

# The International Journal of Conformity Assessment

- **The Cladding Problem: Establishing and Assessing Safe Building Envelopes**  
By Abhishek Chhabra
- **Verification Study of Food Packaging Materials Recoverable Through Composting and Biodegradation**  
By A.V. Chandrajith
- **Mathematical Methods Used to Affirm Fairness, Validity, Reliability, and General Performance of Examinations used for Personnel Certification Complying to ISO/IEC 17024 Standard Requirements**  
By Osman Vural, Ioannis Anastasopoulos, and David S. Nelson
- **Governance of Knowledge in Management Systems Applications**  
By Tolga Aktas
- **Historical Understandings of International Conformity Assessment Systems**  
By William T. Fiske
- **Leveraging Quality Management Systems to Prevent Product Liabilities**  
By Gilbert Gong



The International Journal of Conformity Assessment (IJCA) is an international, peer-reviewed journal for conformity assessment academics and practitioners. IJCA aims to contribute substantially to the fields of testing, inspection, certification, and accreditation by providing a high-quality medium for the dissemination of new knowledge and methods. Among IJCA's targeted audience are conformity assessment scholars, practitioners, and other interested groups and individuals. Main conformity assessment fields covered are testing and calibration laboratories, inspection agencies, product certification bodies, management system certification bodies, organizations offering certification of persons, and training course developers and providers. IJCA welcomes contributions in various aspects of conformity assessment. The editors, while accepting a wide array of scholarly contributions from different disciplinary approaches, especially encourage research that is novel, visionary, or pathbreaking. All submissions must be interesting, relevant to the conformity assessment field, sufficiently rigorous both conceptually and methodologically, and written in a clear, concise, and logical manner. IJCA is published by the International Accreditation Service (IAS), a nonprofit organization based in Brea, California, USA. IAS is an Accreditation Body, a signatory of the MLAs/MRAs of the International Laboratory Accreditation Cooperation (ILAC), the Asia Pacific Accreditation Cooperation (APAC), and the International Accreditation Forum (IAF), and is a full member of the Inter-American Accreditation Cooperation (IAAC). IAS remains neutral regarding any published articles and any related jurisdictional claims.





Each author acknowledges the submission of this manuscript to be published in the IJCA, a journal published by IAS.

Each author acknowledges the voluntary submission of his or her article.

Each author understands that articles will be blind reviewed by a panel of reviewers who will recommend acceptance for publication; acceptance with revisions; or rejection for publication (applicable for submissions to Section A).

Each author certifies that he or she has followed the accepted standards of scientific, creative, and academic honesty and ethics.

Each author confirms that the publishing of this manuscript by the IJCA does not conflict with any other copyrights or legal bindings.

Each author understands that the IJCA maintains the right to reprint any portions of the submitted work.

Each author understands that IJCA is published under a Creative Commons Attribution Noncommercial License (CC BY-NC).

Each author understands that the authors retain the copyright to their work.

## DISCLAIMER

The following disclaimer applies not only to the IJCA, but also to the IJCA Editorial Team, the IJCA Policy Board, the IJCA Editorial Review Board, the International Accreditation Service (IAS), and any other party involved in the preparation of material and content in the journal.

- The IJCA has been created for distribution around the world and is made available to all.
- The IJCA is not responsible for any errors or omissions in the journal articles' results and/or conclusions.
- The IJCA is not responsible for damages of any kind arising out of use, reference to, or reliance on any information contained within the journal.
- The IJCA does not assume any liability or responsibility for the completeness or usefulness of any information provided in the journal.
- The IJCA strives to provide accurate information in all its articles, but makes no claims, promises, or guarantees about the accuracy of information.
- The personal opinions of the authors in any IJCA article does not represent the IJCA's official position on that subject.

Copyright ©2022, International Accreditation Service - IAS. All rights reserved.

No part of the publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the publisher, except in the case of brief quotations embodied in reviews and certain other non-commercial uses permitted by copyright law.

ISSN: 2770-5668



DOI Link

## IJCA Table of Contents

Editorial Team and Review Board.....	page 5
Message from IAS President.....	page 6
From the IJCA Executive Editor's Desk.....	page 7
Section A	
The Cladding Problem: Establishing and Assessing Safe Building Envelopes By Abhishek Chhabra.....	page 9
Verification Study of Food Packaging Materials Recoverable Through Composting and Biodegradation By A.V. Chandrajith.....	page 13
Mathematical Methods Used to Affirm Fairness, Validity, Reliability, and General Performance of Examinations Used for Personnel Certification Complying to ISO/IEC 17024 Standard Requirements By Osman Vural, Ioannis Anastasopoulos, and David S. Nelson.....	page 21
Governance of Knowledge in Management Systems Applications By Tolga Aktas.....	page 40
Section B	
Historical Understandings of International Conformity Assessment Systems By William T. Fiske.....	page 46
Leveraging Quality Management Systems to Prevent Product Liabilities By Gilbert Gong.....	page 51
Competence Versus Qualification in Personnel Certification Programs By Dr. George J. Anastasopoulos.....	page 60

**Policy Board**

- *Craig Joss, Ph.D., PE, Civil & Environmental Engineering, Drexel University and Technical Manager at Applied Infrastructure, LLC, Wilmington, Delaware, USA*
- *David S. Nelson, Ph.D., University of Southern California (USC), CEO QPS LLC, USA*
- *Max Porter, Ph.D., Professor Emeritus at Iowa State University and President of Porter Engineering, USA*
- *Chuck Ramani, Chairman, Built Environment Services & Technology, LLC, USA*
- *Hatem Seliem, Ph.D., Department of Civil Engineering, Helwan University, Egypt*
- *Peter Unger, MSc Engineering, George Washington University and President at International Quality Excellence in Infrastructure Systems (IQEIS), USA*
- *Wang Wah Wong, MSc Food Science and MSc Env Mgt Reading University, UK, and University of Hong Kong, Professor of Practice*

**Executive Editor**

*Dr. George Anastasopoulos*

**Social Media Editor**

*Greg West*

**Editorial Team**

- *Harry Makam – IJCA Secretariat, IAS*
- *Laura Uraine – Editorial Coordinator, IAS*
- *Mahir Vural – Technical Intern, IAS*
- *Alberto Herrera – Editorial Advisor, IAS*

**Editorial Review Board**

- *Prof. Antonio Nanni, Head of Dept. of Civil & Arch. Engineering, University of Miami, USA*
- *Bora Gencturk, Ph.D., Department of Civil Engineering, University of Southern California, Los Angeles, USA*
- *Efstathios P. Efstathopoulos, Ph.D., Department of Medical Physics, Medical School, National and Kapodistrian University of Athens, Greece*
- *Emil Hazarian, Ph.D., Metrologist, College of Extended & International Education, California State University, Dominguez Hills, USA*
- *Evren Meltem Toygar, Ph.D., Department of Mechanical Engineering, Dokuz Eylul University, Turkey*
- *George A. Papadantonakis, Ph.D., Department of Chemistry, University of Illinois at Chicago, USA*
- *George Sakellion, Ph.D., Doctor of Medicine, Anaesthetics and ITU, East Suffolk and North Essex Foundation NHS Trust (ESNEFT) – Colchester General Hospital, UK*
- *Gilbert Gong, Ph.D., Professor at Inha University and President at Global Personnel Certification (GPC), Korea*
- *Lo Yiu Tommy, Ph.D., Department of Architecture and Civil Engineering, City University of Hong Kong and President at Hong Kong Institution of Certified Auditors (HKICA), Hong Kong*
- *Mohamed Zaki, Ph.D., Department of Civil Engineering, Housing and Building National Research Center (HBRC), Egypt*
- *Nuh Yalcin, Ph.D., Gebze Technical University, Physics, Electronics, Material Science and CEO at Consept Europe Kft, Turkey*

- *Panayiotis Tsanakas, Ph.D., Department of Computer Engineering, National Technical University of Athens (NTUA), Greece*
- *Paulo Alexandre Costa Araújo Sampaio, Department of Production and Systems Engineering, University of Minho, Portugal*
- *Samir Chauhan, Ph.D., Massachusetts Institute of Technology (MIT) and President at Sensocal, Inc., USA*
- *Shaurav Alam, Ph.D., Department of Civil Engineering, Construction Engineering Technology, Trenchless Technology Center, Louisiana Tech University, USA*
- *Tony Francis, Ph.D., University of Utah and President at SAW Environmental, USA*
- *Venkatesh Thuppil, Ph.D., St John's National Academy of Health Sciences and CEO/Director at Foundation for Quality, India*

All members of the Editorial Board have identified their affiliated institutions or organizations, along with the corresponding country or geographic region. IAS remains neutral regarding any jurisdictional claims.

ISSN: 2770-5668

Copyright ©2022, International Accreditation Service - IAS. All rights reserved

IAS-3060 Saturn Street, Suite 100,  
Brea, California 92821-1732 U.S.A.  
T: +1 562-364-8201

IAS is a member of the ICC Family of Solutions



+1 562-364-8201  
iasonline.org  
ijcajournal@iasonline.org

## Message from IAS President



This inaugural issue of the International Journal of Conformity Assessment (IJCA) is not only the culmination of IAS's vision to fill a critical gap in our understanding of fast-growing conformity assessment schemes and ecosystems; it is also the beginning of a journey to document and map the interconnected world of conformity assessment standards, the regulatory responses globally to such standards, and the percepts and practices involved in making them effective. This publication is a small step marking an ambitious pathway to navigate this global terrain—not unlike the early “navigators” on a geographical scale!

For those of us who have been involved in the conformity assessment world over the past several decades (I am certainly one of them), it is astonishing how much the language, content, and context of this arena has grown and matured. While the kernels of conformity assessment practices were perhaps embedded from the early days, the current contours of its technical evolution and range of offerings are far-reaching in our global supply chain. The interconnected world of international and national standards bodies, conformity assessment bodies, accreditation bodies, regulators, and users has proliferated to the extent that even serious students of this field cannot see the horizon anymore! The continuing output of conformity assessment standards and the associated work by the global organizations governing their implementation and effectiveness are so prolific now that a senior practitioner in the field noted it's “like drinking from a fire hose; I can barely keep up.”

At IAS, our intent is to encourage each one of you and your colleagues to bring your practices and percepts to a global audience by contributing to this journal. While in today's information-driven society one publication cannot claim to cover the breadth of the subject matter, I believe the IJCA rolls out the canvas to paint the conformity assessment world one stroke at a time. We have assembled a strong set of practitioners who will guide us as members of the IJCA's Policy Board and Editorial Review Board, supported by our editorial team. As further issues are produced, I am certain the messaging and styling of this publication will evolve to engage, enlighten, and excite our unique readership.

Please enjoy this inaugural issue and share it with your colleagues and others in your fraternities. We look forward to your constructive feedback as well.

A handwritten signature in black ink that reads “Raj Nathan”.

Raj Nathan

President, IAS

April 2022

## From the IJCA Executive Editor's Desk



Dear Readers,

The International Journal of Conformity Assessment (IJCA) is an international, peer-reviewed journal for conformity assessment academics and practitioners. IJCA aims to contribute substantially to the fields of testing, inspection, certification, and accreditation by providing a high-quality medium for the dissemination of new knowledge and methods.

In this first issue of IJCA, I would like to highlight our main editorial and publishing policies. This journal, which is published in the English language, fully endorses international rules on publishing and publication ethics. It also adheres to the double-blind, unbiased peer review process—which includes internal reviews by our editors as well as external reviewers and subject matter experts.

We follow principles of international diversity in terms of our editorial board, reviewers, and authors. Therefore, we invite and welcome submissions from around the world. Details of our editorial and publishing policies, instructions to authors, and other necessary information to submit articles for consideration in this publication are available on our website at [www.ijca-journal.org](http://www.ijca-journal.org).

Each IJCA issue will consist of two sections. Section A is dedicated to peer-reviewed publications and scholarly articles while Section B is an informative section that includes general articles, announcements, white papers, etc. This first edition features research articles on quality-control issues in the construction and food packaging industries, validation methods related to personnel certification, and knowledge in management systems applications. There are also two general articles highlighting international conformity assessment systems and the usage of management systems to prevent product liabilities.

We encourage you to submit letters to the editors discussing the results of published studies, systematic reviews, brief reports, and opinions regarding published reports. We believe if we carry out all the above-mentioned responsibilities as authors, reviewers, and editors, we will reach our goal of bringing your research and experience to a broad international audience of conformity assessment professionals and stakeholders.

We look forward to your contributions,

A handwritten signature in black ink, appearing to read 'G. Anastasopoulos', written in a cursive style.

Dr. George Anastasopoulos

Executive Editor, IJCA

April 2022

# Section A

dedicated to peer-reviewed publications and scholarly articles



# The Cladding Problem: Establishing and Assessing Safe Building Envelopes

By Abhishek Chhabra, Market Development Manager, Thomas Bell-Wright International Consultants

## -ABSTRACT-

Worldwide, the negative impacts of fire on cladding materials has increased over the years as buildings grow taller and the complexities of ownership, liability, and responsibilities increase. This paper discusses how the UAE fire code (UAE Fire and Life Safety Code of Practice) has effectively utilized proven conformity assessment standards—specifically ISO/IEC 17025, ISO/IEC 17065, and ISO/IEC 17020—to create robust mechanisms that drastically reduce fire safety hazards for building envelopes.

*Keywords: safe building envelopes, building materials, cladding, fire safety, ISO/IEC 17025, ISO/IEC 17065, ISO/IEC 17020, ISO/IEC 17067, conformity assessment, inspection, fire-rated building materials, UAE Fire and Life Safety Code of Practice*

While the stakeholders of the construction industry have juggled their way into demonstrating quality and safety of the work delivered until now, the cladding fire safety problem has now engulfed governments and financial institutions too. The rate at which gaps are being discovered as major accidents (e.g., Grenfell, Lacrosse tower, Address Downtown Dubai hotel, etc.) is faster than the rate at which skyscrapers around the world are growing.

Though some jurisdictions across the world learned and implemented interconnected mechanisms (linking regulations, building codes, and test standards) after the fires in the 1980s, the loopholes grow faster than the gross domestic products (GDPs) of respective countries.

Evolving conformity assessment guides offer unbiased and robust mechanisms to help establish and assess a (fire) safe building envelope. This paper touches on ISO/IEC 17025, ISO/IEC 17065, and ISO/IEC 17020 and how the jurisdiction of UAE (civil defense) is using these to drastically reduce and control the cladding fire safety problem.

## Cladding System

More than half a century ago, buildings started to move away from structural load-bearing walls to what is now considered modern construction, allowing the structure of a building to grow without load-bearing walls. This freed the height limits that were usually set by constraints of the height of a load-bearing wall, permitting buildings to grow taller and taller. This also gave way to dividing the functions of a load-bearing wall across many materials and systems. Along with the structure, the walls provided weather barriers (air, water, and heat) along with other properties such as acoustics and “fire.” (See Figure 1.) The

shift was great for giving more room to creativity, engineering, architecture, and of course commerce. An example in Figure 2 is one of probably hundreds of possible ways in which a building envelope system undertakes almost all the functions that were earlier fulfilled by walls.

## Growing Challenges

Architects would like a building to blend into existing skylines and sometimes even stand out, demanding them to be unique. Art gets a canvas. Then the building envelope system can be designed well by knowing the climatic conditions, seismic zone, heights, wind loads, etc. Engineers like to design. Before

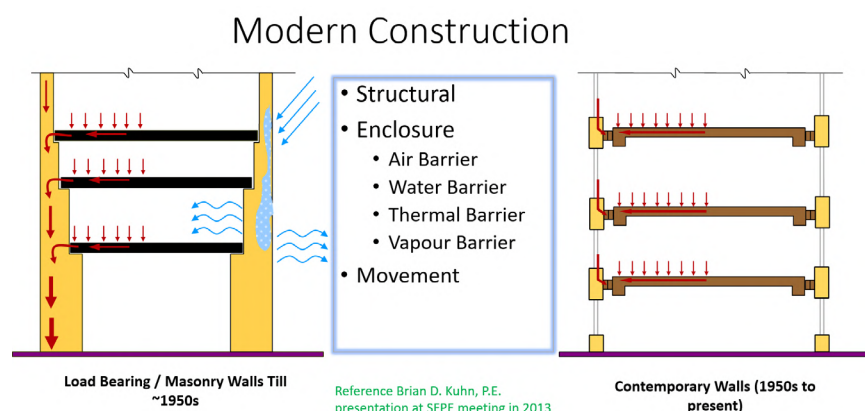


Figure 1: Typical functions taken over by the building envelope

commerce kicks in, the spread of the supply chain expands the possibilities of supplies, creativity, and design options.

So, after the architect's vision is engineered to arrive at a system design that can fulfill functional needs, a new set of engineers who need to "build" it come into the picture. Bombarded by choices of suppliers and their claims, they are held back by budget constraints and constantly nudged to complete the work quickly.

In practice, this is a heady mix. Decisions that are made at the construction site/project without the backing of a proven system can lead to catastrophes.

## Proven Systems

The tug of war between creative differentiation and desiring the comfort of repeatability has led to the progress of humanity. Both have progressed faster in the last five decades compared to earlier years. The increasing availability of devices that measure with higher accuracy, clubbed with the evolution of ever-evolving standards, provides the much-needed tools to establish and assess repeatability of desired

results.

Three published standards that have evolved over the past decades continue to support the concept of **parity** across countless products and services and have driven commerce and economies around the world.

## Testing a Product

The **root** of establishing repeatability is measurement and conducting tests. **ISO/IEC 17025** ("General requirements for the competence of testing and calibration laboratories") enables laboratories to demonstrate that they operate competently and generate valid results. This published standard, last revised in 2017, creates parity across organizations performing testing, sampling, or calibration. The availability and use of this standard has evolved over the last four decades and has helped countless buyers, specifiers, regulators, and manufacturers. By providing the tools to measure consistently, this standard has enabled better quality and safety of countless products and systems evaluated for electrical, mechanical, chemical, thermal, and other behaviors.

## Manufacturing a Product

After establishing the functional equality between entities testing products and materials, there was a need to assess processes that enable repeatable production of products (materials and systems). **ISO/IEC 17065** ("Conformity assessment—Requirements for bodies certifying products, processes, and services") is a published standard that provides the equitable tools to evaluate product conformity. According to the ISO website, "conformity assessment is the collective term for the processes that show a product meets the requirements of something, such as a standard, that is needed in order to meet a regulation or customer expectations." In simple terms, if the insulation material (e.g., mineral wool) needs to prevent cold or hot temperatures outdoors from permeating a building, it needs to be able to do it all the time, irrespective of how, when, and where heat/cold is produced. But manufacturing requires raw materials that need to be tested for given parameters before they are used to make a product. The manufacturer needs repeatable processes, trained manpower, and measuring instruments that establish the certainty. A typical thermal insulation producer needs to establish the repeatability of measurement of parameters such as air permeability, compressive resistance, corrosion resistance, water absorption, fire resistance, and thermal conductivity

Initiated as ISO guide 24 in 1978, the standard in its current form provides the basis of establishing certification programs with varying levels of severity of assessing and establishing

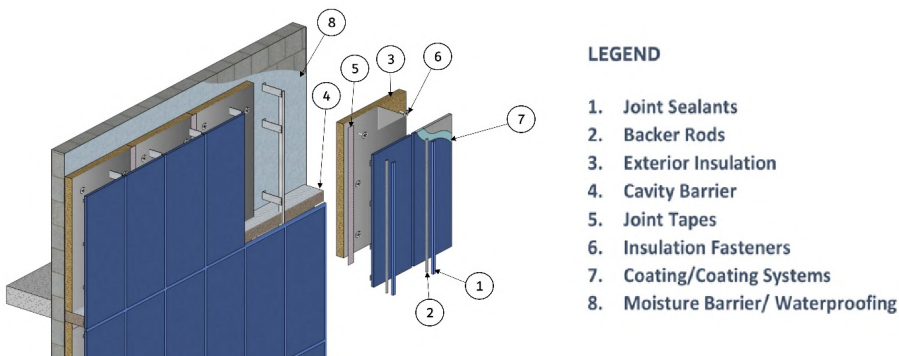


Figure 2: Typical detail of a cladding system showing the spread of different materials

assurance of materials that are bought and sold. Starting from very basic means where a factory's production control system is audited regularly (which includes, of course, the tests conducted for assessing the assured properties), the stringent means of assurance goes to levels where the traceability of samples to be tested is established and regular testing is conducted of randomly selected products, which are either already on the market or have reached the consumer or the construction site.

Used diligently by governments and private service providers, conformity assessment certification programs use unique certification markings on products (such as QR codes and even RFIDs now in some cases) and other features to help buyers track detailed information of the assured properties and manufacturing locations.

Over the decades, products posing a higher safety risk have been regulated using assessment mechanisms written using this standard. This ranges from government requirements set for selling water, to medicines, to local jurisdiction and contractual requirements for assessing fire-rated doors in buildings

## Installation

A large number of products are manufactured and then utilized by consumers, allowing for a direct evaluation of "value of money." Construction is among the very few industries where users of procured materials are not always users of the end result (the building as an example). While the properties of procured materials could be evaluated using testing and

certification, the assurance of their performance depends on correct installation. Just like products manufactured need to be independently tested, and manufacturing processes need to be audited for certification, installation assurance comes from independent inspections. But how do we establish the competence or parity?

**ISO/IEC 17020** ("Conformity assessment—Requirements for the operation of various types of bodies performing inspection") is a conformity assessment standard that specifies requirements for the competence of bodies performing inspections and for the impartiality and consistency of their inspection activities.

Along with several other supporting standards that developed through consensus, the three standards detailed above—ISO/IEC 17025, ISO/IEC 17065, and ISO/IEC 17020—provide parity to the processes used to establish and assess safe building envelopes.

## UAE Fire and Life Safety Code of Practice

As a young country ticking ambitious goals for growth, specifically in tall buildings, the UAE's challenge of establishing safe building envelopes was a steep one. Stakeholders of the fast-paced construction industry—including contractors, material suppliers, and installers across the world—have to be brought together on a common and easily assessable process.

The UAE Fire and Life Safety Code of Practice uses these three standards, placing control mechanisms for various stakeholders involved in the

realization of the building envelope or cladding. The publicly available code also added a new chapter (18) in the latest revision to define and detail the responsibilities of stakeholders, and elucidates the expectations of everyone—ranging from manufacturers to building occupants. It further defines the functions and liabilities of nearly 20 stakeholders including laboratories and certification and inspection bodies.

### Tier One

Along with defining the specific test methods for demonstrating conformity for each material and system (see Figure 3), the code defines certification and listing using ISO/IEC 17065. The code also uses the related ISO/IEC 17067 standard ("Conformity assessment—Fundamentals of product certification and guidelines for product certification schemes") to define the severity level of the certification program. Test reports used to demonstrate conformity can only come from an ISO/IEC 17025-accredited laboratory. The certification and listing evidence forms the basis of applying to the authority having jurisdiction (respective civil defense authority) to get registered as an approved supplier.

### Tier Two

The drawings of a proposed cladding system, when submitted for plan approval, need to assure that only approved suppliers will be used. An authorized fire consultant takes ownership of assessing the existing evidence of assembly tests performed by labs accredited (and recognized) as per ISO/IEC 17025. In case



the available evidence cannot justify the safety of the proposed design, additional mock-up tests (e.g., large-scale tests, as shown in Figure 3) are required. This completes the in-principal approval.

### Tier Three

Installation inspections are conducted at 20%, 40%, 60%, 80%, and 100% completion of the cladding works. Along with the use of ISO/IEC 17020 by specialist fire consultants to assess inspections, the contractors, facade specialists, and material suppliers also assume responsibility of the supplies and workmanship to complete the assurance.

### Safe Building Envelopes

Along with the example of the UAE fire code, it is important to note that processes and procedures often do not get implemented with very high precision. There will always be cracks and gaps in systems that need iterative work and review. But correct usage of such evolving standards that form the backbone of progress will always be lightyears ahead of a process that relies on subjective decision-making. Regulations and specifications that hard-code specific suppliers or non-iterative fixed routines will always become a hinderance as technologies, supply chain factors, materials, and installation methodologies evolve.

MCM and ACP On Non-Fire Resistance rated and Non-Load bearing Exterior wall coverings-Test Requirements		
OCCUPANCY AND TYPE OF BUILDING	TEST 1 MCM/ ACP CORE AND PANEL AS PRODUCT	TEST 2 MCM/ ACP PANELS WITH WALL ASSEMBLY
1. SUPER HIGHRISE BUILDING	i. Core shall be tested to the criteria iii and iv.	v. BS 8414 - 1 Or 2 With pass criteria as per BRE 135
2. HIGHRISE BUILDING	ii. Panel shall be tested with the thickness intended to the criteria iii and iv.	OR
3. MALLS	iii. EN 13501-1 With pass criteria A1 OR A2 s1 d0	vi. NFPA 285 With pass criteria "Pass"
4. THEME PARKS	AND	OR
5. SCHOOLS	iv. ASTM D 1929 MCM/ACP shall have self ignition temperature of not less than 343°C.	vii. FM 4881 With pass criteria "Pass"
6. HOSPITALS		OR
7. ASSEMBLY		viii. ISO 13785-2 With pass criteria "Pass"

Figure 3: Table from the UAE Fire and Life Safety Code of Practice defining the compliance routes and option for ACP/MCM materials

### -AUTHOR BIO-

Abhishek Chhabra currently serves as the market development manager at Thomas Bell-Wright International Consultants, a Dubai-based engineering firm he joined in 2013. Throughout his career, he has advocated for compliance with standards that improve safety and quality across various types of industries. He has worked on several standards and codes development initiatives, including Bureau of Indian Standards (BIS), ASTM International, UAE Fire and Life Safety Code of Practice, and Saudi Standards, Metrology, and Quality Organization (SASO). Additionally, Abhishek speaks often at industry events worldwide and writes frequently for magazines and trade publications. He received a Post Graduate Diploma in Finance Management through SVKM's NMIMS program in India and a Bachelor of Technology degree in electronics and communications from BCET at Punjab Technical University in India.





# Verification Study of Food Packaging Materials Recoverable through Composting and Biodegradation

By A.V. Chandrajith, Ph.D., Managing Director, Wimpey Laboratories

## -ABSTRACT-

Biodegradable packaging has the potential to reduce water usage, solid waste, electricity, and emissions compared to conventional packaging processes. Biodegradable plastics created from renewable sources (cellulose or starch) have novel functionalities and processibilities. The European Union standard EN 13432:2006, "Requirements for packaging recoverable through composting and biodegradation – Test scheme and evaluation criteria for the final acceptance of packaging," solves this issue by explaining clearly how a material becomes biodegradable as well as compostable.

The present work is a verification study of EN 13432:2006 using food packaging materials, which were categorized as biodegradable. Spectroscopic analysis of the samples, inoculum, and compost were performed using Fourier transform infrared spectroscopy and inductively coupled plasma atomic emission spectrometry. The aerobic and anaerobic degradation of the samples was conducted in accordance with ISO 14855-1 and ASTM D5511, respectively. An ecotoxicity study using the compost of the samples was performed as per OECD 208 guidelines. The nature of carbon dioxide evolution and biogas accumulation in biodegradability studies was on par with the ISO as well as ASTM standards. The quality of the compost and the ecotoxicity studies using the samples meet the requirements as stipulated by OECD 208 guidelines. The results proved the material possesses the characteristics recommended by EN 13432:2006; so, the material is undoubtedly biodegradable plastic.

*Keywords: biodegradable packaging, food packaging materials, biodegradable plastics, composting, biodegradation, environment, EN 13432, ISO 14855, ASTM D5511, OECD 208*

## Introduction

Biodegradable packaging has the potential to reduce water usage, solid waste, electricity, and emissions. While this is beneficial for the environment, it also lowers expenses associated with the packaging process. Conventional food packaging materials we use have several drawbacks. However, most of the drawbacks are related to environmental conditions—especially pollution. The destructive impact of single-use plastics originating from petroleum-based sources on the environment remains an urgent crisis. In order to mitigate these issues, it is necessary to switch from single-use plastics to biodegradable plastics. Biodegradable plastics created from renewable sources (cellulose or starch) have novel functionalities and processibilities compared to conventional plastic materials and are seeking attention nowadays. The usage of biodegradable plastics is mounting in the form of food containers, bottles, packaging, etc. The

disposal of packaging materials is predominant in cases of waste management. This means, if the waste is not disposed of properly, it will adversely affect the environment even though the material is biodegradable plastic or bioplastic. Biodegradable materials, bioplastics, biodegradability, and compostability are common terms that are frequently misinterpreted. The European Union standard EN 13432:2006, "Requirements for packaging recoverable through composting and biodegradation – Test scheme and evaluation criteria for the final acceptance of packaging," solves this issue by explaining clearly how a material becomes biodegradable as well as compostable.

The benefits of plastics over metal and paper has gained attention in various packaging applications, especially in the food sector. One key advantage is how plastic packaging enhanced the shelf life of products without using any preservatives.

However, the disadvantages of traditional plastics on human health and environment led to the design of innovative plastic materials that can be recyclable and degradable in environmental conditions without any adverse impact [1,3].

All types of plastics undergo degradation, which may be physicochemical, biological, or both. Degradation occurs as a result of wind, waves, or sunlight, which are examples of physicochemical processes. Oxo-degradable or hydro-degradable plastics are designed in such a way that they undergo degradation via oxidation or hydrolysis. Oxo-degradable plastics are the resultant products of fossil-carbon-derived plastics mixed with some additives such as antioxidants and prooxidants. Photodegradable plastic is a subclass of oxo-degradable plastic, where ultraviolet (UV) light induces the oxidation process. Hydro-degradable plastics are hybrid composites of petroleum-based plastic and a natural polymer such as starch [4,6]. While the degradation of biodegradable plastics is caused by microorganisms such as bacteria, fungi, or enzymes, preferably, plastics degrade via aerobic and anaerobic organisms resulting in carbon dioxide, water, methane, and compost. The majority of commercial biodegradable plastics are converted into compost instead of gaseous products [2,7].

Besides food, most personal care, cosmetic, and domestic products are packaged in plastic containers. Unfortunately, the chemical stability of these polymers—one of the main reasons of its successful application—gives rise to serious environmental and health problems due to the huge amount of plastic waste released into the environment each year. In principle, biodegradable and compostable bioplastics would provide the aforementioned societal benefits while affording, respectively, a lack of harmful residues or valued compost fertilizer. Polylactic acid (PLA), starch, cellulose pulp, polyhydroxyalkanoates (PHAs) such as polyhydroxybutyrate, and polyhydroxyoctanoate are the main biopolymers used to produce today's single-use bioplastics items such as bags, dishes, straws, coffee stirrers, glasses, horticulture pots, mulching film, bin liners, dust sheets, bottles, and packaging items. Today's single-use plastic packaging and lignocellulosic materials are biodegradable and compostable when they meet the requirements of European Union Standard EN 13432, "Requirements for packaging recoverable through composting and biodegradation – Test scheme

and evaluation criteria for the final acceptance of packaging" [8-9].

The present study is a verification study for the specification of EN 13432 and uses a cellophane-based biodegradable packaging material. The sample exhibited was above 92% biodegradation after 48 days of exposure. Moreover, the germination rate of the sample compost demonstrated a more than 93% germination rate. The biodegradability of the test sample meets the criteria stipulated by EN 13432.

## Materials and Methods

Test specimens of biodegradable plastics with a particle size of 250  $\mu\text{m}$  powder were used for the study. AR-grade microcrystalline cellulose 98% was used as the positive control. The compost inoculum of four-month-old, well-aerated compost from the organic fraction of municipal solid waste, sieved on a screen of less than 10 mm, was used for the study. The inoculum was characterized using a pH meter (Eutech Instruments Ion 510 for determining pH), a gas chromatography flame ionization detector (GC-FID) (PerkinElmer gas chromatograph Clarus 590 for determining volatile fatty acids), and Kjeldahl distillation equipment (BUCHI distillation unit K-350 for Kjeldahl nitrogen). The preliminary characterization of the test sample was determined by recording the spectrum using Fourier transform infrared (FTIR) spectroscopy (PerkinElmer Spectrum 65). Aerobic and anaerobic degradability were performed as per ISO 14855-1 and ASTM D5511, respectively [10,11]. The toxic elemental analysis of the compost so obtained after the aerobic degradation study was recorded using inductively coupled plasma atomic emission spectrometry (Shimadzu, ICPE-9820). The ecotoxicity study of the compost was based on OECD 208 guidelines [12], which involved the evaluation of seedling emergence and seedling growth of higher plants following exposure to the test substance in the soil. Seeds of *Brassica juncea* (mustard) and *Vigna radiata* (green gram) were placed in contact with soil treated with the test substance and evaluated for effects 21 days after the 50% emergence of seedlings in the control group.

## Sample Preparation

### For pH measurement

The pH was monitored to treat the sample into neutral in case the sample was in an acidic or basic

condition due its interfere with microorganism activity [13]. The pH of the inoculum was maintained between 7 and 8.5. For determining the pH sample, one part of the sample was mixed with five parts of distilled water. It was mixed by shaking and the pH was measured immediately.

#### For GC-MS analysis

Acetic, propionic, butyric, isobutyric, valeric, isovaleric, and hexanoic acids from the compost material were analyzed using a gas chromatograph equipped with flame ionization detection (splitless injection) and a DB-WAX column (30 m x 0.25 mm x 0.25  $\mu$ m). A DB-WAX column is a polar column used for detecting volatile fatty acids; a polar column is highly recommended due to its great resolution and sensibility. The samples of raw compost (approximately 200-250 g according to sample moistures) were extracted using demineralized water (in a 1:20 ratio representing the mass ratio of solid-phase dry mass to aqueous phase) and agitated for 24 hours in tightly closed brown bottles. The aliquots of liquid phase were then centrifuged at 10,000 rpm, and the supernatants were acidified to pH 2 with oxalic acid. After filtration through a 0.45 ml membrane filter directly into vials, the acidified samples were analyzed using GC-FID to determine the concentrations of individual acids. All samples were extracted and analyzed in duplicate [14].

## Results and Discussion

### Characterization of Packaging Material Using FTIR Spectroscopy

Each packaging material under investigation was identified and characterized prior to testing for the determination of the constituents of the packaging materials. FTIR spectroscopy was employed for the preliminary characterization. When compared with sample spectra available in the library of the equipment, it was found to have a 76.2% match with cellophane. The band at 899.95  $\text{cm}^{-1}$  is characteristic of the glycosidic bond  $\beta$ -(1 $\rightarrow$ 4) cellulose [15]. The range between 1200  $\text{cm}^{-1}$  and 1100  $\text{cm}^{-1}$  is in the region of hemicellulose and cellulose, which attained a maximum value around. A band around 1457  $\text{cm}^{-1}$  corresponds to deformation -CH<sub>2</sub> and -CH<sub>3</sub> groups and an intense peak at 1035  $\text{cm}^{-1}$  corresponds to C-O stretching [16]. Moreover, a band around 1735  $\text{cm}^{-1}$  is characteristic of C=O stretching while the peak at 2921  $\text{cm}^{-1}$  is due to the asymmetrical stretching of -CH<sub>2</sub> and -CH, which denote the characteristics of cellulose [17]. The

broad peak between 3500  $\text{cm}^{-1}$  and 3000  $\text{cm}^{-1}$  is attributed by the sum of the vibration of valence bands of the hydrogen bond of the -OH group and the bands of intramolecular and intermolecular hydrogen bonds. The search spectrum is depicted in Figure 1.

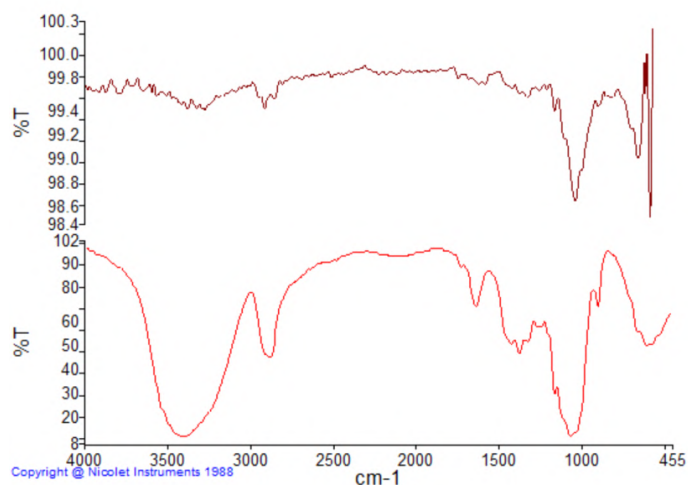


Figure 1: Overlay of FTIR spectra of the sample with the reference sample from the software library

### Aerobic Biodegradability

The ultimate aerobic biodegradability of the test sample was conducted as per ISO 14855-1:2012. Determination of the ultimate aerobic biodegradability of plastic materials occurred under controlled composting conditions ("Method by analysis of evolved carbon dioxide – Part 1: General method"). This test method determines the degree and rate of the aerobic biodegradation of plastic materials on exposure to a controlled-composting environment under laboratory conditions.

The test substances were exposed to an inoculum derived from compost from municipal solid waste. This test method is designed to yield a percentage of the conversion of carbon in the sample to carbon dioxide and the rate of biodegradation. The inoculum possessed an ash content of 60%, pH of 7.6, and total dry solids of 52%. The inoculum should be as free from larger inert materials as possible to make it homogenous.

The samples were exposed to the inoculum in the composting vessels that were incubated in the dark with the temperature maintained at  $58 \pm 2^\circ\text{C}$ . A pressurized air system containing CO<sub>2</sub>-free, H<sub>2</sub>O-saturated air was provided to each of the composting vessels at an accurate aeration rate. CO<sub>2</sub> and O<sub>2</sub>

concentrations were checked daily with a minimum time interval of six hours after the first week for the remainder of the test. The percentage of biodegradability was obtained by determining the percentage of carbon in the test sample that was converted into CO<sub>2</sub> during the duration of the test, and the evolved carbon dioxide was determined by titration. The carbon dioxide that evolved was absorbed by standardized 0.025 N barium hydroxide and the amount of CO<sub>2</sub> was determined by titrating manually with 0.5 N hydrochloric acid using a phenolphthalein indicator. The carbon dioxide produced in each vessel reacted with barium hydroxide and was precipitated as barium carbonate. The amount of carbon dioxide produced was determined by titrating the remaining barium hydroxide with 0.05 N hydrochloric acid to a phenolphthalein end point. Data obtained from the titration was used to calculate the amount of CO<sub>2</sub> produced and percentage biodegradability (Table 1 and Figure 2). The plateau for percentage biodegradability (93%) was obtained for the sample after 42 days.

Table 1. Percentage of Aerobic Biodegradation of positive control and test samples.

Day	% BIODEGRADATION			
	Positive control	Specimen sample1	Specimen sample2	Specimen sample 3
0	0	0	0	0
3	4.45	3.27	3.19	3.21
7	9.88	8.58	8.64	8.68
9	17.11	14.23	14.09	14.13
12	25.23	22.88	22.76	22.93
15	36.11	32.14	31.89	32.03
18	45.88	41.15	40.96	40.85
21	54.23	52.71	51.96	51.82
24	62.12	60.74	59.41	59.12
27	73.88	71.06	70.69	70.96
30	81.75	77.84	76.52	77.89
33	88.82	84.55	83.51	83.96
36	93.23	90.15	90.01	89.95
39	95.56	92.21	91.65	91.68
42	96.11	93.47	92.85	93.01
45	96.23	93.56	92.85	92.99
48	96.27	93.61	93.02	93.45
Mean		93.36		
Standard deviation		0.3051		
RSD		0.3268		

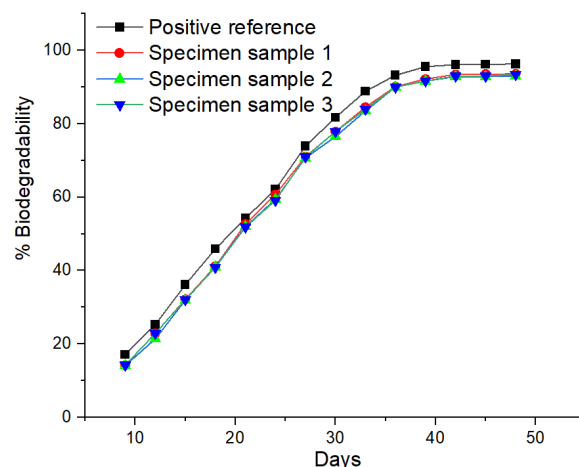


Figure 2. Percentage biodegradation of specimen samples and controlled sample under aerobic composting

### Anaerobic Biodegradability

The ASTM D5511 standard test method for determining anaerobic biodegradation of plastic materials under high-solids anaerobic-digestion conditions considering gas evolution was employed for the determination of anaerobic biodegradability in the present study. The prepared inoculum was subjected to a short post-fermentation of seven days. The pH of inoculum was 7.6, Kjeldahl nitrogen was 1 g/kg, and the volatile fatty acids content was less than 1 g/kg.

The test material was exposed to a methanogen inoculum derived from anaerobic digesters operating at 52±2°C. The test method was designed to yield the percentage of carbon in the test material and its rate of conversion to evolved carbon dioxide and methane (biogas). As these bacteria began to utilize the carbon in the test samples, they generated carbonaceous gas such as CH<sub>4</sub> and CO<sub>2</sub>. These gases were measured, and the results were carefully recorded. If the positive control (cellulose) continues to exhibit bio degradation, then the test is considered valid and the inoculum is considered alive. When the test is run to satisfaction, final gas readings are recorded and the incubation vessels are emptied and the samples are cleaned and weighed; the percentage of biodegradation of the samples is determined based on the conversion of carbon from the test material to carbon in the gaseous phase (CH<sub>4</sub> and CO<sub>2</sub>). After 48 days of incubation under dry (52±2°C), anaerobic-controlled composting conditions using test method ASTM D5511, the reference (positive control) and polymer sample



were gradually biodegraded. The reference sample was degraded 95.38% while the specimen sample showed 92.73% degradation after 48 days (Table 2, Table 3, and Figure 3).

Table 2. Volume of biogas of positive control and test samples.

Day	Volume of Biogas (ml)			
	Positive control	Specimen sample 1	Specimen sample 2	Specimen sample 3
0	0	0	0	0
3	883	846	816	823
7	2216	2107	2095	2075
9	3012	2876	2835	2901
12	4186	3773	3756	3792
15	4988	4736	4710	4756
18	5626	5447	5412	5462
21	6316	6107	6098	6123
24	7283	6942	6901	6952
27	7838	7667	7702	7652
30	8218	8081	8106	8025
33	8536	8387	8356	8364
36	8683	8464	8412	8436
39	8698	8589	8526	8571
42	8786	8654	8654	8671
45	8842	8723	8712	8706
48	8857	8786	8779	8783
Mean		8782.67		
Standard deviation		3.5119		
RSD		0.0399		

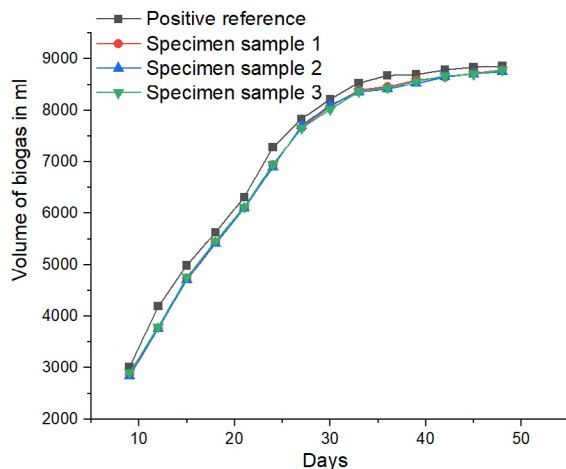


Figure 3: Percentage biodegradation of specimen sample and controlled sample under anaerobic composting.

Table 3. Percentage biodegradability of specimen sample with respect to positive control cellulose.

Group	Inoculum control	% Biodegradation	
		Specimen sample 1	Specimen sample 2
Weight (g)	250	10.0068	10.0126
Total volume (ml)	1556	8857	8782.67
Methane (CH <sub>4</sub> ) (%)	13.7	59.8	56.4
Volume of methane (CH <sub>4</sub> )(ml)	213.17	5296.49	4953.43
Weight of CH <sub>4</sub> (g)	0.15	3.78	3.54
Carbon dioxide (CO <sub>2</sub> ) (%)	19.8	38.3	40.2
Volume of carbon dioxide (CO <sub>2</sub> )(ml)	308.09	3392.23	3530.63
Weight of carbon dioxide (CO <sub>2</sub> )(g)	0.61	6.66	6.94
Total weight of carbon (g)	0.27	4.24	4.12
Theoretical weight of carbon(g)		4.443	4.445
Biodegradation		0.9538	0.9273
Biodegradation (%)		95.38	92.73

### Compostability

At the end of the composting test, the entire contents of the bin were sieved through a mesh the size of 10 mm. The overall compost quality was determined by the analyses performed on the <10 mm fraction. The results of all these analyses are given in Table 4 and Table 5.

Table 4. Chemical characteristics of residue collected after biodegradation (compost).

Test	Test Method	Unit	Substance on Dry Sample	EN 13432:2000 Specification
pH	Probe method	-	8.2±0.1	-
Total solids	Gravimetric	%	23.2±0.05	-
Volatile solids		%	64.3±0.25	Min. 50.
Disintegration		%	92.4±1.2	≥90

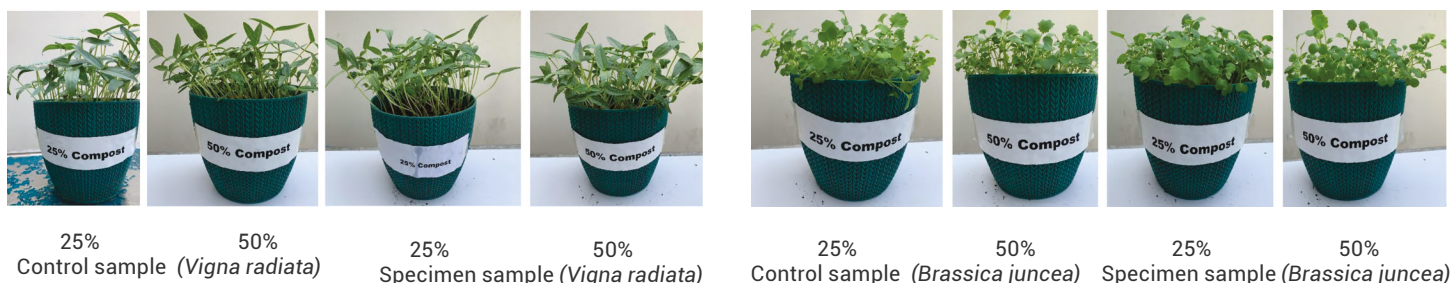
Table 5. Analysis of hazardous/toxic substances present in the residue collected after biodegradation.

Test	Unit	Substance on Dry Sample	EN 13432:2000 Specification
Zinc (Zn)	ppm	<1.0	Max. 150
Copper (Cu)	ppm	<0.05	Max. 50
Nickel (Ni)	ppm	<0.05	Max. 25
Cadmium (Cd)	ppm	<0.05	Max. 0.5
Lead (Pb)	ppm	<0.05	Max. 50
Mercury (Hg)	ppm	<0.05	Max. 0.5
Chromium (Cr)	ppm	<0.05	Max. 50
Molybdenum (Mo)	ppm	<0.05	Max. 1
Selenium (Se)	ppm	<0.05	Max. 0.75
Arsenic (As)	ppm	<0.05	Max. 5
Fluoride (F-)	ppm	<0.1	Max. 100

The sample specimen fulfilled the 90% disintegration requirement stipulated by EN 13432:2000. Moreover, the presence of toxic heavy metals in the compost was on par with the EN 13432:2000 specification. This proved that there was no negative effect on the composting process and on the (physicochemical) quality of the produced compost.

Table 6. Germination rate and biomass of *Brassica juncea* and *Vigna radiata* seeds after 21 days.

Sample	Plant Species	Dose	Germination Rate (%)	Shoot Length (cm)	Root Length (cm)
Positive Control	<i>Brassica juncea</i>	25%	95%	15.68	3.40
		50%	97%	15.90	3.37
	<i>Vigna radiata</i>	25%	98%	15.40	3.39
		50%	98%	15.85	3.33
Specimen Sample	<i>Brassica juncea</i>	25%	94%	10.80	2.10
		50%	95%	11.00	2.70
	<i>Vigna radiata</i>	25%	95%	10.74	2.00
		50%	97%	10.91	2.50

Figure 4. Effect of compost containing specimen sample on *Vigna radiata* and *Brassica juncea* growth.

## Ecotoxicity

At the end of the commercial life of the biodegradable materials, they were expected to degrade into harmless end products. So the testing protocols were developed to characterize the biodegradable plastics and packaging with the inclusion of the assessment of ecotoxicity potential. Using ecotoxicity data, environmentally relevant concentrations of chemicals can be estimated. Phytotoxicity testing using OECD 208 involves the assessment of seedling emergence and seedling growth of higher plants following the exposure to the test substance in the soil. The test is to confirm that the compost of biodegradable plastic does not induce any toxicity in the environment. The experiment was conducted at  $25\pm4^{\circ}\text{C}$ ,  $65\pm10\%$  humidity and 55.74 FC light intensity with 16 hours of light; the results are depicted in Table 6 and Figure 4. In these results, the specimen sample showed an insignificant effect on plant growth, no signs of chlorosis or necrosis, and no visible damage to the plants.

According to EN 13432:2000, the germination rate and plant biomass (on a fresh-weight basis) in the test compost should be more than 90% of those in the corresponding blank compost. These criteria were satisfied for both the germination rate and the plant biomass of both mixtures of the test compost (as per Table 6). Therefore, it can be stated that the requirements of EN 13432:2000 on ecotoxicity were fulfilled.

## Conclusion

By the conditions outlined in EN 13432, the submitted samples disintegrated after 48 days of exposure and showed that 93.36% of organic carbon was converted to carbon dioxide by aerobic degradation. The reference sample was degraded 95.38% while the specimen sample showed 92.73% degradation after 48 days by anaerobic degradation. The germination rate and plant biomass of the sample composts of both plant species were more than 93%, which is on par with EN 13432. The results proved the material possesses the characteristics recommended by EN 13432:2006; so, the material is undoubtedly biodegradable plastic.

## Author Bio

Dr. A.V. Chandrajith is managing director of Wimpey Laboratories, a UAE-based organization, and possesses more than 23 years of technical experience in the fields of food science and consumer products. He has a Ph.D. in chemistry and has published various research papers. A chemist-turned-process-owner, Chandrajith is also a leading auditor of ISO standards, food safety speaker, treatment specialist, Legionella risk assessor, and a technical expert in the testing, inspection, and certification (TIC) sector.



## References

- [1] White A. and Lockyer S. 2020. "Removing plastic packaging from fresh produce—what's the impact?" *Nutrition Bulletin* 45, no. 1 (March): 35-50.
- [2] Filiciotto L. and Rothenberg G. 2021. "Biodegradable plastics: standards, policies, and impacts." *ChemSusChem* 14, no. 1 (January): 56.
- [3] Chiong, J.A., Tran, H., Lin, Y., Zheng, Y., and Bao, Z. 2021. "Integrating Emerging Polymer Chemistries for the Advancement of Recyclable, Biodegradable, and Biocompatible Electronics." *Advanced Science* 8, no. 14 (May): 2101233.
- [4] Kalogerakis, N., Karkanorachaki, K., Kalogerakis, G., Triantafyllidi, E.I., Gotsis, A.D., Partsinevelos, P., and Fava, F. 2017. "Microplastics generation: onset of fragmentation of polyethylene films in marine environment mesocosms." *Frontiers in Marine Science* 28, no. 4 (March): 84.
- [5] Lambert, S. and Wagner, M. 2017. "Environmental performance of bio-based and biodegradable plastics: the road ahead." *Chemical Society Reviews* 46, no. 22: 6855-71.
- [6] Vanoppen, M., Criel, E., Walpot, G., Vermaas, D.A., and Verliefde, A. 2018. "Assisted reverse electrodialysis—principles, mechanisms, and potential." *Clean Water* 1, no. 1 (June): 1-5.
- [7] Thakur, M., Majid, I., Hussain, S., and Nanda V. 2021. "Poly( $\epsilon$  caprolactone): A potential polymer for biodegradable food packaging applications." *Packaging Technology and Science* 34, no. 8 (April).
- [8] Ciriminna, R. and Pagliaro, M. 2020. "Biodegradable and compostable plastics: A critical perspective on the dawn of their global adoption." *ChemistryOpen* 9, no. 1 (January): 8.

[9] European Standards. n.d. "EN 13432:2000 Requirements for packaging recoverable through composting and biodegradation – Test scheme and evaluation criteria for the final acceptance of packaging." <https://www.en-standard.eu/bs-en-13432-2000-packaging.-requirements-for-packaging-recoverable-through-composting-and-biodegradation.-test-scheme-and-evaluation-criteria-for-the-final-acceptance-of-packaging/>

[10] ISO. n.d. "ISO 14855-1:2012 Determination of the ultimate aerobic biodegradability of plastic materials under controlled composting conditions – Method by analysis of evolved carbon dioxide – Part 1: General method." <https://www.iso.org/standard/57902.html>

[11] ASTM International. 2018. "ASTM D5511-18 Standard Test Method for Determining Anaerobic Biodegradation of Plastic Materials Under High-Solids Anaerobic-Digestion Conditions." <https://www.astm.org/d5511-18.html>

[12] OECD. 2006. "Test No. 208: Terrestrial Plant Test: Seedling Emergence and Seedling Growth Test." OECD Guidelines for the Testing of Chemicals, Section 2. <https://doi.org/10.1787/9789264070066-en>.

[13] Cotter, P.D. and Hill, C. 2003. "Surviving the acid test: responses of gram-positive bacteria to low pH." *Microbiology and Molecular Biology Reviews* 67, no. 3 (September): 429-53.

[14] Plachá, D., Raclavská, H., Kučerová, M., and Kuchařová, J. 2013. "Volatile fatty acid evolution in biomass mixture composts prepared in open and closed bioreactors." *Waste Management* 33, no. 5 (May): 1104-12.

[15] Lun, L.W., Gunny, A.A., Kasim, F.H., and Arbain D. 2017. "Fourier transform infrared spectroscopy (FTIR) analysis of paddy straw pulp treated using deep eutectic solvent." *AIP Conference Proceedings* 1835, no. 1 (April): 020049.

[16] Oh, S.Y., Yoo, D.I., Shin, Y., Kim, H.C., Kim, H.Y., Chung, Y.S., Park, W.H., and Youk, J.H. 2005. "Crystalline structure analysis of cellulose treated with sodium hydroxide and carbon dioxide by means of X-ray diffraction and FTIR spectroscopy." *Carbohydrate Research* 340, no. 15 (October): 2376-91.

[17] Colom, X., Carrillo, F., Nogués, F., and Garriga, P. 2003. "Structural analysis of photodegraded wood by means of FTIR spectroscopy." *Polymer Degradation and Stability* 80, no. 3 (January): 543-9.

## Testing Laboratory Accreditation Program



Testing laboratories around the world seek accreditation from the International Accreditation Service (IAS). Our reputation for technical competence and commitment to meeting the needs of customers has helped IAS to become a leader in accreditation services.

### IAS Accreditation:

- Demonstrates compliance with ISO/IEC Standard 17025, the global standard for laboratory competence.
- Enhances recognition and acceptance of accredited facilities in national and global markets.
- IAS accreditation is accepted by regulatory bodies in numerous countries. In the United States, accreditation to ISO Standard 17025 is recognized by federal, state and local governmental bodies.
- IAS is a signatory to the Mutual Recognition Arrangement (MRA) with the Asia Pacific Accreditation Cooperation (APAC) and the International Laboratory Accreditation Cooperation (ILAC), helping increase recognition of test reports.
- IAS offers prompt, personal service, including rapid scheduling of assessments to meet the needs of laboratories.



### For More Information

International Accreditation Service  
[www.iasonline.org](http://www.iasonline.org)



# Mathematical Methods Used to Affirm Fairness, Validity, Reliability, and General Performance of Examinations Used for Personnel Certification Complying to ISO/IEC 17024 Standard Requirements

By Osman Vural, Director, International Personnel Certification Association; Ioannis Anastasopoulos, Department of Mathematics, University of California at Berkeley; and David S. Nelson, PE, Ph.D., President, Quality Psychometric Services

## Abstract

Validating the examinations provided by conformity assessment bodies for personnel certification purposes is a requirement of the international standard ISO/IEC 17024 "Conformity assessment—General requirements for bodies operating certification of persons." A fundamental requirement for each examination developer is to achieve a reliable and fair examination process.

The validity of a test paper depends on the quality of the sections that constitute the examination. Considering that the examination must measure the competencies derived by the relevant job/task analysis, there is a direct connection between the job/task analysis outputs and the respective examination.

There are prescribed steps in linking the relatedness of the examination to the knowledge and skills required for a job. These steps lead to an examination that has been "validated" in that its content accurately measures the necessary knowledge and skills required for the job.

This paper describes the steps to design the ideal examination (personnel evaluation) system, as well as some of the statistical techniques that can be used to analyze items that are necessary to establish question banks.

The scope of this paper is limited to the "job analysis" component of certification schemes and does not extend to requirements for administration of examinations, including performance-based assessments, oral examinations, and remote assessments.

*Keywords: mathematical methods, examination, validation, testing, job analysis, personnel certification, conformity assessment, competence, ISO/IEC 17024*

## Definitions

1. **Fairness** (ISO/IEC 17024:2012, Clause 3.16)  
Equal opportunity for success provided to each candidate (3.14) in the certification process (3.1).
2. **Validity** (ISO/IEC 17024:2012, Clause 3.17)  
Evidence that the assessment (3.8) measures what it is intended to measure, as defined by the certification scheme (3.2). *NOTE: In this international standard, validity is also used in its adjective form "valid."*
3. **Reliability** (ISO/IEC 17024:2012, Clause 3.18)  
Indicator of the extent to which examination (3.9) scores are consistent across different examination times and locations, different examination forms, and different examiners (3.10).
4. **Standard deviation**. Index of variability in a set of numbers. Computationally, this is the square root of the sum of deviations of each score and the mean, divided by the number of data points in the set.
5. **Mean**. The average score for a set of numbers.
6. **Standard error of measurement (SEM)**. This is the estimate of the variance of a person's scores if the person took many tests of a similar size. It is computed with the reliability coefficient of a test and the standard deviation of the set of obtained scores.
7. **Item**. The smallest measurable component of a test that can be scored is called an item.
8. **Item analysis**. Several characteristics of (usually) multiple-choice items that indicate the quality of the item and of the whole test.

### Examination Process Requirements in the ISO/IEC 17024: 2012 standard

The examination process requirements are described in clause 9.3.1 of ISO/IEC 17024:2012. This standard states: "Examinations shall be designed to assess competence based on, and consistent with, the scheme, by written, oral, practical, observational, or other reliable and objective means. The design of examination requirements shall ensure the comparability of results of each single examination, both in content and difficulty, including the validity of fail/pass decisions."

Additional information is also provided in clause 9.3.5 of ISO/IEC 17024:2012, which states: "Appropriate methodology and procedures (e.g., collecting and maintaining statistical data) shall be documented and implemented in order to reaffirm, at justified defined intervals, the fairness, validity, reliability, and general performance of each examination, and that all identified deficiencies are corrected."

### The Function of an Exam (Assessment) in the Personnel Certification Process

A test or examination (informally, exam or evaluation) is an assessment intended to measure a test taker's knowledge, skill, aptitude, or classification in many topics. The goal of the exam is to determine if an individual has sufficient knowledge, skills, and abilities (KSAs) to be professionally competent at an entry-level position in the specified field. An exam may be administered verbally, on paper, on a computer, or in a predetermined area that requires a test taker to demonstrate or perform a set of skills.

There is no general consensus or invariable standard for test formats and difficulty. Often, the format and difficulty of the test is dependent upon the requirements of accreditation or industrial association. Standardized tests are usually used by the personnel certification bodies to determine if a test taker is allowed to practice a profession, use a specific job title, or claim competency in a specific set of skills. It is a direct method of assessment of knowledge, skills, ability, and personal behaviors. (Note: A personnel certification exam has to be designed as a criterion-referenced standardized test or in combination with the criterion-referenced performance-based assessment.)

The assessment types that can be used in personnel

certification programs are as follows:

1. **Criterion-referenced tests** are designed to measure candidate's performance against a fixed set of criteria or industry standards or certification schemes, based on a construct of "minimal acceptable competency." It is possible for all test takers to pass, just like it is possible for all test takers to fail. A criterion-referenced test will use questions that will be correctly answered by candidates who are competent in the specific subject.
2. **Standardized test** are administered and scored in a consistent, or "standard," manner. Standardized tests are designed in such a way that the questions, conditions for administering, scoring procedures, and interpretations are consistent; furthermore, these tests are intended to be administered and scored in a predetermined, standard manner. Any test in which the same test is given in the same manner to all test takers, and graded in the same manner for everyone, is a standardized test. This assessment tool may be formatted as a written test, oral test, or practical skills performance test. The questions can be simple or complex. Standardized tests are designed to permit reliable comparison of outcomes across all test takers, because everyone is taking a test designed to assess the same competencies. Criterion-referenced scoring is used because it is concerned solely with whether or not a particular candidate's answer is correct and complete.
3. **Performance-based assessment** are used to evaluate objective data about a person's knowledge, skill, and attitude; the data is collected from the actual or simulated application site.

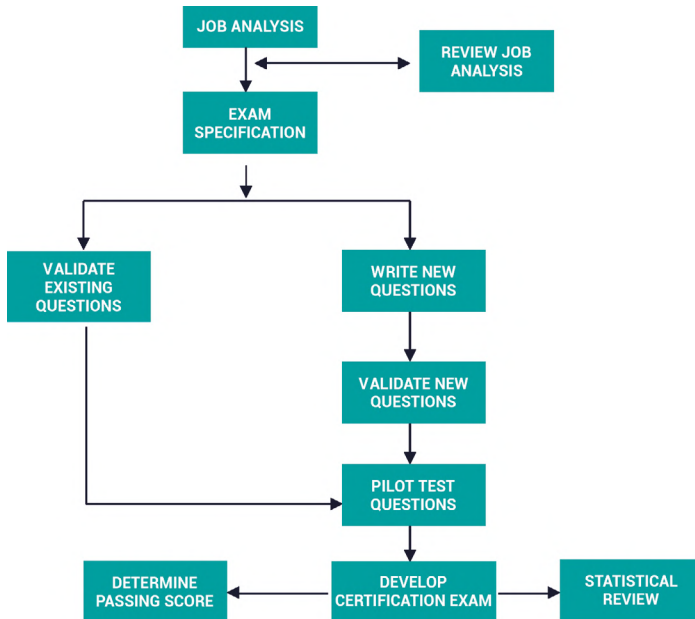
### Fairness

The fairness of an exam refers to its freedom from any kind of bias. The exam should be appropriate for all qualified examinees, without regard for factors that are irrelevant to professional competency such as race, religion, gender, or age. The test should not create a disadvantage for any examinee, or group of examinees, on any basis other than the examinee's lack of knowledge or skills the test is intended to measure.

Item writers should address the goal of fairness as they undertake the task of writing items. In addition,

the items should be reviewed for potential fairness problems during the item-review phase. Any items identified as displaying potential bias or lack of fairness should be revised or dropped from further consideration.

### Exam Validation Process Flow Chart



### Basic Steps in the Exam Validation Process

1. **Job Analysis:** Conducting a job analysis is an essential first step in establishing the content validity of certification exams. Job analysis is the foundation for defining the "certification scheme" (ISO/IEC 17024, Section 8). A job analysis will define the important elements of professional competency through a series of discrete "job tasks" and the associated KSAs required to perform these tasks. Metrics used for ranking the importance of job tasks should consider their "relevance" (relation to professional competency), "frequency" (how often these are done), and "criticality" (significance to professional success and to the protection of public health, safety, and welfare). In this process, job tasks should be eliminated from consideration in an examination when the KSA is adequately assessed by governmental licensing agencies (such as driving skills), and when no valid means of assessing competency in the task is identified. The rationale for eliminating tasks from consideration must be documented. Job analysis information may be gathered by directly observing people currently
2. **Review and Ranking of Job Tasks:** Ranking the importance of job tasks may be accomplished through surveys or through structured focus-group interviews of a representative panel of competent practitioners. One common approach is the "delphi research method," which is leveraged to build consensus and document conclusions. When surveys are used, these should be relayed to a representative group of practitioners (both highly experienced and entry-level) and impacted parties (the employers of certified persons). Job analysis must be periodically reviewed within a certain period of time. If the certification body is not the owner of the certification scheme, it must ensure the owner of the scheme reviews the job analysis.
3. **Exam Specification:** Ratings are used to identify the number of questions to appear on tests for each subject area. The specification (often called a "test blueprint") must clearly link the examination to the job analysis (both tasks and associated KSAs).
4. **Validate Existing Questions:** Existing questions are reviewed by subject matter experts for relevance, accuracy, and style.
5. **Write New Questions:** New exam questions are developed according to the job analysis
6. **Validate New Questions:** All new questions must be reviewed by subject matter experts for relevance, accuracy, and style.
7. **Pilot Test Questions:** Pilot tests allow for volunteers to statistically review each question and the entire test results
8. **Develop Certification Exam (Test Blueprint):** Examination blueprints are compiled from

in the job, interviewing experienced supervisors and job incumbents, and through questionnaires, personnel and equipment records, and work manuals. Workshops are held to identify specific job tasks and capabilities required for successful job performance. During these workshops, subject matter experts verify that the task statements developed are technically correct, unambiguous, and accurately reflect the job. Identification of capabilities must be done on a task-by-task basis, so that a link is established between each task statement and its requisite capability. Job analysis information is central in deciding what to test for and which tests to use.

job analysis results, then validated through committee meetings and workshops. Use and review the pilot test results. Operators, supervisors, and trainers should participate in the workshops.

9. **Determine Passing Score:** The passing score for an exam should be set in accordance with the purposes of the exam. The passing score is defined as the minimum score required to pass an exam to assure that the certificate-holder is professionally competent.
10. **Statistical Review:** Statistically review results of exams to identify problem questions. Questions that perform poorly should be discontinued from current use. These may be relayed back to the examination committee for further review and refinement.

### Details of the Exam Validation Process

It is essential to involve subject matter experts in all parts of the validation process. To qualify as a subject matter expert, a person must have direct, up-to-date experience with the job, and enough experience to be familiar with all of the tasks. Subject matter experts may include operators, supervisors, trainers, or other individuals with specialized knowledge about the job.

The principal steps normally taken for exam validation include:

1. Conduct a job analysis
2. Develop and validate items
3. Develop an exam
4. Establish a passing (cut) score

#### Step 1. Conduct a Job Analysis

Conducting a job analysis is an essential first step in establishing the content validity of certification exams. A job analysis often lists the capabilities (i.e., knowledge, skills, and abilities) required to perform work tasks. Job analysis information may be gathered by directly observing people currently in the job, conducting interviews with experienced supervisors and job incumbents, and through questionnaires, personnel and equipment records, and work manuals.

Workshops are held to identify the specific job tasks and capabilities required for successful job performance. During these workshops, subject matter experts verify that the task statements

developed are technically correct, unambiguous, and accurately reflect the job. Identification of capabilities must be done on a task-by-task basis, so that a link is established between each task statement and requisite capability.

Job analysis information is central in deciding what to test for and which tests to use

#### Step 2. Develop and Validate Items

Exam items are developed from the results of the job analysis so that exams are representative of job tasks. Once the new items are written, they must go through a validation process, which includes:

1. Linking new questions to the results of the job analysis. The purpose of this is to ensure that all questions on the certification exam measure at least one important aspect of an operator's job. During this process, subject matter experts are asked to rate the extent to which the questions reflect specific tasks in the job.
2. Analyzing questions for technical accuracy, style, readability, and possible bias to sub-groups. This is done to determine whether the correct answer is the best answer, confirm the distractors (incorrect answers) are wrong, and verify that the question is free from bias with respect to race, gender, and culture.
3. Reviewing items for job importance. Importance ratings should reflect how well the question distinguishes between effective and ineffective job performance and if the knowledge tested in the question is necessary for competent job performance. The continued relevance of questions that have been validated must be ensured through periodic reviews of the items by subject matter experts. Evaluation of questions should also be conducted through statistical analysis. Of particular importance are the difficulty index (the ratio of examinees that answer each question correctly) and the discrimination index (how well the question distinguishes between the more knowledgeable and less knowledgeable examinees).

#### Conduct the Item Analysis

In this phase, statistical methods are used to identify any test items that are not working well. If an item is too easy, too difficult, fails to show a difference between skilled and unskilled examinees, or is scored incorrectly, an item analysis will reveal it. The two most common statistics reported in an



item analysis are the item difficulty, which measures the proportion of examinees who responded to an item correctly, and the item discrimination, which measures how well the item discriminates between examinees who are knowledgeable in the content area and those who are not.

**Item Difficulty Index ( $p_i$ )** is the level of question difficulty that affects test validity. If the exam is merely composed of difficult or easy questions, the distinction among the applicants cannot be determined clearly. The exam is expected to have an intermediate level of difficulty and this level helps determine the distinction among the applicants. Also, it is used for internal consistency formulas.

It is denoted as:

$$p_i = \frac{n(D)}{N}$$

$n(D)$  : Number of participants that answered an item correctly

$N$ : Number of all participants that take exam

**TABLE 1**

*Evaluation of Item Difficulty Index*

<u>Item Difficulty Index</u>	<u>Item Difficulty Level</u>
Close to 1.00	easy
About 0.50	medium
Close to 0.00	difficult

**For example**, consider an exam with 20 participants that contains multiple-choice questions. If a question had 9/20 test takers answer it correctly, this would then result in an Item Difficulty Index ( $p_i$ ) of 0.45, which would then classify this question as "medium difficulty." If a question, on the other hand, had 19/20 test takers answer it correctly, this would result in a  $p_i$  of 0.9, which would classify it as an "easy difficulty" question.

**Item Discrimination Index ( $r$ )** is the efficiency of test questions used to determine the distinction among the applicants. It expresses the relationship between the overall score and single-question scores. It measures how well an item is able to distinguish between examinees who are knowledgeable and those who are not, or between masters and non-masters. Item discrimination efficiency is to be high for test reliability. When an item discriminates negatively, overall, this means the most knowledgeable examinees are getting the item wrong and the least knowledgeable examinees

are getting the item right. A negative discrimination index may indicate the item is measuring something other than what the rest of the test is measuring. More often, it is a sign that the item has been miskeyed.

When interpreting the value of a discrimination, it is important to be aware that there is a relationship between an item's difficulty index and its discrimination index. If an item has a very high (or very low)  $p$ -value, the potential value of the discrimination index will be much less than if the item has a midrange  $p$ -value. In other words, if an item is either very easy or very hard, it is not likely to be very discriminating.

There are over 20 discrimination indices used as indicators of the item's discrimination effectiveness such as the index of discrimination ( $D$ ), Henryson discrimination index ( $r_{jx}$ ), point-biserial correlation coefficient ( $r_{pbis}$ ), biserial correlation coefficient ( $r_{bis}$ ), etc.

**TABLE 2**

*Evaluation of Item Discrimination Index*

<u>Item Discrimination Index</u>	<u>Item Discrimination Level</u>
0.4 and above	very well
0,30 - 0,39	reasonable
0,20 - 0,29	should be corrected
0,19 and below	very poor, remove from test

Some of the statistical formulas are given below.

**Henryson discrimination index ( $r_{jx}$ )**

It is denoted as:

$$r_{jx} = \frac{\bar{X}(d) - \bar{X}}{S_x} \sqrt{\frac{p_j}{q_j}}$$

$\bar{X}(d)$ : Exam score average of those who answer the item correctly

$\bar{X}$ : Arithmetic mean of the exam scores

$S_x$ : Standard deviation of the exam scores

$p_j$ : Item difficulty index of the item

$q_j$ :  $1 - p_j$

**For example**, consider an exam with 20 participants that contains 45 multiple-choice questions. If the arithmetic mean of the exam scores is 32.85 and the standard deviation of the scores is 6.651, the discrimination index (DI) level of certain questions can then be examined based on the item difficulty index, and the exam score average of correct answers is as follows:

Question Number*	$P_j$	$q_j$	$\bar{X}(d)$	$r_{jx}$	DI Level
1	0.45	0.55	35.78	0.40	Very good
2	0.75	0.25	32.20	-0.17	Very poor
37	0.95	0.05	33.26	0.27	Should be corrected
38	0.75	0.25	34.33	0.38	Reasonable
42	0.8	0.2	33.75	0.27	Should be corrected
43	0.9	0.1	33.27	0.19	Should be corrected
44	0.85	0.15	33.52	0.24	Should be corrected
45	0.8	0.2	34.5	0.49	Very good

\*Questions selected typically from a total of 45.

### Index of discrimination ( $D$ )

When calculating the DI in accordance with the simple method, the respondents are divided into two groups (lower and upper groups) according to the method. First, the total scores are calculated according to the results obtained from the measurement tool and ranked from highest to lowest. The 27% group with the highest success is taken as the upper group and the 27% group with the lowest success is taken as the lower group. The remaining 46% group is excluded from the calculation.

It is denoted as:

$$D = P_u - P_l$$

$P_u$ : proportion of test takers in the upper group who get the item right

$P_l$ : proportion of test takers in the lower group who get the item right

**For example**, consider an exam with 20 participants that contains multiple-choice questions. If a question had 67% of the upper group getting it correct ( $P_u = 0.67$ ) and 33% of the lower group getting it correct ( $P_l = 0.33$ ), then Item Discrimination Index would be 0.33, which would classify the discrimination as reasonable. Meanwhile, if both the upper and lower groups got the question correct ( $P_u = P_l = 1$ ), this would result in an Item Discrimination Index of 0 and imply that said item discriminates very poorly.

### Point-Biserial Correlation Coefficient ( $r_{pbis}$ )

Point biserial in the context of an exam is a way of measuring the consistency of the relationship between a candidate's overall exam mark (a

continuous variable—i.e., anywhere from 0-100%) and a candidate's item mark (a dichotomous variable—i.e., only two possible outcomes). It gives an indication of how strong or weak this correlation is compared to the other items in that exam. In other words, does the way in which candidates answer an item help to indicate whether they are strong or weak candidates?

It is denoted as:

$$r_{pbis} = \frac{M_1 - M_0}{S_n} \sqrt{pq}$$

$M_1$ : mean (for the entire test) of the group that received the positive binary variable (i.e., the "1")

$M_0$ : mean (for the entire test) of the group that received the negative binary variable (i.e., the "0")

$S_n$ : standard deviation for the entire test

$p$ : item difficulty index

$q$ : ( $1 - p$ )

**For example**, consider an exam with 20 participants that contains 45 multiple-choice questions. If the arithmetic mean of the exam scores is 32.85 and the standard deviation of the scores is 6.651, the DI level of certain questions can then be examined based on the item difficulty index, the mean of the group of test takers that answered correctly, and the mean of test takers that answered incorrectly, as follows:

Question Number*	$P_j$	$q_j$	$M_1$	$M_0$	$r_{pbis}$	DI Level
1	0.45	0.55	35.78	30.45	0.40	Very good
2	0.75	0.25	32.20	34.80	-0.17	Very poor
3	0.45	0.55	35.78	30.45	0.40	Very good
25	0.75	0.25	34.33	28.4	0.38	Reasonable
26	0.85	0.15	33.23	30.66	0.136	Very poor
27	0.65	0.35	34.76	29.28	0.39	Very good
37	0.95	0.05	33.26	25	0.27	Should be corrected
38	0.75	0.25	34.33	28.4	0.38	Reasonable

\*Questions selected typically from a total of 45.

### Biserial Correlation Coefficient ( $r_{bis}$ )

A biserial correlation coefficient is almost the same as point biserial correlation, but one of the variables is dichotomous ordinal data and has an underlying continuity.

It is denoted:

$$r_{bis} = \frac{(M_1 - M_0)(\frac{pq}{Y})}{S_n}$$

$M_1$ : mean (for the entire test) of the group that received the positive binary variable (i.e., the "1")

$M_0$ : mean (for the entire test) of the group that received the negative binary variable (i.e., the "0")

$S_n$ : standard deviation for the entire test

$p$ : item difficulty index

$q$ :  $(1 - p)$

$Y$ : Y ordinate of the normal distribution corresponding to the  $p$  value.

### Using Item Analysis on Essay-Type Questions

Personnel certification bodies may want to evaluate their candidates using various types of questions—including essay, modified essay, short answer, and multiple-choice types of questions. Among these, the multiple-choice question (MCQ) is very common and is the preferred type of question used in exams due to the efficiency and reliability of scoring and simplicity of analysis.

One of the most common tools used to assess knowledge is the essay question. These evaluations depend on test and item analysis, which consists of analyzing individual questions as well as the whole test. Although this activity could be done more precisely in objective-type questions, it can also apply to essay, structured essay, and short-answer types of questions.

For item analysis, assessors must determine the intermediate score ranges in accordance with the maximum score that can be given to the essay-type or short-answer question. This involves listing all test takers' marks for individual questions and in accordance with aggregate marks scored, arranging test takers in rank order (with the highest score given on the top), and dividing test takers between the high-ability group (HAG) and low-ability group (LAG).

**For example**, if a question is given five points, each answered question that achieves 5 to 3.5 marks will be considered a correct answer (A). Each answered question that achieves 3 to 2 marks will be considered a near-to-correct answer (B). Each answered question that achieves 1.5 to 0.5 marks will be considered a near-to-incorrect answer (C). Each answered question that achieves 0 marks will

Marks range		5.0 - 3.5	3.0-2.0	1.5-0.5	0	Total no. of considered test takers
Designated sign		A	B	C	D	
No. of HAG test takers	Q1	11	14	0	0	25
	Q2	15	9	1	0	25
	Q3	5	14	5	1	25
	Q4	16	9	0	0	25
	Q5	8	10	7	0	25
	Q6	22	3	0	0	25
	Q7	16	8	1	0	25
	Q8	4	20	1	0	25
No. of LAG test takers	Q1	1	3	12	9	25
	Q2	1	22	2	0	25
	Q3	0	5	15	5	25
	Q4	2	13	8	2	25
	Q5	0	2	20	3	25
	Q6	4	18	3	0	25
	Q7	3	14	3	5	25
	Q8	0	12	11	2	25
Level of Correctness		Correct answer	Near to correct answer	Near to incorrect answer	Incorrect answer	

For all given questions, no. of test taker obtained marks in different range.

be considered an incorrect answer (D). For each question, examiners should count the number of total test takers that obtained marks in the A, B, C, and D categories.

The indices' facility value (FV) and discrimination index (DI) are calculated in the following formulas.

**Facility Value (FV):** This is the number in the group answering a question right. Facility value—also called a difficulty index—measures a question's level of ease or difficulty. The higher the FV, the easier the question.

It is denoted as:

$$FV = \frac{HAG + LAG}{N} \times 100$$

*HAG*: High-ability group

*LAG*: Low-ability group

*N*: Total number of considered test takers

The FV value is expressed as a percentage. Its range is 0-100. Its recommended value is 45-60 and its acceptable value is 25-75.

**Discrimination Index (DI):** This index indicates the ability of a question to discriminate between test takers with higher and lower abilities.

It is denoted as:

$$DI = \frac{2 \times (HAG - LAG)}{N}$$

The *DI* value is expressed as a fraction. Its range is 0-1.0. Its maximum value is 1.0, which indicates an ideal question with perfect discrimination between HAG and LAG. Its value could extend from -1.00 to +1.00. The minus value—also called negative discrimination—means that more test takers in the lower group are answering an item correctly compared to test takers in the higher group.

Recommended value: > 0.25

Acceptable with revision: 0.15-0.25

Discard the question: < 0.15

This item analysis helps to detect specific technical flaws in the questions and provides information for improvement. It also increases the item-writing skills of examiners. There are no clear-cut guidelines in formulating the item analysis. However, regular exercise over this analysis would contribute to a personnel certification body's formulation of appropriate questions.

### Step 3. Develop the Exam

After the job analysis survey is evaluated, the results are used to develop valid certification exams. Specifications for certification exams are based on the results of the job analysis and reflect how often a task, knowledge, skill, or ability is needed in practice and how much impact it has on effective job performance.

### Step 4. Establish the Passing (Cut) Score

The cut score is defined as the minimum score required to pass an exam. Defining the cut score required for certification is one of the most important but difficult aspects of the validation process.

#### Setting the Passing (Cut) Score of an Exam

Standard setting is the process used to select a passing score for an exam. Of all the steps in the test development process, the standard setting phase may be the one most like art, rather than science; while statistical methods are often used in conducting a standard setting, the process is also greatly impacted by judgment and policy.

The passing score (also known as the passing point, cutoff score, or cut-score) is used to classify examinees as either masters or non-masters. An examinee's score must be equal to or greater than the passing point, in order for the examinee to be classified as a master or to pass the test. If an examinee is misclassified, that is referred to as a classification error.

Typically, the passing score is set at a score point on the exam that the judges determine reflects the minimum level of competency to protect the public from harm or to provide minimal competency at the occupational level being assessed. For the standard setting to be conducted successfully, the panel of judges should be carefully selected and then thoroughly prepared and trained for their task.

There are a number of approaches to standard setting, including: informed judgment, conjectural, and contrasting groups methods. All of these methods require the insight of a representative panel of competent practitioners representing appropriate demographics and experience, ranging from those who have recently entered the profession to those who have competently practiced for many years.



## Methods for Standard Setting

**Types of Classification Error:** The passing score for a test should be set in accordance with the purposes of the exam and with consideration to relative risks to the public from incompetent practice. It should not be set arbitrarily, but rather should be carefully determined by a panel of judges who are familiar with the content of the exam as well as the characteristics of the occupation concerned.

Two types of classification errors can occur when the passing score is applied.

One type of misclassification is termed a false-positive (i.e., an error of acceptance). An example of a false-positive error would be an examinee who was not minimally competent, but who passed the test.

The second type of misclassification is termed a false-negative (i.e., an error of rejection). In this type of misclassification, an examinee who actually has the level of competence fails the test.

Depending upon the nature of the exam program, one of these types of errors may be far more problematic than the other. Awareness of these potential consequences may be used to influence the determination of the final passing score, after the panel of judges has made its recommendation. Policymakers of the exam program may adjust that recommended passing point based on other factors, and possibly include operational test score data when it becomes available.

**Informed Judgment Method:** The informed judgment method is a test-based approach. A panel of judges, or stakeholders, reviews the overall test and its content. Based on their holistic reviews, the judges then each suggest a percentage of items on the test that ought to be correctly answered by a minimally competent examinee. This percent-correct score on the total test can be viewed as each judge's recommended passing score. These recommended passing scores from the panel, along with possible additional information, can be used to set the final passing score. The informed judgment method might be difficult to rationally defend when it is used in isolation. However, it may be a very appropriate method for use in combination with other methods, particularly the contrasting groups method.

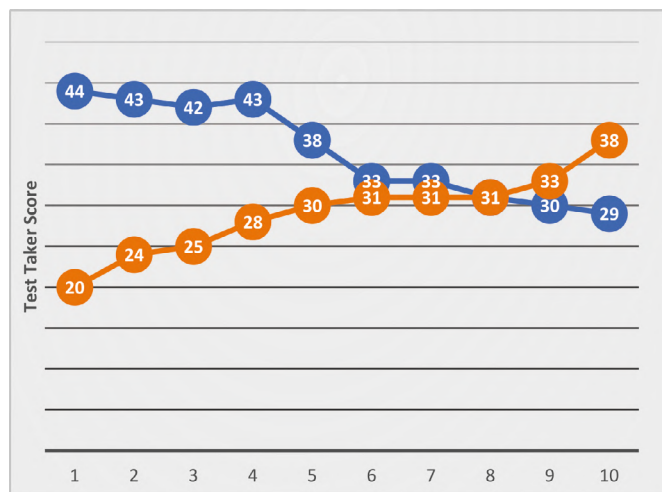
**Conjectural (Modified-Angoff) Method:** The modified-Angoff method is the most commonly used of the conjectural methods, all of which are

item-based approaches to standard setting. A panel of judges is assembled and asked to review the test, one item at a time. For each item, each judge gives an estimate of the probability that a minimally competent examinee would be likely to respond correctly. (Alternatively, the judges may be asked to imagine a hypothetical group of minimally competent examinees and then to indicate the percentage of the group that would be likely to respond to the given item correctly.) When judges are not in agreement regarding the pass/fail standard, those with disparate ratings are given the opportunity to explain their rankings, with the voting process repeated, building consensus. Typically, one or more additional rounds of review are undertaken. These passing scores are then averaged across the individual judges to arrive at the full panel's recommended final passing score.

**Contrasting Groups Method:** The contrasting groups method is an examinee-based approach to standard setting. This method in particular requires that the panel of judges be highly familiar with the target test population. The panel of judges identifies a set of examinees who are clearly non-masters and another set of examinees who are clearly masters; borderline examinees are not included. It is especially important that the non-masters be carefully selected. While the non-master examinees would not yet be considered minimally competent in the occupational area, they should nevertheless be members of the target test population. If, instead, the examinees identified as non-masters are completely unknowledgeable in the exam's content area, the passing score may be set at an artificially low point. After the two groups of examinees have been identified, the test is administered to them. The two resulting test score frequency distributions are plotted on the same continuum. The passing score can be set at the intersection point of the two distributions; or, alternatively, the final passing score can be adjusted somewhat, based on the relative cost of false-positive and false-negative classification errors. While the contrasting groups method can be used independently, it may also be used as a complement to the informed judgment or other standard setting method.

**For example,** consider an exam with 20 participants that contains 45 multiple-choice questions. A list can be created, including the descending order of scores of experienced test takers (pictured below in blue) and the ascending order of scores of other test

takers (pictured below in orange).



Notice that the lists intersect at a score of 31, which can then be used as a cut off score.

### Test Reliability

Test reliability is an index of the consistency of scores produced by the test, with a higher value being desirable. A value of 1.0 indicates a perfectly reliable test. A value of 0.0 indicates the test essentially produces random scores.

The test measures what it claims to measure consistently or reliably. This means that if a person were to take the test again, the person would get a similar test score.

Reliability refers to how dependably or consistently a test measures a characteristic. If a person takes the test again, will he or she get a similar test score, or a much different score? A test that yields similar scores for a person who repeats the test is said to measure a characteristic reliably.

How do we account for an individual who does not get exactly the same test score every time he or she takes the test? Some possible reasons are as follows:

- **Test taker's temporary psychological or physical state.** Test performance can be influenced by a person's psychological or physical state at the time of testing. For example, differing levels of anxiety, fatigue, or motivation may affect the applicant's test results.
- **Environmental factors.** Differences in the testing environment, such as room temperature, lighting, noise, or even the test administrator, can influence an individual's test performance.

- **Test form.** When tests are administered on multiple dates, for security reasons, additional forms of the test may be necessary. It is expected that test forms will be revised at least annually. Test forms must be assembled to the same "test blueprint." Different forms of a test are known as parallel forms or alternate forms. These forms are designed to have similar measurement characteristics, but they contain different items. Because the forms are not exactly the same, a test taker might do better on one form than on another.
- **Multiple raters.** In certain tests, scoring is determined by a rater's judgments of the test taker's performance or responses. Differences in training, experience, and frame of reference among raters can produce different test scores for the test taker.

These factors are sources of chance or random measurement error in the assessment process. If there were no random errors of measurement, the individual would get the same test score. The degree to which test scores are unaffected by measurement errors is an indication of the reliability of the test.

### Types of Reliability Estimates

There are several types of reliability estimates, each influenced by different sources of measurement error. The acceptable level of reliability will differ depending on the type of test and the reliability estimate used.

1. **Test-retest reliability** indicates the repeatability of test scores with the passage of time. This estimate also reflects the stability of the characteristic or construct being measured by the test. For constructs that are expected to vary over time, an acceptable test-retest reliability coefficient may be lower than is suggested in Table 3 below.
2. **Alternate or parallel form reliability** indicates the likelihood of achieving consistent test scores if a person takes two or more forms of a test. A high parallel form reliability coefficient indicates that the different forms of the test are very similar, which means that it makes virtually no difference which version of the test a person takes. On the other hand, a low parallel form reliability coefficient suggests that the different forms are probably not comparable; they may be measuring different things and therefore cannot be used interchangeably.

3. **Inter-rater reliability** applies most often to examinations administered by examiners (vs. objective multiple-choice examinations). Inter-rater reliability indicates the likelihood of achieving consistent test scores when two or more raters score the test. On some tests, raters evaluate responses to questions and determine the scores. Differences in judgment among raters are likely to produce variations in test scores. A high inter-rater reliability coefficient indicates that the judgment process is stable, and the resulting scores are reliable. Inter-rater reliability coefficients are typically lower than other types of reliability estimates. However, it is possible to obtain higher levels of inter-rater reliabilities if raters are appropriately trained.
4. **Internal consistency reliability** indicates the extent to which items on a test measure the same thing. A high internal consistency reliability coefficient for a test indicates the items on the test are very similar to each other in content (homogeneous). It is important to note that the length of a test can affect internal consistency reliability. For example, a very lengthy test can spuriously inflate the reliability coefficient.

### Interpretation of Reliability

The reliability of a test is indicated by the reliability coefficient. It is denoted by the letter "r" and expressed as a number ranging between 0 and 1.00, with  $r = 0$  indicating no reliability and  $r = 1.00$  indicating perfect reliability. .

Generally, you will see the reliability of a test as a decimal, for example,  $r = 0.80$  or  $r = 0.93$ . The larger the reliability coefficient, the more repeatable or reliable the test scores.

**TABLE 3**

*General Guidelines for Interpreting Reliability Coefficients*

<u>Reliability Coefficient Value</u>	<u>Interpretation</u>
.90 and up	excellent
.80 - .89	good
.70 - .79	adequate
below .70	may have limited applicability

One measure of reliability used is Cronbach's alpha. This is the general form of the more commonly reported Kuder-Richardson Formula 20 (KR-20) and can be applied to tests composed of items

with different numbers of points given for different response alternatives. When the coefficient alpha is applied to tests in which each item has only one correct answer and all correct answers are worth the same number of points, the resulting coefficient is identical to KR-20.

Estimates of test reliability are only meaningful when there are a sufficient number of examinations administered, typically requiring data from at least 100 candidates. While newly formed certification bodies may not have access to sufficient data to estimate reliability, it is expected that more mature programs will estimate and consider statistical reliability in their validation processes.

### Kuder-Richardson Method

Kuder-Richardson Formula 20, or KR-20, is a reliability measure for a test with binary variables (i.e., answers that are right or wrong). Reliability refers to how consistent test results are, or how well the test actually measures what it is intended to measure.

The KR-20 is used for items that have varying difficulty. For example, some items might be very easy, while others are more challenging. It should only be used if there is a correct answer for each question—it shouldn't be used for questions where partial credit is possible or for scales like the Likert scale.

**KR20 Scores:** The scores for KR-20 range from 0 to 1, where 0 is no reliability and 1 is perfect reliability. The closer the score is to 1, the more reliable the test.

It is denoted as:

$$KR20 = \left[ \frac{n}{n-1} \right] \cdot \left[ \frac{1 - \Sigma(p \cdot q)}{Var} \right]$$

$n$ : sample size for the test

$p$ : proportion of people passing the item

$q$ : proportion of people failing the item

$Var$ : variance for the test

$\Sigma$ : sum up (add up) \*In other words, multiply each question's  $p$  by  $q$ , then add them all together. If you have 10 items, you'll multiply  $p$  by  $q$  10 times, then you'll add those 10 items together to get a total.

**KR21 Scores:** If all questions in your binary test are equally challenging, use the Kuder-Richardson

Formula 21 (KR-21) method.

It is denoted as:

$$KR21 = \left[ \frac{n}{n-1} \right] \cdot \left[ \frac{1 - (M \cdot (n - M))}{n \cdot Var} \right]$$

$n$ : sample size for the test

$Var$ : variance for the test

$M$ : mean score for the test

**For example**, consider an exam with 20 participants that contains 45 multiple-choice questions. Since all questions in this situation are equally challenging, we would choose to use the KR-21 score. If, however, the summation of the product of people passing and failing each item is 8.0325 and the variance is 42.0275, we could deduce the KR-20 score for this exam to be -0.17, further verifying it is incorrect to use KR-20 in this scenario. Knowing that that mean is 32.85, we could then deduce the KR-21 score to be 0.5299, indicating average reliability of the test.

**Cronbach's Alpha:** This measures reliability, or internal consistency. If you have a test with more than two answer possibilities (or opportunities for partial credit), use Cronbach's alpha instead. Cronbach's alpha is used to see if multiple-question Likert scale surveys are reliable.

It is denoted as:

$$\alpha = \frac{k}{(k-1)} \left[ 1 - \frac{\sum \sigma_j^2}{\sigma_T^2} \right]$$

$k$ : number of items on the test

$\sum \sigma_j^2$ : sum of the "j" item score variances

$\sigma_T^2$ : variance of the total test scores

**For example**, consider an exam with 20 participants that contains 14 questions with more than two answer possibilities (or opportunities for partial credit). If the sum of the "j" item score variances is 42.9 and the variance of the total test scores is 161.4, Cronbach's alpha can be calculated to be 0.7907, which would indicate adequate-to-good reliability.

### Test validity

Validity indicates whether the characteristic measured by a test is related to job qualifications and requirements for entry-level, competent practitioners. Validity gives meaning to the test scores. Validity evidence indicates there is linkage between test performance and job performance.

It is important to understand the differences

between reliability and validity. Validity demonstrates how good a test is for a particular situation; reliability indicates how trustworthy a score on that test will be. Examiners must carefully select a test that is both reliable and valid for each unique situation.

### Methods for Conducting Test Validation Studies

The validity of a certification examination requires analysis of the entire process, including the supporting research for the examination (job analysis and scheme-development) as well as the security and integrity of the process for administering and scoring examinations. A holistic approach is necessary. Because of the diversity of facets that impact validity, statistical indicators of validity of an examination are rarely employed but may be useful.

Broad constructs for analyses for certification examinations are often defined as "face validity," "criterion-related validity," "content-related validity" and "construct-related validity." The simplest of these is face validity—whether or not the examination appears (to examination candidates) to relate to important elements of professional practice. This is a qualitative metric that is important for public acceptance and the reputation of the examination. The remaining constructs include quantitative metrics and are defined as follows:

1. **Criterion-related validation** requires demonstration of a correlation or other statistical relationship between test performance and job performance. In other words, individuals who score high on the test tend to perform better on the job than those who score low on the test. If the criterion is obtained at the same time the test is given, it is called concurrent validity; if the criterion is obtained at a later time, it is called predictive validity.

The criterion-related validity of a test is measured by the validity coefficient. It is reported as a number between 0 and 1.00 and indicates the magnitude of the relationship, "r," between the test and a measure of job performance (criterion). The larger the validity coefficient, the more confidence there is in predictions made from the test scores. However, a single test can never fully predict job performance because success on the job depends on so many varied factors. Therefore, validity coefficients, unlike reliability coefficients,



rarely exceed  $r = 0.40$ .

It is denoted as:

$$\text{Correl}(X, Y) = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma (x - \bar{x})^2 \Sigma (y - \bar{y})^2}}$$

$x$ : exam score of test taker in group 1

$\bar{x}$ : arithmetic mean of the exam scores of group 1

$y$ : exam score of test taker in group 2

$\bar{y}$ : arithmetic mean of the exam scores of group 2

As a general rule, the higher the validity coefficient, the more beneficial it is to use the test. Validity coefficients of  $r = 0.21$  to  $r = 0.35$  are typical for a single test.

**For example**, consider an exam that contains 45 multiple-choice questions with two exam groups of 20 participants each. If the square of the summation of the difference of individual scores to the mean score is 840.55 in group 1 and 779.2 in group 2, and the summation of their products is 792.4, we can derive a correlation of 0.98, which is incredibly beneficial.

**TABLE 4**

*General Guidelines for Interpreting Validity Coefficients*

<u>Validity coefficient value</u>	<u>Interpretation</u>
above .35	very beneficial
.21 - .35	likely to be useful
.11 - .20	depends on circumstances
below .11	unlikely to be useful

- Content-related validation** is a non-statistical type of validity and requires a demonstration that the content of the test represents important job-related behaviors. In other words, test items should be relevant to and measure directly important requirements and qualifications for the job.
- Construct-related validation** requires a demonstration that the test measures the construct or characteristic it claims to measure, and that this characteristic is important to successful performance on the job.

Professionally developed tests should come with reports on validity evidence, including detailed explanations of how validation studies were conducted. If examiners develop their own tests or procedures, they will need to conduct their own validation studies.

### Standard Error of Measurement (SEM)

All examinations are imperfect measures of professional competency. It is important that certification bodies are aware of this and use available statistics to estimate the level of possible errors. For traditional multiple-choice examinations, a statistical estimate of this error is called the "Standard Error of Measurement" (SEM). The SEM is comparable to the statistical estimate "Uncertainty of Measurement" (MU), which is estimated by product-testing laboratories (ISO/IEC 17025).

SEM provides an estimate of the margin of error that is expected in an individual test score because of the imperfect reliability of the test.

The SEM represents the degree of confidence that a person's "true" score lies within a particular range of scores. For example, an SEM of "2" indicates that a test taker's "true" score probably lies within two points in either direction of the score he or she receives on the test. This means that if an individual receives a 91 on the test, there is a good chance the true score lies somewhere between 89 and 93.

The SEM is a useful measure of the accuracy of individual test scores. The smaller the SEM, the more accurate the measurements.

It is denoted as:

$$SEM = S_D \sqrt{(1 - r_{xx})}$$

$S_D$ : standard deviation of tests scores

$r_{xx}$ : reliability or precision of the test

$$r_{xx} = \frac{S_T^2}{S_X^2}$$

$S_T^2$ : variance of the true scores

$S_X^2$ : variance of the observed scores

We use the SEM to calculate confidence intervals around obtained scores..

68 % CI = Score  $\pm$  SEM

95 % CI = Score  $\pm$  (1.96\*SEM)

99 % CI = Score  $\pm$  (2.58\*SEM)

**For example**, consider an exam with 20 participants that contains 45 multiple-choice questions. If the standard deviation of the scores is 6.65128 and the reliability of the test is 0.52988, the calculation for the standard error of measurement is 4.6. This implies that the true scores are as follows: raw score of  $\pm 4.6$  (68% CI), raw score  $\pm 9.02$  (95% CI), and raw score  $\pm 11.87$  (99% CI).

## Author Biographies

**Osman Vural** manages the IAS Accreditation Program for Personnel Certification Bodies (ISO/IEC 17024). He earned a Bachelor of Science degree in civil engineering from Gazi University in Turkey and has worked for many years as a consultant, auditor, and trainer in conformity assessment markets. Osman's expertise extends to the accreditation of testing/calibration laboratories, personnel certification bodies, and management systems certification bodies. Other professional accomplishments include serving as director of the International Personnel Certification Association (IPC) and leading many relevant technical committees.



**Ioannis Anastasopoulos** received his Bachelor of Science degree in mathematics from the University of California at Berkeley, where he is currently pursuing a master's degree in the field of education technology. He has participated in research projects in both the United States and Europe during his course of studies, specifically related to OpenITS, credentialing examinations, and effective implementation of statistical methods and tools. Additionally, Ioannis has co-authored a series of publications related to conformity assessment, testing validation, and education technologies.



**David S. Nelson, PE, Ph.D.**, is the president of Quality Psychometric Services in Alabama and boasts 30-plus years' experience developing and managing state and national licensing and certification programs for construction-sector professions. Registered as a professional engineer in Florida, David earned a Bachelor of Science degree in civil engineering from North Carolina State University and received master's and doctorate degrees in educational psychology from the University of Southern California. Career highlights include serving as vice president of certification and testing for the International Code Council (ICC) and the International Conference of Building Officials. He was also the senior policy advisor of the International Accreditation Service (IAS) and managed the IAS accreditation program for personnel certification bodies (ISO/IEC 17024).

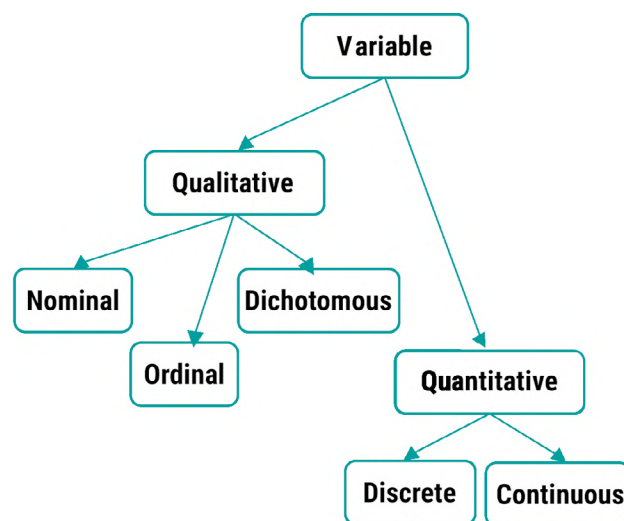


## Annex 1

### Statistical Terms and Definitions That Examiners Need to Know

#### Data

Data are obtained by measurement, counting, experimentation, observation, or research. Data collected by measurement or counting and reporting a numerical value are called **quantitative data**, and data that do not report a numerical value are called **qualitative (categorical) data**.



#### Qualitative (Categorical) Variables

Variables that are not numerical and which do not fit into categories.

##### Nominal Variables

Nominal variables are variables that have two or more categories, but which do not have an intrinsic order.










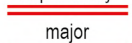

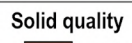

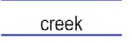
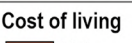
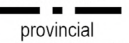
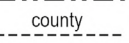
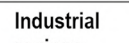








Office	Revenue	Revenue
Whalen	\$4,400.00	\$4,400.00
Hartstein	\$13,000.00	\$13,000.00
Fay	\$6,000.00	\$6,000.00
Raphaely	\$11,000.00	\$11,000.00
Khoo	\$3,100.00	\$3,100.00
Baida	\$2,900.00	\$2,900.00
Tobias	\$2,800.00	\$2,800.00
Himuro	\$2,600.00	\$2,600.00
Colmenares	\$2,500.00	\$2,500.00
Mavris	\$6,500.00	\$6,500.00

##### Ordinal Variables

A categorical variable for which the possible values are ordered. Ordinal variables can be considered "in

between" categorical and quantitative data.

#### Ordinal Data

<b>Point</b>	<b>Airports</b>  international  national  regional	<b>Oil well production</b>  high  medium  low	<b>Populated places</b>  large  medium  small
<b>Line</b>	<b>Roads</b>  expressway  major  local	<b>Drainage</b>  river  stream  creek	<b>Boundaries</b>  international  provincial  county
<b>Area</b>	<b>Solid quality</b>  good  fair  poor	<b>Cost of living</b>  high  medium  low	<b>Industrial regions</b>  major  minor

#### Dichotomous Variables

Dichotomous variables are nominal variables that have only two categories or levels. They have only two possible values (e.g., 0/1, Yes/No, True/False, etc.).

#### Quantitative Variables

A variable that reflects a notion of magnitude—that is, if the values it can take are numbers. A quantitative variable, thus, represents a measure and is numerical.

#### Discrete Variables

Variables for which the values it can take are countable and have a finite number of possibilities. The values are often (but not always) integers.

#### Continuous Variables

Variables for which the values are not countable and have an infinite number of possibilities.

#### Note: Misleading data encoding

In datasets, it is very often the case that numbers are used for qualitative variables. For instance, a person doing statistical analysis may assign the number "0" to the answer "False" and "1" to the answer "True." Despite the numerical classification, the variable answer is still a qualitative variable and not a discrete variable as it may look. The numerical classification is only used to facilitate data collection and data management.

#### Median

The value separating the higher half from the lower half of a data sample

1, 3, 3, **6**, 7, 8, 9

Median = 6

1, 2, 3, **4**, **5**, 6, 8, 9

Median =  $(4 + 5) \div 2$   
= 4.5

#### Arithmetic Mean

The sum of a collection of numbers divided by the count of numbers in the collection. In simple terms, it is known as an "average."

$$A = \frac{1}{n} \sum_{i=1}^n a_i$$

#### Weighted Arithmetic Mean

The weighted arithmetic mean is similar to an ordinary arithmetic mean, except that instead of each of the data points contributing equally to the final average, some data points contribute more than others.

$$\bar{x} = \frac{\sum_{i=1}^n w_i \cdot x_i}{\sum_{i=1}^n w_i}$$

#### Variance

The expectation of the squared deviation of a random variable from its mean.

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

#### Standard Deviation

Standard deviation is a measure of statistical dispersion. "Dispersion" indicates how much of the data is spread out. Specifically, it shows how the data is spread across the mean or average. For example, are all of the scores close to the average? Or are lots of scores way above (or way below) the average score?

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

#### Covariance

A measure of the joint variability of two random variables. In other words, a measure of how much two random variables vary together. It's similar to variance, but where variance tells how a single variable varies, covariance tells how two variables vary together.

$$C(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n-1}$$

#### Correlation

Correlation is a statistical technique that measures the relationship between two variables, such as X and Y, in terms of the units of measurement results for the variables.

### Correlation Coefficient That Can Be Used According to Variable Types

Variable Y/X	Quantitative X	Ordinal X	Nominal X
Quantitative Y	Pearson $r$	Biserial $r_{\text{bis}}$	Point Biserial $r_{\text{pbis}}$
Ordinal Y	Biserial $r_{\text{bis}}$	Spearman rho/ Tetrachoric $r_{\text{tet}}$	Rank Biserial $r_{\text{rrbis}}$
Nominal Y	Point Biserial $r_{\text{pbis}}$	Rank Biserial $r_{\text{rrbis}}$	Phi, L, C, Lambda

### Pearson Product-Moment Correlation Coefficient (PPMCC)

The correlation between sets of data measures how well they are related. It shows the linear relationship between two sets of data. In simple terms, it answers the question: "Can I draw a line graph to represent the data?"

$$r(X, Y) = \frac{n \cdot \sum_{i=1}^n x_i \cdot y_i - (\sum_{i=1}^n x_i) \cdot (\sum_{i=1}^n y_i)}{\sqrt{[n \cdot \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2] \cdot [n \cdot \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2]}}$$

### Point-Biserial Correlation Coefficient ( $r_{\text{pbis}}$ )

This is a special case of Pearson in which one variable is quantitative and the other variable is dichotomous and nominal. The calculations simplify since typically the values 1 (presence) and 0 (absence) are used for the dichotomous variable.

$$r_{\text{pbis}} = \frac{\bar{X}_1 - \bar{X}_0}{S_X} \sqrt{pq}$$

### Phi Coefficient ( $\varphi$ )

A measure of association for two binary variables. It is used for contingency tables when:

- at least one variable is a nominal variable
- both variables are dichotomous variables

Y/X	0	1	Totals
1	A	B	A+B
0	C	D	C+D
Totals	A+C	B+D	N

Contingency table

$$\varphi = \frac{(B \cdot C) - (A \cdot D)}{\sqrt{(A + B) \cdot (C + D) \cdot (A + C) \cdot (B + D)}}$$

### Tetrachoric Correlation Coefficient ( $r_{\text{tet}}$ )

An index reflecting the degree of the relationship between two continuous variables that have both been dichotomized.

$$r_{\text{tet}} = \cos\left(\frac{180^\circ}{1 + \sqrt{\frac{B \cdot C}{A \cdot D}}}\right)$$

## Annex 2 Classical Test Theory

Classical test theory (CTT), sometimes called the true score model, is the mathematics behind creating and answering tests and measurement scales. The goal of CTT is to improve tests, particularly the reliability and validity of tests.

**Reliability implies consistency:** If you take any test five times, you should get roughly the same results each time. A test is valid if it measures what it's supposed to.

### True Scores

Classical test theory assumes that each person has an innate true score. It can be summed up with an equation:  $X = T + E$

Where:

X is an observed score

T is the true score

E is random error

For example, let's assume you know exactly 70% of all the material covered in a statistics course. This is your true score (T). A perfect end-of-semester test (which doesn't exist) should ideally reflect this true score. In reality, you're likely to score around 65% to 75%. The 5% discrepancy from your true score is the error (E).

The errors are assumed to be normally distributed with a mean of zero. Hypothetically, if you took the test an infinite number of times, your observed score should equal your true score.

### Statistics Used in Classical Test Theory

Is your test measuring what it's supposed to?

Classical test theory is a collection of many statistics, including the average score, item difficulty, and the test's reliability.

**1. Correlation:** Shows how two variables, X and Y, are related to each other. Different measures



are used for different test types. For example, a dichotomously scored test (e.g., yes/no answers) would be correlated with point-biserial correlation while a polytomously scored test (one with multiple answers) would be scored with the Pearson Product-Moment Correlation Coefficient.

**2. Covariance:** A measure of how much two random variables vary together. It's similar to variance, but where variance tells how a single variable varies, covariance tells how two variables vary together.

**3. Discrimination Index:** The ability of the test to discriminate between different levels of learning or other concepts of interest. A high discrimination index indicates the test is able to differentiate between levels.

**4. Item difficulty:** A measure of individual test question difficulty. It is the proportion of test takers who answered correctly out of the total number of test takers. For example, an item difficulty score of 89/100 means that out of 100 people, 89 answered correctly.

**5. Reliability Coefficient:** A measure of how well the test measures achievement. Several methods exist for calculating the coefficient, including test-retest, parallel or alternate-form, and internal analysis. Rules of thumb for preferred levels of the coefficient are:

For high-stakes tests (e.g., college admissions):  
> 0.85

For low-stakes tests (e.g., classroom assessment): > 0.70

**6. Sample Variance / Standard Deviation:** Sample variance and sample standard deviation are measures of how spread out the scores are.

**7. Standard Error of Measurement (SEM):** A measure of how much measured test scores are spread around a "true" score.

## Annex 3

### Item Response Theory (IRT)

Item response theory (IRT) is a way to analyze responses to tests or questionnaires with the goal of improving measurement accuracy and reliability. IRT is one way to develop tests that actually measure what they are intended to measure (e.g., mathematical ability, reading ability, historical knowledge).

The first step in IRT is the development of a two-

dimensional matrix, which lists examinees and correct responses. In this matrix, 1 represents a correct answer and 0 is an incorrect answer

	Item 1	Item 2	Item 3	Item 4	Item 5	Mean Proficiency Level (Q)
Person 1	1	1	1	1	1	1
Person 2	0	1	1	1	1	0.8
Person 3	0	0	1	1	1	0.6
Person 4	0	0	0	1	1	0.4
Person 5	0	0	0	0	1	0.2
Mean ID ( $p_i$ )	0.8	0.6	0.4	0.2	0	

A quick look at this table illustrates that Person 1 answered all five questions correctly (100% proficient) while Person 4 correctly answered two questions (40% proficiency). However, proficiency isn't the only factor in IRT theory; the question's level of difficulty must also be considered. Let's say there are two test takers who both score 2/5. The first test taker may have answered two easy questions, and the second test taker may have answered two difficult questions. Therefore, although they both scored 40%, their proficiency is not the same.

Item response theory takes into account the number of questions answered correctly and the difficulty of each question.

There are many different models for IRT. Three of the most popular are:

- Rasch model
- Two-parameter model
- Three-parameter model

Some researchers consider the Rasch model to be completely separate from IRT. This is mainly because the Rasch model uses only a single parameter (called a "threshold"), while general IRT models use three. Another reason is that IRT aims to fit a model to data, while the Rasch model fits data to a model. Despite these differences, both models are used in favor of classical test theory—where the test taker's scores vary from one test to another.

### The Rasch model

In item response theory, a model that specifies only one parameter—item difficulty. This is thought to be a parsimonious way to describe the relation between

an item response and an underlying dimension and is thus preferred in some cases. Also called a one-parameter model.

#### **Two-parameter model**

In item response theory, a model that specifies two parameters affecting an individual's response to a particular test item: (a) the difficulty level of the item; and (b) the discriminating power of the item

#### **Three-parameter model**

In item response theory, a model that specifies three parameters affecting an individual's response to a particular test item: (a) the difficulty level of the item; (b) the discriminating power of the item; and (c) in multiple-choice items, the effect of guessing. The probability of a correct response to the item is held to be a mathematical function of these parameters.

#### **Anchor test**

A set of test items used as a reference point in comparing alternate forms of a test. One alternate form is administered to one group of participants, another is administered to a different group, and the items comprising the anchor test are administered to both groups. Scores on each alternate form are then compared with scores on the anchor test.

## **Annex 4**

### **Scheme Validation Process Flow Chart**

#### **5.1 Scheme analysis**

Scheme Technical Committee (STC)—with the support of specialized experts/consultants—proceeds in a competence analysis. Scheme competences are documented.

#### **5.2 Evaluate academic/training requirements**

STC members evaluate any applicable academic/training requirements of the certification scheme according to all applicable (market/legal/statutory/normative) scheme requirements.

#### **5.3 Evaluate experience requirements**

STC members evaluate any applicable experience requirements of the certification scheme according to all applicable (market/legal/statutory/normative) scheme requirements.

#### **5.4 Evaluate certification maintenance/recertification requirements**

STC members evaluate any applicable certification maintenance/recertification requirements of the certification scheme according to all applicable

(market/legal/statutory/normative) scheme requirements.

#### **5.5 Select and develop tests**

Specialized experts/consultants evaluate the scheme analysis information (competencies requirements) and determine the knowledge, skills, and abilities and the methods for their measurement.

#### **5.6 Set cutting scores and review final test**

STC experts review the test item-by-item. They select the correct answers, are told the keyed answers, and are asked what percent of qualified candidates would pass each item. The STC experts judge which, if any, of the knowledge, skills, or abilities is measured by the test. This is also their final review of the complete test before it is printed. The detailed scheme validation procedure (modified-Angoff model) is provided at the end of this document.

#### **5.7 Edit, compose, and print tests**

Examination department edits, composes, and prints (if required) the tests.

#### **5.8 Write content validation report**

Quality assurance manager writes a content-related validation report. After the STC reviews a draft, the final report detailing the activities undertaken is provided to the certification manager for approval, then it is given back to the quality assurance manager for inclusion in the management review agenda.

## **REFERENCES**

1. AERA (American Educational Research Association), APA (American Psychological Association), and NCME (National Council on Measurement in Education). 2014. *The Standards for Educational and Psychological Testing*. American Educational Research Association.
2. Anastasopoulos, G. 2017. "Setting the Passing Score of an Exam." Presentation at the IPC Annual General Meeting, Montreal, Canada, October 2017.
3. Association of Boards of Certification. 2013. "Validating Your Certification Exam." [http://www.abccert.org/pdf\\_docs/ValidatingYourCertificationExam.pdf](http://www.abccert.org/pdf_docs/ValidatingYourCertificationExam.pdf).
4. Biddle, R. 1993. "How to Set Cutoff Scores for Knowledge Tests Used in Promotion, Training, Certification, and Licensing." *Public Personnel Management* 22: 63-79.

5. Burcu, H., Yüksel, T., and Zafer, K. 2020. "Distractor Analysis Based on Item Difficulty Index and Item Discrimination Index." *Journal of Gumushane University Institute of Science and Technology*.
6. Glen, S. n.d. "Classical Test Theory: Definition." <https://www.statisticshowto.com/classical-test-theory/>.
7. Glen, S. n.d. "Item Response Theory: Simple Definition." <https://www.statisticshowto.com/item-response-theory/>.
8. Henryssen, S. 1971. "Gathering, analyzing, and using data on test items." In *Educational measurement*, 2nd ed., edited by R.L. Thorndike, 130-159. Washington, D.C.: American Council on Education.
9. ISO (International Organization for Standardization). 2012. *Conformity assessment – General requirements for bodies operating certification of persons*. ISO/IEC 17024:2012. <https://www.iso.org/standard/52993.html>.
10. ISO (International Organization for Standardization). 2014. *Conformity assessment – Vocabulary related to competence of persons used for certification of persons*. ISO/IEC TS 17024:2014. <https://www.iso.org/standard/62024.html>.
11. Liu, F. 2008. "Comparison of Several Popular Discrimination Indices Based on Different Criteria and Their Application in Item Analysis." Master's thesis, University of Georgia. [https://getd.libs.uga.edu/pdfs/liu\\_fu\\_200808\\_ma.pdf](https://getd.libs.uga.edu/pdfs/liu_fu_200808_ma.pdf).
12. Professional Testing Inc. n.d. "How Do You Determine if a Test Has Validity, Reliability, Fairness, and Legal Defensibility?" [https://www.proftesting.com/test\\_topics/pdfs/test\\_quality.pdf](https://www.proftesting.com/test_topics/pdfs/test_quality.pdf).
13. Statistics How To. n.d. <https://www.statisticshowto.com>.
14. U.S. Department of Labor Employment & Training Administration. 1999. "Testing and Assessment: An Employer's Guide to Good Practices." <https://wdr.doleta.gov/opr/fulltext/document.cfm?docn=6032>.



# 5th International Conference on Quality Engineering and Management

# SAVE THE DATE

July 14 - 15, 2022

University of Minho, Braga, Portugal

**A better world with Quality!**

**Quality in the digital transformation**

**Contacts:** Paulo Sampaio - Conference Chair | [icqem@dps.uminho.pt](mailto:icqem@dps.uminho.pt)

**Info:** <http://icqem.dps.uminho.pt>

## Abstract

As organizations seek to become increasingly competitive and innovative, knowledge management systems are emerging as critical assets. This article reviews how knowledge is portrayed in management system standards, examining the relationship and how it reflects on the governance of organizations. The framework of ISO standards mentioned in this article is useful for organizations that aim to develop a mature roadmap to business resilience and continuity

**Keywords:** *knowledge management, knowledge management system, information, tacit knowledge, explicit knowledge, ISO 30401, ISO 9001, ISO 27001*

# Governance of Knowledge in Management Systems Applications

By Tolga Aktaş, Management System Auditing Specialist

## Overview of Knowledge and Knowledge Management

The amount of data in our world is increasing—from our personal lives to our professional pursuits, revealing competition in every business field. When processed, data evolves into information. When combined with insights, information evolves into knowledge, which helps organizations make effective decisions and take actions in context. Therefore, organizations can exist and maintain their longevity by controlling knowledge. For this reason, issues related to the protection, sharing, and security of knowledge have come to the fore, discussed and standardized.

Knowledge management (KM), on the other hand, is defined as “management with knowledge” and “a systematic and comprehensive approach to improve results and learning” by ISO 30401, which emerged as an academic discipline especially in the 1990s [1] and was published by the International Organization for Standardization (ISO) in 2018. As stated in ISO 30401, knowledge management “includes optimizing the identification, creation, analysis, representation, distribution, and application of knowledge to create organizational value [2].”

Knowledge is classified differently across many approaches; however, the most widely accepted classification of knowledge is that of Polanyi [3], who classifies knowledge as either tacit (implicit) or explicit. Explicit knowledge is formal and structured and can be codified to be shared. Tacit knowledge is experiential, consisting of lessons learned while executing tasks or projects and insights gained from

continuous problem resolution [4].

Tacit knowledge is the information we carry inside us, in our brains. It is so ingrained in us that we sometimes do not even know that we have it. This makes it exceedingly difficult to share tacit knowledge. For example, we know that car drivers sometimes put themselves on “autopilot” and do not remember the last kilometers they drove. When we ask them how they did that action, they cannot answer. These situations happen to all of us often. We do things without knowing how we do them and have a challenging time explaining the phenomena to someone else. The collective power created by the sum of implicit information is extremely valuable for an organization.

Explicit knowledge is what we can express in words, pictures, documents, or other means. Therefore, to share our knowledge, we must first make it explicit. Let us expand on this thought by circling back to the example about driving. The most awkward thing about teaching someone to drive is that we assume that person knows a lot of things we do. We are surprised when they do not understand what we are saying. It can be difficult to translate all our knowledge of driving into a form that is clear enough to share, that is, in words, text, or pictures [5].

Whether it is tacit or explicit, for organizations, the aim of KM is to generate knowledge from information and convert it into a competitive advantage. Today, ISO 30401:2018 is the main standard that supports an organization's ability to develop a management system that effectively promotes and enables value-creation through knowledge.



## Knowledge Management Systems Standard: ISO 30401:2018

ISO 30401, first published in November 2018, starts with an introduction stating the purpose, importance, and range of KM that is followed by guiding principles and a summary. There are 10 main sections—in line with the high-level structure framework of ISO[6]—followed by three annexes for informative purposes and a bibliography. Here is a brief overview of these contents:

**Section 1 – Scope:** The scope of this standard sets requirements and provides guidelines for establishing, implementing, maintaining, reviewing, and improving an effective management system for KM in organizations.

**Section 2 – Normative References:** This publication has no normative references.

**Section 3 – Terms and Definitions:** There are 30 terms defined briefly in this section to give guidance to users.

**Section 4 – Context of Organization:** This section mandates that an organization must understand its context. An organization shall determine external and internal issues that may affect its knowledge management system (KMS) in addition to needs and expectations of interested parties in its business environment. This analysis helps organizations determine the scope of a KMS, in other words “knowledge domains,” which must be available as documented information. The standard gives three requirements as independent dimensions of a KMS, which include knowledge development, conveyance and transformation, and KM enablers.

According to the standard, the development of knowledge has a life cycle starting with acquiring new knowledge, applying and retaining current knowledge, and handling outdated or invalid knowledge. The main purpose of these activities, along with examples, is shown in Figure 1.

The standard emphasizes that a KMS shall include activities and behaviors supporting different types of knowledge flows such as human interaction, representation, combination, internalization, and learning with given examples. Also, enablers that support KMS objectives are defined as human capital, processes, technology and infrastructure, governance, and KM culture.

Complying with the requirements stated in this section and using the guidelines are the basis of successfully implementing further steps of a KMS.

**Section 5 – Leadership:** This section mandates that top management members within an organization develop, document, and communicate a KMS policy within said organization as well as with interested parties. Demonstrating such commitment is possible by making resources available, directing and leading persons to contribute to the effectiveness of a KMS, and managing changes. For this purpose, organizational roles must be clearly defined with responsibilities, authorities, and competencies per role.

**Section 6 – Planning:** Organizations shall consider the consequences of risks and benefits of opportunities based on context and plan actions to address them. The standard also mandates organizations to determine, document, and communicate objectives that align with the KMS policy. To achieve these objectives, organizations shall have action plans that include a certain time frame, designated responsibilities, and an evaluation methodology of the results.

There are many reasons to implement a KMS (e.g., enhanced communication, better process results, higher profitability, targeted marketing). According to one source, risks can be caused by three barriers: individual, organizational, and technology [7]. Another classifies risks as either human, technological, or operational knowledge risks [8].

**Section 7 – Support:** Organizations shall consider



Figure 1. Knowledge Development According to ISO 30401

resource needs and meet them to achieve their KMS objectives. These resources may include infrastructure, technology, communication, competence, awareness, and documented information. The standard stresses and mandates documented information as evidence of competence. Research [9] shows that organizational barriers are the most prohibitive to successfully implementing a KMS. Therefore, organizations shall design their programs to ensure all required competencies are met or updated and relevant elements are communicated internally (or with external parties), and include activities to raise organizational awareness.

**Section 8 – Operation:** Organizations shall determine and plan KMS processes—including outsourced processes—to meet the requirements of this standard. During implementation, these processes must be controlled according to the established criteria. Documented information is mandatory to ensure all processes are conducted as planned. The knowledge development process given in Figure 1 may be used as a base model. Similarly, according to NASA, the KM process is “the architecture used to acquire and benefit from knowledge resources and capabilities [10].” Considering its business processes, an organization may ask questions to design its KMS process, such as how and from which sources knowledge may be acquired, which is the best strategy to capture knowledge, and what methods will be applied throughout the life cycle. Most importantly, every process must have a goal and envisioned benefit.

**Section 9 – Performance Evaluation:** This section consists of three subtitles. First, organizations must identify, monitor, measure, analyze, and evaluate performance indicators and metrics and then document the results. Second, organizations shall conduct planned internal audits to measure conformance levels in accordance with the standard’s guidelines as well as organizational requirements. The audit program and results must also be documented. Last, management personnel should regularly review the effectiveness of a KMS and document the results.

**Section 10 – Improvement:** Organizations shall have a methodology in place to address nonconformities with root causes and corrective actions as well as strategies for continuous improvement. The standard mandates

documented information for the evaluation of corrective actions.

**Supplemental Materials:** Annex A gives brief information on the range of KM, where Annex B informs on the relationship between this range and related disciplines. Finally, Annex C introduces aspects of KM culture in the organizational culture

## KMS and Other ISO Standards

Although ISO 30401:2018 is the only standard focused on KMS, there are two main standards that relate to knowledge management. In 2011, the technical committee (ISO/TC176/SC2) responsible for the “ISO 9001 Quality Management Systems – Requirements” standard conducted a worldwide survey that revealed a demand to include a KM requirement. The next ISO 9001 update, published in September 2015, included knowledge as a resource requirement, stated in clause 7.1.6 as follows:

“The organization shall determine the knowledge necessary for the operation of its processes and to achieve conformity of products and services. This knowledge shall be maintained and be made available to the extent necessary. When addressing changing needs and trends, the organization shall consider its current knowledge and determine how to acquire or access any necessary additional knowledge and required updates [11].”

To meet this requirement, organizations can either choose to implement ISO 30401 requirements as a whole or use its guidance to integrate knowledge management elements into a ISO 9001 quality management system. This integration should cover process approach, plan-do-check-act cycle (PDCA), and risk-based thinking. Creating a knowledge map by linking an organization’s products and services to identified knowledge categories and linking to resources may be a good starting point.

A simple model may be used for a specific product, service, or organizational activity as illustrated in Figure 2. With the direction of top management, the organization takes considerations into account related to need or update for knowledge, which may be identified through a variety of sources. Once identified, an author prepares all materials (i.e., documents, audio-video sources, web or software applications, etc.) to share available knowledge with interested parties. A review—based on four key principles—is required prior to approval, publication, and communication by organizational

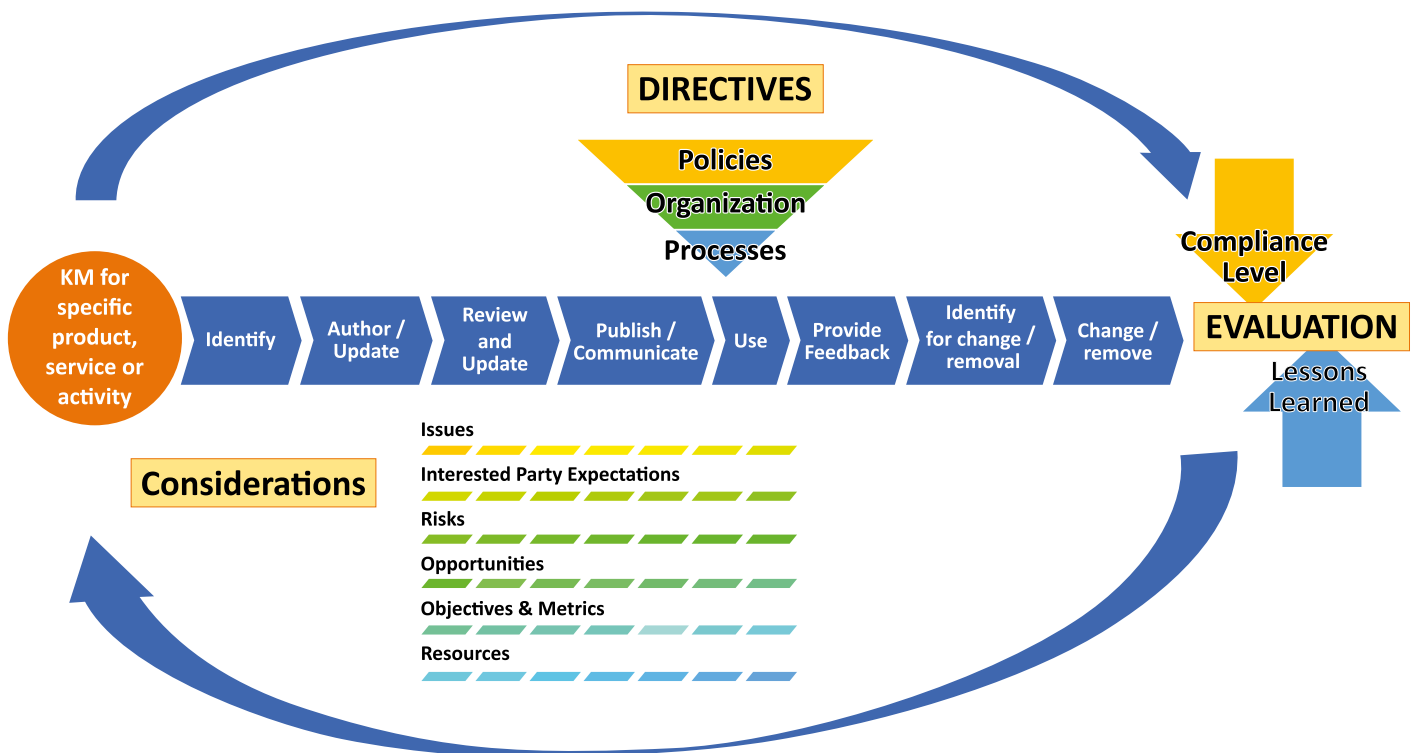


Figure 2. Model for integrating KMS to ISO 9001:2015

representatives. The users of this knowledge may identify areas requiring correction or improvement and provide feedback so that organizational representatives can make any necessary updates. Also, with performance evaluations, the organization has the opportunity to acquire new knowledge. This cycle can be applied to all relevant products, services, or organizational activities if appropriate and applicable.

Another publication, "ISO/IEC 27001:2013 Information Technology – Security Techniques – Information Security Management Systems – Requirements" (ISMS) [12], also has connections with KM. Since data evolves into information and then into knowledge, one must consider the security of information, and therefore knowledge. This standard is based on CIA triad model that includes confidentiality, integrity, and availability. According to ISO 27000:2018 [13]—which provides an overview of ISMS along with relevant vocabulary terms—*confidentiality* requires that information is not made available or disclosed to unauthorized sources (in other words, only authorized individuals or systems can view information). *Integrity* alludes to the accuracy and completeness of information, implying it has not been intercepted or manipulated. *Availability* means information is accessible and

usable on-demand (e.g., a database is available to those who have access privileges). The lack of any of these attributes can result in commercial harm, business damage, or reputation loss.

Another view is that KMS and ISMS applications have a couple of intriguing similarities [14]. Both management systems are dependent on people and both are aimed at the production of public goods [15]. Another similarity is the positive effects of knowledge that is exclusive to a specific organization (e.g., organization-wide promises of higher benefits). ISMS acts as a preventive tool by applying the CIA's triad model—*confidentiality* decreases the risk of knowledge being shared with rivals by assuring exclusiveness, *integrity* safeguards knowledge from manipulation, and *availability* ensures knowledge is accessed on a need-to-know basis. Considering these points, complying with requirements and applying relevant controls given in Annex A of the ISMS standard helps organizations secure their information more effectively.

Other publications such as "ISO 55001:2014 Asset Management – Management Systems – Requirements" [16] or "ISO/TR 13054:2012 Knowledge Management of Health Information Standards" refer to knowledge in the body of

documents. However, further research is required using keywords to make a better query from standard publishers' databases, and is not limited to ISO.

## Conclusion

Due to the rapid development of communication technologies and increasing flow of information, filtering data to acquire usable knowledge is more challenging day by day. Knowledge management is essential for the governance of organizations as well as for their resilience and continuity. International standards are vital tools that can be leveraged for these purposes.

Regarding individual organizational activities, applying micro PDCA cycles based on knowledge management concepts has the potential to increase maturity in management system applications. Furthermore, organizations implementing adequate and quality knowledge management systems will likely have fewer problems and disruptions and be able to adapt to changes more easily. Those that apply mature knowledge management system practices have the opportunity to distinguish themselves as industry leaders.

Looking to the future, discussions of knowledge and knowledge management practices are expected to appear in more standards or similar publications. It is crucial for organizations to proactively devise ways to eliminate disinformation and unnecessary information, as the remaining knowledge is the main resource for decision-making and designing future activities. Ultimately, the degree at which knowledge management systems are implemented—guided by the above-mentioned standards—will be key in determining how organizations rank in their respective markets.

## Author Bio

Tolga Aktaş is a management system lead auditor working as a freelancer on behalf of international accredited assessment companies. He also conducts trainings and workshops on management systems including lead auditor training. He obtained his MBA in management in Turkey. His professional expertise encompasses information security, asset and knowledge management, and business continuity and he provides consultancy services for multinational companies.

## References

- [1] Nonaka, I. 1991. "The knowledge creating company," Harvard Business Review, November-December 1991, 96–104.
- [2] ISO. n.d. "ISO 30401:2018 Knowledge management systems – Requirements." <https://www.iso.org/standard/68683.html>.
- [3] Polanyi, M. 1958. "Personal knowledge: Towards a post-critical philosophy," University of Chicago Press.
- [4] North, K. and Kumta G. 2018. Knowledge Management. Springer Texts in Business and Economics. <https://doi.org/10.1007/978-3-319-59978-6>.
- [5] Barutçugil, İ. 2002. Bilgi Yönetimi. Kariyer Yayınları, 92-93.
- [6] ISO. 2021. "ISO/IEC Directives, Part 1: Procedures for the technical work." <https://www.iso.org/sites/directives/current/part1/index.xhtml>.
- [7] Ruzic-Dimitrijevic, L. 2014. "Risk assessment of knowledge management system." Online Journal of Applied Knowledge Management 3, no. 2: 114-126.
- [8] Durst, S. and Zieba, M. 2018. "Mapping knowledge risks: towards a better understanding of knowledge management." Knowledge Management Research & Practice 17, no. 1: 1-13.
- [9] Knoco Limited. 2020. "Global Survey of Knowledge Management." <https://www.knoco.com/knowledge-management-survey.htm>.
- [10] NASA. 2015. "Knowledge Management Process." <https://appel.nasa.gov/wp-content/uploads/2015/11/Knowledge-Management-Process.pdf>.
- [11] ISO. n.d. "ISO 9001:2015 Quality management systems – Requirements." <https://www.iso.org/standard/62085.html>.
- [12] ISO. n.d. "ISO/IEC 27001:2013 Information technology – Security techniques – Information security management systems – Requirements." <https://www.iso.org/standard/54534.html>.
- [13] ISO. n.d. "ISO/IEC 27000:2018 Information technology – Security techniques – Information security management systems – Overview and vocabulary." <https://www.iso.org/standard/73906.html>.
- [14] Glaser, T. and Pallas, F. 2007. "Information Security and Knowledge Management: Solutions Through Analogies?" <http://dx.doi.org/10.2139/ssrn.1014302>.
- [15] Stanford University. n.d. "Public Goods." Stanford Encyclopedia of Philosophy. <https://plato.stanford.edu/entries/public-goods/> for definition and reference.
- [16] ISO. n.d. "ISO 55001:2014 Asset management – Management systems – Requirements." <https://www.iso.org/standard/55089.html>.



# Section B

an informative section that includes general articles,  
announcements, white papers, etc.

# Historical Understandings of International Conformity Assessment Systems

By William T. Fiske, Director of Technical Affairs, Intertek Testing Services

## Abstract

The nature of conformity assessment (i.e., testing, inspection, and certification) has evolved throughout the 20th and 21st centuries. This article reviews the major changes that have taken place over the more recent half-century, the current situation, and ways in which those changes have led to safer, higher-quality, and less costly electrical and electronic equipment and systems. The author concludes that further internationalization of standards and conformity assessment is both necessary and desirable.

*Keywords: certification, conformity assessment, International Electrotechnical Commission, standards development, testing*

## Introduction

For many decades, product safety certification (listing) was strictly a national matter. Whether by law or by tradition, market access depended on certification by an organization recognized in each country, and in most countries, only one such organization existed. This was duplicative and inefficient.

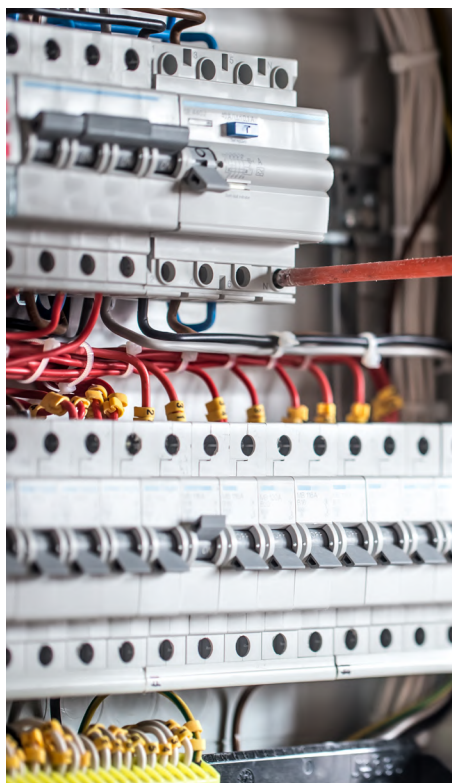
In 1985, the International Electrotechnical Commission (IEC), a global standards organization founded in 1906, entered into an agreement with a loosely associated group of European testing and certification bodies (the CEE) to create a worldwide organization that facilitates trade in electrical and electronic equipment by eliminating duplicate testing in member countries. The new organization was called the IEC System for Conformity Assessment Schemes, for Electrotechnical Equipment and Components (IECEE).

Today, there are four IEC conformity assessment systems. Besides the IECEE, there are IECEx, for equipment used in hazardous (classified) locations, IECQ, a quality assessment system for electronic components, and IECRE, for renewable (marine, solar, wind) energy generating systems. This article explains the basic rules, similarities and differences, and the IEC mechanism for governing them.

Until the early 1980s, every country had its own requirement for national testing and certification of electrical equipment for safety. The cost to manufacturers, not only in direct fees charged by the conformity assessment (testing, inspection, and certification) bodies, but also in time of administrative and technical tasks involved in dealing with the conformity assessment bodies, was considerable, and to some extent, wasteful.

In much of the world, including nearly all European countries, the national certification body was a government entity (the ultimate monopoly). Even in Canada, the United Kingdom, and the United States, where private companies performed the conformity assessment activities, those bodies were de facto monopolies—CSA, BSI, and UL respectively.

It is worth noting that in those three countries, the sole conformity assessment bodies also served as the primary standards-developing organizations (SDOs) for safety of electrical equipment. In essence, those organizations developed, interpreted, and enforced the requirements. Hence, the relationship between manufacturers and conformity assessment bodies was unbalanced.



## Gradual Decentralization Reduces Monopoly Power

In continental Europe, the situation slowly began to improve. In most countries, there had traditionally been home-grown standards as well as national conformity assessment bodies in each country.

The first shift toward a multilateral approach was a pan-European agreement for each country to adopt applicable international standards as national standards, with national differences. The newly harmonized standards were called European norms (ENs). Standards based on standards of the International Organization for Standardization (ISO) were designated as CEN standards, and those based on International Electrotechnical Commission (IEC) standards were designated as CENELEC standards.

In North America, CSA standards and UL standards were, for the most part, technically equivalent; however, the differences between them often made it impossible for a manufacturer to design a product that would conform to both countries' standards without making at least some changes.

Not satisfied with the harmonization or near-harmonization of standards, manufacturers continued to apply pressure on another source of inefficiency, that being the conformity assessment schemes. Requirements for testing in each country meant products were subjected to multiple rounds of testing and certification that provided no added value to the producers or the users. In addition, the national conformity assessment bodies, having no

real competition, had become imperious and inefficient.

Finally taking heed of the complaints, the electrical equipment certifiers in Europe joined the European Commission for Conformance Certification of Electrical Equipment (CEE), an agreement among the bodies to accept one another's test results. This agreement mostly eliminated the duplicative testing in Europe.

Progress in North American certification was also lagging. The Occupational Safety and Health Act of 1970 which created the Occupational Safety and Health Administration (OSHA), put OSHA in charge of workplace safety. This empowered OSHA to set standards for electrical safety as well as other potential workplace hazards. The OSHA general industry standards (29CFR Part 1910 – subpart for electrical safety) named only UL Listed and FM Approvals as permitted sources of product safety certification. This de jure duopoly remained until 1984, when a small private testing company sued the U.S. Department of Labor to open the testing and certification market to other qualified laboratories. Testing and certification bodies accredited by OSHA are known as Nationally Recognized Testing Laboratories (NRTLs).

## The IEC Steps Up as an Advocate of Universal Standards

Since its inception in 1906, IEC has served as a standards body, with only a tenuous connection to product conformity. Looking to grow, the Geneva-based organization worked out

an agreement with CEE in 1985 to globalize conformity assessments of electrical equipment, based on IEC product standards.

This development led to the establishment of the IEC System for Conformity Assessment Schemes, for Electrotechnical Equipment and Components (IECEE CB Scheme). To this day, use of IEC standards is a fundamental requirement of the CB Scheme. Although many countries around the world had adopted IEC standards by 1985, the United States was not among them, so no U.S. body could participate in the system. The situation was similar in Canada, although the latter had begun to harmonize IEC standards as CSA standards. Eventually, the SDOs in both Canada and the U.S. did move forward on harmonizing IEC standards (the U.S. has harmonized far fewer than Canada). With that, their testing laboratories and certification bodies became eligible to join the CB Scheme's mutual recognition process. The U.S. member body, USNC/IECEE, became a CB Scheme member in 1992.

## Other IEC Conformity Assessment Systems Used for Compliance Verification

As time went on, the CB Scheme's success drew the attention of other sectors of the electrical industry, and those sectors formed similar systems for the mutual acceptance of conformity assessment results. Besides the IECEE-CB Scheme, these are:

- International Electrotechnical Commission System for Certification to Standards Relating to

Equipment for Use in Explosive Atmospheres (IECEx System)

- International Electrotechnical Commission Quality Assessment System for Electronic Components (IECQ)
- System for Certification to Standards Relating to Equipment for Use in Renewable Energy Applications (IECRE)

## Breakdown of the IEC CA Systems' Common Elements and Governance

Although each conformity assessment (CA) system has unique elements, there are some aspects common throughout. The most important common elements are as follows.

1. Stakeholder participation: Every national member body must include stakeholder groups, such as regulatory bodies, manufacturers, users, installation and maintenance groups, and conformity assessment bodies, to avoid dominance by any interest group.
2. Use of international standards: While each CA system decides which specific standards will be used, all are IEC or ISO (International Organization for Standardization) standards
3. Peer assessment: All testing laboratories and certification bodies are assessed for competence and quality control by experts from other CA bodies within the system. Peer assessment may well be the feature that makes the IEC conformity assessment systems strong. Each body is assessed by its competitors (what could be more rigorous?), and the

assessment reports are reviewed by all the member bodies before the CA body is accepted into a system. All the systems also feature periodic reassessment

4. Mutual acceptance of results: All the certification bodies in a system agree to accept each other's test reports and certificates of conformity as processed through the system, in issuing their own national certifications. The systems permit minimal verification testing to validate certified equipment; however, that cannot approach anything like a full product evaluation. Regulators accept national certification without favoring testing done in the home country over that done by other participants in the CA system.

All four CA systems are governed by the IEC Conformity Assessment Board (CAB). The CAB sets the basic rules for all the systems, oversees their finances, and renders decisions on issues brought to CAB by the various systems' management committees.

IEC CAB is a group whose members are elected by the four systems' national members. CAB issues documents that apply to all the CA systems on conformity assessment and peer assessment. While the CAB issues basic rules, each system has one or more rules and operating procedures of its own.

## IECEE

The IECEE-CB Scheme employs a very large number of IEC standards, issued by many technical committees. Since there are so many standards in use,

they are organized into categories covering similar products, such as test, measurement, control and laboratory equipment (category MEAS) or household and similar equipment (category HOUS).

The CB Scheme is an ISO Type 1a certification scheme, in that it only covers initial testing (type testing). A certification body that accepts CB test certificates must conduct factory surveillance (follow-up service) if its own certification scheme requires continuing surveillance of certified equipment, and nearly all of them do.

## IECEx

The IECEx system not only covers Ex equipment (equipment intended for use in explosive atmospheres), called the Certified Equipment Scheme, but also certification of personal competence in activities related to explosive atmospheres, such as area classification, installation, and maintenance. There is an IECE scheme for certification of shops performing repair and overhaul of Ex equipment. IECEx is the only IEC CA system that offers a system mark of conformity, although demand for the IECEx mark has been very small to date.

IECEx uses only standards developed by IEC TC 31 and its subcommittees, and there is a permanent liaison between the IECEx Management Committee and TC 31.

The IECEx certified equipment scheme is an ISO Type 5 certification scheme, in which the manufacturer's factory quality system is also under the surveillance of an Ex certification body.



## IECQ

IECQ is a supply chain management scheme that mainly covers quality systems throughout the production and distribution of electronic components. There is particular emphasis on preventing counterfeit components from getting into the stream of commerce. IECQ has conformity assessment schemes for the following:

- Approved processes
- Approved components, products, related materials and assemblies
- ADHP scheme for aerospace, defense, and high-performance (ADHP) component management
- Hazardous substance process management
- Counterfeit avoidance program
- LED lighting

The IECQ system mostly operates out of the public view, as users of the system are nearly always original equipment manufacturers and repair shops.

## IECRE

The IECRE is a certification system for electrical generation equipment powered by renewable energy sources. The system comprises three working groups: Marine Energy, Solar PV Energy, and Wind Energy, all reporting to the Renewable Energy Management Committee (REMC).

The Marine Energy scheme is the newest, and perhaps most interesting, sector of IECRE. It uses only standards issued by IEC TC 114. These standards relate to the conversion of moving

water—wave, tidal and other water current energy—to electrical energy. The sector does not deal with traditional hydroelectric generation and standards developed by TC 4, as of this writing.

For comparison, the Solar PV Energy and Wind Energy sectors use only standards developed by IEC TC 82 and TC 88, respectively.

## Growth of the IEC Certification Systems

Growth is a fairly good indicator of success in any operation, and the IEC CA systems have grown very well.

Today, the IECEE CB Scheme has grown to 54 national member bodies, housing 92 national certification bodies (NCBs) and 533 CB testing laboratories. In 1993, there were 3,501 CB test certificates issued, and 2,673 certificates recognized by other NCBs for certification. For comparison, in 2019, 111,836 certificates were issued and 34,818 recognized.

The IECEx system has grown to 36 national member bodies, housing 60 certification bodies, 34 recognized training providers for personal competence, and 68 Ex testing laboratories. There were only 258 IECEx Certificates of Conformity issued in 2005. In 2019, that number had grown to 4,554.

IECQ, a fairly new system, has 12 national member bodies with 28 certification bodies. The system had issued 9,623 certificates of conformity at the end of 2021.

IECRE, the newest member of the IEC conformity assessment family, has 15 national member bodies. There are 13 certification

bodies, 32 testing laboratories and four inspection bodies. In 2020, 119 IECRE test certificates were issued, including the very first Marine Energy certificate. The IEC conformity assessment systems' success has been recognized by OSHA, which is not known for being open to new approaches to equipment approvals. In 1995, OSHA issued a directive that allows NRTLs to accept CB test reports from other certification bodies in IECEE. That directive specifically prohibited NRTLs from accepting IECEx certificates for hazardous (classified) locations equipment. In 2019, OSHA finally allowed the NRTLs to recognize IECEx test reports, as it had been allowing CB test reports during the preceding 24 years.

Harmonized Standards and Reciprocal Acceptance of Testing and Certification Enhances Product Safety and Quality The IEC conformity assessment systems, created for the specific purpose of facilitating international trade in electrical and electronic equipment and components, have performed as intended.

The manner in which these systems operate has considerably reduced repeated equipment testing across multiple jurisdictions. As a result, the total amount of needless spending by manufacturers, whose costs are ultimately passed down to their customers, cannot be measured reliably, but it is undoubtedly very large in aggregate. In addition to the direct financial gains, the shortened time to market for goods and services benefits the public, which today has more abundant, safe, high-quality product choices.

## Author Bio

William T. "Bill" Fiske serves as the director of technical affairs at Intertek Testing Services in Cortland, New York, where he has been employed since 1977. He received a Bachelor of Science in electrical engineering from Rensselaer Polytechnic Institute and is a registered professional engineer in Louisiana, New York, and Texas. He is a Life Senior Member of IEEE and a member of ASQ, NFPA, and SES.

Bill participates heavily in standards development, being a member of the National Electrical Code (NEC) Correlating Committee and NEC Panel 14 (hazardous locations), and chairs the NFPA Committee on Electrical Equipment in Chemical Atmospheres. In addition, he serves on the USNC/IEC Technical Management Committee, U.S. National Committee of IECEx, IEC TC 31 Maintenance Team 60079-2, and many UL Standards Technical Panels.

Bill received the 1906 Award from the International Electrotechnical Commission (IEC) in 2012, the ANSI Meritorious Service Award in 2014, and the NFPA Committee Service Award in 2020..



# ACT

**ACT Lab** is an ISO/IEC 17025 accredited third-party testing laboratory with a global presence that focuses on certifying market-leading products in the active lifestyle industry.

# Leveraging Quality Management Systems to Prevent Product Liabilities

By Gilbert Gong, Ph.D., President of Institute of Global Certification (IGC) and  
President of Global Personnel Certification Co., Ltd. (GPC)

## Abstract

Manufacturing companies should consider the liabilities for their products because the Product Liability Act is recognized as a global standard and, as such, is an inevitable reality in our industrial world. Accidents, incidents, and failures of products will inevitably occur in the system of mass production, mass distribution, and mass consumption. However, companies should not have to be afraid of incidents developing into lawsuits when distributing products to the market.

The premise of this paper is to share how proactive companies can prepare countermeasures to prevent product liability accidents from happening and/or effectively address any accidents, incidents, or product failures that do occur. At the management level, the top priority should be to prepare Product Liability Prevention (PLP) and Product Liability Defense (PLD) systems.

Implementing an effective quality management can be a good control method to prevent and defend against product liabilities. To maximize the quality management system's capabilities, an organization should be restructured and given more opportunities for training as well as expanded manpower. (All activity results should be recorded for evidentiary purposes in case of a lawsuit.) To enhance product control measures, a testing and measurement system should be introduced under the concept of state-of-the-art technology. If a company cannot handle the testing and measurement of products in-house, it can enlist the help of an external organization.

With effective prevention and defense systems in place for product liability matters, manufacturing organizations stand a better chance of being successful and long-lasting in the global marketplace.

*Keywords: Product Liability Act, product liabilities, quality management system, small and midsize enterprises, product liability litigation, liability response, countermeasures, ISO 9001*

## Introduction

The Product Liability Act is recognized as a global standard and its implementation is an inevitable reality in our industrial world. For companies, product liability accidents should not happen, but if they do, necessary countermeasures should be in place. Leveraging quality management systems should be a top priority to prevent product liabilities.

### 1. Preparing Countermeasures for the Enforcement of the Product Liability Act for Small and Midsize Enterprises.

In the system of mass production, mass distribution, and mass consumption, accidents inevitably occur. There is no way to prevent lawsuits from being filed after such incidents, and companies should not be afraid of dealing with lawsuits when distributing products to the market. The important thing is to prepare comprehensive defensive measures that can help win legal trials. When it comes to product safety best practices, instead of focusing on crafting a product liability defense, small and midsize enterprises should do their utmost to avoid distributing defective products.

If a large compensation lawsuit is filed, and since the manufacturer has already been sued, what a company can do in a trial is quite limited. Since the legal response is a specialized and difficult procedure, it is more efficient to use lawyers with ample experience in product liability litigation in order to legally evaluate the company's position and reflect it in the trial. However, although there is much dependence on lawyers, it is the company—not the lawyer—that makes the final decision in various situations in the trial. In addition, promoting litigation defense measures as part of establishing preventive countermeasures on product liability is very helpful to companies. Therefore, it is most important not only to acquire legal knowledge or accumulate experience for litigation defense, but also to identify matters that companies must implement in relation to product safety through discussions and precedents in court and include them in their product safety activities.



If a lawsuit is filed, it is better to resolve it through settlement rather than a judgment in court; however, unconditional settlements may cause serious damage to the company. Therefore, the enterprise must go on trial with a firm attitude, if necessary, after careful analysis and review of the case. As a result of the company's review, if there appears to be no possibility of winning the trial, it is desirable to settle at an early stage during a lawsuit.

Even if a product liability accident occurs, product liability will not be interrogated if the company actively proves there is no product defect. This is because the Product Liability Act holds the manufacturer liable for damages based on the defect in the product. Proving there are no defects results in the product's safety measures preventing defects in the product's finality. Therefore, the company's product liability prevention plan should not only focus on solving product accidents that have already occurred, but also strive to provide safe products to users (consumers) in preparation of the following topics for the future.

#### ♦ **Improvement of the Accident Information Collection System**

To prevent product liability accidents and achieve a fair and smooth resolution of disputes by narrowing the gap in product awareness and information between consumers and manufacturers, it is necessary to establish a comprehensive management system. This helps with improving safety systems and responding to product liability accidents by introducing and implementing a system for the collection and analysis of product liability accident-related information.

#### ♦ **Maintenance of the Agency that Finds the Cause of the Defect**

Even if the liability requirements are changed from negligence to defects due to the introduction of the Product Liability Act, the victim is still responsible for proving the causal relationships between the existence of defects and damages, which are the biggest issues in product liability disputes.

Scientific and technical analyses of the causes of the accidents are needed to respond to the victim's claim basis for compensation and the prevention of recurring defects. In the use of these cause-finding institutions, above all, support for reducing the cost burden and easy directions for use are required.

In particular, the maintenance of the cause investigation agency can reduce the time and cost of disputes by preventing the abuse of lawsuits by unclear claims and can greatly contribute to preventing the recurrence of accidents that reveal the cause of accidents to manufacturers.

#### ♦ **Improvement of the Non-Judicial Dispute Settlement System**

Taking the introduction of the Product Liability Act, it is necessary to establish a simple and rapid dispute settlement system managed by third parties.

It can minimize the manufacturer's easy damage relief and loss costs by diversifying dispute resolution procedures, leading to quick and easy resolution in preparation for the increase of damage-relief claims incidents.

Product liability disputes may also result in significant monetary damages due to life accidents, but minor physical injuries or small amounts of property damages are expected to account for the majority of the disputes; hence, the current civil trials have many problems considering the time and economic burdens required by the lawsuits.

In addition, if it is difficult to resolve problems through negotiations between consumers and manufacturers, then going directly to trial becomes a concern about the social and economic costs and losses such as excessive costs and the waste of time due to delays in litigations.

#### ♦ **A Proactive Approach to Securing Measures of Compensation**

If the Product Liability Act is introduced and the obligation to compensate for damages due to an unexpected product accident is imposed, this presents a serious business risk to manufacturers that fail to take precautions to perform compensation. Therefore, it is necessary to ensure fullness by organizing systems that can secure various compensation implementation measures such as insurances, guarantees, deposits, deduction systems, mutual company establishments, self-insurances, and mark systems.

#### ♦ **Response to Change Trends in Safety Regulations**

Until now, pre-regulations have been emphasized to promote consumer safety. The introduction of the Product Liability Act is expected to strengthen



post-compensation measures and future changes may promote consumer safeties under autonomous corporate responsibilities.

However, safety regulations are being strengthened on an international level according to the preventive aspects of accidents and the internationalization of product-related standards. To defend development risks, manufacturers are required to respond in a timely manner to changes in product safety-related standards and regulations

## **2. Potential Business Impacts of Product Liability Legislation**

### **(1) Strengthening of product stability**

It can be said that Product Liability Act legislation greatly contributes to strengthening the stability of a company's product.

In the process of developing, designing, planning, purchasing, manufacturing, inspecting, marking, and post-servicing products, whether established or not, various product liabilities such as the existence of product defects play an important role. Therefore, efforts to improve safety are inevitable because it is beneficial to make safe products rather than pay compensation for defective products.

Due to this influence, the product safety is gradually improved, which leads to an increase in sales and enhances the corporate image due to the purchasing preference of consumers who recognize this.

### **(2) Realization of customer satisfaction management (consumer protection)**

From a consumer's point of view, the Product Liability Act facilitates relief from product accident damages, leading to claims, trials, and disputes from consumers who are not compensated for damages due to difficulties in procedural evidence or proof of defect. This number is expected to continue to increase.

Accordingly, companies are expected to develop and design in preparation for disputes or lawsuits caused by defective products, and naturally implement corporate management systems based on the concept of consumer protection and customer satisfaction.

### **(3) Enhancement of business competitiveness**

From a company's point of view, product safety measures are an important concern for corporate management, which spurs competition for the

production and sale of safer products. Resultingly, as consumers purchase products based on factors such as product safety in addition to price and quality, companies that produce products in preparation for product liability naturally strengthen their competitiveness.

Companies should establish and implement product liability prevention measures and product safety measures to actively prevent product accidents, not simply from the perspective of product liability defense.

### **(4) Dispersion of company responsibilities**

Without the Product Liability Act, the company has no choice but to rely on the negligence liability (tort liability); however, when the Product Liability Act is enforced, proving product defects is much easier than proving negligence, so the company's responsibility becomes heavier. Nonetheless, companies can distribute the burden of accident costs through product liability insurance.

### **(5) Prevention of recurring accidents**

In today's consumer society, accidents caused by product safety are inevitable to some extent.

In the U.S., legal systems such as class action laws have led to surging product liability lawsuits and rising insurance premiums—and, consequently, the product liability crisis of corporate bankruptcy. Yet in the case of Korea, which has a similar legal system to Japan, lawsuits and disputes related to product liability are expected to be similar to Japan.

Due to such lawsuits or disputes, companies can take proactive and defensive measures on product safety and prevent the recurrence of accidents.

## **3. Defensive Countermeasures for Product Liability Accidents**

### **(1) The Importance of Product Liability Defense Countermeasures**

From the perspective of product liability accidents and companies' preventative product liability measures, it goes without saying that the pursuit of product safety is of paramount importance.

However, in the modern high-consumption society, even though companies pay maximum attention to prevent accidents, it is difficult to completely prevent product liability accidents. Therefore, it cannot be said that companies have sufficient product liability measures just by taking product liability defense measures.

In anticipation of potential product accidents, a system should be implemented to respond smoothly to claims and measures should be taken to minimize company losses caused by product liability accidents.

One example of a product liability defense measure is an in-house claim response system. If a proper claim response system is established, consumer complaints about product safety are appropriately fed back to the design and manufacturing departments. Building an in-house claim response system is very important because it benefits product liability prevention countermeasures, as well as various methods for developing new products.

While it is true that the customer's trust is temporarily lost when product accidents occur, once the appropriate claim measures are taken, it may be possible to restore trust as well as strengthen and improve customer relationships.

In modern society, where corporate responsibility is becoming more stringent, the attitude of corporations toward society is drawing attention in all aspects. In this sense, it is the responsibility of companies to improve their in-house systems with the perception that an appropriate, rapid claim response is the standard for indicating a company's social attitude.

## **(2) Pre-Accident Product Liability Defense Measures**

### **1) Establishment of an In-House Product Liability Response System**

Clarification of the authority and responsibility of the department in charge of product liability:

In Korea, after the enforcement of the Product Liability Act, there may be a situation in which lawsuits increase significantly like in the United States.

Under the conventional negligence liability, it would have been difficult to prove the requirements for liability for damages, so there have been many cases of giving up filing a lawsuit. After the enforcement of the Product Liability Act, however, a constant increase in the number of claims is inevitable as the liability requirements for damages are changed and eased to meet requirements for no-fault liability.

Therefore, companies should have systems that can more quickly respond to increasing claims by clarifying the authority and responsibility of the department in charge of product liability.

### **2) Matters to be Considered When Reorganizing the Organization**

There are various methods—such as the establishment of a dedicated organization or a temporary organization, or the use of an existing organization—to determine the type of organization in which a product liability department should operate.

It is necessary to review each option individually as this determination must factor in the size of the company, any present in-house organization, and expected number of claims, but it is better to consider the following matters when reorganizing an organization.

#### **Participation of the Technology Sector**

In product liability claims, technical or professional discussions about product safety will be developed frequently. Cooperation of staff related to product design, development, and manufacturing is essential for the operation of the product liability response organization.

Therefore, regardless of the form of the organization, experts in the technology sector should be included in the product liability team.

#### **Establishment of an Enterprise-wide System**

Even if a dedicated product liability organization is installed in-house, it is impossible for this organization to handle all product liability problems.

This is because they can respond effectively only if they obtain the cooperation of sales and after-sales service departments located throughout the country. In addition, there will be many claims from customers against related companies outside the organization, such as retailers and sales agents. To resolve claims smoothly, it is necessary to establish an enterprise-wide system linking related sectors across the country and parties outside the company.

Similarly, it is also necessary to review the placement of full-time persons in charge of activities such as the preparation of accident handling manuals, accident reporting forms, and the education and training of internal and external related systems.

#### **Institution of a Reporting System to Top Management**

A sufficient understanding of a company's top

management team is necessary to promote product safety measures effectively, and the same is true for a smooth claim response.

In particular, in the cases of serious Product Liability claims that developed through litigation, a quick decision by the Top management is required to solve the problem, and a reporting system must be established to receive an immediate judgment from the Top management if necessary.

#### **Support from Outside Experts**

As claim responses progresses, it is desirable to establish a system that can easily obtain support from outside experts, such as lawyers, erudite experienced persons, engineers belonging to public institutions outside the company, or in cases of insurance involvement, staff of insurance companies.

#### **4. Preventive Countermeasures Using the Design Management System of the Quality Management System**

If there is a problem with the design itself, all products will become defective products, and a large amount of liability for damages may occur due to a number of product liability accidents.

Therefore, product safety measures are extremely important during design and development stages. Above all, companies should first try to supply defect-free and accident-free products to the market through safety-conscious designs.

Concretely, it is important to meet various safety standards and rules required at the product design and development stages, and to predict product risk in any way and devise accident prevention measures. Product safety should be reviewed and set at the development and design stages, and the product should be manufactured after it is circulated for each process—such as raw material purchase, outsourcing, and manufacturing—reflecting all design specifications, design drawings, and order specifications.

##### **(1) Design and Development Planning**

Organizations should plan and manage all facets of product design and development. During the planning period for these phases, the organization should determine the following.

1. Stages of the design and development process
2. Appropriate review, verification, and feasibility

verification activities for the requirements of each design and development stage

3. Responsibility and authority for design and development activities

To ensure effective communication and clarity of responsibility, organizations must manage the linkages between the different groups participating in design and development. If applicable, the planning output should be updated according to the design and development progress..

##### **• Management Point**

1) General matters included in the design and development plan:

1. Identification of customer requirements and the market's request of quality identifications when reviewing contracts
2. Design and development feasibility assessment
3. Design expense and product cost price
4. Development schedule and design review, design verification, design verify validity plan
5. Market evaluation
6. Measurement, test method, and acceptance criteria of the product
7. Selection of the designers, reviewers, and verification and validation representatives.
8. Design and update of development plan (or devise renewal regulations)
9. Design input/output documentation and storage methods

2) Organizational and Technical Connectivity

1. Organizational connectivity: Participation in related departments when using manufacturing lines for factory experiments, supporting model production personnel, and reviewing designs.
2. Technical connectivity: Interference in functions, physical properties, and appearance's mutual interference; introduction of new technologies; discovery of new materials; training of service personnel, etc.
3. Establishment of Connectivity
  - Information obtained and delivered
  - Clarification of sending and receiving groups
  - Clarification of the purpose and method of

information delivery

- Delivery of documents and preservation of records

## **(2) Design and Development Input**

Inputs related to product requirements must be determined and records maintained. This should include the following:

1. Function and performance requirements
2. Applied legal and regulatory requirements
3. Information derived from previous similar designs, if applicable.
4. Essentials of other requirements for design and development

The adequacy of these inputs should be reviewed. Also, the requirements should be complete, clear, and not conflict with other requirements.

## **(3) Design and Development Output**

The output of the design and development process shall be provided in a verifiable form compared to the design and development inputs and shall be approved prior to distribution. The design and development output should be as follows:

1. Meet design and development input requirements
2. Provide appropriate information for purchase, production, and service provisions.
3. Include or cite criteria for product acceptance judgment.
4. Determine the characteristics of products that are essential for safe and right use.

## **(4) Design and Development Review**

At an appropriate stage, a systematic review of design and development should be conducted for the following:

1. Evaluation of the ability of design and development results to meet requirements.
2. Recognition of problems and suggestion of necessary measures.

The people participating in such reviews should include those representing functions related to the design and development stages of the subject that is under review. Records of the review, and results of measures caused by the review, should be maintained (see 4.2.4).

## **(5) Design and Development Verification**

Verification must be performed to guarantee that design and development output meets input requirements.

Records of verification and all required activity results should be maintained.

## **(6) Design and Development Validity**

To ensure whether the finished product can meet the requirements for the prescribed or intended use or application, validation of the design and development must be performed in accordance with the planned method.

If applicable, validation must be completed prior to delivery or execution of the product.

Records of the validation results and the results of necessary measures should be maintained.

## **(7) Management of Design and Development Changes**

Changes in design and development should be identified and records maintained. If suitable, verifying changes and their validity should be approved prior to implementation, if applicable.

The review of design and development changes involves evaluating changes in components, and their impacts on delivered products.

Records of reviewing the changes and their results, as well as any measures, should be maintained.

# **5. Manufacturing Process Management**

## **(1) Manufacturing Process Management Plan**

Regarding product liability prevention measures, managing the manufacturing process helps to ensure that safety items considered in design and development stages can be sufficiently reflected when manufacturing.

Workers' workability, normal operation of facilities, and reliability of inspection equipment are directly related to defects that may occur during the manufacturing process, and it is important to establish and continuously manage standardization and management measures for them.

Solutions for defects that may occur during the manufacturing process are as follows:

- 1) Safety Management Concept Input for Each Process

It is necessary to establish a management system



for machinery or production facilities.

Also, a system for securing product safety by linking product certification is required. In particular, this includes system certifications such as ISO 9001, ISO 13485 (medical devices), GMP (pharmaceuticals, cosmetics), HACCP (food, beverage), TL-9000 (telecommunications equipment), IATF-16949 (automotive), AS-9000 (airplane), etc. and CE mark, UL mark, etc.

## 2) Resolution of Defects Due to Non-Uniformity of the Product

Manufacturing defects are problems within the range that can be avoided by quality management activities in process management.

On the other hand, design defects, warnings, and marking defects are essentially different from manufacturing defects.

## 3) Resolution of Defects by Strengthening Inspection Tasks

By using raw materials and parts that have passed the imported inspection, and strengthening inspection tasks in the product production process, the incidence of defects before the final inspection is minimized.

## **(2) Product Liability Accident Prevention Measures at the Manufacturing Stage**

Product liability prevention measures at the manufacturing stage may be affected by several factors.

Each factor—such as raw materials or parts input into the process, manufacturing equipment or machines necessary for product production, and employees directly participating in production, etc.—is to be considered at the manufacturing stage, and also work processes are matters to be considered

Countermeasures for product liability prevention for each factor are as follows:

### 1) Raw Materials and Parts

Selection and training of suppliers, reconfirmation of purchase specifications and import standards, thorough import inspection, and resolution of storage management problems.

### 2) Work Process

Work standardization, thorough process and quality control, thorough management of safety check items, clarification of process change procedures, and development of measures to

prevent foreign substances from entering.

### 3) Manufacturing Equipment and Machinery

Total production maintenance activities and machine abnormalities prevention.

### 4) Employee Training

Training of new employees, safety training, and improvement of product liability awareness.

## **(3) Application of the Quality Control System**

Quality control measures applicable to product liability prevention at the manufacturing stage include production and service provision management, validation of production and service provision processes, identification and traceability, and product preservation. In cases of production with raw materials or parts supplied from customers, separate management of customer assets is required.

### 1) Management of Production and Service Provision

In production, information that defines the characteristics of the product to check product specifications and design considerations and prepare work instructions to align with that.

Production is carried out using appropriate equipment according to work instructions and availability is checked for effectiveness using reliable measuring devices. It is also necessary to check validity continuously through the follow-up management of products that have been released and delivered to consumers.

### 2) Confirmation of Validating Production and Service Provision Processes

To confirm the validation of production and service provision processes, there must be a set standard for reviewing and approving the production process.

There must be approval for production facilities, and qualifications must be recognized about employees who conduct inspection tasks, design and development tasks, etc.

Based on this, production is carried out according to the prescribed methods and procedures, and quality records generated during that process must be managed by setting a storage method, storage period, etc.

The validity of such a process should be checked continuously, and if a problem occurs, appropriate measures should be taken against the applicable process, personnel, facilities, etc.

### 3) Identification and Traceability

The organization should identify the product with appropriate means, if applicable, throughout all stages of product realization. The organization shall identify the condition of the product in relation to monitoring and measurement requirements. If traceability is a requirement, the organization must manage and record the unique identification of the product.

### 4) Customer Assets

The organization should pay attention to customer assets that are under its control or use. The organization must identify, verify, protect, and maintain customer assets used as products or provided for making products. If a customer asset is found to be lost, damaged, or inappropriate to use, it should be recorded and reported to the customer.

### 5) Preservation of Products

The organization must ensure that product suitability is maintained during the internal process as well as until the product is delivered to a designated destination. This should include identification, handling, packaging, storage, and protection. This should also apply to product components.

## 6. Overview of Product Liability Measures for Small and Midsize Enterprises.

It is necessary to establish an appropriate system that can respond to product responsibility in the company from an enterprise-wide perspective, and to improve consciousness that is required from top management to all employees, including lines and staff.

To this end, by introducing the ISO 9001 system, the company's quality management system will be improved to the next level. As a result, it is possible to establish a system that can supply reliable and safe products to the market.

In other words, adopting and implementing ISO 9001 helps to maintain a quality management system that produces products with high stability. Furthermore, product liability measures cannot be considered without introducing such a system. The obvious difference between the two is that ISO 9001 is an arbitrary system, while product liability has compulsory power by law. Therefore, management by ISO 9001 is an important means until a product liability problem occurs. Once a lawsuit is filed for

a defective product, countermeasures based on the product liability law are required.

In particular, it is essential to perform tasks based on the product liability concept from the design stage, which is the initial stage of the product. To prevent design defects, the latest legal standards and technical information of each country should be obtained and adapted.

Companies should also re-review all in-house product safety standards (i.e., design standards, stability evaluation standards). This includes experts in charge of product safety, such as those in charge of product safety propulsion or the stability of the check predecessors that are assigned. At the same time, allow an assuming sense of responsibility by clarifying the duties of the relevant sector.

The design method used by the majority of industry professionals should be adopted, and the reasons for design adoption and change should be carefully reviewed to avoid risks that can occur according to product liability. The design should be evaluated from the viewpoint of stability and the stability should be tested to ensure the results are reflected. Because evaluation result data and test result data are essential matters in litigation, they should be sufficiently checked and managed. Above all, the focus should be on the safety design considering human engineering.

To prevent manufacturing defects, safety design should be done. Safety designs incorporate several important steps—including purchase, manufacturing process, inspection, shipment, distribution, end use, and disposal.

Technology for safety improvement should be researched and developed and an independent department for quality control and stability management should be established.

Safety audits also should be conducted and encompass every aspect of safety management. Best practices in safety management should be informed by data collected during design, development, and manufacturing and by interlocking with the ISO 9001 system.

In addition, the full management of test and inspection equipment such as measurement equipment, and the management system of production equipment, should be re-reviewed to ensure that functions are possible with a focus on safety management.

The fundamental methodology to prevent product liability problems hinges on building and implementing a comprehensive system that can respond rapidly and effectively.

At present, many companies are ISO 9001-certified. Even with an ISO 9001 certificate, it does not mean a company's liability is exempted in the event of a problem caused by product liability. However, if a quality management system or documentation of the quality system is already prepared, and if the work is carried out accordingly, it will be a very important means in implementing measures according to product liability.

Conducting a product liability audit allows a company to gauge whether its organization or quality system is suitable for preventing product liability. Through professional diagnosis by an audit team consisting of lawyers and technicians specializing in product liability, it is possible to learn about product liability prevention and system improvement in preparation for litigation.

Diagnosis will be conducted focusing on whether the product liability organization has been properly

maintained, whether its operation functions organically and smoothly, and whether it is familiar with the types and formats of documents.

Responding to the Product Liability Act using the PDCA (plan-do-check-act) concept, which is based on the above methodologies, is ultimately considered a safer and more effective countermeasure for small and midsize enterprises.

### Author Bio:

Gilbert Gong, Ph.D., serves as president of the Institute of Global Certification (IGC) and as president of Global Personnel Certification Co., Ltd. (GPC), a management system and personnel certification body in Seoul, South Korea. He is also a professor at Inha University in Incheon, South Korea, and an ISO lead auditor and QMS inspector for medical devices. Dr. Gong, who has authored several books and journal articles, received his doctorate in chemistry from the Technical University of Berlin in Germany and a degree in chemical engineering from Chung-Ang University in Seoul.



## Accreditation, Standardization and Conformity Assessment

### Certificate Course

California State University Dominguez Hills (CSUDH) and IAS have developed a four-module course for folks interested in knowing more about accreditation, standardization, and conformity assessment practices.

This course has been developed in coordination with seasoned professionals and subject matter experts in the areas of standardization and conformity. It is delivered via cohort format with class sessions held online. Participants are qualified through participation, projects, and exam.

Those interested in accreditation, standardization and conformity assessment practices, are encouraged to enroll in this program and gain insight about it.

For more information, contact:

Dr. Milton Krivokuca | [mkrivokuca@csudh.edu](mailto:mkrivokuca@csudh.edu)

Prasanth Ramakrishnan | [pramakrishnan@iasonline.org](mailto:pramakrishnan@iasonline.org)



OFFERED REGULARLY, PLEASE CONTACT US  
FOR UPCOMING DATES



## Technical note

# Competence Versus Qualification in Personnel Certification Programs

by Dr. George J. Anastasopoulos, VP, Global Development & Compliance, IAS

Personnel certification has been and will continue to remain a desirable asset for any modern professional. Achieving certification often represents a significant investment in time, effort, and expense. Frequently, candidates must choose between a "competence-based" or a "qualification-based" type of certification program. In most cases, qualification-based personnel certification is easier and less expensive to achieve. But is it actually worth it? What is the difference between competence-based and qualification-based personnel certification programs?

In ISO Standard 19011:2011, "Guidelines for Auditing Management Systems," competence is defined as "the ability to apply knowledge and skills to achieve intended results." Competence-based certification means that the Personnel Certification Body (PCB) is expected to examine a candidate's knowledge, skills, personal attributes, and qualifications specific to the program and/or scope of certification. On the other hand, qualification-based certification relies on an applicant's education and qualifications, rather than being based on measurable competence. The following short dialogue is catalytic to understanding, in a few words, the difference between "competence" and "qualification."

Do you know how to drive a car?

I was trained and acquired a driving license, but I am still not confident in my ability to drive a car.

That means you have the qualifications, but not the competence.

There are college dropouts who are CEOs of Fortune 500 companies because they have competencies, not qualifications. Therefore, having both qualifications and competencies helps immensely,

but people can still excel through competencies rather than qualifications.

Recognizing this fact, the International Organization for Standardization (ISO), through ISO/IEC Standard 17024, "Conformity Assessment - General requirements for bodies operating certification of persons," mandates that the certification of persons should be based on the demonstration of competencies and not the demonstration of qualifications. ISO/IEC 17024 sets the requirements and the framework, at a global level, for the operation of Personnel Certification Bodies. By using ISO/IEC 17024, business, industry, and other key stakeholders have recognized that competency-based certification is the optimum way of achieving confidence in persons certified by PCBs. ISO/IEC 17024 does allow some variation in how competence is demonstrated; consequently, different PCBs may interpret and apply the means for competency assessment in different but technically valid ways.

Still, there are PCBs that insist on offering non-accredited, qualification-based programs on the assumption that qualification equals competence. While that assumption may be correct in some cases and may continue to be acceptable to a range of users, it is less acceptable for those who operate in contexts that require a more rigorous demonstration of competence based on a valid examination. This creates considerable confusion to the market and to certification candidates. And of course, as qualification programs don't satisfy all competence requirements, they are non-accreditable. Another key difference among the competence- and qualification-based programs is the change of emphasis from training to examination.

*[continued to next page]*





*[continued from previous page]*

Qualification-based programs emphasize training while competence-based programs emphasize the results of training by assessing competence through one or more methods of examination that must be valid, reliable, and independent. Competency-based certification programs firstly define the competencies required so that they can be properly examined.

So, is it possible to distinguish a competency-based program from a qualification-based program? The answer is simple: Check for the accreditation of the PCB that provides the certification program to see if it is based on ISO/IEC Standard 17024 requirements. Then, check to see if the PCB's scope of accreditation includes that program. Finally, check if the PCB's accreditation is provided by an Accreditation Body that is a Multilateral Recognition Arrangement (MRA) signatory member of the International Accreditation Forum.



## PERSONNEL CERTIFICATION BODY ACCREDITATION PROGRAM

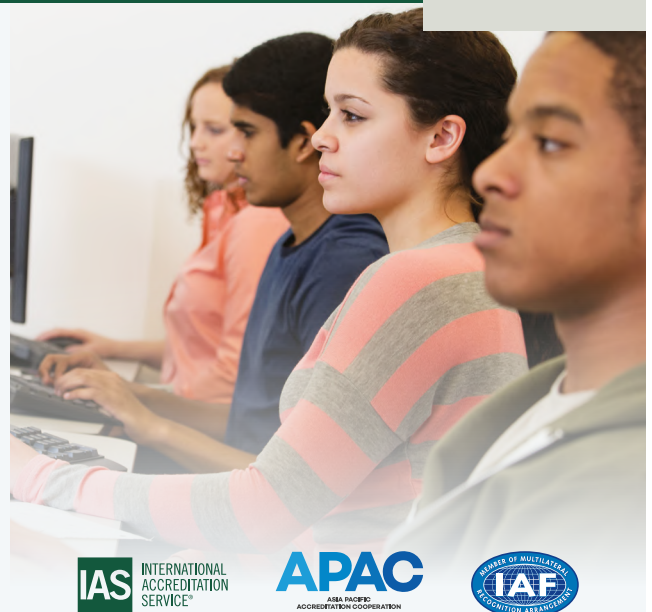
Organizations that provide personnel certification are becoming Accredited Personnel Certification Bodies from the International Accreditation Service (IAS).

IAS accreditation:

- Demonstrates compliance with ISO/IEC 17024.
- Provides verification of industry and/or international standards.
- Helps organizations protect the integrity, and ensure the validity, of individual certification programs.
- Promotes consumer and public confidence in the capabilities and competence of the people who provide specialized services.
- IAS is an MLA signatory to the International Accreditation Forum (IAF), helping to increase acceptance in multiple markets.
- IAS offers prompt, personal service, including rapid scheduling of assessments to meet the needs of laboratories.

**Become a leader in your field!**

Learn more at [www.iasonline.org](http://www.iasonline.org)





## IAS Training Schedule: Apr.-Dec. 2022

Courses are ordered by date. You can navigate to the 2022 IAS Training Schedule page by visiting [www.iasonline.org/training/ias-training-schedule](http://www.iasonline.org/training/ias-training-schedule).

Once on the page, please click on a course title to view the associated **Course Description** from the menu on the left. It will open in a new window and you can review the course at your leisure. Prices for all courses are shown on the **IAS Training Fees** page on the left.

If you wish to purchase participation in an IAS course, use the button that looks like this one [Register](#), but in the right-hand column of the table below, associated with the course name and details of your choice. It will open a link to a secure web page to process your registration and payment.

If you are interested in any IAS training information, let us know by clicking on this **Request Training Information** button [Request Training Information](#) or to the **Request Training Information** link on the left. We will contact you with details regarding your expression of interest within 48 hours. For more information, please contact us at [iasinfo@iasonline.org](mailto:iasinfo@iasonline.org).

START DATE	END DATE	COURSE DESCRIPTION	TIME ZONE
21 Apr	21 Apr	<a href="#">IAS Building Department Accreditation Overview (BDA)</a>	Americas
09 May	11 May	<a href="#">3035 ISO 17034 Reference Material Producer Course</a>	Gulf Region and South Asia
10 May	10 May	<a href="#">3048 Short QMS Courses for Labs - 11</a>	Americas
16 May	17 May	<a href="#">3003 Understanding ISO/IEC 17024 for Personnel Certification Bodies</a>	Americas
16 May	18 May	<a href="#">3002 Understanding ISO/IEC 17021-1 for Management System CBs</a>	Gulf Region and South Asia
17 May	18 May	<a href="#">3005 Understanding ISO/IEC 17065 for Product Certification Bodies</a>	Gulf Region and South Asia
24 May	25 May	<a href="#">3007 Special Inspection Agencies in Building Construction – NYC Focus - AC291</a>	Americas
25 May	26 May	<a href="#">3006 Uncertainty of Measurement for Labs</a>	Gulf Region and South Asia
26 May	27 May	<a href="#">3022 Internal Auditing for all Standards</a>	Americas
31 May	01 Jun	<a href="#">3001 Introduction and refresher to ISO/IEC 17020 for Inspection Agencies</a>	Americas
07 Jun	09 Jun	<a href="#">3002 Understanding ISO/IEC 17021-1 for Management System CBs</a>	Americas
08 Jun	09 Jun	<a href="#">3027 ISO 15189:2012 Understanding the requirements</a>	Gulf Region and South Asia
16 Jun	16 Jun	<a href="#">3004 Decision Rule</a>	Americas
16 Jun	16 Jun	<a href="#">3048 Short QMS Courses for Labs - 12</a>	Americas
05 Jul	06 Jul	<a href="#">3004 Understanding ISO/IEC 17025 for Testing and Calibration Labs</a>	Americas
18 Aug	18 Aug	<a href="#">Fire Prevention Department Accreditation Overview (FPD)</a>	Americas
30 Aug	31 Aug	<a href="#">3004 Understanding ISO/IEC 17025 for Testing and Calibration Labs</a>	Gulf Region and South Asia

START DATE	END DATE	COURSE DESCRIPTION	TIME ZONE
01 Sep	02 Sep	<a href="#">3022 Internal Auditing for all Standards</a>	Gulf Region and South Asia
01 Sep	03 Sep	<a href="#">3002 Understanding ISO/IEC 17021-1 for Management System CBs</a>	Gulf Region and South Asia
06 Sep	07 Sep	<a href="#">3027 ISO 15189:2012 Understanding the requirements.</a>	Gulf Region and South Asia
06 Sep	07 Sep	<a href="#">3001 Introduction and refresher to ISO/IEC 17020 for Inspection Agencies</a>	Gulf Region and South Asia
07 Sep	08 Sep	<a href="#">3007 Special Inspection Agencies in Building Construction – NYC Focus - AC291</a>	Americas
03 Oct	04 Oct	<a href="#">3001 Introduction and refresher to ISO/IEC 17020 for Inspection Agencies</a>	Gulf Region and South Asia
03 Oct	04 Oct	<a href="#">3023 IAF MD Training for Management System CB Personnel</a>	Americas
05 Oct	05 Oct	<a href="#">3004 Decision Rule</a>	Americas
05 Oct	07 Oct	<a href="#">3002 Understanding ISO/IEC 17021-1 for Management System CBs</a>	Gulf Region and South Asia
11 Oct	12 Oct	<a href="#">3006 Uncertainty of Measurement for Labs</a>	Americas
13 Oct	14 Oct	<a href="#">3005 Understanding ISO/IEC 17065 for Product Certification Bodies</a>	Americas
01 Nov	02 Nov	<a href="#">3022 Internal Auditing for all Standards</a>	Americas
01 Nov	03 Nov	<a href="#">3035 ISO 17034 Reference Material Producer Course</a>	Gulf Region and South Asia
02 Nov	03 Nov	<a href="#">3006 Uncertainty of Measurement for Labs</a>	Gulf Region and South Asia
03 Nov	04 Nov	<a href="#">3017 Understanding Risk Approach for Management Systems</a>	Americas
08 Nov	09 Nov	<a href="#">3004 Understanding ISO/IEC 17025 for Testing and Calibration Labs</a>	Americas
10 Nov	10 Nov	<a href="#">3032 Need for full Corrective and Preventive Action</a>	Americas
16 Nov	17 Nov	<a href="#">3003 Understanding ISO/IEC 17024 for Personnel Certification Bodies</a>	Gulf Region and South Asia
29 Nov	01 Dec	<a href="#">3034 ISO IEC 17043 Proficiency Testing Provider Course</a>	Gulf Region and South Asia
30 Nov	01 Dec	<a href="#">3007 Special Inspection Agencies in Building Construction – NYC Focus - AC291</a>	Americas
01 Dec	02 Dec	<a href="#">3023 IAF MD Training for Management System CB Personnel</a>	Americas





# World Accreditation Day 2022



Celebrate Sustainability in  
Economic Growth and  
the Environment with us  
on **June 9<sup>th</sup>**



**#WAD2022**

## Calibration Laboratory Accreditation Program



Calibration laboratories come to International Accreditation Service (IAS) for accreditation because of our technical competence, international recognition, and prompt personal service.

### IAS Accreditation:

- Demonstrates compliance with ISO/IEC Standard 17025, the global standard for laboratory competence.
- Enhances recognition and acceptance of accredited facilities in national and global markets.
- Demonstrates to the marketplace and to regulators that calibration laboratories have met the industry recognized requirements for measurement traceability and undergo a program of periodic monitoring by IAS.
- IAS is a signatory to the Mutual Recognition Arrangement (MRA) with the Asia Pacific Accreditation Cooperation (APAC) and the International Laboratory Accreditation Cooperation (ILAC), helping increase recognition of test reports.
- IAS offers prompt, personal service, including rapid scheduling of assessments to meet the needs of laboratories.

### For More Information

International Accreditation Service  
[www.iasonline.org](http://www.iasonline.org)





INTERNATIONAL  
ACCREDITATION  
SERVICE®

**Trust, Competence, Safety**

3060 Saturn Street, Suite 100  
Brea, California 92821-1732 USA

[www.iasonline.org](http://www.iasonline.org)

Be sure to Like, Follow and Subscribe:

