

RESEARCH

Open Access



Review and interlaboratory comparison of the Oddy test methodology

Ivan Díaz^{1*}, Alba Alvarez-Martin², Josep Grau-Bové³, Sara Norrehed⁴, Barbara Salvadori⁵, Ida Kraševc⁶, Daniel Duran-Romero⁷ and Emilio Cano¹

Abstract

Since the introduction of the Oddy test in 1973, many museums and cultural institutions have put the method in use, developing their own versions and protocols. Currently the 3-in-1 version, temperature at 60 °C and 2 g of tested material are set as common practice; however, other variables of the test are not standardized. The purpose of this study is to examine current versions of the Oddy test, to identify differences in the results derived from variations in the procedures, and ultimately raising awareness within the conservation community to work together towards a standardized protocol. In this article, we review the available information on the methodological differences in Oddy test protocols published in the literature related to glassware cleaning, coupon preparation, reaction vessel setup and rating of materials. Based on the review, and to highlight the many variables that could affect the results of the test, seven European cultural institutions working under the H2020 IPERION HS project performed a comparative 3-in-1 Oddy test by blindly evaluating the same ten materials. Each institution used its own test methodology but some guidelines were advised: (1) Detergents as a cleaning procedure for glassware, (2) P600 sandpaper or micromesh pad close to 1500 to prepare metal coupons and (3) 1:100 as water–air ratio. Despite this, differences between institutions' results were still observed. Some of them are due to the differences in the coupons preparation, either in the sanding pattern or in the edge area. In order to separate the contribution of the experimental setup and the subjectivity of the evaluation in the discrepancies, coupons from all institutions have been rated by a single team of judges with experience in the Oddy Test. Results show that differences in the evaluation criteria play a relevant role in the discrepancies of the results, especially for institutions with less experience in the test. These results highlight the need to further standardize the methodology and criteria for visual assessment. Nevertheless, the Oddy test has been found to be reliable for the identification of materials that produce emissions hazardous for the conservation of cultural assets.

Keywords Oddy test, Methodology review, Standardized protocol, Interlaboratory test, Reproducibility, Preventive conservation

*Correspondence:

Ivan Díaz

ivan.diaz@cenim.csic.es

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

The Oddy test is an accelerated corrosion test widely implemented in cultural institutions as a tool for preventive conservation. Over time, it has shown to be a reliable test for the rejection of potentially hazardous materials (woods, adhesives, textiles, etc.) and, therefore, the selection of suitable materials for the construction of showcases, museum areas and storage facilities. W.A. Oddy published in 1973 [1] the original article of the test that bears his name. Since then, many museum and conservation institutions have integrated this test into their practice, developing their own versions and protocols. Silver and lead coupons were initially placed in individual closed glass flasks together with the testing material, avoiding direct contact. Carbon dioxide was added exclusively for lead testing, periodically avoiding its depletion. After 28 days at 60 °C, a visual inspection was performed and any testing material that did not cause corrosion was considered safe for the respective metal. A blank test, i.e. with no testing material, was used as a reference for comparison. Before the Oddy test similar methods for evaluating confined metals of industrial interest were already in use [2, 3]. The great achievement of W.A. Oddy was to propose first a method focused on metals sensitive to contaminants typical for indoor museum environments and apply it to materials other than wood [4] as a source of danger on display.

Over its long lifetime, the Oddy test has been revised and improved in multiple occasions. The first modification of the Oddy test was performed by W.A. Oddy himself in 1975 [5]. A third metal coupon, copper, was also included in the test. In addition, water was added in the reaction vessel as the tarnishing of silver in the presence of hydrogen sulfide is accelerated by moisture. Initially, a few drops of water were added to the bottom of each flask; however, copper and lead coupons showed corrosion marks in the areas in direct contact with water. Hence, a small vial with moistened absorbent cotton was finally introduced. Blackshaw made further refinements to the test methodology in 1979 [6] to minimize differences between operators. The test was performed in a 250 mL flask or boiling tube. The mass of the testing material was 2 g and 1 mL of distilled water was added to the cotton-filled vial at the beginning of the test. In addition, 0.5 mL of distilled water was added weekly during the course of the test. Scientists from British Museum (BM) carried out in 1993 the first interlaboratory Oddy test [7], in which each participating institution followed its own protocol. The observation of differences in the results prompted the proposal of a standardized protocol, which was assessed in a second interlaboratory comparison [4]. The purity (99.5%), size (10×15 mm) and cleanliness (acetone)

of the metal coupons were standardized, as well as the mass of the testing material (2 g) and the water–air ratio (1:100). Distilled water was placed into a vial instead of wetting the cotton wool. The addition of carbon dioxide was also eliminated, as it was not necessary to cause lead corrosion. The vessel setup was slightly modified: the coupon no longer rested on the bottom of the flask, but hung on a nylon thread trapped by the stopper inserted in a test tube. Finally, reference photographs were prepared to assist in the evaluation of the coupons after the test. After the visual evaluation the materials were classified as permanent (P), temporary (T) and unsuitable (U) use.

In 1999, scientists from the Metropolitan Museum of Art (MMA) [8] published a more practical version of the test: the 3-in-1 Oddy test (Fig. 1a). This version consisted of simultaneously testing the three metal coupons (silver, lead and copper) in the same vessel instead of testing them individually. In addition, an alternative vessel setup was proposed. A threaded glass jar into which an inner beaker was inserted with the metal coupons bent at its rim in a V-shape. A polypropylene lid together with high-vacuum silicone grease ensured a better seal for moisture and the emitted gases.

In 2003, British Museum scientists [9] implemented the 3-in-1 version of the test, but with a different vessel setup (Fig. 1b). A new positioning of the metal coupons was proposed to improve its reproducibility. The three metal coupons were inserted in parallel into the slots cut in a silicone plug, placing the lead in the middle to avoid condensation from contact with the vessel. This was a limitation of the previous setup, so a test tube was used instead of a jar with a beaker inside.

Therefore, in the 30 years following Oddy's first publication and after numerous modifications of the test, a unified protocol has not yet been achieved. Only the use of three coupons in one vessel, temperature at 60 °C and the mass of the testing material at 2 g have been widely adopted as common parameters. Other important factors regarding the reproducibility of the test are volumes and types of vessels (flask, test tube or jar), positioning of the coupons (inserted in plugs, hanged by thread or with the use of a beaker), as well as the addition of water and its ratio (directly into the vessel, with moistened absorbent cotton or added in a vial).

New alternatives for vessel setup have recently been published. For example, the Oddy test protocol of the MMA [10] initially incorporated into its threaded glass jar a platinum-cured silicone stopper with coupons inserted in a triangular formation or a 3D printed nylon holder with coupons wrapped around a strip of the holder. Finally, a stainless steel coupon holder similar to the 3D nylon holder is used in the current protocol [11] to reduce costs and waste (Fig. 1c). Other researchers [12]

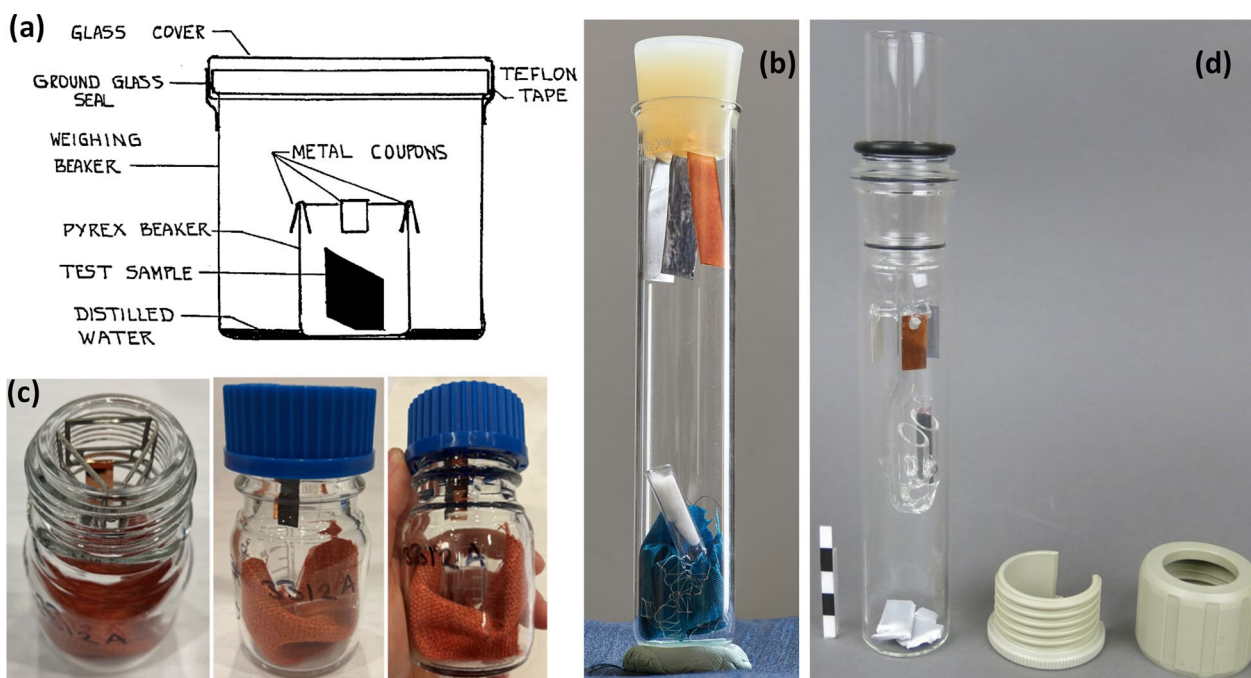


Fig. 1 Photos/illustrations of the various designs used for the Oddy test: First 3-in-1 version of the test (a). Reprinted with permission from Ref. [8]. Copyright (1999) Taylor & Francis Ltd, <http://www.tandfonline.com>. Current set-up of the British Museum (b). Reprinted with permission from Ref. [14]. Copyright (2018) Taylor & Francis Ltd, <http://www.tandfonline.com>. Current set-up of the Metropolitan Museum of Art (c). Reprinted with permission from Alayna Bone, Department of Scientific Research, The Metropolitan Museum of Art, https://www.conservation-wiki.com/wiki/Oddy_Test. MAT-CH set-up (d). Reprinted with permission from Ref. [12]. Copyright (2018) Taylor & Francis Ltd, <http://www.tandfonline.com>.

made a complete redesign of the vessel (Fig. 1d). A glass cylinder with a glass top insert from which hang glass hooks for the three metal coupons and two small vials, one for distilled water and the other one as an option for contaminant absorbing material. Others have proposed minor modifications [13], such as inserting glass hooks in a triangular formation replacing the parallel arrangement of the coupons inserted in the silicone plug.

To date, no fixed amount for the water–air ratio has been established. The volume of water varies from 0.17 to 5 mL and the volume of the test vessels range from 45 to 125 mL. In addition to the above variations, there are also different alternatives for cleaning the glassware, sanding and cleaning the coupons, sealing the vessel, sealing material, etc. [14]. All these modifications are reflected in a survey, carried out in 2014, indicating that at least 19 variations of the Oddy test were used by different U.S. cultural institutions [15]. A selection of these protocols has been tested, including some other Oddy test methodologies for testing materials in contact [16]. Contact test caused more corrosion to coupons compared to the non-contact Oddy test. Therefore, another unified protocol should be standardized in order to be able to test materials that will be in direct contact with works of art. More recently, since 2018,

the Materials Testing & Standards Committee of the AIC Materials Selection and Specification Working Group (MWG) focuses its efforts on promoting the use of protocols with the highest reproducibility between institutions [17]. In this article, we review the available information on the methodological differences of the Oddy test published to date. Four categories have been established to organize and discuss such information: glassware cleaning, coupon preparation, reaction vessel setup and rating of materials. The latter includes alternatives to reduce the subjectivity of the visual evaluation. This allows for quick and detailed comparison of the different methodologies and helps to set variables towards a unified protocol for the Oddy test. To this end, an interlaboratory test has been carried out with the participation of seven European institutions within the framework of IPERION HS (EU H2020, grant agreement No 871034). Work package 5.1: Project 7 (Interlaboratory comparison of Oddy test). The purpose of this study is to examine current and useful good practices for the Oddy test, to identify differences in the results derived from variations in the procedures, and ultimately to create awareness within the conservation community to work together towards a standardized protocol.

Literature review of Oddy test methodologies

Additional file 1: Tables S1, S2, S3 and S4 show respectively the differences in the existing protocols for glassware cleaning, coupon preparation, reaction vessel setup and rating of materials. Most of the information regarding the methodology was provided by the protocols of the European institutions with previous experience in the Oddy test that participated in the IPERION HS interlaboratory comparison. Additional protocols were obtained from the open access AIC Wiki platform [18]. The two institutions that originally contributed the most to the development of the Oddy test, the British Museum (BM) and the Metropolitan Museum of Art (MMA), head the tables in deference. The BM-2004 protocol [19] is included for reference, although this was partially refined by the BM-2017 protocol [14].

Glassware cleaning

There are two general groups among the different cleaning procedures: using detergents or not (see Additional file 1: Table S1). The detergents used such as Decon 90, Mr. Clean, Alconox, Micro-90, PCC-54 enzymatic, Extran AP17 are alkaline in nature with a pH range between 9 and 13. Theoretically, they are suitable for removing organic residues. Some institutions prefer to avoid them due to the risk of leaving residues after the cleaning process. However, some of the above detergent formulations indicate that they are free of organic surfactants and emulsifiers. The alternative to alkaline detergents is to use alkaline aqueous solutions, such as the MMA cleaning procedure. Subsequently this can be more or less exhaustive, i.e. applying acid neutralization baths with intermediate aqueous rinses in which the temperature and time vary or simply multiple aqueous rinses. Other methods, instead of using basic aqueous solutions, use diluted solutions of hydrochloric or nitric acid. Although these solutions are more suitable for removing inorganic residues, they can also be used for the removal of organic residues either because they act on the glass interface or due to their oxidizing nature, such as nitric acid. Finally, among the procedures that do not use detergents, there are also those that avoid the use of both acidic and basic baths. Instead, distilled water is used as solvent, varying or not the temperature to increase solubility, as well as consecutive rinses in organic solvents such as ethanol or acetone. These cleaning procedures are simpler because the reuse of glassware depends on the rating of previous testing material: some institutions do not reuse glassware suspicious of being contaminated by materials that have failed the test, especially for materials such as liquid coatings, adhesives and adhesive tape samples from unsuitable tests.

Coupon preparation

The summary presented in Additional file 1: Table S2 shows that the different protocols use metal coupons with areas ranging from 50 mm² (Green challenge, GC) to 350 mm² (BM). Economic reasons would be a possible explanation for the decrease in size with respect to the BM protocol. In fact, W.A. Oddy [1] initially tested larger coupons of approximately 500 mm². The coupon area does not really affect the thermodynamic tendency of metals to corrode, but may exert a visual corrosion dilution effect for large area coupons when the released contaminant is the limiting reactant in the corrosion reaction. A standard area would not only aid in the comparison with reference photographs [4], but would also eliminate possible visual effects. The thickness of the coupons also varies from one protocol to another; even within the same protocol, the thickness of the three metal coupons may be different. In most cases, the thickness of coupons is around of 0.1 mm. Handling the soft lead is easier if the coupons are thicker and a thickness of up to 0.5 mm is found in existing protocols. Thickness does not seem to affect reproducibility and processing the coupons to reduce thickness should be avoided [4], especially for lead.

The purity of metal coupons varies between 99.5 and 99.998%, only the Auckland War Memorial Museum protocol (AWMM) indicates a slightly lower purity, i.e. $\geq 99\%$. Therefore, most protocols meet the established standard of 99.5% proposed in the 1995 Oddy interlaboratory test [4]. The purpose of high purity coupons is to reduce impurities that can affect the corrosion behavior of the metals, as happened with sterling silver and low purity lead. Nowadays, higher purities can be easily achieved, thus a purity of at least 99.9% should be advised for the Oddy test. The preparation of metal coupons in terms of surface finishing also varies. Most protocols employ fiberglass brushes, following the initial British Museum protocol (BM-2004), although the most recent version (BM-2017) replaced them due to health and safety problems and concerns about their capability to remove all contaminants from the surface [14]. There are two alternatives to fiberglass brushes: micromesh pads and sandpaper. The former is used by the current BM-2017 and MMA protocols, although with different grits, 1800 and 3200 respectively. Sandpaper is only used for the protocols of the Munch Museum (MUM) and the Swedish National Heritage Board (RAA). To ensure a good reproducibility, surface finishes should be equivalent regardless of the methodology applied. On the other hand, although most protocols abrade all three metal coupons, some institutions avoid abrading the lead coupon due to health concerns (Getty Conservation Institute (GCI) and Autry

Museum of the American West (AMAW)). In contrast, in the MMA protocol only the lead coupon is sanded to remove the native corrosion layer, while copper and silver coupons are not sanded and therefore cannot be reused. This surface finishing step might be equalized between institutions if there were a single supplier of the coupons. Regarding the cleaning of the metal coupons, most protocols use very high purity acetone (>99.9%) and the MMA protocol also performs a subsequent cleaning with HPLC-grade isopropanol. In contrast, the GCI protocol submerges silver and copper coupons in Mr. Clean liquid cleaning solution (a commercial mixture of water, surfactants, solvents and preservatives) prior to cleaning with acetone. Lead coupons are not abraded or washed. Finally, a quick drying with a lint- and acid-free tissue is also common, although some protocols avoid this step and coupons are air-dried instead. In the latter case, it is necessary to avoid prolonged drying times because it favors the appearance of corrosion [14] and acetone runoffs that will hinder the evaluation stage.

Reaction vessel setup

Additional file 1: Table S3 groups the data from the different protocols with which the following sections were performed.

Temperature

The temperature in most protocols is set to 60 °C; only the AWMM and Cultural Restoration & Preservation (CRP) protocols apply lower temperatures, between 50–60 °C and 40 °C respectively. The latter also performs the test at room temperature to determine if off-gassing of contaminants occurs at this temperature. Raising the temperature up to 60 °C accelerates the processes that potentially contribute to the emission of VOCs [20], such as diffusion within a material, desorption, evaporation and chemical reactions, and it especially affects the release of less volatile compounds [21]. For example, the vapor pressure of acetic acid at 60 °C increases by about 8 times with respect to 25 °C (from 1117 to 9115 Pa), while formic acid, which is more volatile than the former, increases its vapor pressure at 60 °C by about 6 times (from 3358 to 21,082 Pa) [22]. Temperature also affects corrosion rate kinetics in aqueous media, which is 15.6 times higher at 60 °C compared to 25 °C for a typical activation energy of 65 kJ/mol [23].

Volume of water

Relative humidity is the other parameter along with temperature that makes the Oddy test an accelerated corrosion test. Since these parameters are inversely proportional, a higher absolute humidity is required to reach 100% RH at 60 °C compared to room temperature.

For this purpose, water is added to the Oddy test. However, there is no fixed ratio of water to vessel volume, which mainly affects lead due to its higher sensitivity to moisture. The lead coupons in the blank test may corrode slightly if aqueous condensation occurs [4, 14]. Hence, the water volume should be the minimum necessary to reach 100% RH without causing condensation on the metal coupons. Four protocols (BM-2017, MUM, GCI and Heritage Conservation Centre Singapore (HCC)) differentiate between the water volume added for blank tests and non-hygroscopic materials and that for moisture-absorbing materials. The water volume is lower in the first case, between 0.17 mL (BM-2017) and 1.5 mL (GCI), with vessel volumes also different (50 and 60 mL respectively). The water volume is increased for hygroscopic testing materials, although sometimes in a general way, e.g., greater than 1.5 mL in the GCI protocol. In the rest of the protocols, the water volume is ranging from 0.5 mL to 5 mL regardless of the absorption of moisture by the testing material. Overall, the water volume added is usually less than 1 mL in all protocols. Since the vessel volume varies between 50 and 125 mL, the water–air ratio differs from the ratio initially proposed by Green and Thickett as standard, i.e. 1/100 [4]. Representative examples are the BM-2017 protocol with a water–air ratio of 1/62.5 (water: 0.8 mL; vessel: 50 mL) and the MMA protocol with a ratio of 1/200 (water: 0.5 mL; vessel: 100 mL). On the other hand, relative humidity might also affect the emission rate of mainly polar VOCs from testing materials due to their possible interaction with water molecules [21]. Field studies have shown that the indoor formaldehyde vapor pressure can increase with increased RH [24].

Type of vessel and mass of testing materials

Approximately half of the protocols use a test tube, while the other half a glass jar, i.e., the existing vessel setup in the BM-2017 and MMA protocols respectively. Only the Rijksmuseum (RIJKS) protocol uses a flask as vessel. The volume of the test tubes is around 50 mL, only the IMA protocol uses substantially larger, 75 mL tubes. Heine and Jeberien recently published an alternative vessel, i.e., a flat-bottomed test tube called MAT-CH [12], although its volume is not indicated. On the other hand, the glass jars have three possible volumes, around 50 mL, 100 mL and 125 mL, i.e. generally larger than those of the test tubes. A smaller volume vessel is preferable as it increases the concentration of the volatiles emitted and facilitates a uniform distribution within the vessel. In addition, the vessel type and its volume determines the closure, the disposition of the coupons and the mass of the tested material, especially if it is of low density. The test tubes are closed with silicone stoppers and the

coupons are inserted directly into the stopper in parallel slots. Most specify that the lead be placed in the middle. The National Center for Metallurgical Research (CENIM) protocol inserts glass hooks in a triangular formation into the stopper from which the three coupons are hanged. This is to avoid capillary condensation in the crevices formed by inserting the coupons directly into the silicone stopper. The MAT-CH reaction vessel is the only one that replaces the silicone stopper with a glass insert with a double gasket that ensures a tight seal. The closure of glass jars follows the guidelines of the MMA protocol, i.e. through a screw cap, although some protocols could also use silicone stoppers, since they do not specify it. Most use earlier versions of the MMA protocol, where the metal coupons hang in a U- or V-shape from the rim of a beaker that is inserted into the jar. However, the current MMA protocol inserts a stainless steel coupon holder into the mouth of a 100 mL jar and the coupons are bent 5 to 7 mm from one end and crimped into the holder.

The type of glass used is reported only by some protocols as a commercial brand of borosilicate glass. It is necessary to avoid compositions less chemically stable under Oddy's test conditions, such as soda-lime glass. Smith [25] showed that the latter caused a passing result of the Oddy test by neutralizing possible organic acids released from the Delrin plastic (polyoxymethylene) through leaching alkalis. When borosilicate glass was used, the plastic material was rated as unsuitable.

Regarding the mass of the testing material, most protocols use 2 g. Only two protocols (AMAW and AWMM) report testing different masses, 1 g and 1–2 g respectively. Although only one of the two protocols reports the volume of its vessel (AMAW, 45 mL), it is likely that generically they refer to the limitation of low-density materials for low volume vessels. This limitation is more typical for protocols that use a test tube as a vessel, since its volume is smaller than that of glass jars.

Finally, note that some protocols also report an in-contact version of the Oddy test. This can be performed in the MMA protocol independently of the non-contact Oddy test or simultaneously, but different metal coupons are used. However, the AMAW protocol and an earlier version of the National Museum of the American Indian (NMAI) protocol use the same coupons, in such a way, that half of the coupon should touch the testing material and the other half should not.

Rating of testing materials

The rating of the testing materials is performed indirectly by visually evaluating their corrosive effect on the three metal coupons that act as corrosion dosimeters (Cu, Ag and Pb). Most protocols perform naked eye inspection, although some protocols such as Field Museum of Natural

History (FMNH) and MUM rely on optical microscopy to establish the evaluation (see Additional file 1: Table S4). The procedure is the same, to rate the suitability of the testing material for each metal coupon based on the level of corrosion in comparison to the control coupons. Following the original BM protocol, three categories are established by most protocols: suitable for permanent use (P: no change compared with control), temporary use (T: slight corrosion) and unsuitable (U: obvious corrosion). The overall rating of the tested material is the same as for the most affected coupon. Some protocols may use slightly different, but equivalent terminology, for example: pass/suitable or fail/unsuitable. There is some controversy based on subjectivity regarding the use of a material rated as temporary, ranging from 3 months for the Brooklyn Museum (BKM) and AMAW protocols to 6 months for the BM and MMA protocols. More conservative protocols such as the GCI virtually eliminate this category since it hardly uses materials rated as temporary. On the other hand, the FMNH protocol adds an additional category to the three previous ones, called limited use, indicating that objects composed of lead or calcite, such as shells, which are very sensitive to corrosion by volatile organic acids, should not be exposed to these materials. The Silver Nanofilm Sensor (SNS) protocol establishes a numerical alternative by ranking materials from 1 to 5, i.e., from the least to the greatest change in color associated with corrosion. However, this is finally reduced to the previous rating of three categories, i.e. suitable, permanent and unsuitable. Numerical ranking is useful to address reproducibility studies as previously proposed by Green and Thickett [4]. Five categories were also established: 0 (permanent use), 1 (permanent/temporary), 2 (temporary), 3 (temporary/unsuitable) and 4 (unsuitable). Several operators usually perform the evaluation, but the corrosion of the control coupons is evaluated first. If it is significant, the test is considered not valid. Few protocols adopt the additional measure of the MMA protocol to give validity to the test, i.e. weighing the assembled jar before and after the test. A loss greater than 25% of the water mass is considered a failed test due to poor sealing. The number of replicates tested for each testing material is generally accepted to be at least in duplicate, although sometimes only one replicate is tested. The BM protocol advises to occasionally test duplicate replicas and the AWMM protocol only if possible. The lack of resources, limited time or the scarcity of staff might condition the number of replicates tested as specified by the FMNH and BKM protocols.

Different evaluation methods have been proposed to reduce the subjectivity of the visual inspection of the Oddy test. One of them is to use artificial intelligence

that simulate the behavior of human operators. These are algorithms that can be trained to recognize corrosion patterns associated with the classification of materials for permanent, temporary and unsuitable use [12, 26]. Other algorithms such as k-means clustering have previously been used in digital image processing to quantify the extent of corrosion as a percentage of the total area. Wang et al. proposed the following grading using silver and copper metal nanofilms [27]: P < 20%, T 20–55%, U > 55% for silver and P < 35%, T 35–70%, U > 70% for copper. The SNS protocol based on the above study proposed a further validation for materials that are used to store daguerreotype images consisting mainly of silver, even if they pass the Oddy test, [28]. An additional Oddy test was performed with silver in the form of nanofilm deposited on glass slides. The thickness of 7 nm reproduced the behavior of daguerreotype images and was highly sensitive to corrosive contaminants, therefore the test duration could be reduced to two weeks. As an alternative to visual assessment, there have been some proposals of direct methods for corrosion measurement. Thickett quantifies oxygen depletion during the Oddy test [29]: the higher the oxygen consumption, the higher the corrosion rate. This was in agreement with visual evaluation and mass loss measurements for lead and copper, but not for silver. The CENIM protocol [13] performs standardized electrochemical reduction measurements according to ISO 11844–2 methodology [30] for silver and copper coupons: the longer the reduction time, the higher the corrosion rate. Sometimes, the visual evaluation is confusing and does not agree with the electrochemically quantified corrosion of coupons. Previously, Reedy et al. [31] and Bischoff et al. [32] proposed to replace the Oddy test with electrochemical tests instead of supplementing it. Aqueous extracts were obtained from testing materials and used as electrolyte for electrochemical measurements, such as corrosion potential, polarization resistance or its conversion into corrosion current. This allows quantifying the corrosion rate of metal coupons; however, the aqueous extract does not

reproduce the time condition of the Oddy test. Another alternative is to use analytical techniques such as different gas chromatography-mass spectroscopy (GC–MS) methods to detect VOCs released from tested materials [33]. However, their corrosive effect on silver, copper and lead is not always known. In addition, the few minutes usually taken for the entire experiment hinders the detection of secondary VOCs, i.e., those generated by the degradation of tested materials that may be emitted after several weeks under Oddy test conditions. More limited is the use of classical tests based on wet chemistry analysis for the specific detection of certain corrosive volatiles, for example, the Beilstein test for chlorides and the Purpald and chromotropic acid tests for aldehydes [19]. pH measurements of tested materials, either from aqueous extracts, on surface or with A-D strips, have been also performed by protocols such as the British Museum's to quickly discard materials. Since few materials failed the A-D strip test, this was abandoned as a routine test [14]. Finally, other methods focus on the characterization of corrosion products formed on Oddy test coupons, aiding in the possible identification of corrosive volatiles emitted by the tested materials. The different characterization techniques include X-ray diffraction [34], μ Raman spectroscopy [35], Fourier transform infrared spectroscopy [36], as well as quartz crystal microbalance [37]. Despite the numerous options for scientific analysis of corrosion products and emissions formed during the Oddy test, it is worth remembering that the Oddy test is a tool widely used by institutions without access to that type of instrumentation. Even large institutions have limited resources in terms of staff and time to routinely conduct this type of analysis. Finding ways for standardizing visual inspection therefore remains important.

Experimental

Interlaboratory test

Seven European cultural institutions or research centers (see Table 1) performed a comparative 3-in-1 Oddy test (28 days, 60 °C) under the umbrella of IPERION

Table 1 Institutions participating in the interlaboratory comparison of the 3-in-1 Oddy test, in alphabetic order

Institution	Country
Institute for Sustainable Heritage, University College London (UCL)	UK
Institute of Cultural Heritage of Spain, Ministry of Culture and Sport (IPCE)	Spain
Institute of Heritage Science (ISPC), National Research Council (CNR)	Italy
National Center for Metallurgical Research (CENIM), Spanish National Research Council (CSIC)	Spain
Rijksmuseum Amsterdam (RIJKS)	The Netherlands
Swedish National Heritage Board (RAA)	Sweden
University of Ljubljana (UL)	Slovenia

HS. Five of them had years of experience with the Oddy test, as well as the judges who evaluated the coupons. The other two institutions (II and V), although without previous experience in the Oddy test, were experts in heritage science. Results here are presented anonymously, using Roman numerical notation from one to seven, not corresponding to the order of Table 1. Ten materials were blind tested in duplicate as well as the blank test. The materials were selected and sent by CENIM to each institution, identifying them only with numbers from one to ten. The tested materials and their general composition is shown in Table 2. Material 9 is the only non-solid product before application. Due mainly to the lack of space in the vessels of some institutions (50 mL test tube: 3.4×10 cm), it was not applied as a standard 6×12 cm area coating [14]. Instead, 2 g of the material were applied on aluminum with approximate dimensions of $3 \times 1 \times 0.5$ cm and allowed to cure into a solid. The time elapsed from application to receipt by the participants in the interlaboratory test was at least four weeks. Each institution prepared and weighed 2 g of each solid material for testing, with the exception of material 1 (Ethafoam). Due to its low density, the institutions with reaction vessel volumes less than 135 mL weighed 1 g of this material.

Although each institution applied its own Oddy test methodology, certain constraints were established to limit variables:

- (1) Detergents should be used as a cleaning procedure for glassware, either manually or with a dishwasher.
- (2) To prepare coupons, micromesh pads or sandpaper should be used with grit sizes that provide similar surface finishes (P600 sandpaper or micromesh pad close to 1500). Although micromesh grit does not follow the codification established by the Federation of European Producers of Abrasives (FEPA)

for sandpaper, both are convertible for fine abrasive grits.

- (3) The ratio of water volume to vessel volume was decided to 1/100.

Institution III could not meet constraints 1 and 2 due to its internal operating policy. Instead, new glassware was rinsed with deionized water and dried in the oven overnight. The metal coupons were abraded as originally proposed by the British museum protocol, i.e., with glass brushes.

Regarding the cleaning procedure of metal coupons, all institutions used ultra-high purity acetone as well as a lint-free cloth for drying. Details on glassware cleaning, coupon preparation and reaction vessel setup for each of the participating institutions are shown in Tables 3, 4 and 5, respectively. Finally, each institution evaluated by visual inspection the three metal coupons (silver, copper and lead) and their level of corrosion was compared to the blank test to rate the materials as suitable for permanent (P) or temporary (T) use and unsuitable (U) for use. Institutions should avoid assessing the area close to the insertion of the coupon into the plug when it is the only area affected, as corrosion usually starts from the lower edge of the coupon. The overall rating for each material was decided by selecting the results of the most corroded coupon.

Results and discussion

Table 6 shows the images of the silver, copper and lead coupons from the participant institutions after the Oddy test. Coupons were returned to CENIM and photographs were immediately taken keeping the same illumination and exposure in all cases by using a light box and adjusting the white balance. The rating of each coupon by the respective institutions is also included. Table 7 shows in addition the overall rating of the 10 materials tested by

Table 2 Materials tested in the interlaboratory comparison of the 3-in-1 Oddy test

Tested materials in the Oddy test	
Number 1	Ethafoam (low-density plastic foam made by expanding polyethylene)
Number 2	Tygon S-50-HL (tube made of polyvinyl chloride)
Number 3	Velcro (nylon and polyester are usually used in its manufacture)
Number 4	Heat shrink tubing (made from flexible, flame retardant polyolefin)
Number 5	Plain black rubber
Number 6	Photo Archivakarton Chronos (paper for photographic conservation with a composition of 95–97% Alpha Cellulose)
Number 7	Medium-density fiberboard (MDF, made up of 82% wood fiber, 9% urea-formaldehyde resin glue, 8% water, and 1% paraffin wax)
Number 8	Plywood board (made of birch wood sheets glued with phenolic resin)
Number 9	MS-35 Sealing (hybrid modified silane; MS, polymer of one component. Although no further information is provided by the manufacturer, two possible “chemical skeletons” for modified silanes would be: MS-polyester or MS-polyurethane)
Number 10	Polylactic acid filament for 3D printing

Table 3 Glassware cleaning procedures of metal coupons used by each participating institution in the interlaboratory comparison of the 3-in-1 Oddy test

Participants	Detergent used for glassware cleaning
I	New glassware was used and hand washed with a 2% solution of Decon 90 in deionized water. The test tubes remained submerged in the solution for several hours, ensuring that no air bubbles were trapped in the tubes. Then rinsed in distilled water overnight and subsequently rinsed again in distilled water several more times. Finally washed with technical grade ethanol
II	New glassware was used. It was rinsed with a 0.1% solution of ZVG Washing up liquid HS 50420–010 in tap water. Then rinsed again several times (until no bubbles were formed) with demineralized water, air-dried, rinsed with pure ethanol and finally air dried
III	New glassware was used for this experiment, all Erlenmeyer flask were rinsed with deionized water and dried in the oven overnight
IV	Schott glassware and caps were washed with detergent in warm water (Deconex 20 NS-x, 250 mL per 10 L water , submerged in a tub), rinsed with tap water, deionized water and MQ water. All pieces were ultrasonicated for 15 min in MQ water and dried in an oven at 110 °C for 16 h Silicone stoppers were rinsed with deionized water and left to air dry. Three incisions were cut into each stopper with a scalpel
V	New glassware was used and glassware was manually cleaned with Alconox (1%) detergent. A sponge was dedicated for washing Oddy testing glassware. Glassware was rinsed with deionized water several times and finally washed with ACS acetone. All containers were air-dried completely
VI	Schott glasses, caps, glass vials and silicone stoppers were left to soak in a 5% Decon 90/DI water solution for 4 h, manually washed with a clean sponge, and subsequently rinsed with DI water 3–5 times The Schott glasses, caps, glass vials and silicone stoppers were then wrapped in aluminum foil and placed in an oven at 105 °C to dry overnight (8 h +)
VII	Standard laboratory dishwasher (with Extran AP17 liquid detergent)

Table 4 Surface preparation of coupons performed by the institutions participating in the interlaboratory comparison of the 3-in-1 Oddy test

Participants	Coupon preparation		Additional preparation
	Sandpaper	Micromesh pad	
I	P600 grit	–	A hole of 2 mm in diameter was made in each coupon at a distance of 5 mm from one edge to be hung from glass hooks
II	P600 grit	–	–
*III	Bristle brushes	–	A hole is drilled in each coupon and marked with an F (front) inscribed to the right or left of the hole. The hole is used to knot each coupon to an inert wire from which they hang
IV	–	1500 grit	–
V	P600 grit	–	Each coupon was labelled with numbers by incision with a scalpel on the edge to be pushed in the stopper
VI	–	1500 grit	–
VII	–	1800 grit	–

*The policy of this institution prevented the use of sandpaper and micromesh during the surface preparation of the metal coupons

each institution from the individual evaluation of the three metal coupons.

It should be noted that, although photographs are helpful as reference, they cannot replace direct visual evaluation. Direct visual examination allows to assess features such as the loss of shine or the thickness of the corrosion layer, and distinguish between reflections of the surface and actual degradation layers. All ratings presented in this work are based on direct visual examination.

Differences between institutions were observed in the surface finish of the metal coupons even after completion of the test (see Table 6). This is expected for

coupons sanded with abrasive grit sizes different from those initially proposed as equivalent (P600 sandpaper and 1500 micromesh pad), although could also be attributed to how much pressure is applied during sanding and overall accuracy of the operator. Unsanded horizontal lines can be observed on the copper coupons from institution VII, which come from the forming of the copper sheet in the as-received condition (Fig. 2). Therefore, the 1800 grit size of the micromesh used by this institution may be too fine to uniformly prepare this type of copper coupons. A more consistent alternative could be to adopt the standard method from metallographic preparation laboratories: sanding

Table 5 Information on the reaction vessel setup used by each institution participating in the interlaboratory comparison of the 3-in-1 Oddy test

Participants	Reaction vessel setup			Water	Stopper	Coupons
	Test tube	Jar	Vol.			
I	x		50 mL	0.5 mL	Silicone	The metal coupons were hung on the silicone stopper by means of glass hooks inserted in triangular formation. The position of the coupons is equivalent. Use wide plastic ties to ensure the tightness of the stopper
II	x		150 mL	1.5 mL	Silicone	The metal coupons were inserted in parallel into silicone stopper (lead in the middle)
III	–	–	250 mL	2.5 mL	Glass lid	The coupons were hung one at a time from an inert nylon wire taped to the outside of the Erlenmeyer flask. The tape is removed when the setup is placed in the oven. A 10 ml inert beaker was used to hold the vials containing water
IV		x	135 mL	1.23 mL	Silicone	The coupons were inserted in parallel into silicone stoppers (silver in the middle)
V	x		50 mL	0.5 mL	Silicone	The coupons were inserted in parallel into silicone stoppers (lead in the middle)
VI		x	100 mL	1 mL	Silicone	The three metal coupons were inserted in parallel into silicone stopper. An extra incision was made to insert a stainless steel wire (0.08 mm diam), which was used to secure the silicone stopper and attached coupons at the top of the reaction vessel. The bottle cap was screwed on while holding the wire in place
VII	x		54 mL	0.5 mL	Silicone	The metal coupons were inserted in parallel into silicone stopper (lead in the middle)

After 15 min at 60 °C, the silicone stoppers and screw caps were checked for tightness to ensure an airtight seal.

is done progressively from lower to higher grit, rotating the sanding direction 90 degrees when passing from one grit to another.

The use of glass bristle brushes also does not remove pre-existing deformations on lead coupons; at least within a reasonable sanding time (Fig. 3). Less control over the desired surface finish is also observed with sanding lines appearing in directions other than longitudinal (especially copper and lead). Its limited roughing capacity together with the breakage of the glass bristles into small pieces could explain both facts.

As for the rest of the institutions that did use equivalent abrasive grit, differences in surface finish were also observed due to the sanding procedure. Not all institutions selected a preferred sanding direction. This can be seen in the silver coupons of institutions such as *II* or *VI* compared to the silver coupon from institution *I* (Fig. 4).

Random sanding causes shiny spots as a result of the different pressure exerted on the coupon during sanding. A different sanding pattern is also observed on the edges of the silver coupons of institution *V* and *VII* compared to the rest of the coupon, with different shades of grey appearing, not associated with tarnishing during the Oddy test (Fig. 5). This issue could be avoided by abrading a large surface of metal and then cutting coupons, rather than abrading each coupon individually.

It would be advisable not only to use equivalent sanding grits, but also to establish a preferred sanding direction to avoid possible confusion during coupon rating. Sanding

is performed to remove the existing native corrosion layer on the metal coupons and thus activate their surface before starting the Oddy test. Differences in surface finish could affect the reproducibility of the test between institutions if they were misinterpreted as deterioration during the rating of the metal coupons.

Regarding the evaluation of the testing materials from the participating institutions (see Table 7), the most notable differences came from institution *II*, which rated materials 2, 3 and 6 as unsuitable, in contrast to the other institutions. These differences could be associated with a more conservative evaluation criteria, perhaps due to its lack of experience and training with the Oddy test and/or with the institutional protocol itself. A frequency table, such as Table 8, representing the sum of permanent, temporary and unsuitable ratings per institution could help to differentiate between these.

Institution *II* rated silver coupons as permanent 6 times, the average for all institutions being 8, while the copper and lead coupons only received temporary or unsuitable ratings by this institution (24 times in total). These values are higher than the average obtained, departing from it by at least one standard deviation, a parameter indicating the dispersion of the set of ratings. The copper and lead coupons are specifically responsible for the rating of materials 2, 3 and 6 as unsuitable. Therefore, the discrepancies seem to be due to a more conservative evaluation criteria for these coupons. The principal component analysis (PCA) performed with overall data

Table 6 Silver, copper and lead coupons (from left to right) from each institution (I-VII) after the 3-in-1 Oddy test for the ten materials listed in Table 2

	I			II			III			IV			V			VI			VII		
Mat	P	P	P	P	T	T	P	P	P	P	P	P	P	P	T	P	P	P	P	P	P
1																					
2	P	P	P	P	U	T	P	T	T	P	P	P	P	P	T	T	P	P	P	P	P
3	P	T	P	T	U	U	P	T	T	P	P	P	P	P	T	P	P	P	P	P	T
4	P	T	T	P	T	T	P	T	U	P	P	T	P	P	T	P	P	T	P	P	T
5	U	U	U	U	U	U	U	U	U	T	U	T	P	U	P	U	U	P	P	U	U
	I			II			III			IV			V			VI			VII		
Mat	P	P	P	P	T	U	P	T	P	P	P	T	P	P	T	P	P	T	P	P	P
6																					
7	P	T	U	U	U	U	P	T	U	P	P	U	P	T	U	P	P	U	P	P	U
8	T	U	U	P	T	U	P	T	U	P	P	U	P	T	U	T	U	U	P	P	U
9	P	U	P	T	T	U	P	P	P	P	P	P	P	P	P	P	T	P	P	P	P
10	P	U	T	P	T	U	T	U	U	P	U	U	P	U	T	P	U	U	P	U	U

P means permanent use (no corrosion), T, temporary use (slight corrosion) and U, unsuitable use (large amount of corrosion)

Table 7 Rating of the suitability of the ten test materials, individually for the three metal coupons and overall, according to the 3-in-1 Oddy test of each institution

Participants	Mat. 1			Mat. 2			Mat. 3			Mat. 4			Mat. 5			Mat. 6			Mat. 7			Mat. 8			Mat. 9			Mat. 10					
	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb
I	P	P	P	P	P	P	P	T	P	P	T	T	U	U	U	P	P	P	P	T	U	T	U	U	P	U	P	P	U	T			
		P			P			T			T			U			P			U			U			U			U				
II	P	T	T	P	U	T	T	U	U	P	T	T	U	U	U	P	T	U	U	U	U	P	T	U	T	T	U	P	T	U			
		T			U			U			T			U			U			U			U			U			U				
III	P	P	P	P	T	T	P	T	T	P	T	U	U	U	U	P	T	P	P	T	U	P	T	U	P	P	P	T	U	U			
		P			T			T			U			U			T			U			U			P			U				
IV	P	P	P	P	P	P	P	P	P	P	P	T	T	U	T	P	P	T	P	P	U	P	P	U	P	P	P	P	U	U			
		P			P			P			T			U			T			U			U			P			U				
V	P	P	T	P	P	T	P	P	T	P	P	T	P	U	P	P	P	T	P	T	U	P	T	U	P	P	P	P	U	T			
		T			T			T			T			U			T			U			U			P			U				
VI	P	P	P	T	P	P	P	P	P	P	P	T	U	U	P	P	P	T	P	P	U	T	U	U	P	T	P	P	U	U			
		P			T			P			T			U			T			U			U			T			U				
VII	P	P	P	P	P	P	P	P	T	P	P	T	P	U	U	P	P	P	P	P	U	P	P	U	P	P	P	P	U	U			
		P			P			T			T			U			P			U			U			P			U				

P means permanent use (no corrosion), T, temporary use (slight corrosion) and U, unsuitable use (large amount of corrosion)



Fig. 2 Copper coupons from Institution VII abraded with 1800 micromesh pads showing horizontal unsanded lines after testing material 1 (left) and material 8 (right)



Fig. 3 Lead (left) and copper (right) coupons from Institution III previously abraded with glass brushes showing unsanded lines after testing materials 1 and 7, respectively

in Table 8 would indicate that Institution II is an outlier (Fig. 6). For this analysis, each variable in the dataset was centered by subtracting its mean, and then scaled by dividing by its standard deviation. This ensured that each variable had a mean of 0 and a standard deviation of 1.

PC1 predominantly separates Institution II from the rest, while PC2 separates Institution V. The arrows indicate the loadings of different variables. It can be seen that Institution II is defined by more classifications as U and T. These differences are more pronounced with the PCA conducted (Fig. 7) using the raw data of Table 7. To be able to conduct the PCA with the categorical data with three categories, it was encoded following the one-hot method, which is used to represent categorical data numerically [38]. In this encoding scheme, each category (P, T or U) is represented

by a binary vector, where all elements are zero except for the one corresponding to the index of the category. This was necessary in order to turn the categorical variables into numerical variables that could be used with PCA.

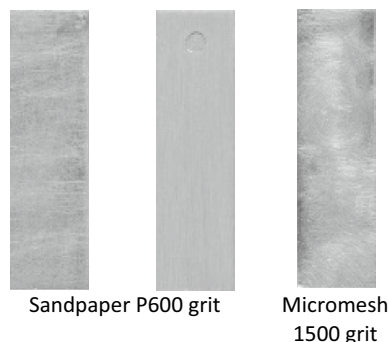


Fig. 4 Silver coupons previously abraded (from left to right: Institution II, I and VI) after testing material 6

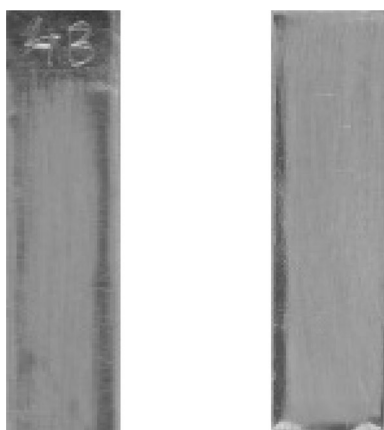


Fig. 5 Silver coupons of the Institutions V and VII after testing material 4

Figure 7 shows only one axis, rather than a complete biplot with variable loadings, because the one-hot encoding method increases the number of variables, rendering the plot too busy for visualization of their individual loadings.

To further explore what makes Institution II different, the contribution of each variable to PC1 can be observed in Figure S1 (supplementary information: file 1). There are no materials that contribute more strongly to PC1 than others; they are due to small differences between all materials. PCA has been used extensively in heritage science as a technique for dimensionality reduction. In the case of metals, it has been used to group bronze artefacts according to the color of their patina [39] or to study relationships between objects according to their elemental composition [40]. PCA is also commonly used to evaluate behavioral patterns, for example, how museum visitors rank certain aspects of their experience [41].

The inter-institutional agreement has been statistically evaluated through the Fleiss’s kappa measure. It considers agreement between institution ratings while accounting for the agreement that could occur by chance, providing a measure of agreement adjusted for random agreement (0 indicates no agreement between institutions beyond what would be expected by chance and 1 perfect agreement). It is commonly used in inter-observer comparison studies, for example, it has been used to see if policymakers agree in digitization priorities for heritage [42].

Table 9 shows Fleiss’s kappa obtained from the evaluations in Table 8. It can be seen that agreement between institutions increases when institution II is removed. The highest agreement occurs in the evaluations rated as unsuitable, followed by permanent. This would show the reliability of the Oddy test to reject hazardous materials, accepting those that are safe. It is in the intermediate or temporary rating where there is the least agreement between institutions.

Discrepancies in the ratings of the institutions could be of two types: (1) associated to a differential or subjective evaluation criterion (which might be affected by inexperience of the evaluator); and 2) those associated to real differences in the degree of corrosion of the metal coupons evaluated by the institutions. Keeping the same evaluator would help to distinguish one from the other. This has been addressed through independent evaluation of coupons from all institutions by two judges with extensive experience in the Oddy test. To help a uniform and coherent evaluation of the different types and extents of corrosion observed in the coupons, the reference photographs of scored coupons from the Metropolitan Museum of Art were used [43–45]. These documents include reference photographs, along with detailed descriptions of the morphology, color and extent

Table 8 Table of frequencies (total and individual per metal coupon) for permanent, temporary and unsuitable ratings made by each institution irrespective of the testing material

Institution	Total			Silver			Copper			Lead		
	P	T	U	P	T	U	P	T	U	P	T	U
I	16	6	8	8	1	1	3	3	4	5	2	3
II	6	11	13	6	2	2	0	6	4	0	3	7
III	13	9	8	8	1	1	2	6	2	3	2	5
IV	21	4	5	9	1	0	8	0	2	4	3	3
V	18	8	4	10	0	0	6	2	2	2	6	2
VI	18	5	7	7	2	1	6	1	3	5	2	3
VII	22	2	6	10	0	0	8	0	2	4	2	4
Average	16.3	6.4	7.3	8.3	1.0	0.7	4.7	2.6	2.7	3.3	2.9	3.9
Standard deviation	5.4	3.1	2.9	1.50	0.82	0.76	3.09	2.57	0.95	1.80	1.46	1.68

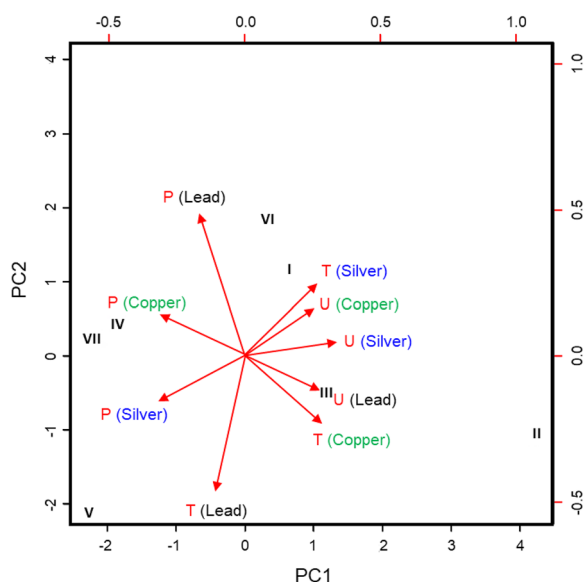


Fig. 6 PCA performed from the overall assessments in columns 4–12 of Table 8. The bottom and left axes of this biplot correspond to the institutions (I–VII). The top and right axes correspond to the variables (evaluations of P, T and U for each metal). These axes indicate the direction and strength of the variables in the space defined by PC1 and PC2. The variable axes help indicate which classifications explain the observed differences between institutions

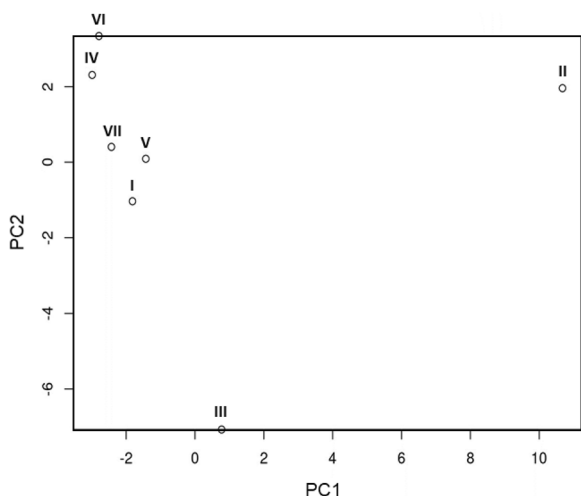


Fig. 7 PCA result with One-Hot Encoded Variables P, T and U

Table 9 Fleiss’s Kappa statistical measures regarding the agreement of the institutions’ overall ratings (P, permanent; T, temporary and U, unsuitable) shown in Table 8

Participants	P	T	U
Institutions I–VII	0.373	0.095	0.551
Excluding II	0.463	0.130	0.706

of corrosion observed, that help assigning consistent ratings to the different degradation phenomena observed in the coupons. This is especially helpful for the borderline situations (between P and T, and between T and U) in which the more simple description by Thickett and Lee [19] leave room for a more subjective decision of the evaluator.

The results of the single evaluation are observed to be more reproducible (Table 10). The discrepancies of Institution II disappear, thus showing that these were not due to differences in the degree of corrosion compared to other institutions, but to an overly conservative evaluation criterion. However, other discrepancies remain due to differences in the extent of corrosion. For example, the greatest disagreement obtained for silver is due to material 5. Silver was mostly rated as unsuitable and yet two institutions (V and VII) did not detect tarnish on their respective silver coupons, rating it as permanent. Both had poor sanding on the edges of the silver coupon, a general example of which was shown earlier in Fig. 5. This might explain why no corrosion was detected. In spite of this, the assessment of silver coupons shows the lowest dispersion between institutions. Its noble nature limits its rating as unsuitable since fewer materials were able to corrode it (see Table 10) and the dark and differential color of its corrosion products facilitates better reproducibility with respect to copper and lead.

Another discrepancy that remains is that of material 9, rated as unsuitable by institution I, in contrast to the permanent rating of most of the institutions. The black corrosion on the lower edge of the copper coupon differs from the dark red of the other institutions, suggesting that the chemical nature of the deterioration would be different. Since the cleaning procedure, surface finish and water–air ratio were restricted, the reason could be due to the preparation of the testing material. Material 9 is a sealant that was applied to a sheet of aluminum foil and left to air cure for several days before being shipped to the institutions. Institution I cut the material as finely as possible with scissors, i.e. between 2 or 3 mm per dimension. This increases the material surface area from which contaminants would diffuse and therefore their diffusion rate would increase. Another reason could be the time elapsed between receipt of the material and its testing and possible changes of the material during this period, including the evaporation of solvents.

On the other hand, interesting information is obtained from the testing of material 10. The volume of the reaction vessel could have controlled the extent of the deterioration on the copper coupon surface (Fig. 8), although all coupons have been rated as U, except institution II who rated it as T.

Table 10 Rating of the metallic coupons of all the institutions (I-VII) by evaluators with experience in the Oddy test

Participants	Mat. 1			Mat. 2			Mat. 3			Mat. 4			Mat. 5			Mat. 6			Mat. 7			Mat. 8			Mat. 9			Mat. 10								
	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb	Ag	Cu	Pb			
I	P	P	T	P	P	T	P	P	T	P	P	T	U	U	T	P	P	T	P	T	U	P	U	U	P	U	T	P	U	U	P	U	U	P	U	U
		P				P		T	P		T				U			P			T						P						T			
II	P	P	T	P	T	T	P	T	T	P	P	T	U	U	T	P	P	T	P	T	U	P	T	U	P	P	T	P	P	T	P	T	T			
		T			U			T	U		T				U		T	U		T	U		U	U		T	T		T	U			U			
III	P	P	P	P	T	T	P	T	T	P	P	U	U	U	T	P	T	T	P	T	U	P	T	U	P	P	T	P	P	T	T	U	U			
											T				U			P									P									
IV	P	P	P	P	T	P	P	P	T	P	P	T	U	U	T	P	P	T	P	P	U	P	T	U	P	P	P	P	U	U	P	U	U			
					P				P					T									P													
V	P	T	T	P	P	P	P	P	P	P	P	T	P	U	P	P	P	T	P	T	U	P	T	U	P	P	P	P	U	U	P	U	U			
		P				T			T																					T						
VI	P	P	P	P	P	P	P	P	P	P	P	U	U	U	P	P	P	P	P	P	U	P	T	U	P	P	P	P	U	U	P	U	U			
					T						T						T						T	U		T										
VII	P	P	P	P	T	T	P	P	T	P	P	T	P	U	U	P	P	T	P	P	U	P	T	U	P	P	T	P	U	U	P	U	U			
					P	P												P					P			P										

As a reference, the ratings of the institutions that disagree with those of the experts are included in gray. P means permanent use (no corrosion), T, temporary use (slight corrosion) and U, unsuitable use (large amount of corrosion)

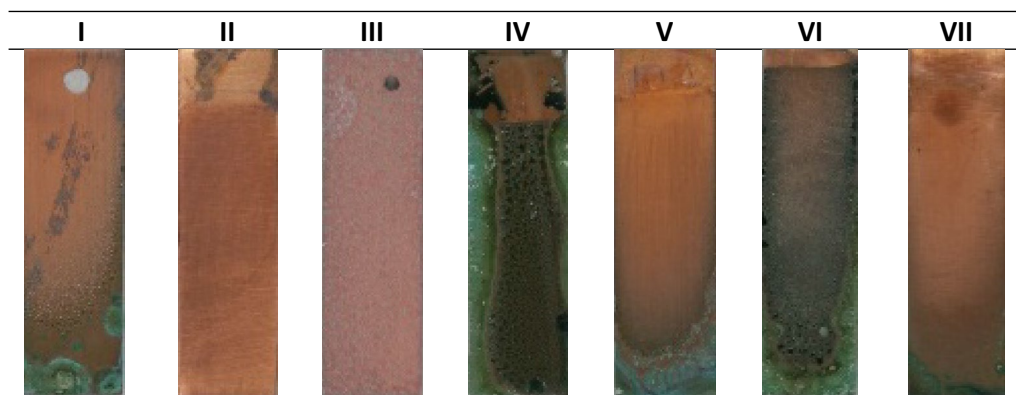


Fig. 8 Close-up of copper coupons from all institutions (I-VII) after testing material 10

The extent of green corrosion is lower in institutions (I, V and VII) with reaction vessel volumes close to 50 mL (test tubes), while it is widespread in institutions (IV and VI) with higher volumes, around 100 mL (glass jars). In both cases, the contaminant diffuses upward from the testing material according to the concentration gradient. However, if the vessel is wide and larger in volume such as glass jars, there would be a greater possibility of lateral interactions reaching the entire surface of the coupon. On the contrary, in small and narrow vessels such as a test tube, the contaminant would interact initially and mostly with the lower part of the coupon. This area would deplete the contaminant, making it less accessible to the rest of the coupon.

Finally, Fig. 9 shows a graphical comparison of the agreement and disagreement counts between the independent evaluation (single evaluator) and that of the set of multiple evaluators corresponding to the seven participating institutions.

The paired counts represent the number of times both evaluations (single and multiple) agreed or disagreed in their assessments of the set of coupons, categorized as P, T, and U. In the majority of cases (162 out of 210 evaluations, 77%) both evaluations coincide in the rating of metal coupons (green bubbles). 47 out of 210 evaluations (22%) show a discrepancy with the closest category (orange bubbles), which in many cases can be related to borderline situations in which the rating to be assigned is dubious. Only 1 out of the

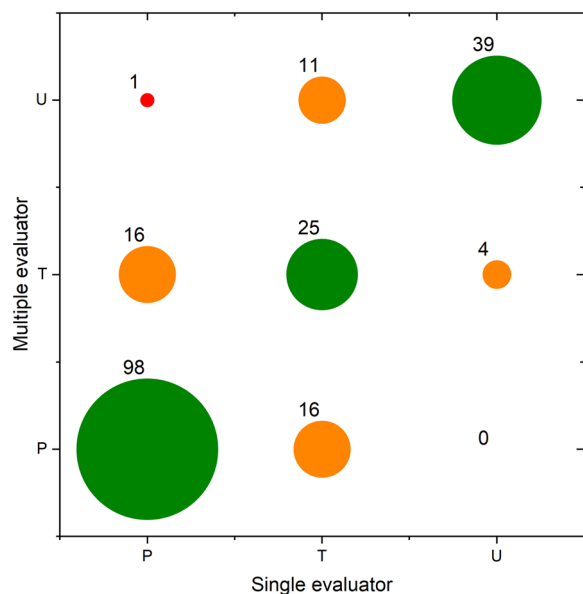


Fig. 9 Paired counts of the ratings of all metal coupons obtained after the single and multiple evaluation. P means permanent use (no corrosion), T, temporary use (slight corrosion) and U, unsuitable use (large amount of corrosion). Green bubbles indicate agreement, orange bubbles indicate disagreement by one category and red bubbles by two categories

210 coupons evaluated shows a relevant difference (from P to U), which would disappear (Fig. 10) excluding the contribution of institution II (outlier from the principal components analysis) for the silver coupons. Figure 10 also shows that the discrepancies with respect to the temporary rating are concentrated in the copper and mainly lead coupons. However, these appear to be defined within the P–T range, excluding Institution II.

Regardless of the discrepancies observed, and although efforts to reduce the subjectivity of the evaluation by means of detailed references [43–45] and /or instrumental measurements [13] are always welcomed, these results show that the Oddy test is a valid test to detect materials that can emit harmful pollutants and should be avoided in the environment of cultural heritage assets.

Conclusions

Although numerous cultural institutions and museum across the world rely on the Oddy test, no consensus has been reached on the protocol to be followed. This

article aims to raise awareness within the conservation community by highlighting the discrepancies that may arise when different non-standardized protocols are conducted to analyse the same batch of materials. Firstly, we present a review of the available information on the methodological differences in Oddy test protocols published in the literature up to date, focusing on the different variables that can affect the results. Review of current practices showed that, although some parameters have been widely adopted (the 3-in-1 procedure, temperature at 60 °C and 2 g of material), others are not yet standardized.

The second part of the study shows an interlaboratory comparison performed by seven European institutions under the umbrella of the IPERION HS project. Some guidelines were advised, such as glassware cleaning, coupon preparation and air–water ratio, to constrain the variables that can affect the results.

The interlaboratory comparison have shown some discrepancies in the ratings assigned to the same material by different institutions. Main discrepancies are found in materials rated as “Temporary”, especially for copper and lead coupons.

Some discrepancies between institutions might be due to non-standardized methodological differences in the protocols. Surface inhomogeneities, arising from a poor edge preparation or differences in sanding pattern, might introduce confounding factors in the evaluation step. Obtaining in the whole surface of the coupon a uniform surface finish with equivalent grit size, preferred sanding direction and applying the methodology of metallographic preparation laboratories (i.e., sanding progressively from lower to higher grit), could ensure a consistent reaction of the metal and help to avoid possible confusion during coupon rating. Establishing a standardized methodology and providing thorough user training would help to reduce possible discrepancies in the results.

Notwithstanding this, our results show that the main differences arise from the evaluation of the coupons. When performed by single experienced evaluators using detailed reference photographs and aspect descriptions, the discrepancies in the ratings are largely reduced, showing the validity of the Oddy test for identifying harmful materials for the conservation of cultural heritage assets.

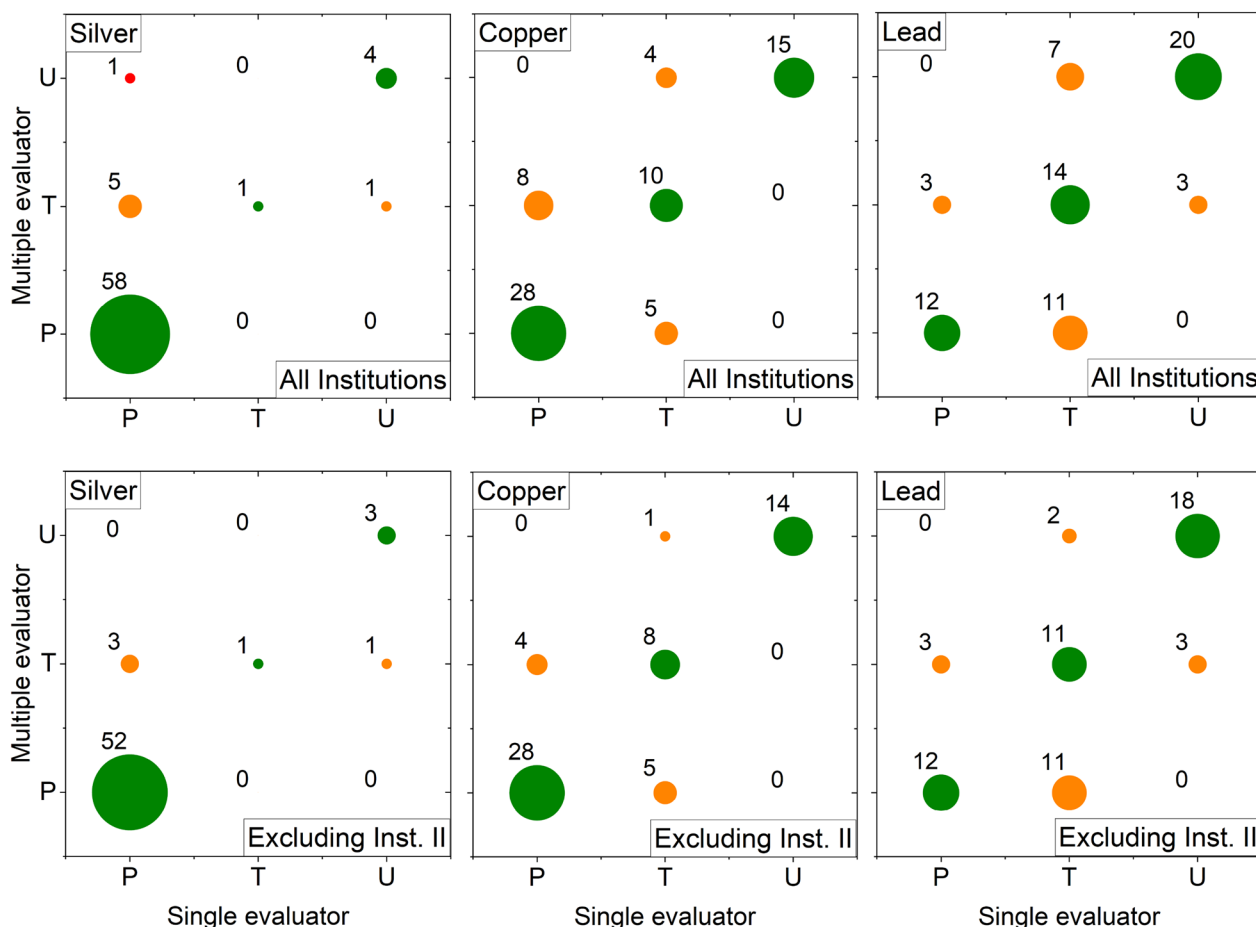


Fig. 10 Paired counts of individual silver, copper and lead coupon ratings obtained after the single and multiple evaluation. P means permanent use (no corrosion), T, temporary use (slight corrosion) and U, unsuitable use (large amount of corrosion). Green bubbles indicate agreement, orange bubbles indicate disagreement by one category and red bubbles by two categories

Abbreviations

- BM-2004 British museum, 2004 protocol [19]
- BM-2017 British museum, 2017 protocol [14]
- MMA Metropolitan Museum of Art [18]
- GC The Green Challenge [18]
- AWMM Auckland War Memorial Museum Protocol [18]
- MUM Munch Museum
- RAA Swedish National Heritage Board
- GCI Getty Conservation Institute [18]
- AMAW Autry Museum of the American West [18]
- CRP Cultural Restoration & Preservation [18]
- HCC Heritage Conservation Centre Singapore [18]
- RIJKS Rijksmuseum Amsterdam
- CENIM National Center for Metallurgical Research, CSIC [13]
- NMAI National Museum of the American Indian
- FMNH Field Museum of Natural History [18]
- BKM Brooklyn Museum [18]

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40494-024-01174-9>.

Additional file 1: Table S1. Different cleaning procedures for Oddy test glassware used in different protocols in the literature. **Table S2.** Differences in the preparation of Oddy test metal coupons in different protocols in the literature. **Table S3.** Different reaction vessel setups of the Oddy test in different protocols in the literature. **Table S4.** Visual evaluation of the Oddy test in different protocols in the literature. Figure S1. Contribution of the different materials to PC1 for Institution II.

Acknowledgements

E. Cano and I. Díaz acknowledge the professional support provided by the CSIC ‘Open Heritage: Research and Society (PTI-PAIS)’ Interdisciplinary Thematic Platform. The authors would also like to thank Irene Roncero Pérez for her participation in the IPCE Oddy test; M. Strlič from the University of Ljubljana for sharing ideas on the Oddy test; Irina Sandu from the Munch Museum and Juliette Remy from the Centre for Research and Restoration of the Museums of France for sharing their respective protocols. A. Alvarez-Martin would like to thanks Sara Creange and Maxime Gerber for their help during the test.

Author contributions

Conceptualization; ID and EC. Funding acquisition; EC. Investigation; ID, EC, AAM, IK, BS, SN and JGB. Methodology; ID, EC, IK, BS, SN, DD and JGB. Validation; ID. Formal analysis; ID. Writing-original draft; ID. Writing-review and

editing; ID, EC, AAM, IK, SN and JGB. Supervision; ID and EC. All authors have read and approved the final manuscript.

Funding

Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This research has been funded by the projects IPERION HS (European Union's Horizon 2020 research and innovation programme, grant agreement No 871034) and TOP-HERITAGE CM (Comunidad de Madrid and European Structural Funds, Ref. S2018/NMT4372).

Availability of data and materials

All data generated or analyzed during this study will be made available at Digital.CSIC, the CSIC open repository (<https://digital.csic.es>).

Declarations

Competing interests

The authors declare no competing interest.

Author details

¹National Centre for Metallurgical Research (CENIM), Spanish National Research Council (CSIC), Avda. Gregorio del Amo 8, 28040 Madrid, Spain. ²Conservation and Science, Rijksmuseum Amsterdam, Museumstraat 1, 1070DN Amsterdam, The Netherlands. ³Data Science for Cultural Heritage, Institute for Sustainable Heritage, University College London, Central House, 14 Upper Woburn Pl, London WC1H 0NN, UK. ⁴Heritage Laboratory, Swedish National Heritage Board, 62122 Visby, Sweden. ⁵National Research Council of Italy, Institute of Heritage Science (CNR-ISPC), Via Madonna del Piano 10, 50019 Sesto Fiorentino (Florence), Italy. ⁶Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, Ljubljana, Slovenia. ⁷Institute of Cultural Heritage of Spain (IPCE), Pintor El Greco 4, 28040 Madrid, Spain.

Received: 15 June 2023 Accepted: 8 February 2024

Published online: 19 March 2024

References

- Oddy WA. An unsuspected danger in display. *Mus J*. 1973;73(1):27–8.
- Clarke SG, Longhurst EE. The corrosion of metals by acid vapours from wood. *J Appl Chem*. 1961;11:435–43.
- Knotková-Čermáková D, Vlčková J. Corrosive effect of plastics, rubber and wood on metals in confined spaces. *Br Corros J*. 1971;6(1):17–22.
- Green LR, Thickett D. Testing materials for use in the storage and display of antiquities—a revised methodology. *Stud Conserv*. 1995;40(3):145–52.
- Oddy WA. The corrosion of metals on display. *Stud Conserv*. 1975;20(sup1):235–7.
- Blackshaw SM, Daniels VD. The testing of materials for use in storage and display in museums. *The Conservator*. 1979;3(1):16–9.
- Green LR, Thickett D. Interlaboratory comparison of the Oddy test. In: Tennent N, editor. *Conservation Science in the UK*. London: James and James Science Publishers; 1993. p. 111–6.
- Bamberger JA, Howe AC, Wheeler G. A variant Oddy test procedure for evaluating materials used in storage and display cases. *Stud Conserv*. 1999;44(2):86–90.
- Robinet L, Thickett D. A new methodology for accelerated corrosion testing. *Stud Conserv*. 2003;48(4):263–8.
- Stephens CH, Buscarino I, Breitung E. Updating the Oddy test: comparison with volatiles identified using chromatographic techniques. *Stud Conserv*. 2018;63(sup1):425–7.
- Buscarino IC, Bone A, Stephens CH, Breitung EM. Oddy Test Protocol at the Metropolitan Museum of Art. AIC wiki's "Oddy test methods chart wiki". https://www.conservation-wiki.com/w/images/8/82/20220511_MMA_Oddy_Protocol.pdf. Accessed 11 May 2022.
- Heine H, Jeberien A. Oddy test reloaded: standardized test equipment and evaluation methods for accelerated corrosion testing. *Stud Conserv*. 2018;63(sup1):362–5.
- Díaz I, Cano E. Quantitative Oddy test by the incorporation of the methodology of the ISO 11844 standard: a proof of concept. *J Cult Herit*. 2022;1(57):97–106.
- Korenberg C, Keable M, Phippard J, Doyle A. Refinements introduced in the Oddy test methodology. *Stud Conserv*. 2018;63(1):2–12.
- Torok E, Wickens JDJ. Reevaluating the Oddy test: an examination of the diversity in protocols used for material testing in the United States. In: *Conservation and exhibition planning: material testing for design, display, and packing*. Washington: Smithsonian American Art Museum & National Portrait Gallery; 2015. p. 33.
- Owens S. Preventive conservation project: a critical comparison of three versions of the Oddy test. [Newark]: University of Delaware; 2016.
- Materials Selection and Specification Working Group. American Institute for Conservation (AIC). Wiki: A collaborative knowledge resource. https://www.conservation-wiki.com/wiki/Materials_Working_Group. Accessed 25 Oct 2021.
- Oddy Test Protocols. American Institute for Conservation (AIC). Wiki: a collaborative knowledge resource. https://www.conservation-wiki.com/wiki/Oddy_Test_Protocols#The_Green_Challenge_Protocol_by_Christian_Hernandez. Accessed 23 Aug 2022.
- Thickett D, Lee LR. Selection of Materials for the Storage or Display of Museum Objects. London: The British Museum. Occasional Paper No. 111, first published 1996, revised edition 2004;1–30 p.
- Lee YK, Kim HJ. "The effect of temperature on VOCs and carbonyl compounds emission from wooden flooring by thermal extractor test method. *Build Environ*. 2012;53:95–9.
- Haghighat F, De Bellis L. Material emission rates: Literature review, and the impact of indoor air temperature and relative humidity. *Build Environ*. 1998;33(5):261–77.
- Yaws CL, Satyro MA. Chapter 1 - Vapor Pressure – Organic Compounds. In: Yaws CL, editor. *The Yaws Handbook of Vapor Pressure (Second Edition)*. [Internet]. Gulf Professional Publishing; 2015. p. 1–314. <https://www.sciencedirect.com/science/article/pii/B9780128029992000015>.
- Sultan AA, Ateeq AA, Khaled NI, Taher MK, Khalaf MN. Study of some natural products as eco-friendly corrosion inhibitor for mild steel in 10 M HCl solution. *J Mater Environ Sci*. 2014;5(2):498–503.
- Markowicz P, Larsson L. Influence of relative humidity on VOC concentrations in indoor air. *Environ Sci Pollut Res*. 2015;22(8):5772–9.
- Smith GD, Snyder C. Something 'oddy' about the Oddy test. In: ICOM Committee for Conservation, ICOM-CC, 15th Triennial Conference New Delhi, 22–26 September 2008: preprints 2008. Vol.II, p.887
- Long ER, Bone A, Breitung EM, Thickett D, Grau-Bové J. Automated corrosion detection in Oddy test coupons using convolutional neural networks. *Herit Sci*. 2022;10(1):150.
- Wang S, Kong L, An Z, Chen J, Wu L, Zhou X. An improved Oddy test using metal films. *Stud Conserv*. 2011;56(2):138–53.
- Hodgkins RE, Centeno SA, Bamberger JA, Tsukada M, Schrott AG. Silver nanofilm sensor for assessing daguerreotype housing materials in an Oddy test setup. *e-PresentationScience*. 2013;10:71–6.
- Thickett D. Oxygen depletion testing of metals. *Heritage*. 2021;4(3):2377–89.
- ISO 11844–2 [Internet]. Corrosion of metals and alloys-classification of low corrosivity of indoor atmospheres-part 2: determination of corrosion attack in indoor atmospheres. 2020. <https://www.iso.org/standard/73254.html>. Accessed 1 Dec 2022.
- Reedy CL, Corbett RA, Burke M. Electrochemical tests as alternatives to current methods for assessing effects of exhibition materials on metal artifacts. *Stud Conserv*. 1998;43(3):183–96.
- Bischoff JJ, Bustamente JA, Reedy CL, Corbett RA, Walton MS, Greene V, et al. From an idea of creativity to a product of reliability: Update of research on electrochemical testing of exhibit and storage materials. In: Greene V, Harvey D, Griffin P, editors. *AIC Objects Specialty Group Session, Volume Ten* [Internet]. Arlington: American Institute for Conservation of Historic and Artistic Works; 2003. p. 11–21 www.conservati-on-us.org.
- Stephens CH, Breitung EM. Impact of volatile organic compounds (VOCs) from acrylic double-sided pressure-sensitive adhesives (PSAs) on metals found in cultural heritage environments. *Polym Degrad Stab*. 2021;193: 109738.
- Shen J, Shen Y, Xu F, Zhou X, Wu L. Evaluating the suitability of museum storage or display materials for the conservation of metal objects: a

- study on the conformance between the deposited metal film method and the Oddy test. *Environ Sci Pollut Res*. 2018;25(35):35109–29.
35. Berger O, Yersin PB, Yersin JMB, Hartmann C, Hildbrand E, Hubert V, et al. Applications of micro-Raman spectroscopy in cultural heritage—examples from the laboratory for conservation research of the Collections Centre of the Swiss National Museums. *Chimia (Aarau)*. 2008;62(11):882–6.
 36. Thickett D, Hockey M. The effects of conservation treatments on the subsequent tarnishing of silver. In: Townsend J, Eremin K, Adriaens A, editors. *Conservation Science*. London; 2003.
 37. Chen J, Zhou M, Yan Y, Cai L. Quantitative characterization of Oddy test with quartz crystal microbalance combined with ionic liquid. *J Shanghai Normal Univ (Natural Sciences)*. 2013;42(3):305–310(6).
 38. García Márquez FP. *Advances in principal component analysis*. Rijeka: IntechOpen; 2022.
 39. Luciano G, Leardi R, Letardi P. Principal component analysis of color measurements of patinas and coating systems for outdoor bronze monuments. *J Cult Herit*. 2009;10(3):331–7.
 40. Fichera GV, Rovetta T, Fiocco G, Alberti G, Invernizzi C, Licchelli M, et al. Elemental analysis as statistical preliminary study of historical musical instruments. *Microchem J*. 2018;137:309–17.
 41. Brida JG, Meleddu M, Pulina M. Understanding museum visitors' experience: a comparative study. *J Cult Herit Manag Sustain Dev*. 2016;6:47–71.
 42. Ortega-Sánchez D, López-Sanvicente AB. Design, content validity, and inter-observer reliability of the 'Digitization of Cultural Heritage, Identities, and Education' (DICHIE) instrument. *Humanit Soc Sci Commun*. 2023;10(1):1–9.
 43. Silver corrosion library. Appendix: Photos of scored coupons, American Institute for Conservation (AIC). Wiki: A collaborative knowledge resource. 2020. https://www.conservation-wiki.com/w/images/8/8d/MMA_Silver_Corrosion_Library.pdf.
 44. Copper corrosion library. Appendix: Photos of scored coupons, American Institute for Conservation (AIC). Wiki: a collaborative knowledge resource. 2020. https://www.conservation-wiki.com/w/images/8/8c/MMA_Copper_Corrosion_Library.pdf.
 45. Lead corrosion library. Appendix: Photos of scored coupons, American Institute for Conservation (AIC). Wiki: a collaborative knowledge resource. 2020. https://www.conservation-wiki.com/w/images/f/f7/MMA_Lead_Corrosion_Library.pdf.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.