Reliability of VOC emission chamber testing – A round-robin test with flooring adhesives and a parquet lacquer

R. Oppl, M. Broege, F. Kuebart, T. Neuhaus, M. Wensing

ABSTRACT The Association for the Control of Emissions in Products for Flooring Installation, Adhesives and Building Materials (GEV) organised a round-robin test in 2017. They wanted to establish a list of recommended testing laboratories on the basis of test results. 33 laboratories from twelve countries received three spiked test products, similar to flooring adhesives and a parquet lacquer. Less variation of results was observed compared to earlier round-robin tests, but the differences between the testing laboratories were still significant. This fact inspired a discussion regarding the analytical challenges. As an example, the parameter „sum of all volatile organic compounds (VOCs) without a target value" includes the non-identified VOCs. This round-robin test showed a relative standard deviation of 100% and more for that parameter, which questions its reliability. The performance of 16 laboratories was rated as good by GEV. Currently, a list of recommended testing laboratories for GEV emissions testing comprises eleven laboratories from two countries. These laboratories performed well in this round-robin test and presented an appropriate accreditation according to ISO/IEC 17025.

Zuverlässigkeit von VOC-Emissionskammerprüfungen – Ein Ringversuch mit Bodenbelagsklebstoffen und einem Parkettlack

ZUSAMMENFASSUNG Die Gemeinschaft Emissionskontrollierte Verlegewerkstoffe, Klebstoffe und Bauprodukte e. V. (GEV) führte im Jahre 2017 einen Ringversuch durch, um auf der Grundlage der eingereichten Prüfergebnisse eine Liste empfohlener Prüfinstitute zu erstellen. 33 Prüflabore aus zwölf Staaten nahmen an dem Versuch teil und erhielten mit flüchtigen organischen Verbindungen (VOCs) angereicherte Prüfmuster, zwei Bodenbelagsklebstoffe und ein Parkettlack. Die Ergebnisse wiesen eine geringere, aber immer noch signifikante Streuung der Prüfergebnisse im Vergleich zu früher auf. Dies führte zu einer Diskussion der analytischen Herausforderungen. So enthält der Parameter „Summe der nicht bewertbaren VOC ohne NIK-Wert" auch die Summe der nicht identifizierten VOC. Dieser Ringversuch zeigte eine relative Standardabweichung von 100 % und mehr für diesen Parameter und stellte damit dessen Zuverlässigkeit infrage. Die Ergebnisse von 16 Teilnehmern wurden von der GEV als gut bewertet. Eine Liste der für GEV-Emissionsprüfungen empfohlenen Prüfinstitute enthält zurzeit elf Prüflabore aus zwei Staaten. Diese Labore hatten beim Ringversuch gut abgeschnitten und legten eine relevante Akkreditierung gemäß ISO 17025 vor.

1 Introduction

The Association for the Control of Emissions in Products for Flooring Installation (Gemeinschaft Emissionskontrollierte Verlegewerkstoffe, Klebstoffe und Bauprodukte e. V. (GEV)) was founded in 1997. Manufacturers of low emitting products for flooring installation, adhesives and building materials needed a reliable tool to assess, classify and label their products with regard to the emissions of volatile organic compounds (VOCs). GEV developed the label EMICODE® with different emissions classes.

Meanwhile, a number of European countries require testing of VOC emissions for specified construction products – e.g. in Germany for a legal technical approval of products that are not required to carry a CE mark. The voluntary label EMICODE® confirms very low emissions of construction products. This is why it is accepted as proof of low emissions by many certifiers of sustainable buildings, such as LEED[1)], BREEAM[2)] and DGNB[3)].

The reliability of the testing method is essential to classify products correctly into the emissions classes of GEV. The complexity of the testing procedure does not make it easy to fulfil this requirement, as was shown by earlier round-robin tests [1; 2]. And the state-of-the-art is developing. This concerns the products to be tested where changes in the composition may lead to new analytical challenges. The same applies to changes in the analytical techniques over time (discard of the flame ionization detector (FID), establishment of mass spectrometry detectors (MS), changes of chromatographic columns, etc.).

Therefore, a frequent evaluation of the testing procedure is highly meaningful. 33 testing laboratories from twelve countries followed a call for voluntary participation in a round-robin test. This should allow the participants to compare their performance

1) LEED: Leadership in Energy and Environmental Design, a rating and certification system for sustainable buildings, developed by U.S. Green Building Council in 1998, in use across the globe.

2) BREEAM: Building Research Establishment Environmental Assessment Methodology, a rating and certification system for sustainable buildings, developed by Building Research Establishment (BRE) in the United Kingdom in 1990, in use across the globe.

3) DGNB: DGNB System, a rating and certification system for sustainable buildings, developed by German Sustainable Building Council (DGNB) in 2007, in use in several countries.

with their competitors when testing for GEV, and to document their ranking.

35 testing laboratories had registered. Two of these abstained from reporting results in the end. 33 laboratories remained:

- 15 laboratories from Germany,
- three laboratories from China,
- three laboratories from Italy,
- two laboratories from Denmark,
- two laboratories from France,
- two laboratories from Switzerland,
- and one laboratory each from Austria, Belgium, Finland, Portugal, Spain and the USA.

Among the 33 participating there were

- 21 commercial testing laboratories,
- six public laboratories (incl. research organisations),
- six factory laboratories.

Some of the laboratories performed only parts of the tests if this corresponded to their special interests.

2 Methodology

2.1 Organisation and preparation

Two flooring adhesives and one parquet lacquer were used for this performance comparison. The samples were specifically prepared for this round-robin test. They were inspired by real formulations, but they included ingredients with higher VOC emissions than normal. The samples were prepared such that significant emissions could be determined after three days storage in a ventilated test chamber, allowing a robust statistical evaluation of the results. Pre-tests had confirmed that the samples could fulfil the purpose. Homogeneity of the testing material had been determined, as well as the stability during the period of testing by the laboratories. Three GEV member companies prepared the test samples. Two other companies distributed these samples to the participants.

2.2 Testing technique

The participants had to test the two flooring adhesives and the parquet lacquer according to the GEV testing method [3] with these additional specifications:

- No pre-conditioning of the adhesives (GEV testing method allows pre-conditioning as an option).
- Three days pre-conditioning of the parquet lacquer in a separate ventilated test chamber (GEV testing method specifies preconditioning as an option).
- Testing of the emissions of VOC, semi-volatile organic compounds (SVOC), formaldehyde and acetaldehyde already after three days storage in a ventilated test chamber.

The GEV testing method [3] specifies details of test specimen preparation, test chamber conditions, air sampling and air sample analyses. Details not specified there shall follow EN 16516 [4]. The tests were performed in a period from early November until mid-December 2017.

2.2.1 Test specimen and test chamber

- Sample preparation for the adhesives: 300 ± 10 g/m² on a glass plate, surface structured with a B1 notched trowel.
- Sample preparation for the parquet lacquer: 150 ± 2 g/m² on a glass plate, with even surface.
- Loading of the test chamber: 0.4 m²/m³.
- Storage of the test specimen in the test chamber at 23 ± 1 °C, 50 ± 5 % relative humidity of the supplied air, 0.5 ± 0.025/h air change rate.

2.2.2 Air sampling, analyses, reporting of results

VOCs and SVOCs were sampled from test chamber air with adsorption tubes which contained the polymer Tenax TA®. These tubes were analysed by gas chromatography (GC) with a mass selective detector (MSD) after thermal desorption (TD).

Formaldehyde and acetaldehyde were sampled with cartouches or adsorption tubes. These contained silica gel that had been impregnated with 2,4-Dinitrophenylhydrazin (DNPH). The tubes or cartouches were analysed by high pressure liquid chromatography (HPLC) with a UV or a diode array detector after extraction with acetonitrile.

In most cases, the laboratories performed the test chamber air measurements as double or multiple determinations. However, the participants were asked to report only one test result – the one that would have been reported if this had been a real test for a client. Therefore, this investigation did not consider the variation of the test results within each testing laboratory. The participants reported their test results into an online entry mask.

2.3 Statistical evaluation

The company QuoData performed a statistical evaluation of the reported test results according to ISO 13528 [5] in January 2018. This evaluation generally assumes a normal distribution of the data. While this could be confirmed for some data, such as the values for total volatile organic compounds (TVOCs), this was not the case for some other results (*R* values, sum of the non-identified VOCs, and the emissions of some individual VOC substances such as acetic acid, 2-methylisothiazolinone etc.).

A more robust evaluation procedure, the *Q*/Hampel method as specified in ISO 13528, allowed an evaluation in these cases. This method just requires a unimodal distribution. QuoData applied a kernel density estimation (ISO 13528, clause 10.3 [5]), using a proprietary software solution, for all data within this GEV round-robin test. There were no critical anomalies, i.e. none of the evaluated data series showed a bimodal or even multimodal distribution.

The reported results were presented as described in the following.

2.3.1 Assigned value (substitute target value)

There is no objective true value available for the tested samples. Therefore, a consensus value is calculated from the reported test results as robust, weighted arithmetic mean value. This is called „assigned value“ in ISO 13528 [5] and it is taken as a substitute for the true value.

Individual results that deviate from the mean value by more than 1.5 times the standard deviation are included with less weight in the statistical evaluation – the larger the deviation, the less weight is given. As an example, test results deviating from the mean value by more than 4.5 times the standard deviation does not have any influence on the evaluation. For more details, see ISO 13528, Annex C.5.3 (Hampel estimator). If a laboratory did not report a specific parameter, it could not be determined whether that laboratory saw insufficient amount of that substance to report it, or whether it did not detect the substance at all. Therefore, no value was included in the evaluation in those cases at all. Consequently, the assigned value is not significantly influenced by missing or strongly deviating results and can be used as a substitute target value.

2.3.2 Standard deviation for proficiency assessment

No universal and method specific standard deviation is known for this testing method. Therefore, the „standard deviation for proficiency assessment" was calculated from the reported test results as robust reproducibility standard deviation according to the Q method (ISO 13528, Annex C.5.2. [5]).

Again, strongly deviating results are included with less weight and do not have any significant impact on the overall evaluation. Should the homogeneity test show a non-homogeneous distribution of the emissions across a total testing sample, then the standard deviation for proficiency assessment is expanded by the standard deviation caused by the inhomogeneity.

This explains why the standard deviation of the testing method does not always equal the reported standard deviation of the test results.

2.3.3 Standard and expanded uncertainty of the assigned value

The robust standard deviation is multiplied by 1.25 and divided by the square root of the number of the participants in the round-robin test (see ISO 13528). This gives the „standard uncertainty of the assigned value" which is generally assumed to be the best estimator for round-robin tests. The „expanded uncertainty" (95 percentile) of the assigned value is given by multiplication by 2. Test results within this uncertainty of the assigned value are regarded to be close to the target value.

2.3.4 *z* score

This evaluation assumes that the reported test results follow a normal distribution, or at least a unimodal distribution. If the test results follow a normal distribution, then it can be expected that

- 68% of the test results are within one standard deviation of the target value,
- 95% of the test results are within two standard deviations of the target value,
- 99.7% of the test results are within three standard deviations of the target value.

The *z* score is the difference of a single test result from the target value, divided by the standard deviation for proficiency assessment. This value can be larger or smaller than 0: A *z* score larger than 0 means, the test result is above the target value. And a *z* score smaller than 0 means, the test result is below the target value.

- *z* score of +1 to -1

The deviation of a test result from the target value is smaller than 1 standard deviation, i.e. within the 68% confidence interval. This is generally regarded as a satisfactory result.

- *z* score of +1 to +2, or -1 to -2

The deviation of test result from the target value is larger than one standard deviation, but smaller than two standard deviations, i.e. within the 95% confidence interval. This is generally regarded as a satisfactory result as well.

- *z* score of +2 to +3, or -2 to -3

The deviation of a test result from the target value is larger than two standard deviations, but smaller than three standard deviations, i.e. within the 99.7% confidence interval. This is generally regarded as a questionable result.

- *z* score of +3 or larger or -3 or smaller

The deviation of a test result from the target value is larger than three standard deviations. This is generally regarded as an unsatisfactory result.

All *z* scores refer to a single test parameter and a single sample in one testing laboratory. ISO 13528 does not specify any overall *z* score for a testing laboratory as a total.

2.3.5 Limitations of the statistical evaluation

This purely statistical evaluation relies on the reported test results of the participating laboratories. However, it may happen that the reported data do not allow a meaningful evaluation, even if the statistical procedures could formally be applied. Further, it should be noted that there is no objective true target value available. It cannot be excluded that a larger group of laboratories delivers wrong results either systematically or at random. Therefore, a statistical evaluation should always be followed by a technical assessment and a plausibility check. This includes knowledge of ingredients that can be expected to be emitted during the chamber tests, as well as analytical challenges as discussed in chapter 4.

3 Results

Not all laboratories reported all requested parameters. Non-reported values were not replaced by substitute values, but just ignored in the evaluation.

3.1 Homogeneity and stability of the emissions

Eurofins Product Testing A/S performed homogeneity tests for selected VOCs on behalf of GEV. The VOCs with the highest emission rates in preliminary tests were selected. No homogeneity tests were conducted for sum parameters such as TVOC and *R* value. Five randomly chosen test items were analysed in single determination at the beginning of the round-robin testing period (early November 2017) and evaluated statistically by QuoData. As the heterogeneity standard deviation was not larger than 0.3 times the standard deviation for proficiency assessment, the test material was considered sufficiently homogeneous (see Annex B.2.2 in ISO 13528 [5]).

Another test sample was analysed at the end of the round-robin testing period (mid December 2017) in single determination. QuoData performed a statistical assessment of the stability of the testing material over time with a *t* test, taking into account the variance of the results of the homogeneity testing. 2-Ethylhexanol showed statistically significant deviations between early and late testing. But an evaluation of the reported test results, taking into account the respective testing date, did not confirm any statistically significant trend and any instability. Accordingly, the assessment of the laboratory test results by means of *z* scores was carried out without any restriction.

3.2 TSVOC and TVOC results

The value for total semi-volatile organic compounds (TSVOCs) was not evaluated statistically because the results of most testing laboratories were below the required determination limit of 5 µg/m³.

The TVOC value was calculated in two different ways. In both cases, the contribution of acetic acid to TVOC was excluded from the calculation as specified by the GEV testing method. Several investigations had shown that neither the GEV testing method nor the EN 16516 method provides reliable results for this compound, see also clause 8.2.1 of EN 16516 [4].

The TVOC value according to the GEV testing method [3] and EN 16516 (clauses 8.2.6.2 and 10.6.5) [4] was determined in toluene equivalents, i.e. all single VOC substances were calculated with the response factor of toluene and summed, if above the reporting limit of 5 µg/m³ (see **Table 1** and **Figure 1**).

Table 1. Results – TVOC value (in toluene equivalents) according to GEV und EN 16516 - 10.6.5, but without acetic acid.

Test sample	Weighted mean in µg/m³	Expanded uncertainty of target value in %	Relative standard deviation in %	Participants with				Number of results/ All participants
				z score +1 to -1	z score +1 to +2 and -1 to -2	z score +2 to +3 and -2 to -3	z score above +3 and below -3	
Adhesive K1	270	17	38	24	8	0	1	33/33
Adhesive K2	180	17	38	24	8	0	1	33/33
Parquet lacquer PL	450	13	30	25	4	3	1	33/33

z scores between +2 and -2 are regarded satisfactory, between +2 and +3 or -2 and -3 as questionable, larger than +3 or smaller than -3 as unsatisfactory.

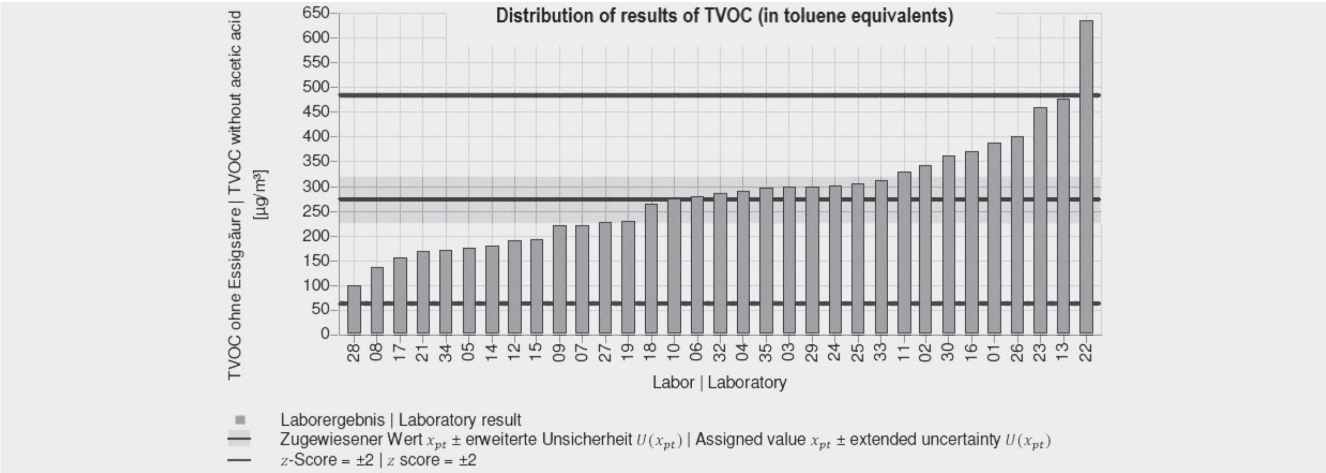

Figure 1. Distribution of results of TVOC (in toluene equivalents) without acetic acid for adhesive K1.

Source: QuoData: GEV Round-Robin Test 2017 – Final report of results, 30 January 2018

The sum of the VOCs according to EN 16516, clause 10.6.8 [4], also called $TVOC_{SPEZ}$ value by AgBB, was determined such that

- all single VOC substances with a German LCI limit value (AgBB 2015) [6] (LCI = lowest concentration of interest) were calculated with their respective substance specific response factor, and
- all single VOC substances with a German LCI value and all non-identified VOCs were calculated with the response factor of toluene,

and (different to EN 16516) all values were summed for which the result in toluene equivalents was above the reporting limit of 5 µg/m³ (see **Table 2** and **Figure 2**).

Table 2. Results – $TVOC_{SPEZ}$ value, sum of VOCs, according to EN 16516 - 10.6.8, but without acetic acid.

Test sample	Weighted mean in µg/m³	Expanded uncertainty of target value in %	Relative standard deviation in %	Participants with				Number of results/ All participants
				z score +1 to -1	z score +1 to +2 and -1 to -2	z score +2 to +3 and -2 to -3	z score above +3 and below -3	
Adhesive K1	300	16	36	20	10	0	0	30/33
Adhesive K2	300	19	41	22	8	0	0	30/33
Parquet lacquer PL	690	17	37	24	3	2	1	30/33

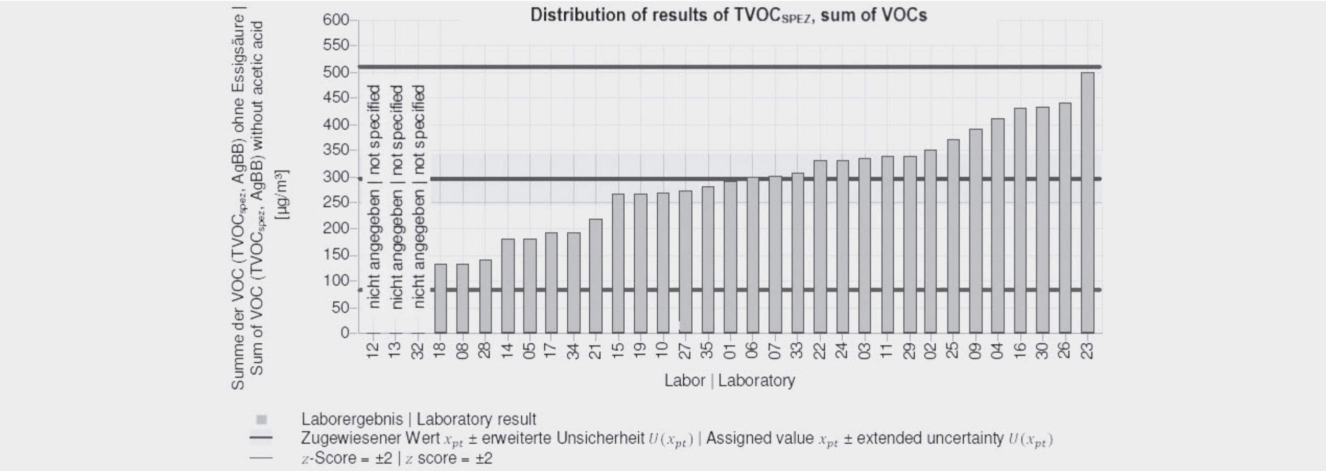

Figure 2. Distribution of results of TVOC$_{SPEZ}$, sum of VOCs, without acetic acid for adhesive K1.
Source: QuoData: GEV Round-Robin Test 2017 – Final report of results, 30 January 2018

See chapter 4.2 for a critical comparison of the TVOC value and the TVOC$_{SPEZ}$ value.

3.3 Results of other sum parameters

3.3.1 R value without acetic acid

Individual VOC substances had to be quantified if the respective test result in toluene equivalents was at least 5 µg/m³ (reporting limit). Substances with a German LCI limit value (AgBB 2015) [6] had to be calibrated and quantified with their respective substance specific response factors. Then the *R* value was calculated as sum of all quotients of each individual substance test result and its respective German LCI value (see **Table 3**). Results of acetic acid have not been included in that calculation due to the low reliability of their determination, see chapter 4.2.

Table 3. Results – *R* value without acetic acid according to GEV with German LCI values (AgBB 2015).

Test sample	Weighted mean in µg/m³	Expanded uncertainty of target value in %	Relative standard deviation in %	Participants with				Number of results/ All participants
				z score +1 to -1	z score +1 to +2 and -1 to -2	z score +2 to +3 and -2 to -3	z score above +3 and below -3	
Adhesive K1	0.6	18	40	24	7	1	0	32/33
Adhesive K2	0.3	41	97	24	3	0	5	32/33
Parquet lacquer PL	8.7	24	55	24	5	1	2	32/33

See chapter 4.3 for a critical discussion on the significance of a statistical evaluation of the *R* value.

3.3.2 Sum of the non-identified VOCs (in toluene equivalents) and number of non-identified VOCs

The sum and the number of non-identified VOCs reported by the participants showed large differences.

Adhesive K1

- 20 laboratories reported two to three non-identified substances.
  - 14 laboratories reported a sum of non-identified substances from 7 to 30 µg/m³.
  - Six laboratories reported much larger sums, up to 140 µg/m.
- Seven laboratories did not report any non-identified substances at all.

Adhesive K2

- 19 laboratories reported one to four non-identified substances.
  - Twelve laboratories reported a sum of non-identified substances from 4 to 30 µg/m³.
  - Seven laboratories reported much larger sums, up to 130 µg/m³.
- Seven laboratories did not report any non-identified substances at all.

Parquet lacquer PL

- 15 laboratories reported one to three non-identified substances.
  - Four laboratories reported a sum of non-identified substances from 10 to 35 µg/m³.
  - Four laboratories reported much smaller sums, < 10 µg/m³.
  - Seven laboratories reported much larger sums, up to 750 µg/m

- Seven laboratories did not report any non-identified substances at all.

The relative standard deviations of the results were at or above 100%.

3.4 Results for single substances

The participants had not been informed about the substances to be analysed. Each participant had to identify and quantify the individual VOCs that were emitted under the specified testing conditions. Individual VOC substances had to be quantified if the respective test result in toluene equivalents was at least 5 µg/m³ (reporting limit as specified in GEV testing method [3]).

Table 4. Results of individual substances, calibrated substance-specifically, for adhesive K1.

Substance	Weighted mean in µg/m³	Expanded uncertainty of target value in %	Relative standard deviation in %	Participants with				Number of results/ All participants
				z score +1 to -1	z score +1 to +2 and -1 to -2	z score +2 to +3 and -2 to -3	z score above +3 and below -3	
2-Ethyl-1-hexanol (LCI # 4-10)	65	12	27	20	9	0	0	29/33
Acetic acid (LCI # 9-1)	480	25	64	19	5	0	1	25/33

Table 5. Results of individual substances, calibrated substance-specifically, for adhesive K2.

Substance	Weighted mean in µg/m³	Expanded uncertainty of target value in %	Relative standard deviation in %	Participants with				Number of results/ All participants
				z score +1 to -1	z score +1 to +2 and -1 to -2	z score +2 to +3 and -2 to -3	z score above +3 and below -3	
2-Ethyl-1-hexanol (LCI # 4-10)	21	10	21	20	5	1	1	27/33
1-Butanol (LCI # 4-6)	30	24	59	21	6	0	1	28/33
Propylene glycol (LCI # 6-1)	77	19	39	18	8	0	1	27/33

Table 6. Results of individual substances, calibrated substance-specifically, for Parquet lacquer PL.

Substance	Weighted mean in µg/m³	Expanded uncertainty of target value in %	Relative standard deviation in %	Participants with				Number of results/ All participants
				z score +1 to -1	z score +1 to +2 and -1 to -2	z score +2 to +3 and -2 to -3	z score above +3 and below -3	
Triethylamine (LCI # 12-11)	390	25	49	20	3	1	1	25/33
Dipropylene glycol mono methyl ether (LCI # 6-12)	210	22	61	21	3	0	1	25/33
Butylhydroxy-toluol (BHT) (LCI # 5-2)	15	20	42	20	7	2	0	29/33

The results for formaldehyde and acetaldehyde were not evaluated statistically because almost all testing laboratories delivered either no results for these substances, or results below the reporting limit of 5 µg/m³ (see **Tables 4** to **6**).

Several individual substances were not detected by all testing laboratories, for example 2-methyl-4-isothiazolin-3-one (LCI # 12-10) in the adhesive K1. 15 out of 33 laboratories delivered a result with substance specific calibration, with a weighted mean value of 27 µg/m³. There were 19 reports in toluene equivalents with a weighted mean of 10 µg/m³. The low mean value of the results in toluene equivalents indicates that probably several other laboratories determined a result, but without reporting it, because it was below the reporting limit of GEV (5 µg/m³ in toluene equivalents). If all participants had used the reporting limit of EN 16516 (5 µg/m³ determined with the specific response), then more laboratories would have reported a result and a statistical evaluation might have been better possible.

For this reason, this parameter could not be used for an assessment of the performance of the laboratories in this round-robin test.

4 Assessment and discussion of the results

4.1 Acetic acid

It is well-known that it is a challenge to determine acetic acid with the presented methodology. The adsorption tubes filled with Tenax TA® have only a low adsorption capacity for acetic acid and are quickly saturated. That is the reason why EN 16516 [4] states in a note to clause 8.2.1:

„A few VOCs like acetic acid are not quantitatively analysed under the conditions specified in this method. In this case, an alternative

sorbent or series of sorbents or alternative conditions can be used to confirm a test result.“

Alternative testing methodologies for acetic acid are specified in VDI 4301 Part 7 [7].

The present round-robin test used only test results obtained by adsorption on Tenax TA®. In line with the GEV testing method, the contribution of acetic acid to TVOC and *R* value was excluded from calculation.

4.2 TVOC and $TVOC_{SPEZ}$ results (without acetic acid)

The results showed a lower variation of the TVOC results compared to previous round-robin tests of the GEV. Nevertheless, the highest and the smallest TVOC value (in toluene equivalents) differed by a factor of 6 to 9. The relative standard deviation from the weighted mean was 30 to 40%.

More than 70% of the participants delivered satisfactory results with a *z* score between +1 and -1 for the TVOC value in toluene equivalents, and 60 to 70% for the sum of VOCs, the $TVOC_{SPEZ}$. Both were determined without the contribution of acetic acid.

The TVOC value in toluene equivalents is used historically and globally to assess the total emissions of products. Meanwhile, authorities and many labelling organisations in German speaking countries prefer to use the sum of VOCs for that purpose, the $TVOC_{SPEZ}$. This is justified by stating that the $TVOC_{SPEZ}$ value is closer to the real concentrations than the TVOC value in toluene equivalents. EN 16516 allows both ways of calculation. The significance of both parameters is discussed in the following clause.

4.2.1 TVOC and $TVOC_{SPEZ}$: Background

VOC came into the focus of indoor air quality research in the 80s and 90s of the last century. Health complaints when occupying certain buildings (the so-called „Sick Building Syndrome“) and the presence of VOCs in indoor air often were linked, and the VOCs were assumed to cause the problem.

A variety of products manufactured with addition of e.g. monomers, solvents, or additives was identified as a source for these VOCs, which could be emitted into indoor air during the use-phase of such a product. Most times, the emissions include a mixture of many substances rather than only few single compounds.

The early analysis of VOCs in those times mainly had to answer two questions: 1. Is it possible, in a simple way, to characterize such a mixture of many substances with a sum parameter, and 2. Can the significance of such emissions for human health be indicated by such a sum parameter?

This led to the introduction of the so-called TVOC approach. The main analytical challenge then was to combine different individual substances with very different analytical behaviour into one analytical value. This should be done in a way that different measurements can be compared with each other, without false results relative to the assumed true value. The main health-related challenge was that substances with very different toxicological properties were considered and assessed together in one single value.

A prerequisite for any comparison of TVOC values is a coordinated and uniform analytical protocol. The ECA Report 18 [8] proposed an intermediate procedure to determine TVOC where the whole signal in the gas chromatogram was integrated and calculated in toluene equivalents. Even at that time it was known that determining TVOC values in that way can substantially underestimate the real concentrations. This TVOC procedure was then specified in the testing standard ISO 16000-6 [9] where it was explicitly stated that this TVOC determination is half-quantitative only.

EN 16516 [4] specifies a procedure where only test results of substances above 5 µg/m³ are calculated in toluene equivalents and summed to the TVOC value. Test results of minor traces show a much higher uncertainty and this way of determination was intended to exclude the contribution of such small traces with their high analytical uncertainty.

In this case, any such determined TVOC value can also deviate strongly from the assumed true value, often underreporting the values. The risk of false and too low findings with a TVOC value in toluene equivalents increased during the last 20 years, as many of the VOCs in use today show a much lower analytical response compared to toluene, i.e. they deliver a lower measurement signal. On top of that, it has not yet been proven that TVOC values in toluene equivalents can be compared between different analytical devices when applying the current state-of-the-art technique, GC/MS.

The ECA Report 19 [10] proposed another analytical procedure. As many VOCs as possible should be identified. At least the ten largest signal peaks in a chromatogram should be quantified and summed with their original responses (sum of the identified VOCs, S_{id}). The other VOCs and the remaining unidentified substances would be determined in a simplified manner in toluene equivalents and summed (sum of the non-identified VOCs, S_{un}). The sum of S_{id} and S_{un} then is the TVOC value.

The approach to combine a quantification with original responses and a use of toluene equivalents was developed further by the German AgBB since 2004 [6]. At this time, 180 individual substances with a German LCI value are identified by GC/MS and quantified with their original responses. All other VOCs are determined in toluene equivalents. The sum of all is the TVOC value. This procedure already is close to the assumed true value. This way of TVOC calculation is specified in EN 16516 [4] as well, there called sum of VOCs. AgBB calls this parameter $TVOC_{SPEZ}$.

The impact of the differences is illustrated in **Figure 3**. The sum of VOCs, the $TVOC_{SPEZ}$, was compared with the TVOC in

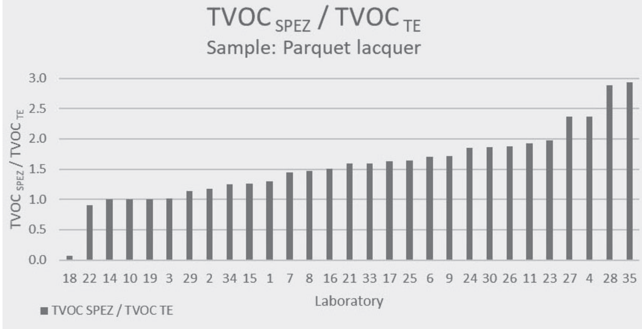

Figure 3. Ratio of $TVOC_{SPEZ}$ to TVOC in toluene equivalents (TE).
Source: Reinhard Oppl

toluene equivalents (TE) for the parquet lacquer. The $TVOC_{SPEZ}$ was very similar to the $TVOC_{TE}$ for some laboratories (the ratio was around 1). But in most cases the $TVOC_{SPEZ}$ was much higher, and the ratio of $TVOC_{SPEZ}$ to $TVOC_{TE}$ was in the range of 1.2 to 1.9, in few cases even almost 3.

4.2.2 TVOC and $TVOC_{SPEZ}$: Conclusions

A TVOC value with all its limitations may be a significant parameter to survey total emissions where there is a similar composition of the emitted mixture of VOCs, e.g. for development purposes or for factory control, but without any ambition of health control. In the case of product evaluation, a TVOC value (calculated by whatever procedure) only can be an indicator whether „more“ or „less“ is emitted in total. This is of special relevance for rating systems that do without evaluation of the individual VOCs, such as BREEAM and Green Star[4]. The same applies to rating systems that have very short lists of target VOCs, such as the French VOC emissions label and CDPH[5].

GEV uses the TVOC in toluene equivalents as well, to compare products regarding the total emissions. While small analytical differences can have huge impact on the *R* value (see section 4.3) the TVOC is more useful in levelling out any differences in how the testing laboratories measure individual substances. The test results of different testing laboratories then are more comparable (see section 3.2).

But whatever procedure is used to determine a TVOC value, the statement of ECA Report 19 [10] remains true that there is no proof that a TVOC value alone can indicate how significant emissions are for human health and well-being. *Andersson* et al. [11] came to the same conclusion after evaluating published studies on a relationship between exposure to TVOC and human health. During the last 20 years there has been fundamentally no change in this finding. Any health-related assessment of VOCs always should relate to an evaluation of single substances after best possible analytical determination.

As a consequence, the US rating system LEED in its newest version 4.1 [12] skipped any evaluation of products on the basis of a TVOC value. LEED requires disclosure of the TVOC level, but it does not depreciate a product because of a high TVOC value. The product rating relies on test results of individual substances only.

4.3 *R* value

The health impact of a potentially hazardous substance depends on its substance specific toxicity and on the exposure to this substance. Most times a mixture of VOCs is present. There is no toxicological cause-and-effect scheme available to describe interactions between the VOCs in such mixtures that is generally accepted and easy to handle. In a simplified approach, an additive effect is assumed, ignoring any possible reciprocally intensifying (synergistic) or attenuating (antagonistic) effects between the substances. Instead, the substance specific toxicity equivalents are evaluated separately and then summed.

LCI values have been specified for a variety of VOCs. LCI values are part of the product evaluation by several rating systems, such as the EMICODE testing method [3] and the German AgBB scheme [6]. Currently, the German list includes 180 LCI values for individual substances. A harmonized EU list [13] includes 140 LCI values. These have been specified on the basis of a harmonized European procedure. The addition of the health effects is assumed for all VOCs with a concentration of at least 5 µg/m³ each. Then it is specified that *R*, i.e. the sum of all R_i of the individual substances, must not exceed a limit value of 1.

$R_i = C_i / LCI_i$ with C_i = Concentration of an individual substance in air

R = Sum of all R_i = Sum of all quotients

This procedure has the ambition to allow the best possible assessment of potential hazards of a mixture of substances in air, considering the toxicity of the individual substances. It provides a reliable assessment factor (*R* value) for the risk potential of a tested product.

Another source of uncertainty arises if substances with an LCI value far below the reporting limit of 5 µg/m³ are detected (examples see **Table** 7).

Table 7. Substances with very low LCI values (AgBB [6]).

Substance	LCI value, AgBB 2018 in µg/m³
2-Butenal, all Isomers (Crotonaldehyde)	1
Glutaraldehyde	1
5-Chlor-2-methyl-4-isothiazolin-3-one (CIT)	1
2-Methoxyethanol; Ethylene glycol monomethyl ether	3
1,2-Dimethoxyethane	4

This shall be illustrated for the substance CIT, having an LCI value of 1 µg/m³:

- If the analysis of CIT after 28 days provides a test result of 4.4 µg/m³ this value will be rounded to one significant digit according to EN 16516, and be reported as 4 µg/m³. CIT then will not be included in the calculation of the *R* value, as the result is below 5 µg/m³, and its contribution to the *R* value is zero.
- Another test result of 4.5 µg/m³ would be rounded and reported as 5 µg/m³. Consequently, CIT will be included in the calculation of *R* with a contribution of 5.

This example shows how small variations of a test result might have a huge impact on the calculation of the *R* value – even within the normal uncertainty range of emissions testing results.

Furthermore, some participants included substances below the reporting limit of the GEV testing method in the calculation of the *R* value. This occurred if the test result was above 5 µg/m³ with the original response, but below 5 µg/m³ when calculated in toluene equivalents. This procedure is in line with EN 16516, but not with the GEV testing method. These differences led to inconsistent *R* value test results.

So, it was not surprising that the *R* value test results did not follow a normal distribution. Nevertheless, the evaluation with the Q/Hampel method showed that more than 70% of the par-

[4] Green Star: A rating and certification system for sustainable buildings, developed by Green Building Council Australia in 2003, in use in Australia, New Zealand and South Africa.
[5] CDPH: California Department of Public Health, Standard Method for the Testing and Evaluation of Volatile Organic Chemical Emissions from Indoor Sources using Environmental Chambers.

ticipants delivered satisfactory values with a *z* score between +1 and -1 for the *R* value without the contribution of acetic acid.

As a consequence, if VOCs with very low LCI values are emitted, the *R* value should not be included in a rating of a round-robin test without special considerations.

4.4 Sum of the non-identified VOCs and sum of the VOCs without a target value

The parameter „sum of all VOCs without a target value" includes two items – all identified VOCs without an LCI limit value, and all non-identified VOCs. There was no requirement in this round-robin test to report the sum of all VOCs without a target value. But the reported sum of the non-identified VOCs was evaluated by QuoData and showed a very high uncertainty with relative standard deviations of 100% or more. This parameter is inaccurate, and the results did not follow a normal distribution.

Authorities and several labelling organisations require to report the „VOCs without a target value" mostly in German speaking countries. This is criticized by other countries and by industry, among others, because of the inaccurate determination of the non-identified VOCs. This parameter may lead to a discrimination of a product with emissions of VOCs without an LCI value – independently of the actual significance of these VOCs for human health.

In addition, a laboratory may fail to identify a VOC with an LCI value, and then attribute it erroneously as a VOC without an LCI value, even though accredited testing laboratories should be able to identify and quantify all substances with an LCI value. Such cases are another issue leading to unjustified discrimination of products.

A laboratory that identifies more substances than other laboratories will report no or few non-identified VOCs. This is a sign of good quality. But a purely statistical evaluation even could assign a bad *z* score as this laboratory then deviates substantially from the other laboratories.

4.5 Results of individual VOCs

The testing method covers a large variety of VOCs with very different chemical properties. EN 16516 and the GEV testing method are appropriate for some of these VOCs, but less applicable to other ones. 50 to 70% of the participants produced satisfactory results with a *z* score between +1 and -1 for the evaluated individual substances with significant emissions. Worst results were observed for propylene glycol and for acetic acid. The results of acetic acid did not follow a normal distribution.

The identification and quantification of glycols, glycol ethers and glycol esters can dominate the assessment of a product where these emissions occur. EN 16516 [4] – the basis of the GEV testing method – specifies only a general testing method to allow determination of emissions from a large variety of construction products. The testing laboratory has to adapt its own analytical procedures for the detected compounds that are visible in the chromatogram. The determination of certain glycols is rather difficult. As an example, propylene glycol shows a broad shape of the signal peak. This complicates the quantification, as partial overlap with peaks of other signals may occur, which cannot be further resolved.

Many technical products containing glycols will show several isomers in the same retention time range and with very similar mass traces in the mass spectrum. This can be illustrated by two substances found in the tested parquet lacquer, DMM (dipropylene glycol dimethyl ether – # 6-39 on the LCI list) and DPM (dipropylene glycol monomethyl ether – # 6-12 on the LCI list). Isomers of both substances appear in a similar retention time range in the gas chromatogram. In the mass spectrum, both substances show signal peaks at m/z 45, m/z 59, m/z 73. Specific peaks for DMM and DPM identification are the peaks at m/z 117 or m/z 103, respectively. But errors in quantification may occur in the case of overlap with mass traces from other signal peaks.

An additional challenge is the fact that the composition of the VOC mixtures is unknown for most cases of emissions testing. Information on real ingredients would be very helpful to allow correct identification and quantification, but is usually not available to the testing laboratory.

4.6 Round-robin tests as performance assessment of testing laboratories

Round-robin test data can be used for the assessment of the qualification of testing laboratories. One essential criterion is the number of reported satisfactory test data (i.e. with a *z* score between +2 and -2, or even between +1 and -1), but this assessment should be based only on those parameters for which a statistical evaluation is meaningful. These were in this round-robin test:

- TVOC (EN 16516) without the contribution of acetic acid,
- Sum of VOCs ($TVOC_{SPEZ}$, AgBB) without acetic acid,
- *R* value without acetic acid (with limitations, see chapter 4.3),
- Individual substances with significant emissions, determined with substance specific calibration.

There is no objective true target value available, as it cannot be excluded that a larger group of laboratories delivers the wrong results systematically or at random. Therefore, a statistical evaluation should always be followed by a technical assessment and a plausibility check.

Another criterion is how many of the expected individual substances have been detected. An assessment of the performance of a testing laboratory can include information outside the round-robin test as well, such as the degree of experience with testing for the EMICODE, testing capacity and delivery time.

16 participating laboratories fulfilled the requirements of GEV in this round-robin test. An accreditation according to ISO/IEC 17025 [14] is very important for a positive rating of a testing laboratory, if this accreditation clearly includes the GEV testing method or EN 16516, as this accreditation requires basic quality performance. On this basis, GEV offers a list with recommended testing laboratories on its homepage (www.emicode.com). Currently the list comprises eleven laboratories from two countries.

5 Conclusions and outlook

The results of the GEV round-robin test 2017 showed a smaller, but still significant variation of the results of TVOC and of individual substances, compared with previous round-robin tests of the GEV. Evaluation of the variation in results inspired a discussion on the analytical challenges. As an example, the parameter „sum of all VOCs without a target value" includes the non-identified VOCs. This round-robin test showed a relative standard deviation of 100% and more for the determination of that parameter. Those test results only can indicate the order of magnitude.

The performance of 16 laboratories was rated as good by GEV. Currently, a list of recommended testing laboratory for GEV emissions testing comprises eleven laboratories from two countries. These laboratories performed well in this round-robin test and presented an appropriate accreditation according to ISO/IEC 17025.

ACKNOWLEDGEMENTS

The authors express their gratitude

- to the Association for the Control of Emissions in Products for Flooring Installation, Adhesives and Building Materials (GEV) and its Technical Council, and to the involved GEV member companies, for their support of this round-robin test,
- to the involved employees of the testing laboratories for running the chemical testing,
- to *Kirstin Frost*, *Steffen Uhlig* and *Kirsten Simon* and other involved employees of QuoData for the statistical evaluation of the results of this round-robin test,
- for critical supervision and commenting of this work to
  - Dr *Thomas Brokamp*, Bona, Limburg, Germany.
  - Dr *Mechthild Haveresch-Kock*, Sika Deutschland, Rosendahl, Germany.
  - Dr *Martin Spleiss*, PCI Augsburg, Augsburg, Germany.
  - *Klaus Winkels* and *Natascha Zapolowski*, GEV, Düsseldorf, Germany.

Literature

[1] *Oppl, R.*: Reliability of VOC emission chamber testing – progress and remaining challenges. Gefahrstoffe – Reinhalt. Luft 68 (2008) Nr. 3, S. 83-86.

[2] *Windhövel, U.; Oppl, R.*: Praktische Überprüfung des Konzepts zur gesundheitlichen Bewertung von Bauprodukten (Practical evaluation of the approach of health-related assessment of construction products, in German). Gefahrstoffe – Reinhalt. Luft 65 (2000) Nr. 3, S. 81-89.

[3] Association for the Control of Emissions in Products for Flooring Installation, Adhesives and Building Materials (GEV): GEV – Testing Method, Determination of Volatile Organic Compounds for Evaluation of Emissions Controlled Products for Flooring Installation, Adhesives, Construction Products, as well as Lacquers, Finishes and Oils for Parquet, for Mineral Floorings and for Resilient Floorings, 2019, available in the Internet at www.emicode.com/home/.

[4] EN 16516: Construction products: Assessment of release of dangerous substances – Determination of emissions into indoor air. Berlin: Beuth 2017.

[5] ISO 13528: Statistical methods for use in proficiency testing by interlaboratory comparison. Berlin: Beuth 2015.

[6] AgBB-Scheme: Committee for Health-related Evaluation of Building Products AgBB: Health-related Evaluation Procedure for Emissions of Volatile Organic Compounds (VVOC, VOC and SVOC) from Building Products, 2018 and 2015. www.umweltbundesamt.de/en/topics/health/commissions-working-groups/committee-for-health-related-evaluation-of-building#textpart-1

[7] VDI 4301 Part 7: Measurement of indoor air pollution – Measurement of carboxylic acids. Berlin: Beuth 2018.

[8] ECA – European Collaborative Action (1997). Report No 18 Evaluation of VOC Emissions from Building Materials. EUR 17334 EN. Office for Official Publications of the European Community, Luxembourg.

[9] ISO 16000: Indoor air – Part 6: Determination of volatile organic compounds in indoor and test chamber air by active sampling on Tenax TA® sorbent, thermal desorption and gas chromatography using MS or MS-FID. Berlin: Beuth 2011.

[10] ECA – European Collaborative Action (1997). Report No 19 Total Volatile Organic Compounds (TVOC) in Indoor Air Quality Investigations. EUR 17675 EN. Office for Official Publications of the European Community, Luxembourg.

[11] *Andersen, K.; Bakke, J. V.; Bjørseth, O.; Bornehag, C.-G.; Clausen, G.; Hongslo, J. K.* et al.: TVOC and health in non-industrial indoor environments: Report from a Nordic Scientific Consensus Meeting at Långholmen in Stockholm, 1996. Indoor Air 7 (1997) No. 2, p. 78-91.

[12] LEED v4.1: https://www.usgbc.org/node/12025195?return=/credits/commercial-interiors/v4.1

[13] EU LCI list: https://ec.europa.eu/growth/sectors/construction/eu-lci_en

[14] EN ISO/IEC 17025: General requirements for the competence of testing and calibration laboratories. Berlin: Beuth 2017.

Reinhard Oppl,
VOC-Consult, Wedel, Germany.

Martina Broege,
EPH, Dresden, Germany.

Dr Frank Kuebart,
eco-INSTITUT Germany, Köln, Germany.

Thomas Neuhaus,
Eurofins Product Testing A/S, Galten, Denmark.

Dr. Michael Wensing,
Fraunhofer-Institut für Holzforschung Wilhelm-Klauditz-Institut WKI, Braunschweig, Germany.