Educational Aspects Of Proficiency Testing Programs

The operation and use of Proficiency Testing Programs

Proficiency testing programs (PTPs) have long been recognized as a tool that *accreditation bodies and regulatory authorities* (regulators) use to determine whether laboratories perform tests with acceptable results. PTPs can have additional purposes. This paper argues that PTPs should be perceived and overtly promoted as tools for laboratories to use as part of their quality assurance and continuous improvement efforts, and not simply a tool of regulators to pass or fail a laboratory. That is, PTPs have an educational component which, if realised, can lead to tangible improvements in laboratory performance. Since this is one of the underlying benefits of accreditation, there is an inherent link between the aims of laboratory accreditation and that of educationally- driven PTPs.

The process employed by traditional PTPs is outlined in Figure 1 below.

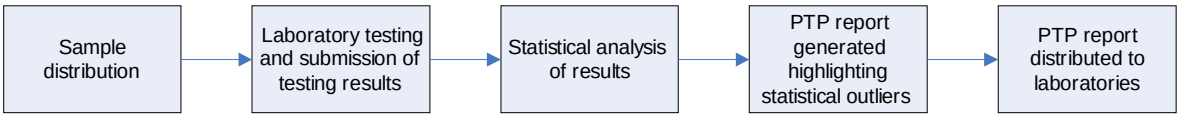


Figure 1: Traditional PTP process

In traditional PTPs, an acceptable result is often only determined as being one which satisfies a statistician in a regulatory body. Statisticians often work in isolation to industry requirements and avoid contentious or challenging areas. For example the measurement of parameters that approach limits of detection or the obsession of measuring “paired” sets of samples for Youden plots. There may be minimal input from technical advisors in terms of the evaluation of laboratory results.

Historically, educational aspects have generally been a by-product of PTPs and have not been considered as an integral component of the PTP framework. The reports coming from traditional programs have centred mainly on the results, and while they may have provided some additional information, the reports have not, in general offered feedback to laboratories or industry stakeholders, such as standards writing organisations, other than whether a given laboratory’s reported results are in line with the bulk of other participants.

To obtain educational and progressive PTP it is desirable that all relevant stakeholders utilizing laboratory data contribute to the structure and development of PTPs. Ultimately, the acceptability of the laboratory’s PTP results should be defined only after clarification as to whether the samples being assessed were appropriate to the industry served by the laboratory in question.

Indeed, fundamentally, the PTP should deliver what an industry requires. If this approach is conducted in a collaborative atmosphere the science of the analytical processes tested by the PTP will improve, as will overall improvements in laboratories’ performance in similar types of PTP.
The authors can cite examples where laboratories have been able to gain significant improvements in laboratory testing purely due to the fact that the design of the PTP was intended to include educational aspects. Please refer to the appendix for case studies.

PTPs that concentrate on the “results”, as typified by regulators, imply that it is only possible to either “pass” or “fail”. Participants in these types of programs also often find themselves in “an examination” situation, and are probably unlikely to handle the proficiency samples in their “normal” manner. Cases have been cited to the authors where only the “best analyst” is allowed to

handle the PTP samples in “traditional programmes”and in such cases, reagents are made and tested especially. The whole exercise merely demonstrates the results for these special circumstances, and does not necessarily reflect the quality of results reported in a day to day situation. In such cases, a PTP is incompatible with the quality assurance and continuous improvement philosophy cited earlier in this document.

While it is acknowledged that accreditation and regulatory checks are necessary in various circumstances, result-focussed PTP’s are of little benefit to either the laboratory or the PTP provider, other than to indicate the “best result” the laboratory can achieve. Sometimes scientific and statistical boundaries need to be challenged in order for the science to improve.

Proficiency Testing Programs as the ultimate quality assurance and continuous improvement tool.

Laboratories generally exist for a purpose, and that purpose is to test samples submitted by customers. (Customers can belong to the same organisation, for example production testing, or they can be external to the laboratory, as when private testing facilities seek samples from the public.) Customers have specific reasons to require testing of the laboratories, but invariably, the testing is specific to one or more areas of industry. The laboratory is responsible for ensuring that the test results meet their customers’ needs. A given PTP may or may not be appropriate for the routine tests the laboratory carries out, but historically, regulators have judged laboratory performance only on PTP performance. This is not appropriate, as a PTP result can only provide a snap shot of the real laboratory situation. Additionally, there is an assumption that the samples in the PTP are typical of the samples normally handled by the laboratory.

If a PTP is to be used as a quality assurance and continuous improvement tool, then participant laboratories need a true picture of their testing quality, together with information that allows them to identify and realise continuous improvement opportunities. This function is not for regulators to govern, but rather to assess at routine audits along with other aspects of the laboratory’s quality system. (NB: with exceptions that will be discussed later).

The value of PTP’s would be enhanced if the structure of a PTP could be designed to:

a) accommodate the requirements of the regulator when appropriate. i.e. punitive action be initiated by continued poor performance of a laboratory accredited for a particular test; (meaning, that a regulator can revoke a laboratory’s accreditation on assessment, or at earlier times when brought to the attention of the regulator. The reasons for this could be continuous poor performance in PTP, or because the PTP and/or other quality assurance tools are not being used in accordance with the requirements of the regulator. The role of the regulator is to assess the laboratory’s actions in the context of their accredited activities, and not to base judgement and action solely on PTP results.)
b) encourage new laboratories or laboratories seeking accreditation for a particular parameter to use PTP as a process of validating their new methodologies, or
c) encourage innovation, new and emerging techniques, perhaps in parallel with traditional methodologies and instrumentation,

This can be practically achieved for PTPs by allowing a laboratory to have the opportunity to demonstrate how they have dealt with their PTP reports in an “outcomes - focussed” manner. It must also be borne in mind that laboratories have many other quality requirements outlined in ISO/IEC 17025 that could easily be met by well designed and managed PTPs.

Hence, a well designed PTP will have many of the following features:

- Questionnaires submitted by the PTP providers to encompass methodology, testing details and testing approach in addition to reporting the results. This allows for a thorough analysis of results and provision of specific information regarding the effects of various aspects of the methodology.
- Testing of more than one sample in a single round. This allows for the development of a sound program design covering different analyte levels or other variables such as sample matrix. Laboratories can also then determine whether they have satisfactory repeatability/reproducibility for their routine capabilities.
- Designing a series of rounds in a scheme that allow laboratories to demonstrate all aspects of their capabilities. (A scheme can cover high and low levels of analytes, and also samples yielding typical and atypical reactions both due to the matrix effects and unrelated to a matrix. A laboratory in such a scheme may have outlying results for a small number of analyses. If the scheme is well designed, the laboratory will be able to determine the circumstances under which their results tend to be satisfactory, whether their analysis tends towards random or systematic errors, and which type of sample matrix testing they may need to develop further. A single round of PTP is unlikely to yield such information, and the snapshot result provided from a single round of PTP results will not allow the same assessment of capabilities as participation in a designed scheme.)
- Excess samples from PTPs can be used as Reference Samples or Certified Reference Materials if samples can be adequately categorised and made traceable to a stated reference.
- Allowing more than one analyst to participate in the program as individuals. This would assist participating laboratories to determine whether their “differences in results” are analyst based or reagent/equipment based. The PTP should make provision for multiple results for each test being evaluated.
- Reports include sufficient data analysis to provide reasonable information for laboratories to complete corrective actions. It should be noted that laboratories often participate in PTPs that are not strictly suitable for the testing they normally perform, but participate because the program is either mandated by a regulator, or because no other more suitable program is available. It is known that certain testing methodologies are suitable for some sample types, accuracy requirements and analyte levels, but less so for others. A laboratory “competing” with others using different methods may yield a “bad” report, but this may be purely due to their method (which might be quite suitable for their own testing).
- Reports should be provided in a timely manner to allow follow up activities to be carried out while the operator still has good recollection of the testing activity.
- PTPs should be operated with sufficient frequency to allow the lab to repeat a “blind test” after carrying out corrective actions.
- There should be sufficient resources available from the PTP provider to render assistance to participants that find themselves “lost”.
- Results from PTPs may provide guideline values on what is a reasonable estimate for the "top down" approach in the Measurement of Uncertainty (MU) at the laboratory level. MU for specific sample types and can be linked to different processes associated with measurement and can include items or variables likely to influence a measurement e.g. reagents and instrumentation. This approach to the estimation of MU provides a pragmatic and consistent approach to the participants of a PTP and hence the value of MU as an industry measure can have has some real significance. Current process used to calculate MU can vary significantly within an industry group and it's value is questionable and often misused. For example, two laboratories producing data of a similar quality for a particular parameter may use different methods to estimate MU and therefore derive significantly

different MU for that parameter. These differences could be instrumental in gaining commercial advantage in tendering for contractual work.

- The program should cover as much as possible of the testing process, including, sample receipt, storage, log in and preparation reporting, calculations and checking of results.
- Follow-up workshops that focus on the results and outcomes of PTPs is a powerful way to advance the lessons learnt and provides a non-threatening forum to engage the thoughts of stakeholders. The authors of this paper and other PTP providers have seen first hand the enthusiasm of workshop participants in engendering new thoughts and processes in improving PTPs. This often occurs in a socially enhanced atmosphere of anonymity, especially when workshops are extended beyond a single day.

Regulator focus should be on the acceptability of routine testing work. If laboratories participate in PTPs with sufficient frequency, they should begin to be treated as "routine" by laboratory staff. The outcome of PTPs should be circulated to the staff, not with threats, but in an open and constructive manner. Further, the attitude of regulators can assist in promoting PTP as an activity that is expected to be seen as routine.

While PTP results can provide some objective evidence that "all is going well" with laboratories, the management system of the participant laboratories should cover how PTP reports are treated. The main objective of the assessment of PTPs by regulators should really be to see what has been done with the reports, as well as on the results themselves. The corrective actions resulting from a PTP should not have greater emphasis than corrective actions arising for any other reason.

Authors
Dan Wruck, Queensland Health Scientific Services (Brisbane, Australia)
Ingrid Flemming, IFM Quality Services (Sydney Australia)

Revision:20060921

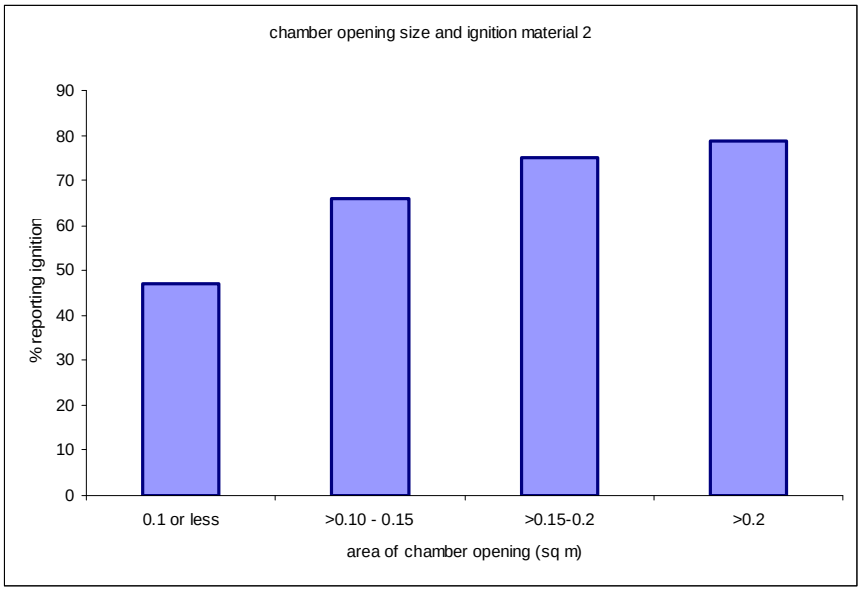
Appendix

Case Study 1

Electrical Testing Laboratory obtained a "fail" for an ignition test on a plastic using a needle flame apparatus.

The questionnaire in the proficiency program asked for detailed information about the apparatus and the housing of the apparatus. Some of these questions were general and did not constitute part of the cited standard.

The analysis of the data isolated laboratories that had met the requirements of the standard and compared more general aspects of their practices in addition to actual "choices" allowed in the standard. A relationship between the area of the opening in the housing and the proportion of laboratories obtaining ignition was found.



The technical advisers for this program reviewed the data and recommended that in future, laboratories should minimise the area of the opening to make test results more consistent.

The laboratory in question yielded to this advice and closed the opening as far as possible, noting that the expected results for ignition were now obtained on re-test of samples.

A program focussing on the result would not have assisted this laboratory to improve their performance.

Further, the latest amendment of the standard for this test now recommends the above practice.

Case Study 2

Electrical Test laboratory obtained a fail for heating test of a transformer using the change in resistance method of analysis.

The cited method uses the principle that the temperature of an inaccessible coil of wire can be determined by calculation measuring the change of resistance of the wire as it heats. The resistance is measured over a period of time after disconnection of the circuit and the cooling curve is extrapolated back to time = 0 to obtain the resistance at the point of disconnection of the circuit, which is used in the calculation. Typically, the curve assumes the shape of radioactive decay.

The laboratory reported statistically unacceptable results.

The program had been designed to collect the key raw data generated in determination of the results together with thermocouple measurements on parts of the sample. The program report allowed the laboratory to compare the shape of their cooling curve with those of other laboratories, and determined that both their input power and their measurement of resistance was comparable to other laboratories. (These factors obviously being critical in the measurement.)

However, they could not as easily assess the zero point of their curve because their first measurement of resistance occurred after 30 seconds, while most other laboratories could obtain their first resistance measurement within 3 seconds.
The equipment they had, while extremely accurate, was not suitable for this type of determination, as it took too long to obtain a stable reading. They changed equipment and have yielded comparable results since that time.

A program focussing on the result would not have assisted this laboratory to improve their performance.

Case Study 3

Food microbiology testing facilities failed to detect *Salmonella* in their samples.

This program consisted of testing multiple samples, and many test organisms additional to *Salmonella* are included.

Salmonella, while a common food poisoning agent, is seldom found in foods without suitable pre-testing steps to enhance their recovery. The traditional test involves first enriching all the organisms in the sample to increase the numbers present, then performing a selective enrichment in growth media containing agents that inhibit some of the more common flora, thus increasing the relative proportions of *Salmonella* in the sample. After these steps, the enriched sample is cultured onto selective growth agars, which, by the characteristics of the colonies formed, facilitate recognition of the organism. This process takes several days. Different national standards employ this principle of testing, but recommend various combinations of primary and secondary enrichment media. (Meaning that there are a large number of choices, even within one national standard.) Manufacturing companies often are unable to employ traditional methods due to shelf life restrictions and testing facilities then must employ a rapid method. (These are usually enzyme based or antibody based methods) Hence, there are literally hundreds of ways to look for *Salmonella* in foods. Further, there are thousands of strains of *Salmonella*, and not all strains behave the same way, meaning that any method used will not detect all *Salmonella*. It is impossible to validate every possible method of detecting *Salmonella* for every food sample coming into the laboratory. This is particularly true for commercial consulting laboratories who cannot predict that samples will be arriving that day.

The proficiency program was designed to allow all testing methods to be used in order that more information could be gained. A detailed analysis of methods and media was presented in the report. In all, 30% of the participant laboratories failed to detect *Salmonella* in one or more of the samples.

- Since multiple samples and tests were involved in the proficiency testing program, correct enumeration of other organisms in the samples could confirm successful sample transit and storage.
- Laboratories could compare their detection of *Salmonella* from samples with low numbers of competing flora, with those where higher levels of competing flora existed. As a result, some laboratories were able to identify that their enrichment techniques were not optimal, for the sample type, particularly when competition existed. This raised awareness of the need to perform parallel tests with spiked samples of the same food, particularly when the content is unknown.
- Some laboratories using rapid tests failed to detect *Salmonella*, and this was linked to their primary sample set up.
- One particular brand of rapid test was identified as not being able to detect two of the strains of *Salmonella* included. Laboratories in this situation then needed to establish the likelihood of these particular strains in their regularly tested foods. Two participants replied that they were changing brands of rapid test kit as a result.
- One of the *Salmonella* included was very slow to produce hydrogen sulphide (a typifying trait of *Salmonella*). One of the common growth agars that specifically look for H2S production therefore did not easily display the target organisms. Laboratories in this situation were then able to ensure that their second media focussed on other traits of *Salmonella* growth in addition to H2S production.

It should be noted that the participants, while there was a high “fail” rate, appreciated the information and constructive reporting fashion.

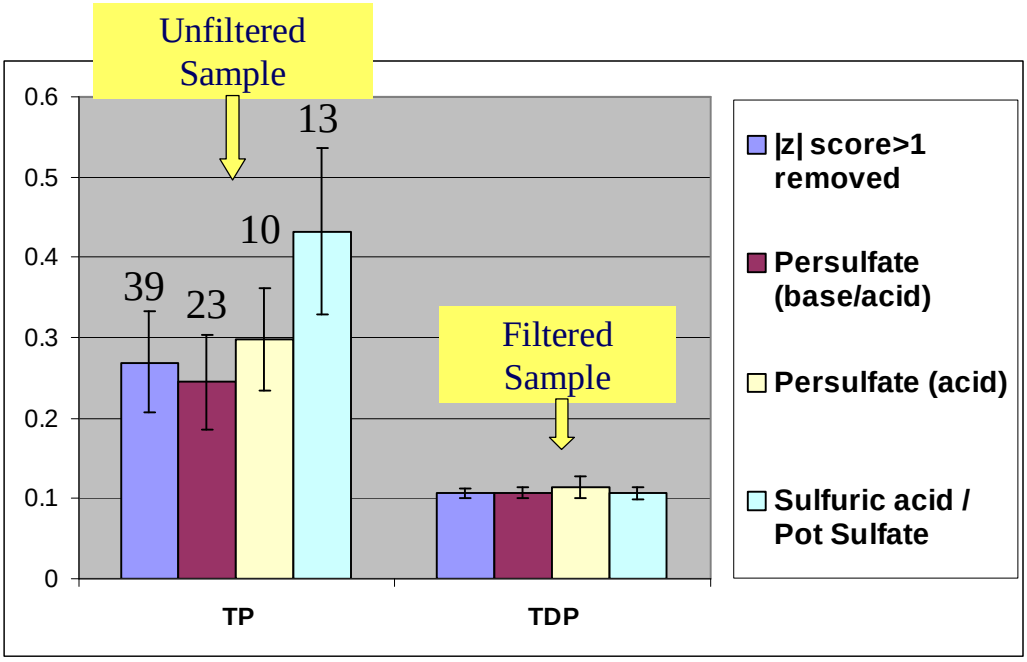
A program focussing on the result would not have assisted these laboratories to improve their performance.

Case Study 4

New Digestion Procedure Underestimates Phosphorus Loads in Waterways.

In the past 10 years the digestion procedure used to measure total phosphorus (TP) has moved away from a Kjeldahl based method to a persulfate based method. Both these methods can provide equivalent results in pristine or waters containing organic matter. However, when waters contain high particulate loads of soil origin, variable results are obtained. This situation commonly occurs in floods when large loads of sediments are transported down rivers. The consequences can be devastating and lead to the death of seagrass and corals. In other situations, managers use TP data to monitor improvements/degradation of waterways over periods of time.

The following figure graphically demonstrates data obtained from a major Australian river. All techniques obtained equivalent results when measuring total dissolved phosphorus (TDP) from the filtered portion of this water. Results from the unfiltered sample varied considerably based on which digestion procedure was used. Recoveries of 60% were obtained by the persulfate (base/acid) method relative to the sulfuric acid/potassium sulfate (Kjeldahl based) method.



Assumption:

- There has been an improvement in nutrient loads over past 10 years - it appears nutrient concentrations have declined – management practices are effective.

Reality:

- There has been a change in digestion procedures from Kjeldahl type to persulfate digestion procedures in the last 10 years.
- Data bases currently DO NOT differentiate how TP is measured

The design of the PT program was critical in quantitating the differences in digestion techniques. A program focussing on the result only, would not have identified issues relating to changes in analytical methodology.