INTERNATIONAL STANDARD

ISO/IEC 33003

Second edition
2015-03-01

Information technology — Process assessment — Requirements for process measurement frameworks

Technologies de l'information — Évaluation du processus — Exigences relatives au cadres de mesure du processus

Reference number
ISO/IEC 33003:2015(E)

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/IEC JTC 1, *Information technology*, SC 7, *Software and systems engineering*.

Introduction

This International Standard provides requirements for process measurement frameworks that support and enable the assessment of process quality characteristics, from conceptualization to empirical validation. In process measurement frameworks, measurement of a process quality characteristic produces a composite measure (e.g. process capability levels of ordinal scale in ISO/IEC 33020). Examples of process quality characteristics that are constructs (theoretical concepts) include process capability, process security, process agility, and process safety. The main users of this International Standard are developers of process measurement frameworks and process assessment models. Conformity to this International Standard ensures that any process measurement framework is developed with reliable structures or elements which will generate quality composite measures.

This International Standard is part of a set of International Standards designed to provide a consistent and coherent framework for the assessment of process quality characteristics, based on objective evidence resulting from implementation of the processes. The framework for assessment covers processes employed in the development, maintenance, and use of systems across the information technology domain and those employed in the design, transition, delivery, and improvement of services. The set of International Standards, as a whole, addresses process quality characteristics of any type. Results of assessment can be applied for improving process performance, or for identifying and addressing risks associated with application of processes.

This International Standard provides requirements for the development of process measurement frameworks, such as ISO/IEC 33020. These can then be used to define process assessment models, conformant to ISO/IEC 33004, that can be employed for process assessments conformant with ISO/IEC 33002. The overall architecture and content of the series is described in ISO/IEC 33001.

Several International Standards in the ISO/IEC 330xx family of standards for process assessment are intended to replace and extend parts of the ISO/IEC 15504 series of Standards. ISO/IEC 33001, Annex A provides a detailed record of the relationship between the ISO/IEC 330xx family and the ISO/IEC 15504 series.

Information technology — Process assessment — Requirements for process measurement frameworks

1 Scope

This International Standard sets out the requirements for process measurement frameworks for use in process assessment. The requirements defined in this International Standard form a structure which

a) establish the requirements for process measurement frameworks in the context of process assessment,

b) establish the requirements for the validation of process measurement frameworks for use in process assessment, and

c) establish requirements that are applicable to any process measurement frameworks to develop composite measures across domains.

This International Standard is applicable to the development of process measurement frameworks for any process quality characteristic across all application domains.

Annex A presents a map of terminologies used in this International Standard. Annex B provides an explanation of construct specifications. Annex C reviews statistical validation methods. Annex D provides some methods including references that can be utilized in implementing the requirements for process measurement frameworks. These Annexes will be moved to a guide for constructing process measurement frameworks to be developed as part of the set of International Standards.

NOTE ISO/IEC 33020 is a process measurement framework for assessment of process capability based on this International Standard.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies

ISO/IEC 15939:2007, *Systems and software engineering — Measurement process*

ISO/IEC 33001:2015, *Information technology — Process assessment — Concepts and terminology*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 33001, ISO/IEC 15939, and the following apply:

3.1
aggregation method
method that combines a set of measurement values to create a composite value

Note 1 to entry: Aggregation methods are based on compensatory or non-compensatory models.

3.2
compensatory model
MCDM model in which a composite measure is composed of individually weighted terms and where criteria (also refer to attribute terms) with a high value can compensate for those of a low value in proportion to each weight

Note 1 to entry: A compensatory model suggests that improving the more important measures (those with a higher weighting) is more likely to increase or improve the overall composite value than improving the less important ones. This model assumes that the weight (influence level) of criteria remains the same regardless of the measured level of the criteria.

3.3
composite measure
variable derived from a set of operations of a construct's multi-item measures defined according to construct specification (either reflective or formative) that is the way in which the latent variable representing the construct of interest is linked to its measures

3.4
composite value
value from a composite measure

Note 1 to entry: A composite value can be from an ordinal, interval, or ratio scale.

3.5
construct
concept such as the abstract idea, image, underlying theme, or subject matter that one wishes to measure using process assessments

Note 1 to entry: In process measurement frameworks, constructs (also refers to latent constructs) are theoretical concepts such as the process quality characteristics and process attributes.

Note 2 to entry: The meaning that one assigns to a construct is called theoretical definition, which should explain its meaning, as well as discuss its distinct dimensions (facets).

3.6
dimension
distinct components that a multidimensional construct encompasses

3.7
formative construct
construct that is formed from its observed measures in the relationship between a construct and its measures

Note 1 to entry: The construct is a consequence of its measures and each measure is a determinant of the construct.

3.8
latent variable
variable representing a unidimensional construct

Note 1 to entry: There should be a separate latent variable for each dimension of a construct and a minimum of one measure per latent variable.

3.9
MCDM
Multiple-Criteria Decision Making or Multi-Attribute Decision Making
making preference decisions (e.g., evaluation, prioritization, and selection) of available alternatives characterized by multiple criteria

Note 1 to entry: A criterion in MCDM corresponds to measure.

Note 2 to entry: An MCDM with one alternative is the same as the development of a composite measure.

3.10
measurement model
the implicit or explicit relationship between a latent variable and its (multi-item) measures

Note 1 to entry: The relationship between a reflective (formative) construct and its measure(s) is called a reflective (formative) measurement model.

3.11
multidimensional construct
construct that consists of a number of unidimensional constructs.

Note 1 to entry: Each dimension of a multidimensional construct is called unidimensional and is represented by one latent variable. Each dimension can have multiple measures. In a multidimensional construct, for example, the meaning of capability when it is defined as the common factor underlying its process attributes is different from the case when capability is defined as a simple sum of its process attributes. The former is called a reflective multidimensional construct and the latter is formative. A multidimensional construct can span an indeterminate number of levels.

3.12
non-compensatory model
MCDM model that does not allow criteria to compensate for each other in proportion to their weights

Note 1 to entry: Strongly positive or negative terms influence the overall composite value disproportionately, although the weight stays the same. There are various non-compensatory models depending on the evaluation policy, the purpose of the composite measure, and/or the measurement scale.

3.13
reflective construct
construct that is viewed as the cause of measures in the relationship between a construct and its measures

Note 1 to entry: Reflective construct is an underlying factor of the variation of its measures.

3.14
scale
ordered set of values, continuous, or discrete, or a set of categories to which the attribute is mapped

Note 1 to entry: The type of scale depends on the nature of the relationship between values on the scale. Four types of scales are commonly defined:

Nominal — the measurement values are categorical. For example, the classification of defects by their type does not imply order among the categories.

Ordinal — the measurement values are rankings. For example, the assignment of defects to a severity level is a ranking.

Interval — the measurement values have equal distances corresponding to equal quantities of the attribute. For example, cyclomatic complexity has the minimum value of one, but each increment represents an additional path. The value of zero is not possible.

Ratio — the measurement values have equal distances corresponding to equal quantities of the attribute where the value of zero corresponds to none of the attribute. For example, the size of a software component in terms of LOC is a ratio scale because the value of zero corresponds to no lines of code and each additional increments represents equal amounts of code.

[SOURCE: ISO/IEC 15939:2007]

3.15
unidimensionality
existence of a single trait or construct underlying a set of measures

4 Requirements for process measurement frameworks

This clause defines the requirements for developing process measurement frameworks. Guidance in this International Standard is limited to providing a better understanding of these requirements. Figure A.1 provides a mapping of the relationships between some terms used in this Clause.

NOTE Guidance on achieving conformance to these requirements, including examples and methods, will be provided in a guide for constructing process measurement frameworks to be developed as part of the set of Standards.

4.1 Conceptualization

4.1.1 Requirements

a) A measurement framework shall identify and address a single process quality characteristic;

b) A process quality characteristic in a process measurement framework shall be defined on the basis of a multidimensional construct;

c) A process quality characteristic in a process measurement framework shall be defined as a set of process attributes;

d) Each process attribute shall define a property of the process quality characteristic;

e) Each process attribute that is not directly measurable shall be considered as a construct;

f) Process attributes in a process measurement framework shall be defined as either reflective or formative.

g) The measurement framework shall document the policies and assumptions underlying its use and application;

4.1.2 Guidance

The process of identifying and clarifying concepts is called conceptualization. A concept is an idea or image that unites phenomena of interest (e.g., traits, behaviour traits) under a single term. It is a summarizing device to replace a list of specific traits. Most process quality characteristics (e.g., process capability) are not observable but are theoretical concepts called constructs.

The composite measures (e.g., process capability level) used in process measurement frameworks are defined on the basis of a construct composed of process attributes. A measurement framework may be structured into a series of levels of achievement.

When a process attribute is not directly measurable, it may also be defined as a construct. The set of process attributes for any construct may be either reflective or formative.

Participation of experts and stakeholders can increase the validity of the process quality characteristic and its process attributes; aspects of validity are discussed in C.3.

A multidimensional construct can be depicted with a path diagram including a set of dimensions and their relationships. Use of a path diagram improves the understandability of model scope and structures.

4.2 Construct definition

4.2.1 Requirements

a) The construct definition shall define the meaning of the process quality characteristic and its process attributes in a process measurement framework;

b) The construct definition shall clarify the specification of the process quality characteristic and its process attributes as dimensions;

c) The construct definition shall provide a guide for the operationalization of the process quality characteristic and its process attributes;

d) The construct definition shall state the scales of composite measures such as categorical (e.g., a series of ordinal values such as capability level) or numeric;

e) At least one of the process attributes shall comprise the achievement of the defined process purpose and process outcomes for the process; this is termed the process performance attribute;

4.2.2 Guidance

Although a process quality characteristic or process attribute should convey an intuitive understanding of what it represents, interpretation may vary according to the observer. Thus, a definition is required to explain and provide the meaning of a construct. This is called the construct definition.

Clarification of a construct implies that for example the definition of the process quality characteristic as specified super-ordinate fully covers all of process attributes on the basis of construct specification, where process attributes as sub-ordinates are its distinct dimensions. A latent variable can be assigned to a unidimensional construct in the model. Statistical methods related to dimensionality are introduced in C.1.

4.3 Operationalization

4.3.1 Requirements

a) All process attributes shall be defined according to their construct specification;

b) Achievement of process attributes shall be verifiable through objective evidence.

4.3.2 Guidance

When a process attribute is directly observable through formal assessments, self-reports, surveys (including questionnaires and interviews), observations, or other empirical means, it is a base measure that is functionally independent of other measures. If a process attribute is measured with its several sub-constructs or measures, it can be considered as a construct. Four or more base measures are recommended to measure a construct and perform a set of statistical tests (including model validation and construct specifications) in reflective specification.

NOTE Refer to Clause 6.3.4 of ISO/IEC 33004 for assessment indicators that are utilized for process attribute rating.

4.4 Construct specification examination

4.4.1 Requirements

Construct specifications of the process quality characteristic and its associated process attributes shall be examined through operationalization and with rationale.

4.4.2 Guidance

There are two kinds of construct specifications that refer to the way in which the latent variable representing the construct is linked to its measures (i.e., the relationship between a unidimensional construct and its measures): reflective and formative measurement models. A process quality characteristic or process attribute can be viewed either as underlying factors or indices produced by observed measures. The former is referred to as reflective (effect) constructs or reflective measurement models, and the latter formative (causal) constructs or formative measurement models.

The objective of a reflective measurement model is to measure a single property by using multiple measures, whereas a formative model attempts to summarize multiple properties with a single composite value. In Annex B, these two specifications can be represented as Figure B.1 (a) and Figure B.1 (b), respectively.

Decision rules to examine reflective or formative, construct specification, are summarized in Table 1. These decision rules can be applied to the process quality characteristic and its associated process attributes. They can be assessed *a priori* statistical validation of construct specification. Annex B provides the construct specification in detail.

Table 1 — Decision rules to examine reflective or formative measurement model

Decision rule	Reflective measurement model	Formative measurement model
Characteristics of measures of the construct	• Measures are manifestations of the construct. • Measures share a common theme. • Measures should be interchangeable. • Measures should have the same or similar content. • Excluding a measure should not alter the conceptual domain of the construct. • Measures are expected to co-vary with one other.	• Measures are defining characteristics (aspects) of the construct. • Measures need not share a common theme. • Measures need not be interchangeable. • Measures need not have the same or similar content. • Excluding a measure may alter the conceptual domain of the construct. • Measures need not co-vary with one another.
Direction of causality between construct and measures	• The direction of causality is from the construct to its multi-item measures. • Changes in a measure should not cause to changes in the construct.	• The direction of causality is from measures to the construct. • Changes in the construct should not cause changes in the measures.

In some instances, the relationships depicted in Figure B.1 (Annex B) can have a higher-order level, i.e., conceptual definitions of constructs are often specified at a more abstract level, which sometimes include multiple reflective and/or formative first-order dimensions. The definition of a higher-order model should be theory-driven in a reflective measure model. Statistical analyses should be used to support or validate the definition.

4.5 Rating process attributes

4.5.1 Requirements

a) The process attributes shall be rated;

b) A measurement scale, i.e., nominal, ordinal, interval, or ratio, shall be defined for the process attributes;

c) A measurement method shall be identified that objectively assigns a value to each process attribute.

4.5.2 Guidance

Some assessments can generate the ratings of for example a process quality characteristic or process attributes for individual process instances assessed. On the other hand, others providing an overall picture without ratings can simultaneously assess a set of process instances under the same context as a process. Rating of process attributes can be based on formal assessments, self-reports, surveys (including questionnaires and interviews), observations, or other empirical means. Thus, a measurement scale for rating base measures should be consistent with the granularity of assessment. Occasionally, rating in self-reports or surveys is on the base of perception rather than objective evidence. Rating scale for the process quality characteristic and its process attributes should be addressed with rationale, consistent with the construct specification (refer to 4.3).

A well-established documented assessment process for rating process attributes provides credible measurement results. The approach to rating the process attributes shall be defined in the documented assessment process, and may depend on the class of the assessment, based on the assessment objectives. Thus, for this purpose, a documented assessment process will guide the process for establishing, planning, performing and evaluating assessment under an integrated assessment scheme. If there is consensus in the community, a validated documented assessment process can be adopted after examining its conformity with measurement purposes.

4.6 Aggregation

4.6.1 Requirements

Aggregation derives a composite value or rating by combining a set of measurement values.

a) All aggregations required within the measurement framework shall be identified;

b) Aggregation methods shall be specified;

c) Aggregation methods shall be statistically valid.

d) Aggregation methods shall utilize consistent measurement scales;

e) Aggregation methods shall be consistent with the measurement framework policies and assumptions;

f) Aggregation methods shall be consistent with construct specifications.

4.6.2 Guidance

The scale of composite measure for the process quality characteristic or process attribute should be stated in accordance with its construct specification. The number of aggregation required depends on the structure of a multidimensional construct of process quality characteristic in 4.1. Principally, the point of aggregation can be the hierarchical order of constructs such as process quality characteristics and its process attributes. Each process quality characteristic level on the scale is defined in terms of the achievements of a set of process attributes.

A MCDM with one alternative can also be regarded as the aggregation method to derive a value of composite measure. An aggregation method may be based on compensatory or non-compensatory models depending on the construct specification, evaluation policy, the purpose of the composite measure, and/or the measurement scale. A formative model with no measurement error can be considered as a compensatory type MCDM which aggregates different aspects or dimensions into a composite value.

An aggregation example related would be the combination of a set of process attribute ratings to a level of process capability, assuming a formative specification. In a mutidimensional construct such as the process quality characteristic and its process attributes, aggregation is used to determine a process capability level from a set of process attribute ratings. In addition, if process attribute rating is performed for each of multiple process instances, aggregation methods should be provided.

A rating scale of a process quality characteristic or process attribute represents the extent of its achievement. The scale expressed as an ordinal scale can be transformed from an interval or ratio scale to provide anchor points for the rating. For example, the rating scale may be applied to express the extent of achievement of a process attribute for a process instance in a specific organizational context, or to express the extent of achievement of a process attribute across multiple process instances within the defined organizational unit scope.

Consistency in measurement scale implies that lower level transformation, from a higher measurement level to lower level, is possible such that, (i) a ratio scale can be transformed to an interval, ordinal, or nominal scale, (ii) an interval to an ordinal or nominal, and (iii) an ordinal to a nominal. However, the inverse direction is not allowed.

A composite value of the reflective construct can be computed by averaging or summing the values of measures if associated assumptions are satisfied. Those methods can also be applied to the aggregation of sub-constructs to obtain a composite value of higher-level construct in the multidimensional construct.

The presence of outliers should be examined, and highly skewed measurement values should be transformed, if necessary. True outliers may be removed from the aggregation. If measures have a different range of values, normalization is required before any data manipulation because of differences in measurement units. An appropriate normalization method should be used with respect to both the theoretical basis and data properties.

4.7 Sensitivity analysis

4.7.1 Requirements

Sensitivity analysis aims to examine the robustness of the composite value. The kinds and methods of sensitivity analysis depend on rating and aggregation method in process measurement frameworks.

a) Sensitivity analysis shall be performed for measurement scales of process attributes;

b) Sensitivity analysis shall be performed for aggregation methods;

c) Sensitivity analysis shall be performed for weights, if applicable.

4.7.2 Guidance

The robustness of the composite value can be evaluated by uncertainty analysis or by sensitivity analysis. Uncertainty analysis examines how uncertainty in input factors such as measurement values propagates through the structure of the composite measure and affects the composite value. Sensitivity analysis examines the extent to which each individual source of uncertainty contributes to the output variance. Sensitivity analysis can be performed on the basis of process attributes.

Including weights in a composite measure, for example most compensatory models, requires sensitivity analysis for weights, where a weight assignment method should be specified. Non-compensatory MCDM models do not require weights.

5 Requirements for the validation of process measurement frameworks

5.1 Requirements

a) Plans for reliability and validity of process measurement frameworks shall be established at the beginning of standardization. These plans shall include post-standardization activities;

b) Claims on reliability and validity of process measurement frameworks shall be consistent with construct specification;

c) Consistency (also refers to equivalence) as a reliability measure shall be examined for process attributes, if reflective;

d) Validities shall be examined for the process quality characteristic and its process attributes in a process measurement framework;

e) Construct specification shall be empirically examined for the process quality characteristic and its measures in a process measurement framework;

f) External measures (e.g., goals, criteria, and/or achievements) of a process measurement framework under development shall be documented for validity investigation.

5.2 Guidance

The quality of the process quality characteristic and its process attributes can be examined by using empirical methods such as reliability estimation (especially if reflective) and validity tests. Process measurement frameworks state their reliability and validity claims and how those claims shall be corroborated. Statistical validation of requirements specified in this Clause can be provided by a separate document or an Annex of a process measurement framework.

5.2.1 Reliability

A general definition of the reliability of a measure is the variance of the true (latent variable) variance divided by the total measure variance. Reliability concerns the degree of repeatability (stability) and consistency (equivalence) of a measure in terms of its ability to capture latent variables. Repeatability implies that “repeated assessments [at two different points in time, of the same process to the same or alternative instrument by the same assessor] should produce results that can be accepted as being identical. Consistency (equivalence) focuses on multiple measures of a construct measured at a single point in time, where each measure is considered a separate but equivalent measure of the underlying concept. C.2 briefly introduces statistical methods for estimating reliability.

A satisfactory level of reliability depends on how assessment results from a process measurement framework are used. For instance, in applied settings where important decisions are made with respect to the composite value, a high value of consistency (e.g., 0.9) is usually recommended as the minimally acceptable value.

In reflective constructs, unidimensionality is a required condition for the reliability analysis and construct validity.

5.2.2 Construct validity

The quality of the process quality characteristic and process attributes can be examined by using empirical methods such as reliability estimation (if reflective) and validity tests. The validation of a process measurement framework is a procedure for determining whether there is objective evidence that the process quality characteristic and process attributes what they are intended to measure, and that they are useful for their intended purposes. Some validation methods can be performed during standard development as indicated by Clause 4. However, this clause addresses *post hoc* statistical analyses to validate process measurement frameworks during trials and/or after publication.

If the process quality characteristic and its process attributes in a process measurement framework are not correctly operationalized, measured, or statistically validated, any composite measure may be weak or inappropriate. Thus, the process quality characteristic and its process attributes in a process measurement framework should be linked to its statistical validation, although statistical tests are not the purpose of International Standards. Threats to the validity should be addressed when evaluating the validity of process measurement frameworks. Validity tests depend on construct specifications.

Data of external measures can be objectively or subjectively collected. They are used for examining the predictive validity. Construct validities are briefly explained in C.3.

5.2.3 Construct specification

Construct specifications (also referred to as specification models) can be statistically tested to determine whether the relationship between process quality characteristic and its process attributes is formative or reflective. A simulation study can be performed if necessary. All aggregation should meet the rationale of construct specification.

NOTE Confirmatory tetrad analysis, (addressed in B.3), can be used to statistically test construct specifications.

6 Verifying conformity of process measurement frameworks

This clause is concerned with the mechanisms that may be used to verify that the requirements of this International Standard have been fulfilled.

Conformity to the requirements of this International Standard may be verified by:

— self-declaration (first party);

— a second party;

— a third party.

The party performing verification shall obtain objective evidence that the process measurement framework fulfils the requirements set forth in Clause 4. Objective evidence shall be provided to demonstrate the integrity and consistency of the process measurement framework.

Annex A
(informative)

A terminology map

Figure A.1 presents a map of terms defined in Clause 3 and includes requirements and guidance defined in Clause 4 (the arrows present reading direction). A process quality characteristic is explained, described, and organized by a multidimensional construct that consists of a set of unidimensional constructs. Each of the unidimensional constructs assigns meaning to phenomenon of interest. A construct is identified and clarified via conceptualization and embodied by construct definition.

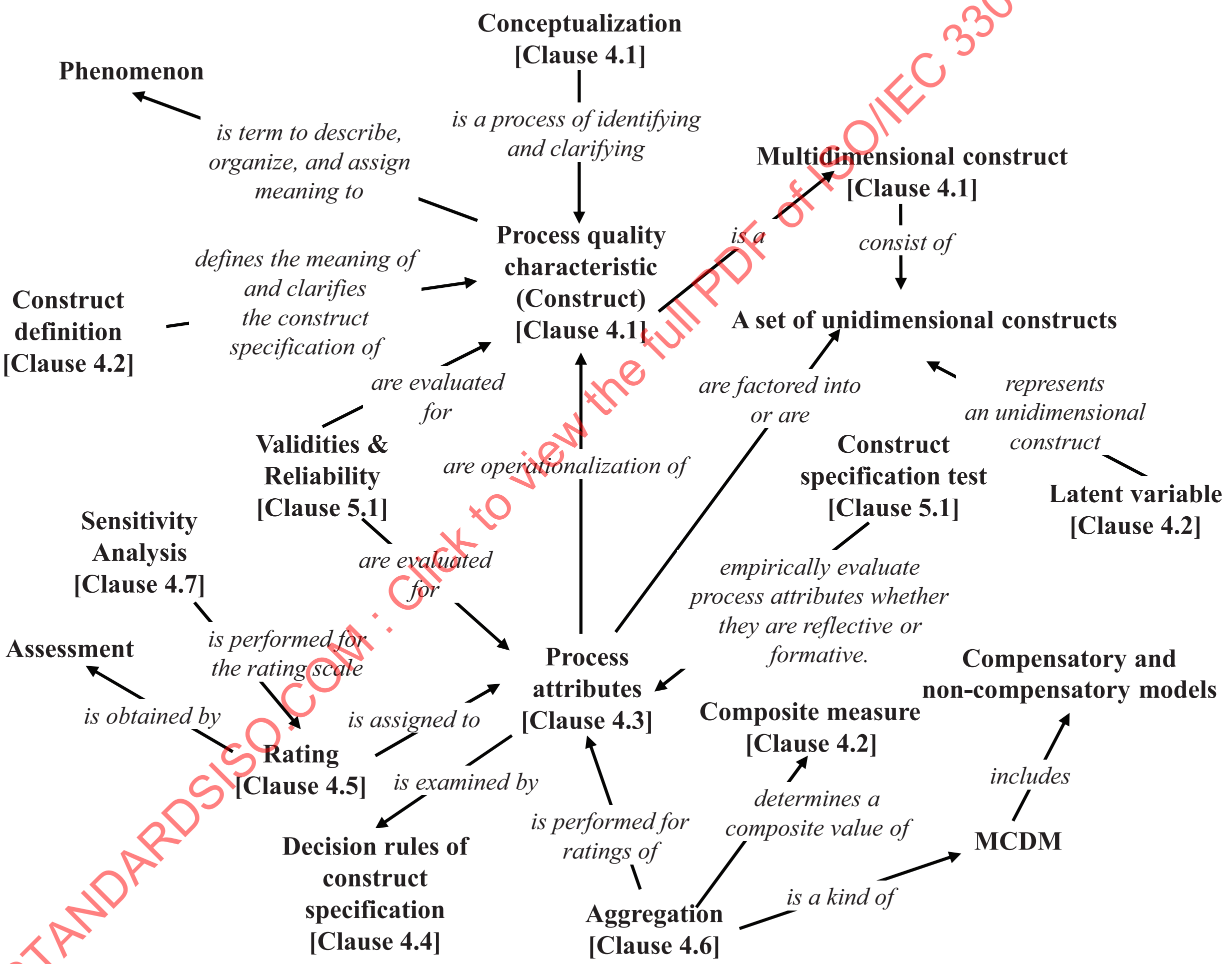

Figure A.1 — Map of terminologies

The process quality characteristic is operationalized as a set of process attributes, which are specified as either reflective or formative. Decision rules provide guides on the definition of process attributes. If process attributes are defined as reflective measures, factor analysis can support to determine a set of unidimensional constructs. On the other hand, process attributes defined as formative may treat each process attribute as a dimension. Each dimension of process attributes can be represented by a latent variable. Construct specification tests can empirically evaluate process attributes to determine whether they are reflective or formative. A rating obtained by assessment is assigned to each of process attributes.

Ratings assigned to process attributes are aggregated to derive a composite measure which is a value (numerical or ordered level). The aggregation method is a kind of MCDM model (compensatory or non-compensatory). The aggregation method selected may be influenced by construct specifications (i.e., reflective or formative), evaluation policy (e.g., compensatory or non-compensatory), the purpose of the composite measure, and/or the measurement scale.

Sensitivity analysis can be performed for measurement scales of process attributes, aggregation methods, and weights, if applicable. The quality of the process quality characteristic and its process attributes can be examined by using empirical methods such as reliability estimation and validity tests.

Annex B
(informative)

Construct specification: Reflective or formative

This annex addresses construction specification. The following describes three models, which may be used for developing composite values (e.g., process quality level) via the aggregation of measurement values (e.g., a set of the rating of process attributes). The two specifications as a basis of explanation can be represented as Figure B.1 (a) and Figure B.1 (b), respectively.

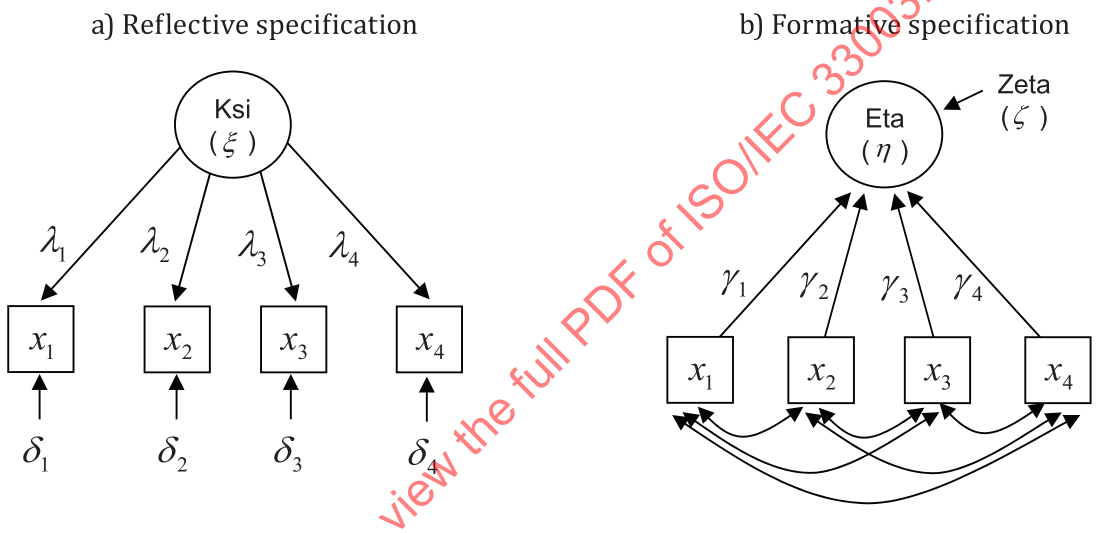

Figure B.1 — Relationship between a construct and its measures.

In the above Figure B.1 (a) lambda (λ) is a loading parameter and denotes the correlation between construct Ksi (ξ) and measure x (delta (δ) is an error term); Figure B.1 (b) gamma (γ) is a loading parameter for the measure x (zeta (ζ) is a disturbance term).

The causal direction has important implications for process attributes in the process context. If a process quality characteristic is assumed to be reflective, then the goal of actions should be one from which broad benefits flow naturally and are reflected in the process attributes. Interventions focused solely on individual process attributes may divert resources from more useful activities. Interventions in a formative construct should be focused on specific areas related to process attributes that constitute the process quality characteristic. In this case, improving one process attribute does not imply improvements in other measures. The decision rules for construct specification are summarized in Table 1.

B.1 Reflective model

As shown in Figure B.1 (a), a construct theoretically defined is an abstract concept that can be measured by using indirect multi-item measures. Constructs can be viewed as causes of measures. That is, measures reflect or manifest the extent to which a construct is changed, and variations in a construct result in variations in measures. A shift in the construct expects all measures to shift in the same direction because the measures reflect the same underlying construct. Thus, a high correlation between any two measures may be expected.

Furthermore, because measures have the same or similar content and they are supposed to be sampled from the same conceptual domain, reliable measures are interchangeable, and excluding one measure should not alter the conceptual domain of the construct. When measuring psychological constructs such as personality traits and attitudes, reflective measures are recommended.

The relationship between a construct and its measures is represented by a set of equations in which each measure depends on a latent variable as follows:

$$x_i = \lambda_i \xi + \delta_i,$$

where x_i is the i[th] reflective measure that depends on the latent variable ξ; the coefficient λ_i is the expected impact of a one-unit difference in ξ on x_i; and the random error term δ_i is the measurement error. The decision rules for reflective construct are summarized in Table 1.

B.2 Formative model

In the formative construct in Figure B.1 (b), measurement values are viewed as causes of a construct and the construct is a composite variable formed or induced by a combination of its measures. Measures characterize a set of distinct causes that are not interchangeable. Each measure captures a specific aspect of the construct domain. Thus, omitting a measure may alter the conceptual domain of the construct, i.e., it may adversely influence content validity. Because measures represent different facets of the construct domain, they need not be highly correlated. High correlations between formative measures can influence the stability of measurement coefficients and make it difficult to separate the distinct impact of individual measures on the construct. Construct measures of activities or behaviors are usually considered formative constructs.

A formative construct can be represented as follows:

$$\eta = \gamma_1 x_1 + L + \gamma_q x_q + \varsigma,$$

where η is the construct being estimated by its formative measure x_i; the coefficient γ_i denotes the effect of measure x_i on the latent variable η; the disturbance term ζ denotes the effect of measures omitted in the model on η. The decision rules for formative construct are summarized in Table 1.

B.3 Formative model with no error (composite measure)

A formative construct can be represented without the error term in Figure B.1 (b), i.e., the disturbance term ζ is assumed to be zero. Then, this works as an MCDM process and denotes a composite measure determined by the combination of a set of measures *xs* weighted by the importance or priority of those measures. Its relationship can be rewritten as follows:

$$C = \gamma_1 x_1 + \mathrm{L} + \gamma_q x_q,$$

where C represents the weighted composite of *xs*. In the equation, if all $\gamma_i x_i$ terms are known, then C is not a latent variable but a composite value. The validation procedure for a composite measure overlaps the formative model described above. This standard clearly indicates any differences between formative construct and composite measure if appropriate.

Annex C
(informative)

Some statistical validation methods

If data from the use of a process measurement framework is available, Clause 5 requires statistical analysis in order to provide evidence of conformity to the applicable requirements. This annex illustrates examples of the statistical methods for meeting the requirements of Clause 5.

C.1 Dimensionality

A statistical test of dimensionality is applicable only to reflective specifications. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) can be used to determine the number of dimensions underlying a set of measures and test the unidimensionality of each dimension[1][2].

C.1.1 Exploratory factor analysis

EFA is used to explore the dimensionality of a measurement instrument by finding the lowest number of interpretable factors needed to explain the correlations among a set of measures. EFA, as indicated by the term “exploratory,” does not specify the structure of the linear relationship between the observed variables and factors. In EFA, the structure of the factor model or the underlying theory is not known. Data are used to identify the number of factors and the quality of measures. Thus, EFA can be viewed as a technique to aid in theory building. This is valid only for reflective measurement models.

C.1.2 Confirmatory factor analysis

CFA deals specifically with measurement models, i.e., the relationship between observed measures and latent variables or factors. In CFA, the analyst must provide the number of underlying factors based on theory. CFA provides an evaluation of method effects and an examination of the stability or invariance of the factor model over time. Moreover, CFA should be conducted before the specification of structural equation model.

C.2 Reliability

Reliability estimation of constructs is applicable only to reflective specifications. There are various reliability estimation methods such as test-retest, alternative-form, split-half, and internal consistency (Cronbach’s alpha)[6]. These four are usually categorized into repeatability (stability) and consistency (equivalence) by the basic strategies used to evaluate reliability. Repeatability (stability) implies that “repeated assessments [at two different points in time, of the same process to the same or alternative instrument by the same assessor] should produce results that can be accepted as being identical”[36]. It is estimated by using the test-retest and the alternative-form methods. Consistency (equivalence) focuses on multiple measures of a construct measured at a single point in time, where each measure is considered a separate but equivalent measure of the underlying concept. The split-half and internal consistency methods are used to measure this consistency.

If assessors are presented with the same evidence, they ideally will produce exactly the same ratings. In practice, however, the subjective nature of ratings makes it highly unlikely that there will be perfect agreement in all cases. Inter-rater agreement, sometimes referred to as external reliability, is defined as the extent to which assessments in the same process with the same standards by two independent teams of assessors produce the same results. The Cohen Kappa coefficient[14] has been used as a measure of the reliability of process assessments. However, because of its paradox, an index of observed agreement is proposed in process assessment[41].

C.3 Construct validity

Construct validity denotes the degree to which operationalization accurately reflects its construct. Following describes construct validities such as face, content, predictive, concurrent, convergent, and discriminant. These validities can be applied to reflective measurement models. Face and content validities are generally applicable to formative models as well.

C.3.1 Face validity

Face validity addresses measures from operationalization to determine whether they appear to be a good translation of the construct. That is, it involves determining whether measures appear to cover intended topics clearly and unambiguously, and if the measures reflect the opinion of experts. Face validity involves a critical review of measures after they are developed.

Face validity should be examined, for examples, in the definitions of a process quality characteristic, its process attributes, outcomes, practices, and achievement.

C.3.2 Content validity

Content validity examines whether measures, as operationalized, captures the construct for which the latent variable stands. Content validity depends on the extent to which a measurement instrument reflects a specific domain of content in terms of the number and scope of the individual measures it contains. The theoretical definition of a construct states the domain and dimensions of the concept. For content validity, the design and development of a process measurement framework should follow rigorously defined development procedures.

Face validity should be examined, for examples, in the definitions of a process quality characteristic, its process attributes, outcomes, practices, and achievement.

C.3.3 Predictive validity

Criterion validity compares a measure with some standard variable that it should be associated with if it is valid. Predictive validity concerns a future criterion (Y) that is correlated with a relevant measure (X). The higher the correlation between X (e.g., the level) and Y (e.g., performance), the more valid the measure is this particular criterion. The validity coefficient can vary depending on the criterion and the degree of error associated with it, even though the measurement characteristics remain the same.

Concurrent validity is assessed by correlating a measure with the criterion at the same point in time. The difference between concurrent validity and predictive validity rests on the point in time at which the two measures are administered.

For predictive and/or concurrent validity, the achievement level of a process quality characteristic should be associated with external measure(s) defined in 5.1 d). An example is to test a question “Does increasing the capability level improve the ability to meet the schedule commitment?”

C.3.4 Convergent validity

Convergent validity refers to the degree to which multiple methods for measuring a variable provide the same results. Each individual measure can be viewed as a different method for measuring the same construct. This is an analysis of a question “Are reflective measures highly associated with its construct.”

C.3.5 Discriminant validity

Discriminant validity assesses the extent to which a construct and its measures differ from another construct and its measures. For a valid measure, the variance in the measure should reflect only the variance attributable to its intended latent variable and not to other latent variables. The discriminant validity of a set of constructs can be evaluated after the convergent validity of individual constructs is established. An example is to question whether an ordinal scale of capability level is a distinct construct.

C.4 Construct specification

Confirmatory tetrad analysis (CTA) is used to statistically test construct specification (reflective or formative specifications). The term confirmatory implies that a model is specified in advance. The term tetrad refers to the difference between the product of a pair of co-variances and that of another pair[3]. CTA can be applied to some under-identified models and non-nested models that cannot be tested using a conventional approach. Further, CTA does not require numerical minimization and thus avoids the associated convergence problems present in other estimation approaches. CTA can be performed by using the CTA-SAS routine[35][44].

Annex D
(informative)

Methods for implementing the requirements for process measurement frameworks

This annex provides a summary of methods for implementing the requirements for process measurement frameworks. This table does not provide a full list. The phases denote the steps that the development of process measurement framework should follow.

	Definition	Methods/references
Phase 1: Develop concepts/constructs (reflective and formative)		
Theoretical concept (framework)	A big picture showing the constructs and their relationships in a manner consistent with theory and/or previous research on process characteristics or properties to be measured by a process measurement framework.	• Law et al.[10] • Johnson et al.[9] • Maxwell[11] • Miles and Huberman[12]
Theoretical definition of constructs	A theoretical definition states the domain and dimensions of a construct and classifies construct specifications as reflective or formative. Formative issues in SEM are discussed by Edwards[19].	• Maxwell[11] • Bollen[1]
Phase 2: Perform operationalization (develop outcomes and/or practices as measurement instruments) (reflective and formative)		
Face validity	The extent to which a construct is accurately translated into the measurement instrument. That is, the extent to which the measurement instrument appears to measure what it is intended to measure).	• Content validity ratio by Lawshe[15] • Cohen's kappa[14] • Nunnally and Bernstein[42]
Content validity	The extent to which the outcomes and/or practices in a measurement instrument represent the domain, i.e., a check of the operationalization against the relevant content domain for the construct.	• Content validity ratio by Lawshe[15] • Cohen's kappa[14]
Mental experiments/decision rules in operationalization	Mental experiments or decision rules for determining reflective and formative.	• Decision rules (Jarvis et al.[20], Petter et al.[21]) • Measurement instrument development (Diamantopoulos[17], Diamantopoulos and Winklhofer[18], Edwards and Bagozzi[18], Rijsdijk et al.[26])
Phase 3: Conduct a confirmatory tetrad test to determine whether a construct is reflective or formative		
Confirmatory tetrad test (CTA)	A statistical test for determining whether a construct is reflective or formative.	• Ting,[44] Bollen et al.,[34] Bollen and Ting,[35] Hipp et al.[40]
Phase 4: Examine the unidimensionality of constructs (the statistical test can be applied only to reflective)		
Unidimensionality	The extent to which observed measures (e.g., process attributes) is closely related to one another and represents a single concept (Hattie[8]). A composite value calculated by the unweight sum of item ratings can be used as an estimate of the corresponding construct under unidimensionality (Gerbing and Anderson[13], p. 186).	• Annex B • Principal components factor analysis (PCFA) (Brown[2], Gerbing and Anderson[13]) • Confirmatory Factor Analysis (CFA) (Brown [2])
Phase 5: Validate constructs (only for reflective)		