
EVALUASI PENGGUNAAN NILAI FORMULASI PADA UJI PROFISIENSI AIR PERMUKAAN

Evaluation of The Use of Formulation Value in Surface Water Proficiency Testing

Yohanes Susanto Ridwan dan Willy Cahya Nugraha

Loka Penelitian Teknologi Bersih LIPI, Komplek LIPI Gd. 50, Jl. Sangkuriang Bandung, Jawa Barat, Indonesia
e-mail: yo.susantoridwan@gmail.com

Diterima: 28 Juni 2019, Direvisi: 7 Oktober 2019, Disetujui: 10 Maret 2020

Abstrak

Penggunaan nilai konsensus pada uji profisiensi (UP) saat ini merupakan praktik umum di Indonesia. Evaluasi kinerja yang didasarkan pada nilai konsensus sangat tergantung pada kualitas data-data dari laboratorium peserta dan kadangkala gagal menilai kinerja peserta. Dalam pekerjaan ini, uji profisiensi dilakukan menggunakan nilai formulasi sesuai dengan panduan ISO / IEC 17043 dan ISO 13528. Dari 5 parameter yang diujikan, 4 parameter (Fe, Cu, Cr(VI), pH) menggunakan formulasi sedangkan 1 parameter (NO₃-N) menggunakan nilai konsensus sebagai nilai acuan. Diketahui bahwa penggunaan nilai formulasi efektif digunakan sebagai dasar penilaian kinerja peserta. Bahkan untuk parameter Cu dan Cr(VI), nilai formulasi dapat menghindarkan provider dari kebuntuan pengolahan data akibat data-data yang sangat bervariasi. Peserta yang mendapatkan penilaian kinerja memuaskan untuk parameter Fe, Cu, Cr(VI), dan pH adalah berturut-turut sebesar 81%, 79%, 71%, dan 67% dari jumlah peserta; yang diperingatkan sebesar 7%, 6%, 8%, dan 16%; yang berkinerja tidak memuaskan sebesar 12%, 15%, 21%, dan 17%. Hal ini mengkonfirmasi bahwa skema UP menggunakan nilai formulasi memiliki kemampuan baik untuk mengevaluasi kinerja laboratorium. Evaluasi bias terhadap hasil UP menunjukkan bahwa tidak ada bias yang signifikan. Disimpulkan bahwa laboratorium yang berpartisipasi memiliki kesesuaian yang baik terhadap nilai formulasi, sehingga memberikan tingkat kepercayaan tambahan terhadap kualitas pengukuran dalam bidang analisis ini. Untuk parameter uji NO₃-N, nilai formulasi tidak tersedia sedangkan nilai konsensus pun tidak dapat ditentukan akibat bervariasinya data-data peserta sehingga penilaian kinerja tidak dapat dilakukan. Diusulkan alternatif upaya yang dapat dilakukan untuk mengatasi masalah ini agar nilai formulasi tetap dapat ditetapkan dan digunakan sebagai nilai acuan.

Kata kunci: air permukaan, uji profisiensi, nilai acuan, nilai formulasi, nilai tertelusur

Abstract

Use of consensus values in proficiency testing (PT) are common in Indonesia. Performance evaluation based on consensus often failed to evaluate performance of participating laboratories. In this work, PT was made use of formulation according to ISO/IEC 17043 and ISO 13528. Out of 5 tested parameters, 4 of them (Fe, Cu, Cr(VI), pH) were used formulation while 1 parameter (NO₃-N) was used consensus as assign value. Based on the statistical results, it was confirmed that use of formulation was effective to evaluate laboratories performance. In addition, for Cu and Cr(VI), use of formulation value was successfully made provider avoided risk of being unable to made performance evaluation because of the high level of variation of results. Percentage of participants having satisfied result for Fe, Cu, Cr(VI), and pH were 81%, 79%, 71%, and 67%; warning were 7%, 6%, 8%, dan 16%; and unsatisfied were 12%, 15%, 21%, and 17% respectively. This results confirmed that PT scheme using formulation was effective to make performance evaluation. Post evaluation for the bias of PT showed that there were not any significant bias in the PT round. It could be concluded that the participating laboratories have a good agreement with formulation value, hence provide more confident to the quality of the laboratories. For NO₃-N, formulation value was not available and also consensus could not obtained due to high variation of results, hence, performance evaluation has not been able to performed. In this paper, the suggestion of alternative solution is presented.

Keywords: assign value, formulation value, proficiency testing, surface water, traceable value

1. PENDAHULUAN

Validitas hasil pengujian merupakan hal yang sangat penting karena hasil uji laboratorium

seringkali digunakan oleh para pemangku kepentingan dalam membuat keputusan strategis yang mempunyai dampak besar di bidang ekonomi, lingkungan, kesehatan,

keamanan, dll. Oleh karena itu upaya pengendalian mutu merupakan hal mutlak yang tidak dapat dihindari oleh laboratorium pengujian yang berfokus pada mutu. Peningkatan intensif jumlah laboratorium pengujian yang terakreditasi oleh Komite Akreditasi Nasional (KAN) Indonesia dalam dua dekade terakhir telah mempromosikan dan memperkuat penerapan pengendalian mutu internal maupun eksternal di laboratorium pengujian di Indonesia. Uji profesiensi (UP) memainkan peran penting karena menyediakan sarana pengendalian mutu eksternal bagi laboratorium dimana kinerja dari satu laboratorium dibandingkan dengan laboratorium lainnya. Sejalan dengan intensifnya upaya pemerintah dalam hal pengendalian lingkungan maka aktifitas pengujian lingkungan sangat masif dilakukan oleh laboratorium-laboratorium dinas, kementerian, dan lembaga milik pemerintah, perguruan tinggi, dan swasta. Upaya pemantauan kinerja laboratorium melalui kegiatan UP di bidang pengujian lingkungan merupakan alat penting yang sangat dibutuhkan untuk memonitor kinerja laboratorium.

Artikel ini memaparkan hasil evaluasi dari penyelenggaraan UP di bidang pengujian lingkungan, khususnya air permukaan yang diselenggarakan oleh Pusat Penelitian Kimia LIPI pada tahun 2010, dimana pada saat itu telah digunakan nilai formulasi sebagai nilai acuan. Parameter yang diujikan adalah besi (Fe), tembaga (Cu), krom heksavalen (Cr(VI)), nitrogen dari nitrat (NO₃-N), dan pH. Konsentrasi dari kelima parameter uji tersebut ditetapkan dengan mengacu pada baku mutu PP 82 tahun 2001. Pada kegiatan UP tersebut, 4 parameter diolah menggunakan nilai formulasi sebagai acuan, yaitu untuk Fe, Cu, Cr(VI), dan pH. Sedangkan untuk NO₃-N digunakan nilai konsensus dikarenakan nilai formulasi tidak dapat ditentukan secara akurat.

Program UP yang dilakukan di Indonesia setidaknya hingga tahun umumnya menggunakan nilai konsensus. Berdasarkan Laporan-laporan hasil UP KAN yang diselenggarakan periode 2005-2010, 100% pengolahan datanya menggunakan nilai konsensus. Ini mencakup program UP KAN mulai dari UP KAN VIII/2005 (air limbah), KAN IX/2006 (air minum dalam kemasan), KAN X/2007 (tepung ikan, bubur bayi tepung maizena, mikrobiologi), KAN XI/2008 hingga

KAN XIII/2010 (mikrobiologi) tercatat menggunakan nilai konsensus. Demikian juga program UP LIPI sebelum 2010 untuk komoditi tepung beras, tepung kacang hijau, dan sereal juga menggunakan nilai konsensus. Penggunaan nilai formulasi dalam program UP air permukaan membuat UP ini memberikan kontribusi tersendiri bagi upaya peningkatan kualitas pemantauan kinerja laboratorium pengujian di Indonesia karena menggunakan nilai acuan yang independen dan tidak bergantung pada hasil-hasil peserta.

2. TINJAUAN PUSTAKA

Dalam rangka mengendalikan pencemaran dan menjaga kelestarian lingkungan, pemerintah telah mengeluarkan berbagai regulasi yang membatasi pencemaran lingkungan baik di badan air, tanah, maupun udara. PP 82 tahun 2001 tentang baku mutu air permukaan (akuifer, mata air, rawa, sungai, danau, waduk, muara), Permen LH No 5 tahun 2014 tentang baku mutu air limbah industri, Kepmen LH 13 tahun 1995 tentang baku mutu emisi sumber tidak bergerak, PP 41 tahun 1999 tentang baku mutu udara ambien, merupakan beberapa contoh dari banyak regulasi pemerintah yang dijadikan pedoman dalam pengendalian pencemaran dan pengelolaan lingkungan. Masifnya kegiatan pengujian yang dilakukan oleh laboratorium-laboratorium di Indonesia perlu didampingi oleh kegiatan uji profesiensi sebagai alat monitoring kinerja laboratorium secara sektoral ataupun nasional.

Kegiatan uji profesiensi sangat penting untuk menunjang keberhasilan implementasi sistem jaminan mutu nasional (Tholen, 2011). UP juga merupakan elemen penting untuk memastikan validitas pengujian yang dilakukan oleh banyak laboratorium yang berbeda-beda (Ponomareva & Shpakov, 2012). Badan akreditasi juga dapat menggunakannya sebagai indikator kinerja laboratorium di bidang pengujian tertentu. Manfaat lain dari perbandingan antar laboratorium adalah dapat memberikan informasi historis kinerja satu laboratorium pada jangka waktu tertentu, yang akan bermanfaat bagi peningkatan kualitas laboratorium itu sendiri (Sykes, 2012).

Partisipasi dalam skema UP diperlukan dan bermanfaat bagi semua laboratorium

pengujian. Banyak skema UP telah dilakukan di seluruh dunia dalam berbagai bidang pengujian, seperti di bidang analisis lingkungan (Asiah, 2017), analisis air (Kim *et al.*, 2010), sifat fisik (Hegazy *et al.*, 2010), serum manusia (Carter *et al.*, 2010), klinis (Tamiru *et al.*, 2015), mikrobiologi (Smith, 2015), dan makanan (Dajay *et al.*, 2019). Di negara berkembang, di mana sejarah penerapan sistem manajemen mutu belum terlalu lama dibandingkan negara maju, keberadaan skema UP reguler sangat signifikan dan bermanfaat. Namun, ada beberapa kendala yang dihadapi oleh *provider* UP dalam mengorganisasi kegiatan UP, seperti keterjangkauan biaya partisipasi, transportasi sampel UP, dan masalah bea cukai untuk UP regional dan internasional (Koch, 2018). Di bidang analisis tertentu, jumlah peserta sangat terbatas dan ini menambah kesulitan lain pada skema UP tersebut (Wong, 2011). Penyedia UP juga harus menyadari potensi kolusi dan pemalsuan data oleh peserta, karena akan mengurangi kepercayaan para pemangku kepentingan (Briggs, 2018; Mianes and ten Caten, 2017).

Provider UP biasanya akan berurusan dengan data mentah yang berlimpah serta pemrosesan data dalam rangka untuk melakukan evaluasi kinerja, oleh karena itu, penerapan prinsip dan teknik statistik tidak dapat dihindari. Metode statistik memainkan peran penting dalam hasil UP (Hegazy *et al.*, 2010). Dua variabel penting UP adalah nilai acuan (X_{UP}) dan target standar deviasi (σ_{UP}). Ada beberapa cara untuk mendapatkan nilai acuan sebagaimana dijelaskan dalam ISO 13528, yaitu: 1) nilai formulasi, dimana sampel UP merupakan sampel sintetis sehingga nilai acuan diturunkan dari perhitungan gravimetrik dan/atau volumetrik dari standar primer, 2) nilai tersertifikasi, dimana sampel UP merupakan bahan acuan bersertifikat, 3) nilai referensi, dimana sampel UP merupakan sampel alami yang ditetapkan nilainya dengan didampingi bahan acuan bersertifikat, 4) konsensus laboratorium pakar, dimana nilai konsensus dari sekumpulan laboratorium pakar digunakan sebagai acuan, dan 5) konsensus peserta, dimana nilai acuan diambil dari nilai konsensus hasil-hasil laboratorium peserta UP.

Penggunaan nilai konsensus peserta pada UP telah menjadi skema yang populer di

Indonesia, serta di banyak negara berkembang lainnya (Beckert & Fischer, 2018 ; Darie & Burian, 2018), namun, sifat ketertelusurannya merupakan yang terendah di antara nilai acuan lainnya. Penggunaan nilai formulasi akan berfungsi lebih baik dalam mengevaluasi kinerja laboratorium peserta (Cheng *et al.*, 2014 dan De Bièvre, 2012). Di wilayah Asia-Pasifik, sekarang ada peningkatan kebutuhan akan nilai acuan yang tertelusur untuk digunakan dalam skema uji profisiensi (Kaarls *et al.*, 2017).

3. METODE PENELITIAN

3.1. Persiapan sampel UP

Sampel uji profisiensi terdiri dari 2 jenis, A dan B. Sampel A merupakan larutan sintetik dengan komposisi matriks menyerupai matriks air sungai (dari segi mineralnya) sebagaimana tersaji dalam Tabel 1, dan diperkaya (*spiked*) dengan ion logam Fe, Cu, dan nitrat dalam suasana asam pada konsentrasi tertentu untuk menstabilkan analit di dalamnya. Sedangkan larutan B diperkaya dengan Cr(VI) pada suasana basa pada pH tertentu.

Tabel 1 Komposisi matriks sampel UP.

Komponen	Konsentrasi	Satuan
Na	1,1	mg/L
K	0,3	mg/L
Ca	2,2	mg/L
Mg	0,65	mg/L
Fe	0,1	mg/L
Mn	0,02	mg/L
Al	0,01	mg/L
Cl	0,8	mg/L
SO ₄	1	mg/L

Konsentrasi analit target (X_{UP}) pada sampel uji profisiensi A dan B, beserta nilai ketidakpastiannya ($\mu(X_{UP})$) disajikan dalam Tabel 2. Nilai standar deviasi uji profisiensi (σ_{UP}) diambil dari standar deviasi reproduibilitas yang diturunkan dari model Horwitz.

Tabel 2 Nilai formulasi paramater uji pada sampel A dan sampel B (baris abu-abu).

Parameter Uji	X_{UP}	$\mu(X_{UP})$	σ_{UP}
Fe (mg/L)	1,94	0,031	0,2809
Cu (μ g/L)	20,8	0,27	6,0099
Cr(VI) (μ g/L)	25,3	0,25	7,1348
NO ₃ -N (mg/L)	-	-	9,6059
pH	8,60	0,60	0,130

Untuk memberikan panduan yang jelas bagi peserta, satu set petunjuk teknis disiapkan sebagai acuan laboratorium peserta dalam hal penanganan sampel UP, pengujian, pelaporan hasil, tanggal dan mekanisme pelaporan hasil, serta menyediakan form-form untuk merekam dan melaporkan kondisi sampel, identitas metode uji yang digunakan, dan hasil uji.

3.2. Uji homogenitas dan stabilitas

Dalam rangka uji homogenitas, 10 sampel acak dianalisis secara duplikat, yang akan menghasilkan 10 pasang data (simplo dan duplo, Tabel 3). Pengolahan statistik dilakukan untuk mengekstraksi variansi data dalam duplikat (*within sample*) dan variansi data antar sampel (*between sample*), sehingga inhomogenitas, S_s , dapat diperkirakan. Sampel dianggap memiliki homogenitas yang cukup ketika $S_s \leq 0,3 \times \sigma_{UP}$, di mana pada kondisi ini inhomogenitas tidak akan mempengaruhi evaluasi kinerja peserta. Prinsip yang sama diterapkan untuk uji stabilitas, di mana nilai rata-rata yang diperoleh dalam uji homogenitas, Y_{Hom} , dibandingkan dengan nilai rata-rata yang diperoleh dalam uji stabilitas, Y_{Stab} (biasanya dilakukan dengan mengukur setidaknya 3 arsip sampel dalam secara duplikat setelah periode waktu tertentu). Perbedaan absolut antara keduanya akan menjadi dasar untuk mengevaluasi stabilitas menurut persamaan $|Y_{Hom} - Y_{Stab}| \leq 0,3 \times \sigma_{UP}$, di mana jika kondisi tersebut dipenuhi maka ketidakstabilan sampel tidak akan mempengaruhi evaluasi kinerja peserta, yang berarti sampel tersebut layak.

3.3. Skema UP

Peserta UP terdaftar terdiri dari laboratorium dari instansi pemerintah, perguruan tinggi, dan swasta. Tercatat jumlah laboratorium yang mengikuti UP untuk parameter Fe, Cu, Cr(VI), NO_3-N , dan pH adalah berturut-turut 84, 70, 69, 75, dan 88 laboratorium.

Setelah uji homogenitas, diselenggarakan pertemuan perdana untuk mendistribusikan sampel UP kepada para peserta, serta untuk memperkenalkan skema UP, jadwal

pelaksanaan, petunjuk teknis mengenai metode analitik, tanggal analisis, dan tanggal pelaporan hasil.

Skema untuk evaluasi statistik UP dilakukan berdasarkan ISO 13528: 2005. Nilai acuan, X_{UP} , adalah nilai yang digunakan sebagai acuan untuk mengevaluasi bias hasil uji dari masing-masing laboratorium peserta, sedangkan σ_{UP} digunakan untuk mengukur seberapa besar bias bias tersebut dengan menggunakan metode penilaian standar (*z score*). Data-data yang dikumpulkan dari peserta diseleksi terlebih dahulu baik secara visual untuk mengamati pola distribusi data yang abnormal, ataupun menggunakan uji *Grubbs* untuk menyaring data-data pencilan.

Dalam skema UP ini, nilai-nilai yang berasal dari formulasi diambil sebagai X_{UP} , sedangkan σ_{UP} diambil dari standar deviasi reproduibilitas yang diturunkan dari model Horwitz. Tabel 2 menyajikan kompilasi dari nilai-nilai acuan (X_{UP}) dan standar deviasi target untuk semua parameter uji (σ_{UP}), serta nilai estimasi ketidakpastian baku dari nilai acuan $\square(X_{UP})$. Estimasi ketidakpastian baku dilakukan dengan memperhitungkan sumber-sumber ketidakpastian pada tahap preparasi secara gravimetrik dan volumetrik dari analit. Evaluasi kinerja laboratorium dilakukan dengan menggunakan *z score*, yang diperoleh dengan menghitung rasio antara perbedaan hasil masing-masing laboratorium dengan nilai acuan, dengan standar deviasi target, seperti yang dinyatakan dalam persamaan $z_i = (x_i - X_{UP}) / \sigma_{UP}$. Berdasarkan *z score*, kinerja masing-masing laboratorium dapat dievaluasi. Interpretasi konvensional *z score*, menurut ISO/IEC 17043: 2010 adalah sebagai berikut: a) hasil yang memberikan $|z| \leq 2,0$ dianggap memuaskan, b) hasil yang memberikan $2,0 < |z| < 3,0$ dianggap berstatus diperingatkan, dan c) hasil yang memberi $|z| \geq 3,0$ dianggap tidak memuaskan. Laboratorium yang memiliki kinerja tidak memuaskan diwajibkan melakukan investigasi terhadap penyebab masalah yang dihadapi dan memulai tindakan korektif untuk memulihkan sistem pengujian mereka.

Tabel 3 Pengolahan data uji homogenitas untuk paramater Fe (mg/L) dan Cu (µg/L).

Kode Botol	Simplo ($X_{t,1}$)		Duplo ($X_{t,1}$)		X_t		$(X_t - X_r)^2$		w_t^2	
2	1,959	19,297	1,968	21,081	1,9635	20,1892	0,0010	3,2741	0,0001	3,1819
11	1,949	21,081	1,949	21,081	1,9490	21,0811	0,0003	0,8419	0,0000	0,0000
23	1,947	21,081	1,932	21,081	1,9394	21,0811	0,0001	0,8419	0,0002	0,0000
37	1,913	22,865	1,908	21,081	1,9104	21,9730	0,0004	0,0007	0,0000	3,1819
42	1,983	21,081	1,980	21,081	1,9816	21,0811	0,0025	0,8419	0,0000	0,0000
47	1,915	23,081	1,906	24,865	1,9104	23,9730	0,0004	3,8980	0,0001	3,1819
56	1,942	22,297	1,935	24,865	1,9382	23,5811	0,0000	2,5041	0,0001	6,5924
64	1,903	21,081	1,913	24,865	1,9080	22,9730	0,0005	0,9493	0,0001	14,3170
69	1,901	21,081	1,908	24,865	1,9044	22,9730	0,0007	0,9493	0,0001	14,3170
75	1,901	21,081	1,913	21,081	1,9068	21,0811	0,0006	0,8419	0,0001	0,0000
G	10	10						$S W t^2$	0,0008	44,7721
Xr	1,931	21,999						$S x^2$	0,0007	1,6604
Sw	0,006	1,4962						$S s$	0,0270	0,7356
								$0.3 \cdot \sigma_{UP}$	0,0843	1,8030

Keterangan: Area sel tanpa warna adalah untuk Fe dan warna abu-abu untuk Cu.

4. HASIL DAN PEMBAHASAN

4.1. Homogenitas dan stabilitas sampel

Uji homogenitas dan stabilitas dilakukan segera setelah selesainya preparasi sampel UP, dan dilakukan terhadap analit Fe dan Cu sebagai perwakilan dari kelima analit. Beberapa seri pengukuran telah dilakukan dalam rangka uji homogenitas tersebut.

Pengolahan statistik dilakukan untuk menghitung standar deviasi antar samples, S_s , sebagaimana dapat dicermati pada pengolahan data uji homogenitas untuk Fe dan Cu pada Tabel 3. Dari pengolahan data tersebut diperoleh nilai S_s sebesar 0,027 dan nilai ini lebih kecil dibandingkan dengan kriteria homogenitas ($0,3 \times \sigma_{UP}$) sebesar 0,084, sehingga disimpulkan bahwa sampel UP mempunyai tingkat homogenitas yang memadai untuk keperluan UP. Dengan cara yang sama dilakukan uji homogenitas untuk Cu yang memberikan nilai S_s sebesar 0,7356 dan masih lebih kecil dari kriteria homogenitas ($0,3 \times \sigma_{UP}$) sebesar 1,80.

Sekitar satu bulan (tepatnya 42 hari) setelah preparasi sampel, uji stabilitas dilakukan dengan melakukan pengujian duplikat terhadap 3 arsip sampel. Untuk analit Fe, selisih dari nilai rata-rata uji homogenitas dan stabilitas, $|Y_{Hom} - Y_{Stab}|$ adalah sebesar 0,061 dan nilai ini masih di bawah kriteria stabilitas sebesar 0,084.

Demikian juga untuk Cu, dimana nilai $|Y_{Hom} - Y_{Stab}|$ diperoleh sebesar 0,004 yang masih jauh lebih kecil dari kriteria stabilitas ($0,3 \times \sigma_{UP}$) sebesar 1,80. Disimpulkan bahwa sampel UP menunjukkan tingkat homogenitas dan stabilitas yang memadai dan layak untuk keperluan UP.

4.2. Hasil evaluasi kinerja

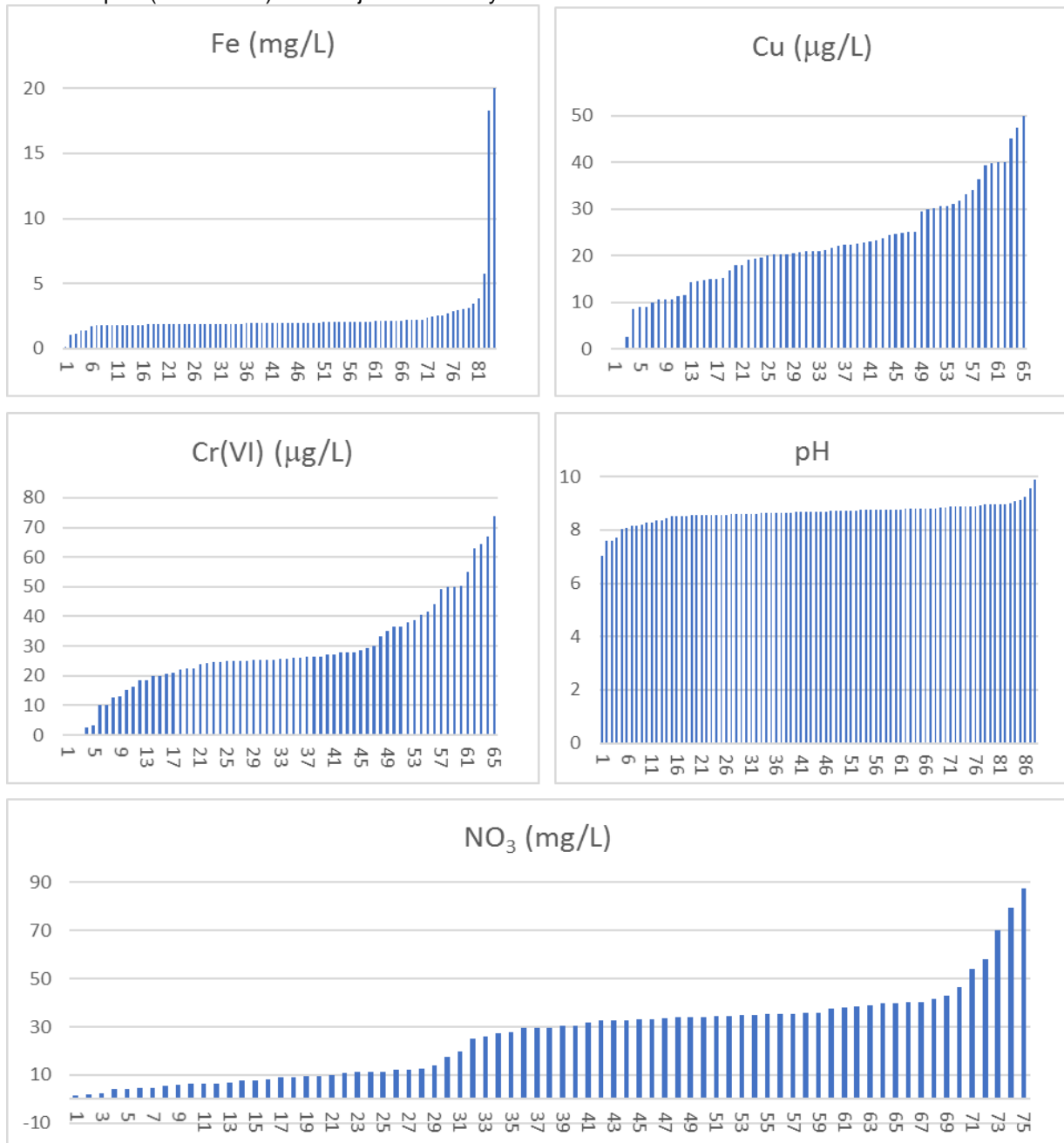
Untuk parameter Fe, dari 84 laboratorium yang berpartisipasi, sebagian besar (68 lab) memberikan kinerja yang memuaskan, 6 laboratorium mendapat status diperingatkan, dan 10 tidak memuaskan (data-data yang terkena seleksi uji Grubbs dihitung ke dalam kelompok tidak memuaskan). Pada parameter Cu, dari 70 peserta terdapat 5 lab yang memberikan hasil "di bawah limit deteksi" sehingga datanya tidak diikutsertakan dalam pengolahan data lebih lanjut. Dari 65 lab yang tersisa, terdapat 51 lab mendapatkan hasil memuaskan, 4 diperingati, dan 10 tidak memuaskan.

Pada parameter Cr(VI), dari 69 peserta, 4 lab memberikan hasil "di bawah limit deteksi". Dari 65 lab yang tersisa, terdapat 46 lab mendapatkan hasil memuaskan, 5 diperingati, dan 14 tidak memuaskan. Pada parameter pH, dari 88 laboratorium, terdapat 59 lab mendapatkan hasil memuaskan, 14 diperingati, dan 15 tidak memuaskan. Pola distribusi data-data hasil peserta untuk kelima paramater uji

dapat dicermati pada Gambar 1. Perlu dicatat bahwa label pada sumbu X dari gambar tersebut hanya merupakan angka urutan dan tidak berkorelasi dengan kode lab peserta, karena tampilan grafis tersebut disajikan untuk memberikan gambaran mengenai pola distribusi data dari keempat parameter tersebut.

Distribusi data yang dihasilkan untuk parameter Fe dan pH (Gambar 1) menunjukkan adanya

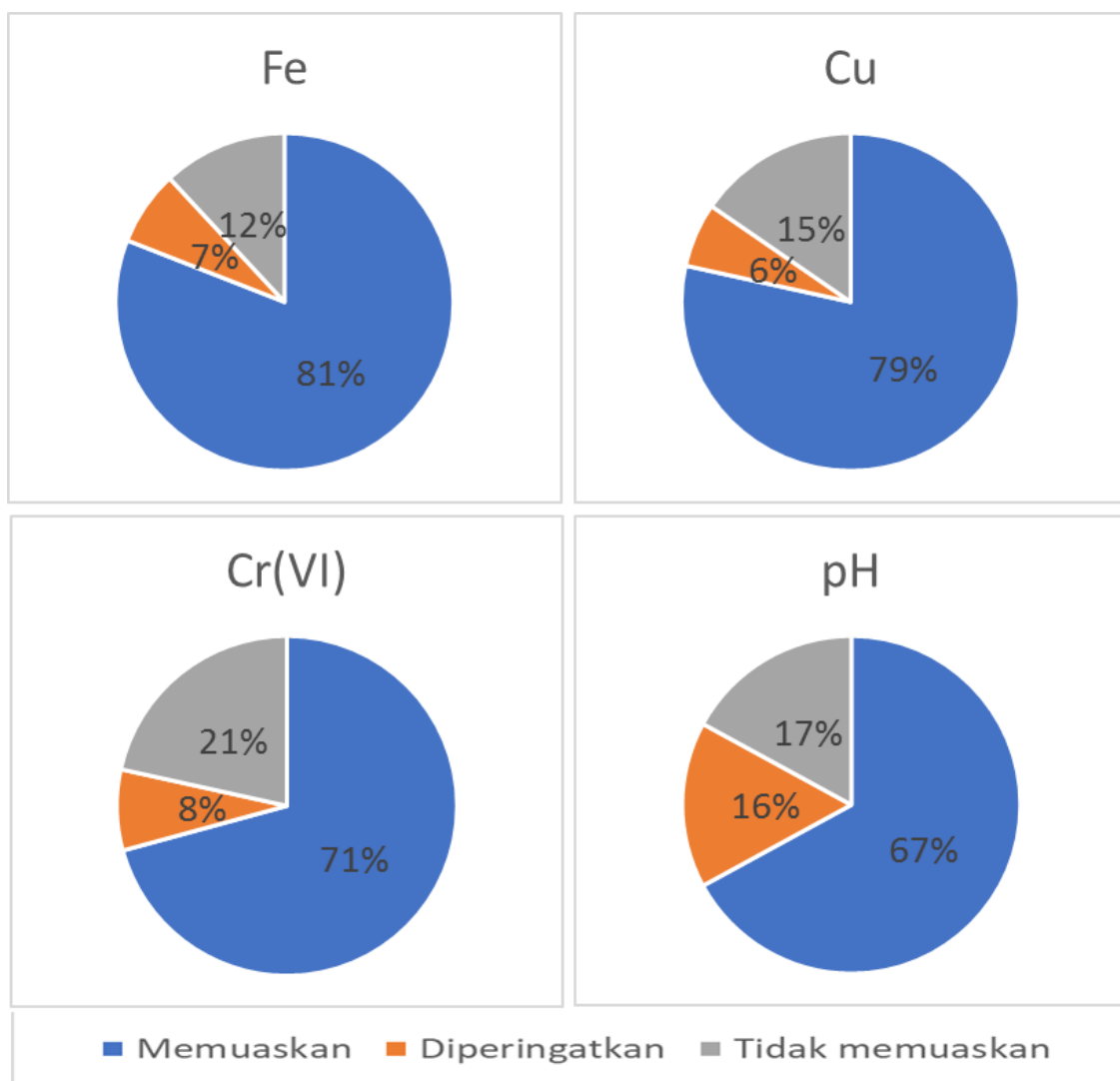
pola yang mendatar yang menunjukkan bahwa data-data peserta cukup seragam, walaupun terdapat nilai-nilai ekstrim bawah dan ekstrem atas di bagian tepi kurva. Dalam kasus ini, baik penggunaan nilai konsensus ataupun nilai acuan dapat secara efektif digunakan sebagai dasar penilaian peserta.



Gambar 1 Tampilan grafis data-data peserta yang telah diurutkan untuk kelima parameter uji

Hal yang berbeda terjadi pada parameter Cu, Cr(VI), dan NO₃-N dimana dapat diamati bahwa pola distribusi data yang dihasilkan cenderung menaik tanpa ada pola mendatar yang menunjukkan bahwa data-data peserta cukup beragam. Pada situasi ini penggunaan nilai konsensus akan menemui kesulitan untuk mendapatkan konsensus sehingga penilaian kinerja peserta UP gagal dilakukan. Karena penyelenggara UP menggunakan nilai formulasi untuk Cu dan Cr(VI), maka penentuan kinerja dapat dilakukan secara obyektif dan tidak bergantung dari konsensus peserta.

Untuk pengujian NO₃-N, data-data hasil uji dari peserta sangat bervariasi sebagaimana dapat diamati pada Gambar 1. Pengolahan lebih lanjut terhadap data-data tersebut dilakukan untuk mengamati pola distribusi data dan diketahui bahwa data-data tersebut mempunyai 2 populasi yang berbeda seperti tampak pada Gambar 3, yang menunjukkan bahwa nilai konsensus tidak dapat ditentukan. Sayangnya, nilai formulasi untuk NO₃-N pun tidak tersedia sehingga penilaian kinerja laboratorium peserta untuk parameter ini tidak dapat dilakukan (data tidak dapat diolah).



Gambar 2 Presentase kinerja peserta UP untuk 4 parameter uji.

Secara keseluruhan program UP ini telah berhasil menjadi alat penilai kinerja laboratorium peserta untuk parameter Fe, Cu, Cr(VI), dan pH.

Persentase kinerja laboratorium untuk keempat parameter ditampilkan secara grafis pada Gambar 2. Dapat dilihat bahwa kinerja pada

pengujian Fe secara merupakan yang terbaik dibanding yang lainnya. Namun demikian penilaian kinerja tidak dapat dilakukan untuk NO₃-N karena nilai konsensus tidak diperoleh dan nilai formulasi pun tidak tersedia. Harap dicatat bahwa jika penyelenggara hanya mengandalkan nilai konsensus saja, maka UP hanya akan berhasil menilai kinerja peserta untuk parameter Fe dan pH, sedangkan untuk Cu, Cr(VI), dan NO₃-N penilaian kinerja akan gagal dilakukan.

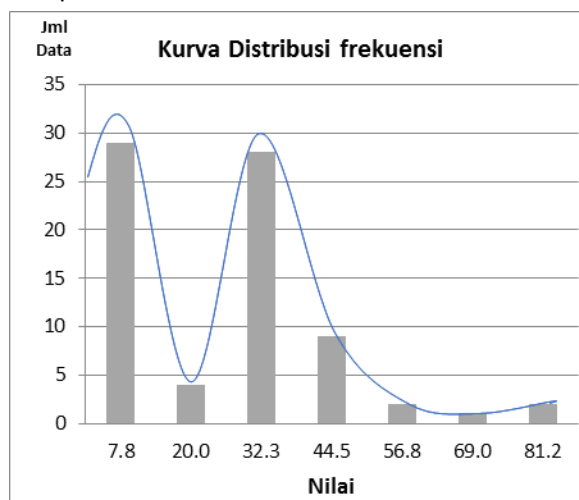
Hasil-hasil uji untuk parameter NO₃-N tidak dapat diolah karena data-data hasil uji peserta sangat bervariasi dan tidak dapat diperoleh konsensus. Hasil UP KAN VIII/2005 untuk parameter ini juga pada saat itu tidak dapat menilai kinerja peserta karena tidak diperoleh nilai konsensus. Bahkan untuk jumlah peserta yang cukup banyak seperti UP ini (peserta untuk NO₃-N sebanyak 75 laboratorium), nilai konsensus juga tidak diperoleh. Selain dari beragamnya jenis metode uji yang digunakan peserta, yaitu *UV screening method*, metode brucine yang diikuti pengukuran spektrofotometri UV/Vis, metode reduksi cadmium, elektrode selektif ion, kromatografi ion, serta metode spektrofotometer UV/Vis yang menggunakan *test kit*. Hal ini tentu memberikan kontribusi yang cukup signifikan terhadap bervariasinya data-data peserta. Selain dari itu, perhatian yang cukup harus diberikan kepada jenis *measurand* yang dilaporkan oleh peserta UP, mengingat NO₃ dan NO₃-N merupakan dua *measurand* yang berbeda. Kekeliruan dalam melaporkan hasil sebagaimana yang diminta provider (dalam hal ini NO₃-N) tentu akan memberikan kontribusi pada variasi data-data peserta.

Penggunaan nilai acuan yang seharusnya dapat mengatasi masalah ini ternyata juga mengalami kendala. Nilai formulasi untuk NO₃-N sulit untuk ditentukan dengan akurat karena penggunaan garam-garam nitrat dan asam nitrat dalam penyiapan matriks sampel turut menyumbang konsentrasi NO₃ secara keseluruhan sehingga konsentrasinya sulit ditetapkan secara pasti. Alternatif yang dapat dipertimbangkan adalah dengan mengestimasi rentang nilai perkiraan dari NO₃-N dan menghitung ketidakpastian bakunya. Kemungkinan nilai ketidakpastian yang dihasilkan akan secara relatif cukup besar sehingga perlu diperhitungkan dalam penetapan

nilai standar deviasi target (σ_{UP}) yang akan lebih besar. Penilaian kinerja kemudian dapat dilakukan dengan teknik *z' score* (ISO 13528). Hal ini berarti bahwa kriteria penilaian kinerja peserta menjadi lebih longgar. Meski demikian hal ini masih lebih baik dibandingkan kegagalan dalam menilai kinerja peserta karena menggunakan nilai konsensus.

4.3. Evaluasi UP

Evaluasi dilakukan setelah pelaksanaan UP untuk mengevaluasi bias dari putaran UP secara keseluruhan, dengan mempertimbangkan nilai acuan, X_{UP} dan ketidakpastiannya, $\mu(X_{UP})$, serta nilai rata-rata *robust* (*robust average*, RA) dari nilai-nilai peserta, X_{RA} , dan ketidakpastiannya, $\mu(X_{RA})$. Karena X_{UP} adalah nilai formulasi, maka $\mu(X_{UP})$ dapat dengan mudah diestimasi dari ketidakpastian dari kemurnian bahan standar, serta dari preparasi gravimetrik dan volumetriknya. Sementara X_{RA} diperoleh dari Algoritma A, yaitu teknik statistik yang digunakan untuk mendapatkan parameter *robust* dari suatu kumpulan data.



Gambar 3 Kurva distribusi frekuensi NO₃

Ketidakpastian dari nilai rata-rata *robust average*, $\mu(X_{RA})$, dihitung menggunakan standar deviasi robust yang diperoleh dari Algoritma A, $\sigma_{(robust)}$, menurut persamaan $\mu(X_{RA}) = 1,25 \times \sigma_{(robust)} / \sqrt{p}$, di mana p adalah jumlah nilai peserta.

Selanjutnya, nilai acuan X_{UP} dibandingkan dengan nilai konsensus yang diestimasi dengan nilai *robust average* X_{RA} . Selisih absolut antara keduanya, $|X_{UP} - X_{RA}|$, kemudian dibagi dengan ketidakpastian baku dari selisih tersebut,

$\sqrt{(\mu(X_{UP})^2 + \mu(X_{RA})^2)}$, yang akan menghasilkan suatu rasio tertentu.

Jika rasio yang diperoleh lebih dari dua maka disimpulkan ada bias yang cukup besar yang muncul dalam kegiatan UP tersebut (ISO 13528, 2015). Dalam hal ini, di mana nilai formulasi digunakan sebagai nilai acuan, kemungkinan bias dapat berasal dari bias dalam metode pengukuran yang digunakan oleh peserta, atau dari bias umum dalam hasil laboratorium. Hasil analisis bias dari kegiatan UP disajikan pada Tabel 4. Semua rasio jauh di

bawah 2, yang menunjukkan bahwa perbedaan antara nilai acuan dan nilai *robust average* yang berasal dari konsensus laboratorium peserta tidak berbeda secara signifikan. Dapat disimpulkan bahwa tidak ada bias yang signifikan dalam putaran UP tersebut. Selain itu dapat disimpulkan lebih jauh bahwa para peserta memiliki akurasi yang cukup memadai dalam pengukuran parameter uji yang diujikan di kegiatan UP tersebut. Khusus untuk NO₃-N evaluasi bias tidak dapat dilakukan karena nilai formulasi tidak ada.

Tabel 4 Evaluasi bias uji profisiensi

Parameter	X _{UP}	X _{RA}	S _{Robust}	μ(X _{RA})	μ(X _{UP})	X _{RA} -X _{UP}		Ratio (A/B)	Bias
						(A)	(B)		
Fe	1,94	1,99	0,1771	2,0289	0,031	0,05	2,0292	0,147	Tidak
Cu	20,8	21,58	9,354	94,267	0,270	0,780	94,2683	0,000	Tidak
Cr(VI)	25,3	26,482	9,556	96,304	0,250	1,182	96,3040	0,009	Tidak
pH	8,6	8,684	0,2068	2,425	0,600	0,084	2,4981	0,009	Tidak

5. KESIMPULAN

Evaluasi kinerja laboratorium peserta menggunakan nilai formulasi merupakan sesuatu yang memberikan nilai tambah tersendiri di tengah-tengah penggunaan nilai konsensus yang umum dilakukan saat ini. Dalam UP air permukaan ini manfaat penggunaan nilai formulasi sebagai nilai acuan sangat dirasakan, khususnya untuk parameter uji Cu dan Cr(VI) dimana distribusi data peserta cukup tersebar. Jika provider menggunakan nilai konsensus maka bukan tidak mungkin akan sampai pada kesimpulan bahwa "hasil tidak dapat diolah" yang artinya pemborosan sumber daya waktu dan biaya. Dari sisi penyelenggara UP sumber daya yang telah dikeluarkan berupa biaya pengadaan bahan dan alat dalam rangka persiapan sampel UP, biaya penyelenggaraan kegiatan, transportasi sampel, pengolahan data, dll. Sedangkan di sisi lain laboratorium peserta pun setidaknya telah mengeluarkan dana untuk partisipasi dalam kegiatan UP.

Penggunaan nilai formulasi sebagaimana disarankan dalam ISO 13528 dan ISO/IEC 17043 secara meyakinkan telah mampu digunakan untuk mengevaluasi kinerja

laboratorium-laboratorium peserta secara independen dan tidak bergantung pada hasil-hasil peserta.

Selain itu, evaluasi telah dilakukan untuk mengevaluasi bias pada kegiatan UP, dimana dari hasil evaluasi tersebut tidak teramati adanya bias yang signifikan. Hasil ini tentu saja memberikan informasi tambahan akan kualitas laboratorium pengujian di bidang analisis ini dan dapat meningkatkan keyakinan para pemangku kepentingan.

Pada UP ini untuk parameter NO₃-N nilai formulasi tidak dapat ditentukan secara akurat sehingga upaya tambahan perlu dilakukan untuk mengatasi hal ini. Satu alternatif solusi yang diusulkan adalah dengan menggunakan nilai yang diestimasi dari perkiraan rentang nilai formulasi (berdasarkan estimasi sumbangan NO₃-N dari bahan-bahan pembuatnya) sebagai nilai acuan. Ketidakpastian dari nilai acuan yang diperoleh kemungkinan akan relatif besar dan perlu diperhitungkan dalam penetapan standar deviasi target. Selanjutnya penilaian kinerja dapat dilakukan dengan menggunakan metode *z' score*.

DAFTAR PUSTAKA

- Asiah (2017). Mapping of environmental laboratory competency through proficiency testing program for heavy metal in clean water. *Ecolab*, 11(2), 53-104.
- Beckert, S.F., and Fischer, G.E. (2018). Interlaboratory comparison of roughness measurement: application of algorithm a of iso 13528: 2015 in determining the designated value and the standard deviation interlaboratory comparison of roughness measurement. *Journal of Physics: Conference series*, 1065, 082007.
- De Bièvre, P. (2012). A reference value in a proficiency testing scheme should be as metrologically credible as possible. *Accreditation and Quality Assurance*, 17(5), 561-562.
- De Bièvre, P. (2016). Measurement results in a Proficiency Testing programme need to be evaluated against independent criteria. *Accreditation and Quality Assurance*, 21(2), 167-169.
- Briggs, P. (2018, August). Collusion and falsification of results in proficiency testing. In *Journal of Physics: Conference series* (Vol. 1065, No. 21, p. 212001). IOP Publishing.
- Carter, G. D., Berry, J. L., Gunter, E., Jones, G., Jones, J. C., Makin, H. L. J., Wheeler, M. J. (2010). Journal of steroid biochemistry and molecular biology proficiency testing of 25-hydroxyvitamin d (25-ohd) assays. *Journal of Steroid Biochemistry and Molecular Biology*, 121(1-2), 176-179.
- Cheng, S., Mesquida, C., Gui, E. M., Cheow, P. S., Lee, T. K., & Teo, T. L. (2014). Providing a traceable assigned value in a proficiency testing programme on the determination of benzoic acid in orange juice. *Accreditation and Quality Assurance*, 19(6), 439-443.
- Dajay, L. C., Portugal, T. R., Climaco, J. C., Parcon, M. R. V., Udarbe, M. A., Placio, R. E. E., & Adona, P. E. (2019). Establishment of proficiency testing programs in the Philippines. *Accreditation and Quality Assurance*, 24(1), 65-71.
- Darie, M., & Burian, S. (2018). Proficiency testing for determination of water content in toluene of chemical reagents by iteration robust statistic technique proficiency testing for determination of water content in toluene of chemical reagents by iteration robust statistic technique. *IOP Conference Series: Material Sciences and Engineering*, 362, 012023.
- Hegazy, R., Mohamed, M. I., & Abu-Sinna, A. (2010). A comparative study of statistical methods used in analyzing the proficiency testing results of yield stress. *Mapan*, 25(2), 107-113.
- ISO 13528 (2015). Statistical methods for use in proficiency testing by interlaboratory comparison, 2nd edition.
- ISO/IEC 17043 (2010). Conformity Assessment – General requirements for proficiency testing, 1st edition.
- Kaarl, R., Mackay, L., Samuel, A., Sin, D. W. M., Mok, C. S., Wong, Y. L., & Yip, Y. C. (2017). Laboratory capacity building through the use of metrologically traceable reference values in proficiency testing programmes. *Accreditation and Quality Assurance*, 22(6), 321-334.
- Koch, M. (2018). Changes to proficiency testing in developing countries over the last 10 years. *Accreditation and Quality Assurance*.
- Laporan Uji Profisiensi Komite Akreditasi Nasional UP KAN VIII, Air Limbah, 2005
- Laporan Uji Profisiensi Komite Akreditasi Nasional UP KAN IX, Air Minum Dalam Kemasan, 2006
- Laporan Uji Profisiensi Komite Akreditasi Nasional UP KAN X, Tepung Ikan, Bubur Bayi, Tepung Maizena, 2007
- Laporan Uji Profisiensi Komite Akreditasi Nasional UP KAN X, Mikrobiologi, 2007
- Laporan Uji Profisiensi Komite Akreditasi Nasional UP KAN XI, Mikrobiologi, 2008
- Laporan Uji Profisiensi Komite Akreditasi Nasional UP KAN XII, Mikrobiologi, 2009

- Laporan Uji Profisiensi Komite Akreditasi Nasional UP KAN XIII, Mikrobiologi, 2010
- Laporan Uji Profisiensi Komite Akreditasi Nasional UP KAN XIII, Tepung Ikan, Bubur Bayi, Tepung Maizena, 2007
- Mianes, R. L., & ten Caten, C. S. (2017). Organisation of proficiency schemes by testing and calibration laboratories. *Accreditation and Quality Assurance*, 22, 119-123.
- Ponomareva, O. B., & Shpakov, S. V. (2012). Testing the proficiency of analytical laboratories by means of inter-laboratory comparison tests—an important element in assurance of the uniformity of measurements. *Measurement Techniques*, 54(12), 1427-1432.
- Smith, R. (2015). Analyst interpretation of results: experience from microbiology proficiency testing. *Accreditation and Quality Assurance*, 20(4), 325-327.
- Kim, Y. H., Lee, J. S., Song, K. B., Kang, Y. Y., Shin, S. K., Jung, G. T., ... & Kim, K. H. (2010). A review of proficiency testing scheme in Republic of Korea: in the field of water analysis. *Accreditation and quality assurance*, 15(4), 251-254.
- Sykes, M. (2012). Proficiency testing for the improvement of analytical practice. *Accreditation and Quality Assurance*, 17(4), 467-471.
- Tamiru, A., Boulanger, L., Chang, M. A., Malone, J. L., & Aidoo, M. (2015). Field assessment of dried Plasmodium falciparum samples for malaria rapid diagnostic test quality control and proficiency testing in Ethiopia. *Malaria Journal*, 14(11),1–8.
- Tholen, D. (2011). Metrology in service of society: the role of proficiency testing. *Accreditation and quality assurance*, 16(12), 603-605.
- Wong, S. (2011). Performance evaluation for proficiency testing with a limited number of participants. *Accreditation and Quality Assurance*, 12, 539–544.

