecimento entre as relações de EP, validação de métodos e estimativa da incerteza de medição; ausência de critérios objetivos para seleção de variáveis em testes de homogeneidade e estabilidade em EP; verificar a influência de procedimentos estatísticos robustos de análise de dados em EP e a relação dos mesmos com a distribuição de probabilidade do conjunto de dados dos laboratórios participantes das rodadas de comparação.

A partir de então, foram definidos os demais objetivos específicos voltados à aplicação prática da pesquisa: (ii) Analisar a relação existente entre a realização de EP, validação de métodos e a estimativa da incerteza de medição de ensaios e calibrações realizadas em laboratórios; (iii) Investigar a relação entre o desenvolvimento de EP e conhecimentos utilizados na gestão dos projetos das comparações interlaboratoriais; (iv) Propor e aplicar uma sistemática para identificar parâmetros representativos para execução de testes de homogeneidade e estabilidade conduzidos em EP; (v) Analisar a influência do tipo

de distribuição de probabilidade na avaliação de desempenho de laboratórios quando se trabalha com valor de consenso baseado nos dados dos participantes das rodadas de EP.

Para desenvolver a análise da relação existente entre EP, validação de métodos e a estimativa da incerteza de medição foi realizada uma pesquisa qualitativa com oito especialistas brasileiros e quatro internacionais (dos Estados Unidos, Portugal, França e China). O método utilizado abordou a técnica *web-Delphi* e resultou em um diagrama de relações que apresentou de forma explícita a conexão entre os fatores estudados. O diagrama desenvolvido facilitou o entendimento do pesquisador sobre a interface dos temas validação de métodos e incerteza de medição com os EP. Percebeu-se que os EP podem contribuir significativamente para melhoria da exatidão e precisão dos resultados e que de forma indireta contribuem para competitividade do laboratório no mercado, bem como para a manutenção dos seus serviços, além de poderem ser utilizados como apoio na validação de métodos e estimativa da incerteza de medição.

Na investigação da relação entre o desenvolvimento de EP e conhecimentos utilizados na gestão dos projetos das comparações interlaboratoriais, elaborou-se um estudo de caso com os provedores de EP acreditados do Brasil, o qual identificou as áreas mais carentes no desenvolvimento de projetos, bem como a relação da norma dos EP (ABNT NBR ISO/IEC 17043) com outras referências na área de gestão de projetos (PMBOK e ABNT NBR ISO10.006). As áreas que mais demandam melhorias são a gestão de riscos e custos. Também foi constatado que existe uma integração de áreas da gestão de projetos com as áreas de desenvolvimento de EP, em especial nas áreas de qualidade, recursos humanos, comunicação, aquisição e gestão do escopo.

Na proposta desenvolvida para identificar parâmetros representativos para execução de testes de homogeneidade e estabilidade conduzidos em EP foram utilizados dados reais de EP desenvolvidos pela Rede Metrológica RS em três diferentes áreas (vinhos, meio ambiente e carvão). Utilizou-se Análise de Componentes Principais para apoio na seleção das variáveis mais representativas dentro de cada EP. Foi criada uma equação para seleção de variáveis, combinando o peso dos autovalores obtidos com as cargas de cada item analisado. O score gerado pela equação foi avaliado criticamente por especialistas das áreas relacionadas com o EP, para validação do procedimento. A proposta foi considerada adequada para os três EP do estudo de caso, pois o procedimento desenvolvido permitiu melhor seleção dos analitos

críticos nos testes de homogeneidade e estabilidade, uma vez que a representatividade da variável escolhida foi priorizada com o uso da Análise de Componentes Principais e com a validação posterior de especialista.

Como uma contribuição final da tese, desenvolveu-se um estudo para analisar a influência do tipo de distribuição de probabilidade na avaliação de desempenho de laboratórios em EP quando se trabalha com valor de consenso baseado nos dados dos participantes. Esta pesquisa abordou dados reais de um EP acreditado pela CGCRE segundo a norma ABNT ISO/IEC 17043, sendo provido pela Rede Metrológica RS. O estudo analisou 2.034 dados, oriundos de 27 parâmetros, que foram complementados com simulação de Monte Carlo. Os resultados indicaram que os procedimentos estatísticos robustos que usam mediana e IQN e Média e Desvio Robusto (segundo algoritmo C proposto pela norma ISO13528) são equivalentes e não apresentam diferenças significativas no resultado geral da avaliação de desempenho dos laboratórios participantes em um EP. A pesquisa também identificou que o tipo de distribuição de probabilidade dos dados, que é correlacionado com o coeficiente de curtose, tem influência significativa na análise de desempenho dos laboratórios. Distribuições bimodais não são adequadas para determinação de valores de referência com base no grupo de participantes do EP, pois não fornecem resultados de desempenho confiáveis, mesmo com procedimentos robustos de análise de dados. Por outro lado, os procedimentos robustos de análise de dados se demonstraram adequados mesmo em distribuições assimétricas, o que é relevante, pois nem sempre é possível obter uma distribuição simétrica usando dados de uma rodada de EP.

Ao longo do desenvolvimento da pesquisa, houve algumas limitações. A revisão sistemática não considerou teses e dissertações sobre o tema e também não foram inseridos relatórios de EP de provedores nacionais e internacionais nos critérios de busca e seleção de conteúdo para ser analisado criticamente. O diagrama de relações desenvolvido no segundo estudo de caso contou com a participação de especialistas de organismos de acreditação de diferentes países, que cobriu os acordos de cooperação que envolve as Américas (relacionadas com o IAAC), a Europa (relacionada com o EA) e a Ásia (relacionada com o APLAC). Entretanto, a participação do especialista do APLAC, se comparada com os demais participantes da pesquisa, foi menor (apenas um especialista, sendo que 9 são do IAAC e 3 do EA), o que pode proporcionar uma tendência na aprovação do diagrama.

A pesquisa relacionada com a gestão de projetos nos provedores de EP teve a participação de 100% dos provedores acreditados na época que foi desenvolvida, entretanto os provedores atuam em áreas diferentes, o que pode causar certa diversidade nos tipos de projetos desenvolvidos por cada instituição. Houve dúvidas de alguns participantes ao responder o questionário e também o pouco conhecimento sobre gestão de projetos de alguns participantes dificultou a interpretação de alguns pontos da pesquisa.

Os dois últimos estudos de caso desenvolvidos foram aplicados na área de ensaios, não tendo sido validados ou testados para EP na área de calibração. As definições normativas para EP de calibração são menos complexas e mais simples, como a própria revisão sistemática demonstrou. As principais demandas de melhorias e situações mais críticas usualmente estão relacionadas com a área de ensaios, o que justifica o fato do enfoque da aplicação dos estudos de caso citados acima.

Dentro do contexto da proposta desta tese e reconhecendo as limitações da pesquisa, considera-se que os objetivos (geral e específicos) foram atendidos de forma adequada. A seção a seguir apresenta as propostas de trabalhos futuros, que foram identificadas ao longo da pesquisa.

3.2 SUGESTÕES DE TRABALHOS FUTUROS

A pesquisa é uma atividade que, à medida que se aprofunda, abre novas demandas. O conhecimento gerado também gera as novas lacunas.

Como sugestões de trabalhos futuros, propõe-se uma revisão sobre as práticas de avaliação de desempenho utilizadas por diferentes provedores de EP nos diferentes blocos dos acordos do ILAC (IAAC, EA, APLAC). Também recomenda-se o desenvolvimento de uma sistemática para gestão de custos e riscos específica para empresas que trabalham com EP.

Outra possibilidade de pesquisa futura identificada seria desenvolver uma proposta de protocolo para execução de testes de homogeneidade e estabilidade com os dados dos participantes das rodadas dos EP, desenvolvendo um teste por consenso. Este fato proporcionaria uma redução de custos, pois não seria necessário subcontratar um laboratório

específico para esta atividade. Além disso, 100% dos parâmetros analisados no EP seriam testados em relação à sua homogeneidade e estabilidade.

Por fim, propõe-se o desenvolvimento de um estudo para uso de dados históricos de EP objetivando gerar modelos matemáticos de previsão da variabilidade (desvio padrão) dos dados com base no valor de concentração obtido em uma rodada de comparação interlaboratorial na área de ensaios.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR ISO/IEC 17011**. Avaliação da Conformidade – Requisitos gerais para os organismos de acreditação que realizam acreditação de organismos de avaliação de conformidade. Rio de Janeiro, 2005.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR ISO/IEC 17025**. Requisitos gerais para competência de laboratórios de ensaio e calibração. Rio de Janeiro, 2005.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR ISO/IEC 17043**. Avaliação de conformidade — Requisitos gerais para ensaios de proficiência. Rio de Janeiro, 2011.

ASSOCIAÇÃO REDE DE METROLOGIA E ENSAIOS DO RIO GRANDE DO SUL (RMRS). **Requisitos sobre a participação de laboratórios em ensaios de proficiência**. Revisão 5. 2014.

ASSOCIAÇÃO REDE DE METROLOGIA E ENSAIOS DO RIO GRANDE DO SUL. **Relatório do Programa de EP em resíduos de Agrotóxicos**. MOD 1. Revisão 4. Porto Alegre, 2011.

BRASIL. MAPA. Ministério da Agricultura, Pecuária e Abastecimento. **Instrução Normativa nº 57**, de 11 de dezembro de 2013 - Estabelecer os critérios e requisitos para o credenciamento e monitoramento de laboratórios pelo MAPA. Disponível em: http://www.agricultura.gov.br/portal/pls/portal/!PORTAL.wwpob_page.show?_docname=1400717.PDF Acessado em 22 de janeiro de 2015.

BRASIL. ANVISA. Agência Nacional de Vigilância Sanitária. **Resolução – RDC n.º 12**, de 16 de fevereiro de 2012 – Dispõe sobre a Rede Brasileira de Laboratórios Analíticos em Saúde (REBLAS) - Estabelecer os critérios de habilitação de laboratórios. Disponível em: http://portal.anvisa.gov.br/wps/wcm/connect/17853e804b571c30bb15bbaf8fded4db/RDC12+de+2012.pdf?MOD=AJPERES Acessado em 22 de janeiro de 2015.

BREZIS, M. Warning: false sense of certainty from the illusion of accuracy and precision in measurements. **Accreditation and Quality Assurance Journal**. Vol. 16. 599–602p. 2011.

BRIDWELL, H. DHINGRA, V, PECKMAN, D. ROARK, J. LEHMAN, T. Perspectives on Method Validation: Importance of Adequate Method Validation. **Quality Assurance Journal**. Vol. 13, 72–77p. 2010.

COORDENAÇÃO GERAL DE ACREDITAÇÃO (CGCRE). **NIT DICLA 26** - Requisitos sobre a participação dos laboratórios de ensaio e de calibração em atividades de ensaio de proficiência. Revisão 8. 2011.

GIL, A. C. **Como Elaborar Projetos de Pesquisa**. São Paulo: ATLAS, 1991.

GUST, JEFF C. A discussion of stability and homogeneity issues in proficiency testing for calibration laboratories. In: **4th ILAC PTCG** – Vienna PTCG. 2007.

HOWERTON, D; KROLAK, J; MANASTERSKI, A; HANDSFIELD, J. Proficiency Testing Performance in US Laboratories. **Arch Pathol Lab Med Journal**—Vol 134, May 2010.

ILAC PROFICIENCY TESTING CONSULTATIVE GROUP (IPTCG). **Discussion Paper on Homogeneity and Stability Testing.** Diponível em http://www.ilac.org/documents/PTCG_HandS_April0808.pdf . Acesso em 14 de dezembro de 2011.

KISETS, D.. Performance indication improvement for a proficiency testing. **Accreditation Quality Assurarnce Journal**. Vol. 10. 461–465p. 2006.

LAKATOS, E; MARCONI, M. **Fundamentos de metodologia científica**. São Paulo: Atlas, 1993.

MOURA, M. **Mapeamento das demandas e ofertas por programas brasileiros de ensaios de proficiência no contexto da acreditação de laboratórios e provedores pela coordenação geral de acreditação (CGCRE).** Dissertação de Mestrado em Metrologia e Qualidade. INMETRO. 2014.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). **Directory of National Accreditation Bodies - NIST GCR 05-877**. United States of America, 2005.

PIZZOLATO, M. CATEN, C. JORNADA, J, A, H. A influência do sistema de gestão de laboratórios nos resultados dos ensaios de proficiência da construção civil. **Gestão e Produção**. v. 15, n. 3, p. 579-589, 2008.

PIZZOLATO, M. **Mapeamento da estrutura global que fornece confiança as medições: análise da inserção brasileira.** Tese de Doutorado do Programa de Pós Graduação em Engenharia de Produção (PPGEP). Universidade Federal do Rio Grande do Sul. 2006.

ROSA, D.; PIZZOLATO, M.; ALBANO, F. M.; SARTORI, T. . Analysis of nonconformyties of testing and calibration laboratory. In: **Metrologia 2011- Metrology as Factor of Quality, Innovation and Competitiveness.** Natal - RN. Book of abstracts, 2011.

TAWFIK, H. FATAH, A. ISO/IEC 17025 Accreditation: Between the Desired Gains and the Reality. **Quality Assurance Journal**. Vol. 13. 21–27p. 2010.

THOLEN, D. Metrology in service of society: the role of proficiency testing. **Accreditation and Quality Assurance Journal**. 1-9 p. 2011.

THOMPSON, M. Test for sufficient homogeneity in a reference material. **Royal Society of Chemistry Journal**. No 1. Jul 2004.

WONG, S (2011). Performance evaluation for proficiency testing with a limited number of participants. **Accreditation and Quality Assurance Journal**. 16:539–544. doi:10.1007/s00769-011-0816-8