Diplomarbeit

Enterprise Risk Management Maturity Assessment – Conceptualization and Operationalization of a Formed Maturity Scale

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines

Diplom-Ingenieurs

unter der Leitung von

Univ.-Prof. Mag. Dr. Walter S.A. Schwaiger, MBA

(E330 Institut für Managementwissenschaften, Bereich: Finanzwirtschaft und Controlling)

eingereicht an der Technischen Universität Wien

Fakultät für Maschinenwesen und Betriebswissenschaften

von

Michael E. Brandstätter

01227256

Nötsch 282

A-9611 Nötsch

Wien, im Mai 2021

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Ich habe zur Kenntnis genommen, dass ich zur Drucklegung meiner Arbeit unter der Bezeichnung

Diplomarbeit

nur mit Bewilligung der Prüfungskommission berechtigt bin.

Ich erkläre weiters Eides statt, dass ich meine Diplomarbeit nach den anerkannten Grundsätzen für wissenschaftliche Abhandlungen selbstständig ausgeführt habe und alle verwendeten Hilfsmittel, insbesondere die zugrunde gelegte Literatur, genannt habe.

Weiters erkläre ich, dass ich dieses Diplomarbeitsthema bisher weder im In- noch Ausland (einer Beurteilerin/einem Beurteiler zur Begutachtung) in irgendeiner Form als Prüfungsarbeit vorgelegt habe und dass diese Arbeit mit der vom Begutachter beurteilten Arbeit übereinstimmt.

Wien, im Mai 2021

Michael E. Brandstätter

Danksagung

An dieser Stelle möchte ich meinen Eltern danken, dass sie mir alle Möglichkeiten geboten und mich meinen Bildungsweg frei wählen haben lassen. Das Wissen, jederzeit ein Sicherheitsnetz zu haben hat mir sehr geholfen und dass sie während meiner Reise durch den Studien-Dschungel niemals nervös geworden sind war ein besonderer Bonus.

Außerdem möchte ich mich bei Herrn Prof. Schwaiger besonders bedanken. Die Ausarbeitung meiner Bachelorarbeit, welche er schon betreute, war die interessanteste Aufgabe des Studiums und führte schließlich zu meiner Anstellung als Projektassistent am Institut, welche wiederum in der vorliegenden Masterarbeit gipfelt. Ich danke ihm für diese spannende Zeit und dass er mir den ersten Karriereschritt ermöglicht hat. Wie Reifegradmodellierer wissen, ist der erste Schritt immer der Wichtigste, denn auf ihm bauen alle weiteren Schritte auf.

Zuletzt möchte ich mich bei meiner Partnerin bedanken, die mich als Freundin nicht nur moralisch unterstützt hat, sondern als Studienkollegin und Lernpartnerin auch inhaltlich eine Stütze im Studium war. Ich möchte mir nicht vorstellen, wie die Studienzeit ohne sie und ihre Eltern verlaufen wäre.

Abstract. To soundly assess the quality of implemented information systems (IS) in MIS research, a measurement construct has to be designed in the first place. Dominated by the social sciences, construct measurement research predominantly takes a realist approach in the conceptualization of the measurement where indicators are identified as reflections of the underlying, latent factor (e.g. IS implementation quality) (Nunnally and Bernstein, 1994; Babbie, 2006; Likert, 1932). Subsequently the measurement model is statistically calibrated and validated with the risk that theoretically meaningful model properties are dropped to achieve high validity scores (Petter et al. 2007, MacKenzie et al., 2011; Rossiter, 2002).

In this work, a constructivist approach is applied in the creation of a formed maturity scale. Here, the rational is that the latent factor is formed by its defining attributes (Rossiter, 2002). By combining modeling principles of the C-OAR-SE procedure (Rossiter, 2002) to conceptually define the rated object (Enterprise Risk Management system) and attribute (maturity), with Capability Maturity Model Integrated (SEI, 2010) design principles, the conceptualization and operationalization of an expert knowledge based formed ERM maturity scale is presented under the umbrella of the Design Science Research Methodology (Geerts, 2011; Peffers et al., 2008).

Keywords: Maturity Model, ERM, Scale Development, Formative Indicators, CMMI, Design Science Research Methodology, Construct Measurement Research, MIS

Kurzfassung. Um die Qualität der in Unternehmen implementierten Informationssysteme (IS) bewerten zu können bedingt es eines allgemein akzeptierten Messmodells, welches typischerweise mit Construct Measurement Research Methoden erstellt wird. Dieser wissenschaftliche Zweig wird jedoch von den Sozialwissenschaften dominiert (Nunnally and Bernstein, 1994; Babbie, 2006; Likert, 1932), welche zumeist Realismus-basiert von der Prämisse ausgehen, dass sich der zu messende latente Faktor in beobachtbaren Indikatoren (reflektiv) widerspiegelt. In der statistischen Kalibrierung und Validierung solcher Modelle werden häufig theoretisch sinnvolle Modellkomponenten zugunsten höherer statistischer Validitätsscores vernachlässigt (Petter et al. 2007, MacKenzie et al., 2011; Rossiter, 2002).

In dieser Arbeit wird die Erstellung einer Reifegradskala präsentiert, welche Konstruktivismus-basiert aus ihren definierenden (formenden) Attributen gebildet wird (Rossiter, 2002). Es wird gezeigt, dass Elemente aus der C-OAR-SE Methode (Rossiter, 2002), zur Konzeptualisierung des zu bewertenden Objektes (Enterprise Risk Management System) und Attributes (Reifegrad der Implementierung) mit Designmethoden des Capability Maturity Model Integrated (SEI, 2010) kombiniert werden können um eine ERM Reifegradskala zu konzeptualisieren und operationalisieren, welche ausschließlich auf Expertenwissen basiert. Diese Fusion sich ergänzender Modellierungsansätze wird nach der strukturierten Design Science Research Methodology (Geerts, 2011; Peffers et al., 2008) präsentiert.

Schlagwörter: Reifegradmodell, Unternehmensweites Risikomanagement-System, Skalenentwicklung, Formatives Messmodell, CMMI, Design Science Research Methode, Managementinformationssysteme

Content

1 Introduction

1.1 Prologue

Let's assume you are a government official in the middle of an economic crisis. Assume further, that the situation is so bad that all businesses except financial institutions need some form of government bailout in order to survive. A federal stimulus package has already been signed by the Prime Minister and although it is the largest package in history, it is by far not big enough to rescue all businesses. Thus, a decision mechanism has to be found how to distribute the money in an ethical and sustainable way. An equal distribution of funds across all businesses is not an option, since the resulting amounts wouldn't be enough to keep the businesses solvent. Long and hard discussions in parliament resulted in the decision that no one industry should be treated as more important than another (since the relative shares of the industries in the market reflect their implied importance for society). At the same time, it is well known that certain businesses are very bad run and wouldn't be able to survive in normal times anyway. After consultation of all relevant experts, the Prime Minister, as the head of the crisis team, announced that the money will be distributed according to the quality of the implemented Enterprise Risk Management System in businesses. It has been brought to him that Enterprise Risk Management facilitates the capacity for businesses to adapt, survive and prosper (COSO-ERM, 2017). The crisis team thinks this is exactly the capacity that is needed in businesses, since the crisis will not end in the foreseeable future.

After this big call was made, the experts went through all prominent Enterprise Risk Management (ERM) guidelines such as COSO II, ISO 31000 and the Three Lines of Defense Model only to realize that they offer valuable guidelines on how to implement an ERM system, but not how to efficiently assess the quality of an implemented system. However, the government needs a valid and reliable assessment tool for ERM quality and they tasked you to come up with a conceptual as well as an operational model for such a tool.

To enable the ERM quality assessment and further the distribution of funds in a timely manner, the crisis team decided that the assessment should be deployed as a web-based self-assessment. The alternative appraisal technique of sending officers with checklists to all business was deemed to be not viable. However, since Chief Executives may be prone to lie in order to get their businesses funds needed for survival, high attention must be put to develop a self-assessment that is based on objective evidence, i.e. on facts that can be checked in an on-site audit at a later time. Vague and to a certain degree subjective measurement items that allow for different interpretations must be strictly eliminated.

To give you a kick-start into your construct measurement endeavor an expert from the Technical University advises you to study the Capability Maturity Model (SEI, 1993). He said that the complex measurement task of ERM quality reminds him of the situation the US Defense Dept. faced in the 1980s when they needed to assess the quality of the software development process in companies in order to determine which company is best equipped to develop military software. Their solution was to develop a model comprising five levels of process maturity…(Humphrey, 1988)

1.2 Development of a Formed Maturity Scale

The prologue should help to outline the objectives of this work. Even if it is first and foremost for scientific purposes, the requirements described above are pretty much the same.

At the heart lies a construct measurement task, or to be more specific, an attribute of an object shall be measured (cf. Rossiter, 2002). The object could e.g. be an ERM system and the attribute would be its implementation quality or maturity, as in fact the attribute will be referred to in this work. Furthermore, the measurement should be conducted as a self-assessment, hence the rater entity is the business itself (cf. Rossiter, 2002), e.g. in the form of an all-knowing CEO.

Such an objective is in no way new in maturity modeling (de Bruin et al., 2005; Enkel et al., 2011; Monda and Giorgino, 2013; Schumacher et al., 2016) but the conceptualization as well as the operationalization of most maturity models in the literature falls short of reaching the full potential of maturity models as a (1) comparative model, where differences in ratings can be unambiguously interpreted and as a (2) prescriptive model that serves as a roadmap for improvement (de Bruin et al., 2005; SEI, 2010). Furthermore, these "modern" maturity models deviate to a great extent from the design principles of the original Capability Maturity Model (CMM, (SEI, 1993)) and its successor the Capability Maturity Model Integrated (CMMI, (SEI, 2010)) without justification in terms of construct measurement research and scale development.

The practices of construct measurement research and scale development will be used in the conceptualization and operationalization of the aspired measurement model. A decisive decision at the very start of this process is the determination of the epistemic relationship between construct and items, i.e. the nature and direction of the relationship between the theoretical construct and its measurable empirical phenomena (Bisbe et al., 2007). This is so important because it sets the philosophical measurement perspective and must be justified on conceptual grounds (Bisbe et al., 2007). Depending on the overall measurement objective a reflective relationship between construct and measurement items (realist approach) or a formative relationship between construct and measurement items (constructivist approach) may be appropriate (Box 1). However, as will be pointed out in multiple places in this work, especially in chapter 2, traditional construct measurement research practices are tailored to reflective models and are not appropriate for formative measurement (Edwards, 2011). Instead, the traditional construct measurement practices have to be evolved as in the C-OAR-SE (Construct definition, Object classification, Attribute classification, Rater identification, Scale formation, and Enumeration and reporting (Rossiter, 2002)) procedure to accommodate formed measurement models, where statistical model validation is dropped in favor of expert judgement.

This work demonstrates the development of a formed Enterprise Risk Management maturity scale. The development is conducted via the Design Science Research Methodology (DSRM, chapter 4) (Geerts, 2011; Peffers et al., 2008) which serves as the umbrella under which the abstract modeling concepts of construct measurement research, scale development and maturity modeling are applied to the specific ERM domain. For the sake of reaching maximum clarity in the elaborations, the exact sequence of steps in the DSRM process (Figure 1) will not always be followed exactly in this presentation. Deviations, however, will be highlighted and explained where appropriate.

Nevertheless, before some sort of developmental work can start, as a prerequisite the concepts of construct measurement research (chapter 2) and scale development (chapter 3) need to be discussed.

Since the objective is to develop a *formed* maturity scale, construct measurement research literature will be used to define a formed measure and distinguish it from the more popular reflective measures prevalent in the social sciences, which is the primary application field of construct measurement research (Nunnally and Bernstein, 1994; Babbie, 2006; Likert, 1932). To be more specific, elements of the C-OAR-SE procedure will be used to conceptually define the object (ERM system) and attribute (maturity) for the concrete instance of the formed maturity scale. A central step in C-OAR-SE is to determine the nature of the object and attribute that shall be rated, which is dependent on the overall objective of the measurement and includes determining the epistemic relationships of the model. Based on the specific objectives outlined in the prologue, it will be shown that ERM is best classified as a formed object, comprising multiple components and the attribute maturity is best classified as a formed attribute, where again the complex attribute's different components add up to what the attribute means (Rossiter, 2002).

Once it is clear that the concept is of such complexity (formed object and formed attribute), the aggregation of attribute scores to an overall *scale* score for the object (or for each object component) must be conducted very carefully. Different to index scores, where usually a summation of compensatory attribute scores is derived (Rossiter, 2002), a scale score of formed attributes is typically the result of noncompensatory attribute scores, where a conjunctive rule (e.g. a minimum level for each component) is needed in the aggregation (cf. Rossiter, 2012). More details on the distinction between indexes and scales will be presented in chapter 3.

Only one model is known to the author that represents a prescriptive measurement scale based on formed attributes. Surprisingly it is the origin of all maturity models, the CMM (SEI, 1993) and its successor the CMMI (SEI, 2010). Maturity models have emerged in all kinds of domains (cf. de Bruin *et al.*, 2005), but the deviation in design from the original CMM is noteworthy. Even in guidelines on how to develop maturity models, the design principles that assure scale property of the model are thrown out of the window (see de Bruin *et al.*, 2005). Maturity models are often conceptualized based on the construct measurement research principle of the social sciences where they have reflective measurement in mind. But unwittingly these principles were applied to formative models which resulted in bad maturity indexes, at best, not fit for purpose of the intended assessment.

Especially the conceptualization of the CMMI takes some time to comprehend completely and maybe time constrains hindered maturity model developers to fully go through its documentation. It's worth mentioning that the existence of the CMM and CMMI is well known in the maturity modeling domain, since every paper lists at least one of them in its references. Although the conceptual model of the CMMI is comprehensively documented, it is never explicitly specified in the light of construct measurement research. This could be another reason why its design principles are not followed by other maturity model developers. But the last point is almost certainly the cause why the CMMI design is not discussed in construct measurement research, where there seems to be an ongoing discussion if and how such a model (a multi-dimensional formative (formed) maturity scale) can be constructed (Petter et al., 2007; Edwards, 2011; MacKenzie et al., 2011).

After the theoretical foundations are discussed in chapter 2 and chapter 3 and the methodology of this work is presented in chapter 4, the CMMI-design principle for maturity models is analyzed and discusses it in the light of construct measurement research (chapter 8). The analysis should help in future maturity model development tasks since it (1) highlights the shortcomings of modern, Likert-scale based maturity models (chapter 7), (2) presents the CMMI-design principle to construct formed maturity scales not only in a parsimonious way, but also in the scientifically standardized terms of construct measurement research and (3) demonstrates the CMMI-design principle in the construction of a formed maturity scale for Enterprise Risk Management systems (chapter 9). Furthermore, the operationalization of the ERM maturity model as a web-based self-assessment is presented as proof of concept.

As shall be shown, the operationalization of the maturity assessment deviates largely from the operationalization of the CMMI model, where thorough on-site audits are performed for appraisals. For the specific purpose outlined in the prologue and for most scientific endeavors, however, a web-based assessment is the more viable option than expensive and time-consuming on-site audits.

2 Construct Measurement Research

A measurement task starts with the definition of the concept that should be measured (Babbie, 2006). In the case at hand, this is the ERM system. The scientific literature in the ERM domain (i.e. the object domain) supplies all the building blocks needed to conceptually define the object. But for complex systems this can easily result in 100+ measurement items, since the ERM literature serves as a very broad pool of requirements for ERM systems. The problem, however, is that the relationship between the requirements is not always well defined. For the desired measurement, a conceptual model is needed that somehow orders all the requirements with respect to the attribute of the object. But the ordering of requirements may not automatically result in an ordinal sequence. Some requirements may have a nominal relationship with respect to the attribute. An example for two ERM system requirements with a nominal (orthogonal) relationship could be:

- Risk Information shall be taken into account in financial decision making.
- Risk Information shall be taken into account in operational decision making

For an arbitrary company, it is impossible to say that one requirement is more important in terms of ERM maturity. Neither it makes sense to postulate that one is kind of a prerequisite for the other. Hence, the object domain alone is not enough to conceptually define a measurement model of an object's attribute.

One the other hand of the measurement task, there is construct measurement research. This scientific stream deals with the conceptualization of a construct. I.e., the definition of *a specific, agreed-on meaning for a concept for the purposes of research. This process of specifying exact meaning involves describing the indicators we'll be using to measure our concept and the different aspects of the concept called dimensions.*(Babbie, 2006, p. 125)

Construct measurement research is dominated by the social sciences, where for example measures of certain character traits or attitudes in people are constructed (Likert, 1932; Nunnally and Bernstein, 1994; Babbie, 2006). Nevertheless, construct measurement research also finds broad application in Management and Information Systems research (Diamantopoulos, 2005; Bisbe et al., 2007; Petter et al. 2007; Rossiter, 2008; Edwards, 2011; MacKenzie et al., 2011).

In the social sciences the conceptualization process is normally guided by the paradigm that the construct is a theoretical creation that is based on observations, but cannot be observed directly (cf. Babbie, 2006, p. 124). In other words, the construct (e.g. religiosity) is a latent variable (character trait) that manifests itself in certain observable (!) behavioral patterns. The observable behavioral patterns would subsequently be conceptualized as the indicators to measure the latent variable. Babbie's (2006, p. 125) definition of indicators in his book "The Practice of Social Research" underlines this dominant conceptualization approach in the social sciences that the indicators are merely reflections of the construct: [An indicator is] *an observation that we choose to* *consider as a reflection of a variable we wish to study. Thus, for example, attending religious services might be considered an indicator of religiosity.*

Coming back to the measurement task of ERM maturity, reflective measurement would mean that it is assumed that there is a ERM system somewhere hidden in a company, and its quality level is e.g. reflected in different publicly available Key Performance Indicators. While this may be an interesting investigation, especially in times where businesses are not willing to disclose information via surveys, some effort is required to proof the validity of such a measure.

Especially for Management research another form of epistemic relationship between construct and indicators is an interesting alternative for the conceptualization. In formative measurement models, the construct is defined as nothing else than the combination of all defining indicators.

In fact, the ERM literature can be seen as a collection of the main building blocks of an ERM system. While an ERM system is not something that naturally exists in companies, it is rather a theoretical creation formed by all necessary risk management activities. An ERM quality assessment based on determining the presence or absence of these building blocks should naturally represent a theoretically valid measurement based on expert knowledge (cf Petter et. al., 2007; MacKenzie et al., 2011; Rossiter, 2002).

Reflective Measurement: indicators are the outcomes of a latent variable
Formative Measurement: indicators are the causes of a latent variable
(Edwards, 2011)

Box 1: Reflective vs. Formative Measurement

However, in construct measurement research the interplay between domain knowledge and scale development is typically of the form that domain expertise is only used to identify potential indicators of the construct. In the next step, the measurement model is calibrated and validated based on empirical observations (Diamantopoulos and Winklhofer, 2001, Petter et. al., 2007; MacKenzie et al., 2011; Edwards, 2011), where theoretically valid indicators might be dropped from the measurement model only to achieve high validity scores (Rossiter, 2002)[1].

First, this procedure is problematic from the outset for formative measurement models, since it results in measurement scores that are not interpretable due to non-unidimensional indicators (Edwards, 2011). And second, formative measurement models which are used to measure an attribute of a formed object – that is, an object having different components (Rossiter, 2002) – should be identified, calibrated and validated using expert judgement only (Rossiter, 2002). In this process, experts determine the main components of the construct, which can be seen as different dimensions that need to be measured separately.

[1] A more detailed elaboration on empirical identification of measurement models can be found in chapter 12.1 in the appendix.

Given the volume of ERM frameworks and dedicated papers in the literature, as well as the changing focus on different aspects of ERM in the publications, there is no question that people's interpretations of the object differ, which is an indication that ERM should be seen as an abstract formed object (Rossiter, 2002).

Hence, reflective measurement can be eliminated as a viable alternative for the measurement task. Formative measurement presents the right perspective in terms of the epistemic relationship between construct and measurement items, but comes with the burden of empirical identification (Diamantopoulos and Winklhofer, 2001, Edwards, 2011). Thus, the measurement task at hand has to go a step further and adopt elements of the C-OAR-SE procedure (Rossiter, 2002) in terms of object and attribute classification. That is the conceptualization of ERM as an abstract formed object, developed and validated on domain/expert knowledge only.

3 Scale Development

Once the decision regarding the epistemic relationship between the construct and its indicators is made in favor of formative measurement and the object is well defined, the conceptualization needs to be driven further with careful consideration of the overall objective of the assessment tool.

3.1 Scale vs. Index

Remember that a comparative measure is needed by the government in order to determine which business is better suited to survive in the economic crisis and thus worth rescuing. At the same time, the measure should be descriptive in the way that a score can be directly associated with a state that the ERM system is in, so that it can be explained to businesses for which reason one is doing better than the other. In other words, the assessment result should be unambiguously interpretable (Edwards, 2011).

Hence, it is not enough to rank order all business in terms of some quality score of their implemented ERM system. Different than an index score of "18 out of 30" – which can be attained in multiple ways – a scale score should only be attainable in one particular way, or from one particular state of the system, respectively. These specific states, which map one-to-one to the respective scale scores need to be defined by particular patterns of indicator scores, i.e. the states of the system are defined by performing an indicator grouping. The grouping of indicators is needed, since in the end *indexes and scales (especially scales) are efficient data-reduction devices: They allow us to summarize several indicators in a single numerical score, while sometimes nearly maintaining the specific details of all the individual indicators* (Babbie, 2006, p. 153). With this property in mind, it is easy to see their natural appeal to top level executives and, of course, the Prime Minister. Since the *terms index and scale are typically used imprecisely and interchangeably in social research literature* (Babbie, 2006, p. 153) the defining differences should be elaborated at this stage to highlight the benefits of a measurement construct possessing scale characteristic.

Indexes and scales are both ordinal measures of variables (thus they rank-order the units of analysis). However, while indexes are constructed by accumulating scores assigned to responses to individual indicators, scales are constructed by assigning scores to certain patterns of responses. In scales, it is recognized that some indicators represent a relatively weak degree of the variable, while others represent a relatively strong degree of the variable (cf. Babbie, 2006, p. 154). Hence, in scales the relationship between the indicators of the construct is taken into account. Logical response patterns are predetermined and respondents are scored based on which response pattern they most closely resemble (cf. Babbie, 2006, p. 154). Depending on the nature of the object and attribute under investigation, a scale or an index may be the more suitable form of composite measure.

If, for example a monitoring measure for the growth rate (attribute) of the German economy (object) should be constructed, a share index as an aggregation of weighted share prices may be appropriate, where the weighting does reflect the (proposed) relative contribution of an individual share to the overall measure, but does not reflect the relationships between individual shares.

On the other hand, if a measure for math skills (attribute) of pupils (object) is needed, it would be natural to develop a scale with indicators for

1) "addition"
2) "subtraction"
3) "multiplication" and
4) "division".

The logical response patterns would only allow for answers that include lower order indicators when higher order indicators are present. To further underline the difference between indexes and scales, the example presented in Babbie (2006, p. 155) regarding an index and a scale for political activism is presented in Figure 24 in the appendix.

3.2 Maturity as a Formed Attribute

Similar to the evolution of math skills as an attribute of pupils, the implementation quality of information systems in businesses can be conceptualized as an evolving attribute, which aids the development of a true scale measure. Therefore, the term maturity as a means of competency, capability, level of sophistication (de Bruin *et al.*, 2005) should be the right term to refer to this attribute. I.e. with "a more mature system" a system that is superior in every aspect compared to a less mature system is meant. This wording serves to (1) express exactly what is wanted from the aspired scale scores: they should represent levels of maturity of the assessed system; and (2) overtly bring the elaborations from the scientific perspective of scale development to the more specific stream of maturity modeling.

Nevertheless, the attribute maturity shall be classified more precisely here, again adopting the C-OAR-SE classification schema (Rossiter, 2002). Elaborations on the exact attribute type are necessary because the attribute classification is not straightforward since it depends on the broader context of the construct.

The simplest case would be to have a concrete attribute, where raters agree unanimously what the attribute is. In this case, only one measurement item for each component of the object would be necessary. Since (ERM) maturity, even if the defining components (dimensions) of the ERM object are well defined, may highly likely result in different interpretations amongst raters without further clarification, the requirements for a concrete attribute are not met and so another attribute type has to be considered.

Formed attributes allow the attribute maturity to be formed (similar to formed objects) as the product of its components. These components are attributes themselves, i.e. second-order attributes. The component of the component, i.e. a specific item of the attribute maturity must be concrete in order to enable consistent ratings (cf. Rossiter, 2002, p.314). A good example would be to define sub-attributes of the formed attribute maturity as concrete, binary measurement items.

The combination of object and attribute to the overall concept can be thought of as a two-dimensional classification schema, where the main object components (of formed objects) represent one dimension as the entities that shall be rated. These entities are projected onto an attribute, which represents the orthogonal dimension. Ratings eventually might be the result of responses to one attribute (concrete) or multiple components of the attribute (formed). Figure 18 depicts such a high level (first order) schema of the ERM maturity model as a combination of an abstract formed object (ERM with sub-dimensions) with formed attributes (maturity levels).

In fact, the maturity levels represent a grouping of attributes that is isolated from the C-OAR-SE conceptualization. The attribute grouping to maturity levels is the result of expert judgements and serves to facilitate true scale characteristics of the measurement construct. It is an enhancing property that maturity models naturally offer to derive scale scores from formed measurement models, as will be shown for the CMMI (SEI, 2010).

3.3 Prescriptive Maturity Scales

In the fictive scenario of the prologue, companies would only be too happy to participate in an survey-esque quality assessment, since their survival depends on government funds. In the scientific reality, however, other incentives have to be found to motivate businesses to disclose information about their risk management activities. An assessment based on a formed maturity scale not only produces comparative (how is one company doing compared to another) and descriptive assessment results (presenting the as-is situation), it also allows for prescriptive assessment results (what exactly needs to be done to reach the next maturity level, i.e. the model represents a roadmap for improvement) (de Bruin et al., 2005).

Many "maturity models" found in the literature do produce maturity scores that rather resemble an index (de Bruin et al., 2005; Enkel et al., 2011; Monda and Giorgino, 2013; Schumacher et al., 2016). Typically, they identify all necessary indicators to measure the concept. Next, they introduce a scoring schema via which each indicator is scored during an assessment (more often than not, the scoring is operationalized via Likert-type questions, this important point will be discussed later) and at the end the indicator scores are irreversible aggregated into a single overall figure. A valid question regarding the outcome of such an assessment would be: "So, Company A has a score of 18 out of a possible 30 … what does it mean?". Without going through the assessment process again, the only possible interpretation is: "Well, it is more than Company B which has 17 and less than Company C with 19" (for a discussion on the interpretability of fomative index scores see Edwards, 2011).

To aid the interpretability of maturity scores, the inherent scale property of true maturity models must be made use of. The scale property comes from the characteristic of the original CMM, that the maturity levels represent a natural development path, describing continuous evolutionary improvements (Humphrey, 1988). A measure like a maturity score loses scale property whenever some kind of (weighted) average over indicator scores is abstracted, unless (!) the indicators are in a reflective relationship to the concept. Why this is will be explained later, but for now it is important to remember that for the measurement task of the Prime Minister, reflective measurement is not desirable for reasons described earlier and statistical operations to aggregate indicator scores are not desirable since the resulting aggregate loses expert knowledge and its theoretical interpretability.

To sum up, the preceding discussion should help to frame the problem of developing a maturity scale based its forming components and attributes. The relevant scientific streams and most important model characteristics have been stressed. Moreover, a fictive scenario helped to colorfully describe the wider context on how such a maturity scale may be used as an assessment tool. Important restrictions stemming from the intended usage of the assessment results were highlighted, which need to be considered at the conceptualization phase already. The major criterions for the aspired measurement model shall be summarized here:

- Businesses shall be rank ordered with respect to the maturity level of their ERM system.
- Differences between businesses on different maturity levels shall be interpretable instantly and unambiguously.
- Maturity levels shall be interpretable instantly and unambiguously.
- The maturity measurement shall be deployed as a web-based self-assessment.
- The validation of the model shall be based on theoretical reasoning only.
- The reliability of the model shall be enhanced via very precise, evidence-based attribute specifications, i.e. a respondent shall be lead through the assessment by leaving no room for interpretation or personal preference whatsoever
- The model shall be applicable to all industries except financial institutions, this needs to be especially considered when trying to combine different ERM aspects.

Box 2: Assessment Tool Requirements

A model that satisfies all these criterions is missing in the ERM domain (Beasley et al., 2005; Monda and Giorgino, 2013; Lundqvist, 2015), more general maturity modelling domain (de Bruin *et al.*, 2005; Enkel et al., 2011; Schumacher et al., 2016) and is even argued against in the construct measurement research domain (Edwards, 2011). Nevertheless, advocates of formative measurement can be found in the management sciences (Diamantopoulos and Winklhofer, 2001; Petter et al., 2007; MacKenzie et al., 2011), even though they try to construct formative models with methods more appropriate for reflective measurement. The big exception is Rossiter (2002) who promotes formed measurement constructs based on theoretical considerations without the need for empirical calibration and validation[2].

[2] Henceforth, in line with Rossiter (2002), the term attribute shall be used to refer to the items defining formed attributes. The term indicator shall only be used when referring to items of reflective models. For general elaborations, the perspective neutral form item will be used.

4 Create Measurement Artefacts via Design Science Research

This work is associated with a research effort undertaken by the Institute of Management Science at the Technical University of Vienna. Research Projects sponsored by the Funk Foundation[3] (Hamburg) aimed at developing a web based ERM maturity assessment which could be applied in a use case, e.g. such as the one outlined in the prologue. The development of the maturity assessment followed the scientific method of design science research. *Whereas natural sciences and social sciences try to understand reality, design science attempts to create things that serve human purposes* (Simon (1996, p. 55) as cited in Peffers *et al.* (2008, p. 4)). As such, design science research focuses on the question of how things should be in order to attain goals or how to change existing situations into preferred ones (cf. Simon, 1996; Geerts, 2011). Design science research creates artefacts (e.g. concepts, models, methods, instantiations) that help to solve an unsolved problem or to solve an already solved problem in a more effective or efficient way (Geerts, 2011). The artifact that resulted from the research effort undertaken by the Institute of Management Science was the ERM maturity assessment, which's conceptualization and operationalization are central to this work. However, to grasp the whole process of the design science research methodology (DSRM), which is depicted in Figure 1, it is important not to look at this work in isolation, but to see it in the context of the whole research effort.

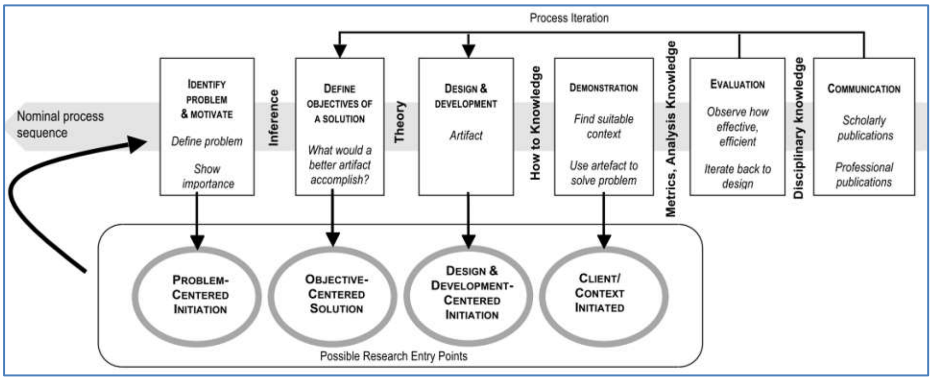

Figure 1: Design Science Research Methodology (DSRM) Process Model (Figure 1 in Peffers *et al.* (2008))

Going with the story of the prologue, the trigger for the design science exercise is problem centered, since – step 1 – the need and importance of an ERM assessment tool are clearly defined. The objectives of a solution – step 2 in the DSRM process – are specified in Box 2. Design and development principles – step 3 – of domain independent maturity assessments will be comprehensively elaborated in the next chapters and represent the core contribution of this work. For demonstration purposes – step 4 – the research project's ERM maturity assessment will be presented as proof of concept. The evaluation – step 5 – will be centered around other maturity models found in the

[3] The Funk Foundation | Risk Management and Cultural Projects - (funk-stiftung.org)

literature, in order to show their shortcomings and/or valuable characteristics that need to be embedded in the aspired maturity model.

To better lead the reader through the complex elaborations and to tell an interesting story the exact sequence of the DSRM steps will not be followed in this presentation. For example, the evaluation step will be the next stage presented in this work, when the DSRM process will be re-entered in chapter 7, after details about the accompanying scientific fields of construct measurement research and maturity modelling are presented in chapter 5 and chapter 6, respectively. This is done because the pitfalls should be clear for the "first (and only) iteration" of design and development in this work. Nevertheless, it should be mentioned that the ERM maturity assessment had to go through multiple loops of iteration during the research project until the design was satisfactory.

The last step in the DSRM process is communication, which is exactly what this paper represents and why it is important to know the context of the whole design science research undertaking.

5 Conceptualization and Operationalization

5.1 Defining the Terms

The interrelated processes of conceptualization, operationalization, and measurement allow researchers to move from a general idea about what they want to study to effective and well-defined measurements in the real world (Babbie, 2006, p. 149).The role of conceptualization, i.e. the precise specification of the construct of interest in terms of its meaning, its potentially different facets (dimensions) and measurement variables (indicators/attributes, or the perspective-neutral form item) has already been touched in the introduction. It was pointed out that not only the identification of the items itself, but also the determination of the epistemic relationship between construct and item is important since the two possible directions of causality represent inherently different measurement perspectives.

The operationalization of the measurement, on the other hand, *specifies precisely how a concept will be measured – that is, the operations that will be performed. An operational definition* […] *has the advantage of achieving maximum clarity about what a concept means in the context of a given study*. (Babbie, 2006, p. 128) The operationalization of one measurement variable, i.e. one item of the construct, would for example involve specifying the levels composing the variable[4]. An operational choice that has to be made in that sense is, if the item should be operationalized as a binary variable, i.e. the measurement is only concerned to determine the presence or absence of the item. Another very popular alternative in construct measurement research is to operationalize items as ordinal variables (Likert, 1932). In this case the measurement is intended to capture different degrees of intensity of the variable. To give another example of an operationalization alternative, e.g. in quantitative research a third type of item operationalization may be an attractive choice, namely ratio variables (Altman, 1968).

Operational definitions of measurement models go one step further though, than the pure specification of item levels. They also specify the measurement procedure that will be used to measure variable's level (Babbie, 2006). Since the measurement model, which is subject to this study, should be deployed as a web-based self-assessment, the measurement procedure involves answering written questions. The wording of the questions is thus part of the construct's operationalization, and depends on the item type (e.g. binary vs. ordinal vs. ratio).

The following sub-chapters will show how the process of conceptualization and operationalization can be applied in the management sciences to derive at a sound measurement model. Especially the choices necessary to specify the epistemic relationship between the theoretical construct and its items, as well as the operationalization of the items to specify the measurement procedure will be discussed in depth. This discussion

[4] In this context the levels of a variable can be seen as the different response options to a question. In the wording of Babbie (2006) the question would represent an indicator and the different response options would represent the attributes of the indicator. This wording, however, clashes with the attribute definition of Rossiter (2002) and is thus avoided in this work.

will provide the terms and definitions used later to discuss maturity modelling in the light of construct measurement research.

5.2 Construct Measurement Research in the Management Sciences

As already mentioned, construct measurement is a topic dominated by the social sciences and adopted to research into management constructs (e.g. MIS, MACS) (Bisbe et al., 2007; Petter et al., 2007; MacKenzie et al., 2011). Especially Bisbe et al. (2007) stresses the need for a proper conceptual and operational specification of the construct in theory-based research. To better illustrate the relationship between conceptualization and operationalization of a construct, the Predictive Validity Framework (PVF) is presented in Figure 2 (Libby et al., 2002; Bisbe et al., 2007). It shows the way how to come from a theoretical construct such as ERM with all its apparent merits postulated by prominent ERM frameworks (ISO 31000, 2009; IIA-3LoD, 2013; COSO-ERM, 2017) to an empirical construct such as the ERM maturity model presented in this paper, using the methods of construct measurement research.

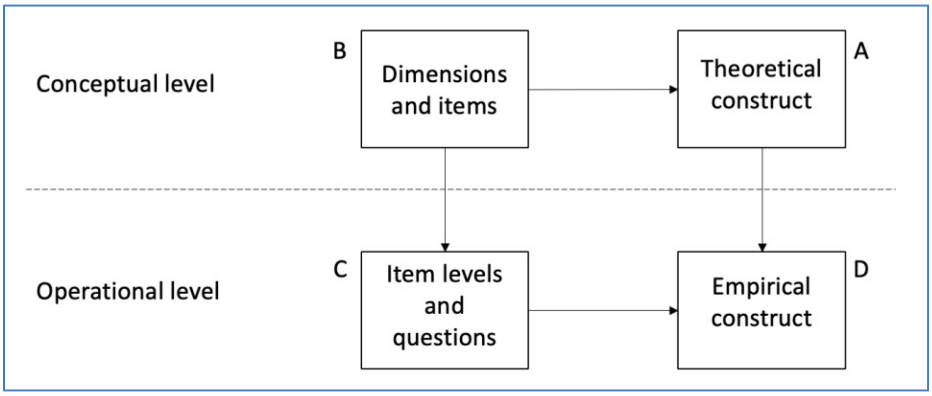

Figure 2: Predictive Validity Framework (Libby et al., 2002; Bisbe et al., 2007)

5.2.1 Use Cases of the Empirical Construct

Before the specific operations performed in the conceptualization and operationalization of a theoretical construct are discussed, the overall objective of construct measurement shall be framed, i.e. the potential use cases of the empirical construct should be presented.

The empirical construct can be used as a tool on its own, such as the aspired ERM maturity assessment tool for the Prime Minister. But the scientific endeavor does not have to stop there. For many studies the empirical construct only serves as a vehicle to perform the main inquiry, namely some form of causal analysis between the construct and other metrices. From a statical perspective, two causal analysis purposes can be distinguished using the empirical construct:

1. Reflective Purpose: The empirical construct can be used as an outcome proxy to statistically identify it's drivers. (Beasley, Clune and Hermanson, 2005)
2. Formative Purpose: The empirical construct can be used as a driver proxy and its impact on another metric can be statistically assessed. (Göstl, 2017)

Referring to the story of the prologue, the causal relationship postulating that ERM maturity is a driver of sustainable firm performance is taken as given, but if circumstances wouldn't be so critical and the distribution of funds less urgent, the proof of the causal relationship between ERM maturity and firm performance would be subject to a follow-up study before funds are allocated on this basis[5].

Since the empirical construct (as the result of the construct measurement effort) itself is not the central scientific contribution in many studies, but only one block in a structural equation model (SEM) that analyzes relationships between different constructs, it is sometimes not getting enough attention in terms of careful conceptualization and operationalization (e.g. Gordon et al. (2009); Lundqvist (2015)). It is pointed out by Bisbe et al. (2007) though, that conceptual misspecification of the empirical construct undermines the value of the resulting explanatory model.

To round up the discussion of use cases of conceptually sound empirical constructs, the following two notes should be added:

- Note 1: To avoid the circular reasoning problem, it is very important to mention that for whichever further purpose the empirical construct is used, it must be made sure that items which were used to construct the construct in the first place are not used in any causal analysis afterwards.
- Note 2: The terms formative and reflective in the potential further analysis purposes are representative for the (new) conceptual model of the follow-up study (i.e. the explanatory model), but are not dependent on the formative or reflective measurement perspective chosen to derive the empirical construct in the first place. Still, for the taken measurement perspective in the follow-up study, the definitions for formative and reflective measurement elaborated earlier remain valid.

5.3 From the Theoretical Construct to Measurement Items

As shown in Figure 2 the conceptualization starts with the definition of the theoretical construct (Box A) and continues with the specification of the construct's dimensions[6] and items (Box B). For simplicity, the elaborations should focus on one-

[5] To save resources, one might be tempted to say the causal relationship between ERM maturity and sustainable firm performance should be investigated before efforts are put into the construction of an extensive ERM maturity assessment. But since a sufficiently valid and reliable proxy of ERM maturity is missing in the literature, there is no value in analyzing the relationship based on a purely constructed ERM maturity measure.

[6] *does the construct have more than one conceptually distinguishable facet, aspect* […]? (MacKenzie et al., 2011, p. 300)

dimensional constructs at this point, i.e. we are only concerned to specify a theoretical constructs in terms of its items. A construct's items are identified based on domain knowledge (which is especially the case for formed attributes) and/or based on statistical grounds. However, item identification is not at the center of attention of this work. Together with the item identification, a very important step in the conceptualization is to determine the direction of the relationship between constructs and items (Diamantopoulos and Winklhofer, 2001; Bisbe et al., 2007; Petter et al. 2007; Edwards, 2011; MacKenzie et al., 2011).

One option is to view the construct as an underlying factor (latent variable) causing variation in the item/indicators. Hence the assumption is that an object's score on an indicator is a function of the object's real score on the latent variable plus error. Each indicator is therefore viewed as an imperfect *reflection* of the latent variable (see MacKenzie et al., 2011). It is important to notice, that in reflective measurement unidimensionality is assumed, consequently each indicator must capture the entire construct. Aggregating the indicators to abstract the score on the latent variable is therefore essentially an attempt to reduce the measurement error that would accrue if the construct would be measured via one indicator only (Edwards, 2011).

Formative or formed measurement, on the other hand, assumes that meaning is emanating from the items/attributes to the construct in a definitional sense. Hence, attributes are not reflections of the construct, but instead they combine to *form* the construct (see MacKenzie et al. 2011). This allows the conceptualization of constructs that are combinations of different components, where each component is covered by an individual attribute.

In the introduction, numerous reasons were elaborated why the aspired ERM maturity model should be constructed as a formed measurement model. When the characteristics of true maturity models will be discussed in the remainder of this work, even more arguments will be given why it is beneficial to have a formed model, always keeping the requirements outlined in Box 2 in mind. Nevertheless, a short demonstration for the need of a formed, theory based ERM maturity model shall be presented here.

For clarity, the demonstration should focus on one aspect of an ERM system only, the Risk Management Process as defined by the International Organization for Standardization (Figure 3).

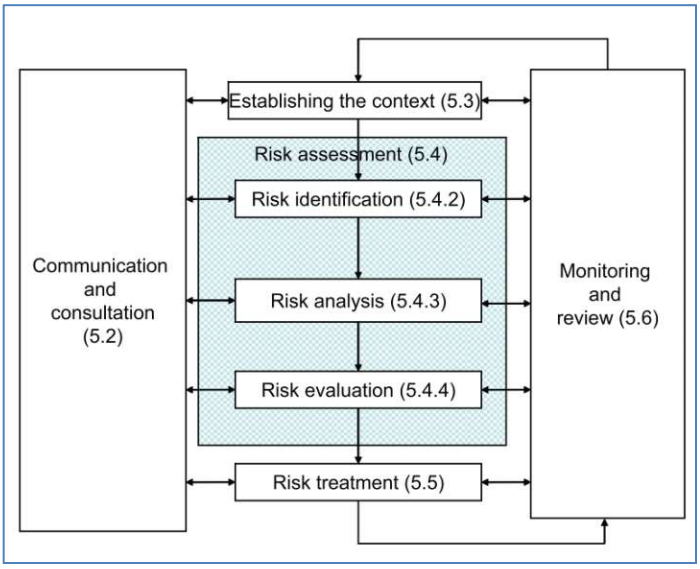

Figure 3: Risk Management Process (Figure 3 in ISO 31000 (2009))

ISO 31000 (2009) defines the core of the risk management process, called risk assessment as a succession of three tasks:

- Risk identification (ISO 31000, 2009, p. 17):

The organization should identify sources of risk, areas of impacts, events (including changes in circumstances) and their causes and their potential consequences. The aim of this step is to generate a comprehensive list of risks […]. *Comprehensive identification is critical, because a risk that is not identified at this stage will not be included in further analysis.*

- Risk analysis (ISO 31000, 2009, p. 18):

Risk analysis involves developing an understanding of the risk. Risk analysis provides an input to risk evaluation and to decisions on whether risks need to be treated, and on the most appropriate risk treatment strategies and methods.

- Risk evaluation (ISO 31000, 2009, p. 18):

The purpose of risk evaluation is to assist in making decisions, based on the outcomes of risk analysis, about which risks need treatment and the priority for treatment implementation.

To create a measurement construct for risk assessment according to the ISO 31000 (2009) definition, clearly a formed measurement perspective has to be adopted. In the provided definitions it is clear to see that each of the three steps (risk identification, risk analysis and risk evaluation) is necessary to accomplish an appropriate risk assessment. The items are neither interchangeable (e.g. substituting risk identification for another item like "risk definition") nor is each item intended to capture the whole concept of risk assessment, both requirements for a reflective measurement perspective. The example is also proof of the point Petter et al. (2007, p. 633) make when they argue that *…with formative measures, dropping one of the measures would affect the meaning of the construct since the construct is defined by these measures*.

As a small teaser of what to expect in one of the next chapters involved with discussions about the nature of maturity models, the attention should be drawn to the box "monitoring and review" in Figure 3. The figure indicates that a monitoring and review process is somehow associated, but not part of the risk assessment process. In fact, the monitoring is an extension of the risk assessment. A monitored risk assessment process is therefore a concept possessing all properties of a pure risk assessment process, but extended for a monitoring layer. This constellation exhibits exactly how maturity levels work. A concept (e.g. a system) in a lower maturity level is defined by certain properties (measured via the identified items), and the system in a higher maturity level maintains all the properties, but is enriched by something else (captured by one or more additional items), making it a superior system in every sense.

To close the argument for formative conceptualization of the aspired ERM maturity assessment at this point, it should be mentioned that a formed model is compatible with a constructivist, operationalist or instrumentalist interpretation in an ontological sense (MacKenzie et al., 2011), which is perfectly in line with the methods of design science research, which focuses on the creation of artefacts that serve practical purposes (Geerts, 2011). Referring to multiple sources, Edwards (2011, p. 372) annotates that *formative measurement is consistent with a constructivist position* […] *in which constructs are viewed as elements of language in theoretical discourse and are not ascribed any real existence independent of their measurement* […]. *Formative measurement might also be framed in terms of operationalism or instrumentalism, such that constructs are merely latent variables that serve as analytical devices for combining measures…*[7].

5.4 Operationalization of Measurement Items

One of the big challenges in the conceptualization of a theory-based measurement model is to structure the whole body of domain knowledge found in the literature, while always keeping a careful eye on the intended meaning of the concept. The counterpart of this challenge in the operationalization is to carefully define the measurement procedures, while always take the intended application of the empirical construct into consideration. The main requirements for the application of the measurement model, with

[7] The following statement, that formative measurement is akin to data reduction in Principal Component Analysis (PCA) in Edwards (2011) is not supported by the author of this work. In fact, the author would much rather liken reflective measurement to PCA.

respect to the operationalization were already specified in the introduction (Box 2) and are presented again in Box 3 for recollection.

- The maturity measurement should be deployed as a web-based self-assessment.
- The reliability of the model shall be enhanced via very precise, evidence-based attribute specifications, i.e. a respondent shall be lead through the assessment by leaving no room for interpretation or personal preference whatsoever

Box 3: Main Operational Requirements

The operationalization, remember, is concerned with the specification of an items levels (e.g. gender used as an item would be composed of the levels female and male (Babbie, 2006)) and the procedure how to determine which level of the item is present (e.g. posing a question). This operation is represented by Box C in Figure 2.

The first bullet point in Box 3 leads to the conclusion that the measurement itself is conducted by answering written questions on the computer. Hence, the measurement is not conducted via an interview, where ambiguities can be followed up in a natural conversation, nor is the measurement conducted via an on-site audit, where points on a check list can be investigated in depth. Therefore, the validity and reliability of the self-assessment has to be assured via a precise operationalization of evidence-based items, where it must be possible to demand proof of each answer. If there is ambiguity or space for subjective interpretation in the question, assessment results are immediately compromised with no chance for clarification.

A question is valid if it is measuring what it is intended to measure and reliable if it gives the same result dependently, e.g. a factual question is answered in the same way by independent entities. A colorful analogy regarding validity and reliability can be found in Babbie (2006, p. 148) and is depicted in Figure 4.

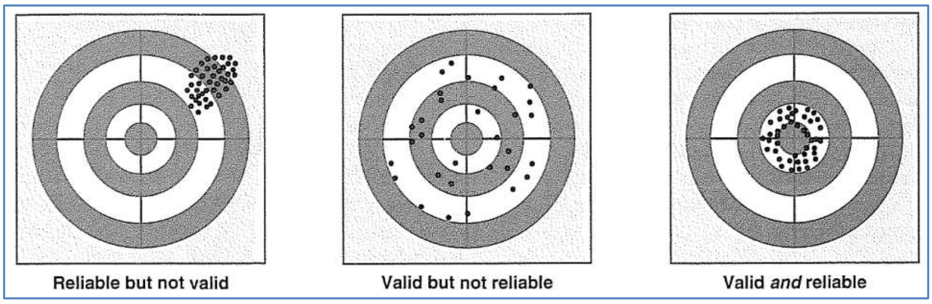

Figure 4: Analogy of validity and reliability (Figure 5-2 in Babbie (2006, p. 148))

5.4.1 Operational choices – Likert-type questions in formed measurement

Since the ERM maturity assessment is of qualitative nature, viable item operationalizations are rather binary variables and ordinal variables, as opposed to ratio variables which were also mentioned earlier.

It can be seen that many maturity models and formative indexes found in the literature lean towards ordinal variables as the operationalization of choice (Diamantopoulos and Winklhofer, 2001; de Bruin et al., 2005; Lundqvist, 2015; Schumacher et al., 2016). As a consequence, they are tempted to measure the ordinal item via Likert-type questions. These are questions that use response categories such as “strongly agree” > “agree” > “disagree” > “strongly disagree” (Babbie, 2006). Here, the item’s levels are reflected in the relative intensity of the response categories. Originally, the question format was developed with the aim to improve the measurement quality in social research through the use of standardized response categories (cf. Babbie, 2006, p. 171). However, after reading the preceding sentence and with the objectives of the measurement task at hand in mind, alarm bells should be ringing. First, because it was already elaborated that the social sciences mainly deal with reflective measurement models. Second, one of the main inquiries in the social sciences is the measurement of attitudes or opinions. In fact, the work in which Likert-type questions where introduced is titled: “A Technique for the Measurement of Attitudes” (Likert, 1932).

Without going too much into the specifics of social research, it can easily be seen that there is merit in developing an attitude (scale) measure as a reflective measurement model. Attitudes are subjective by definition. Hence, to reliably rank-order people in terms of their attitude towards a certain concept, it may not be sufficient to conceptualize the concept via one indicator only, since different entities may comprehend the specific indicator differently and thus make the measurement error prone. To the contrary, multiple indicators must be used to cover the different manifestations of the construct and give the responding entities a better chance to get down to the intended meaning of the underlying construct. Still, all manifestations have to be representative for the whole underlying concept. This is the essence of reflective measurement, where a construct is conceptualized via a sample of different, interchangeable indicators where each indicator covers a different manifestation of the latent construct (Edwards, 2011). To illustrate this, the operationalization of two indicators of the Imperialism Scale presented in Likert (1932) are shown in Table 1. The concept Imperialism seems to refer to the US military strength, but it should be pointed out that Likert made no attempt to conceptually define the meaning of the construct[8], the focus was solely on the operationalization technique.

[8] *The different scales presented here have been given their respective names merely for convenience in referring to them. The names given them seem the most plausible, but to avoid any "jingle fallacy" it should be recognized that the scales measure merely what the different statements included in them involve.* (Likert, 1932, p. 15)

Table 1: Example of Likert-type questions

Score		Response option
All men who have the opportunity should enlist in the Citizens Military Training Camps.		
Score		**Response option**
1	○	Strongly Approve
2	○	Approve
3	○	Undecided
4	○	Disapprove
5	○	Strongly Disapprove
The united states should have the largest military and naval air fleets in the world.		
Score		**Response option**
1	○	Strongly Approve
2	○	Approve
3	○	Undecided
4	○	Disapprove
5	○	Strongly Disapprove

It can be seen that both indicators are representative for the whole concept, but show different manifestations of the concept Imperialism. For such a measurement construct, where each indicator is a scale on its own, an aggregation of (weighted) scores over all indicators is legitimate since it serves its intended purpose of error reduction and the resulting overall score preserves the scale property of the indicators[9].
Would the concept be specified via formed attributes, where each attribute represents a different aspect of the concept, (1) the chosen attributes would need to represent the main components construct and (2) a weighted aggregation of attribute scores would immediately result in an overall index score.

In the next chapter, the rationale behind maturity models in the sense of the original CMM will be presented, before the Design Science Research process is re-entered as promised with an evaluation of current maturity models and their deviation in design from the CMM design principles. There the preceding discussion of the use of Likert-type questions as a means of maturity model operationalization will be picked up once more, in a more specific way. To round up the abstract elaboration about ordinal variables and Likert-type operationalization choices, and to make another point against their usage in maturity modelling, two statements in the methodological section of (Likert, 1932, p. 44) should be stated here, which are in stark contrast to the operationalization objectives of this work:

- *It is essential that all statements be expressions of desired behavior* [sic] *and not statements of fact.*

[9] *In this study, however, each statement becomes a scale in itself and a person's reaction to each statement is given a score. These scores are then combined by using a median or a mean.* (Likert, 1932, p. 24)

- *Each statement should be of such a nature that persons with different points of view, so far as the particular attitude is concerned, will respond to it differentially.*

In support of the here aspired evidence-based operationalization of assessment questions, Likert (1932, p. 44) states that *two persons with decidedly different attitudes may, nevertheless, agree on questions of fact. Consequently, their reaction to a statement of fact is no indication of their attitudes.* This is exactly the behaviour needed for the aspired ERM maturity assessment.

6 Maturity Level Measurement

After a colorful introduction into the context and the outline of the objectives of this work, the design of an ERM maturity assessment with bespoken properties, it was argued that the Design Science Research Method is the right methodological umbrella to conduct the analysis of design-principles for a conceptually sound maturity model, operationalized as a valid and reliable web-based self-assessment.

Of the six-step DSRM process, the "problem identification and motivation" step as well as the "define objectives of a solution step" where already conducted in the introduction. In order to present the following three steps ("design & development", "demonstration", "evaluation") the concepts of construct measurement research where extensively elaborated in the preceding chapter. This was important, since a central contribution of this work is to "translate" the CMMI-design principles to the wording of the scientific stream of construct measurement research. Especially the discussion of the value of formative or formed models in construct measurement research should be enriched, since it seems that the CMMI-design is a proof of concept that is needed for the advocates of formative measurement (Diamantopoulos and Winklhofer, 2001; Petter et al., 2007; MacKenzie et al., 2011) and if there is no need to determine the construct statistically, it should also serve as an answer to the critics (Edwards, 2011).

However, before the pending three iterative steps of the DSRM process are re-entered at the "evaluation" step regarding present maturity models, the nature of maturity levels needs some further elaborations to see why maturity models are in general the right choice to bail out the Prime Minister.

6.1 Fundamental Characteristics

In 1984 the US Defense Dept. formed the Software Engineering Institute (SEI) at Carnegie Mellon University to *establish standards of excellence for software engineering and to accelerate the transition of advanced technology and methods into practice*. (Humphrey, 1988, p. 73)

In one SEI project, the objective was to provide some way to characterize the capabilities of software development organizations. The result was a software-process maturity framework which can be used to *assess an organizations capabilities* and *identify the most important areas for improvement* (Humphrey, 1988).

In this maturity framework, five maturity levels for software-processes where conceptualized on the basis of actual historical development paths that represent continuous evolutionary improvements from ad-hoc (maturity level 1: Initial) to best practice (maturity level 5: Optimizing) processes (Figure 5).

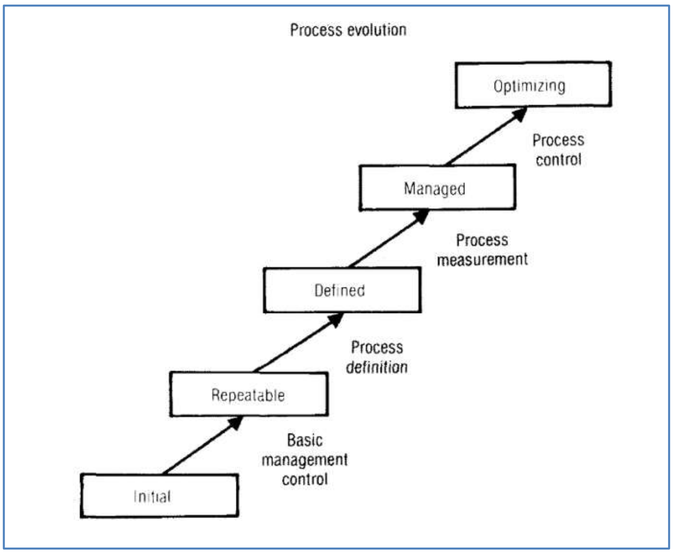

Figure 5: The five levels of process maturity (Figure 1 in (Humphrey, 1988))

The conceptual model of the maturity framework, which was evolved to the CMM (SEI, 1993), constitutes a scale where the levels of the variable process maturity are defined as discrete ordinal stages of the continuous process development process. It is important to understand that each higher maturity level builds on the basis of competences in lower maturity levels and so can't exist without them[10].

In the conceptualization of the maturity model, items are assigned to each maturity level. These items shall represent objective evidence that the requirements for the respective maturity level are met. To finally derive the empirical construct as a scale measure, this design principle (i.e. the assignment of items to maturity levels), however, requires that the relationship between all the items (in a dimension) are taken into account. Thus, maturity levels represent nothing other than groups of items that have a specified relationship to the items in the neighboring maturity levels. This characteristic is in line with the definition of a scale, where relationships between the items are taken into account and scores are assigned to specific combinations of items (see Babbie, 2006, p. 154).

6.1.1 Multi-dimensional Constructs

Up until this point, the discussions focused on one-dimensional constructs. Coming to the application of the practices of construct measurement in maturity models, however, the discussion must be broadened and include multi-dimensional constructs, which are nothing else than the formed objects in the wording of Rossiter (2002).

[10] Remember: Processes (objects) in higher maturity levels are always superior in every aspect to processes in lower maturity levels. This is an important characteristic of maturity models that can also be found in scales, but not in indexes.

As already mentioned, in reflective measurement uni-dimensionality is assumed, therefore a reflective model can by definition have one dimension only (Edwards, 2011). The situation is different for formative constructs, since it was already established that formed objects themselves represent constructs with different components that need to be rated/measured separately.

In maturity modelling, multi-dimensional constructs have to be considered, when there might be progress in different aspects of the concept can't be attributed to the same development path. The dimensions may well effect each other, but progress in each dimension can be achieved independently. In this work, a separation of the dimensions in the maturity model is promoted. Thus, multi-dimensional maturity assessments shall result in a maturity profile, i.e. a scale score for each dimension without aggregation to an overall maturity score for such multi-dimensional concepts. There is a severe risk that the aggregation to an overall score would represent an index score that can't be unambiguously interpreted. In terms of the objectives of maturity level measurement from an assessed entity's point of view, assessment results may be benchmarked against a target profile. This approach is for example incorporated in the continuous representation of the CMMI, as will be shown later.

6.2 Objectives of Maturity Level Measurement

Maturity level measurement, based on a sound maturity model and accepted as a standard in the respective domain serves different purposes on the micro (i.e. perspective of the assessed entity) and macro perspective (i.e. perspective of the assessment's sponsor).

First, the assessment presents a snapshot of the as-is situation, e.g. it serves a descriptive purpose. For the assessed entity this information is fundamental for any improvement activities since it highlights respective strength and weaknesses.

Second, since the model is clearly specified in terms of the items for each maturity level, information about necessary improvements for the next maturity level can easily be abstracted. As the maturity levels reflect a natural development path, reaching the next higher maturity level is the only logical way to progress. Thus, the model itself represents a roadmap for improvement and therefore serves a prescriptive purpose.

From a macro, e.g. scientific perspective, measurement models are often created to (1) get insights to score distributions on an empirical construct (Box D in Figure 2) in certain populations and compare scores across different populations as well as (2) identify the effects (e.g. firm performance (Gordon, Loeb and Tseng, 2009)) or drivers (Beasley, Clune and Hermanson, 2005; Lundqvist, 2015) of the empirical construct.

In most cases these scientific analyses are dependent on survey data and unfortunately there is an increasing reluctance from firms to participate in this kind of scientific endeavor. In order to still be able to collect the data, firms have to be incentivized to participate and the mere presentation of the score benchmarked against the survey population may not be enough. For reasons outlined above, maturity assessments bear much more value for the assessed entity than other (formative index) measurements do, where norms have to be developed for representative samples in order to be able to give some interpretations to the results (MacKenzie et al., 2011).

After discussing the fundamentals and objectives of maturity level measurement, as intended by the SEI, with first references to construct measurement research, the following chapters should first highlight deviations from design principles of the CMM in the conceptualization of many maturity models found in the literature and subsequently give a detailed presentation of the CMMI-design. According to the DSRM process, it will be evaluated if exemplary, non-CMM compliant maturity model designs are fit for their intended purpose, as well as for the purpose of the Prime Minister's ERM maturity assessment.

7 From Maturity Indices to Formed Maturity Scales

Since the advent of maturity modeling with the Capability Maturity Model (CMM) the concept has been widely adopted, while it also has been developed by the SEI themselves to the Capability Maturity Model Integrated (CMMI). But the maturity model design has evolved in different directions.

The mainstream is to conceptualize maturity models as an aggregation of ordinal items, where each item is operationalized as a Likert-type question so that each response option is directly assigned to a maturity level. There the logical progression of the maturity levels as stages in the development process is used to specify the item as an ordinal variable. This technique makes it easy to map responses to maturity levels (de Bruin *et al.*, 2005), but suffers either from misinterpretation risks by respondents if the question is double-barreled or too vague, or from interpretation problems of the resulting overall maturity level if it is the result of an aggregation of formed attribute scores. I.e. the overall maturity level represents an index value as opposed to a scale value (Edwards, 2011).

The CMMI takes an objective evidence-based approach to maturity modeling. In this approach, the logical progression of the development process is used to conceptualize the resulting maturity levels (as an ordinal variable). In a second step, the maturity levels are specified in terms of necessary and sufficient requirements. These requirements are binary attributes in the sense that their presence or absence alone decides if the assessed entity has reached the respective maturity level in its development or not. This method helps to split each maturity level into its basic elements, which can then be measured independently, not least to mitigate the misinterpretation risk of respondents. To further increase validity, CMMI appraisals (similar to maturity assessments) are set up to seek objective evidence for the presence or absence of the attributes in the assessed entity. The evaluation of the maturity level in the end is only a matter of determining the highest maturity level for which all attributes were measured as present. Via this method results can be unambiguously interpreted, since the maturity level was clearly specified from the outset and is not a product of an arbitrary combination of formed attribute scores.

Depending on the specified objective of the research the mainstream or the CMMI approach may be appropriate for maturity level measurement. The implied modeling decisions of both approaches will be elaborated in this and the subsequent chapter. Situations will be presented where the mainstream approach ends in the accidental creation of a maturity index, rather than a maturity scale where it is not possible to derive at meaningful measurement results that serve the purpose of the research. It will be concluded that to fulfill the objectives of the ERM maturity assessment (Box 2), maturity models have to be conceptualized in terms of their forming attributes while retaining scale characteristics.

7.1 Maturity Level as Reflective Proxy Variable

In research projects where the focus is on determining factors associated with high maturity levels (as discussed in chapter 5.2), time and budget constraints may exist which prevent researchers from allocating significant resources to the conceptualization of maturity models. In this cases, often reflective indicators (sometimes even a single indicator) are used as a proxy maturity. An example for this can be found in Beasley et al. (2005) where factors associated with the extend of ERM implementation where analyzed. For this analysis, a single variable was used to capture the stage of ERM deployment (Table 2).

Table 2: Example maturity assessment question adopted from Beasley et al. (2005)

To what extend is Enterprise Risk Management implemented in your organization?		
Score	**Response option**	
5	○	Complete ERM is in place
4	○	Partial ERM is in place
3	○	Planning to implement ERM
2	○	Investigating ERM, but no decision made yet
1	○	No plans exist to implement ERM

In this research the implementation stage of ERM was conceptualized as a reflective construct, which can be interpreted and is appropriate given the research objective. The indicator clearly covers the whole concept of ERM implementation and in order to reduce measurement error they could have added some additional questions with similar response options (see Table 3) which could have been aggregated without losing scale properties.

Table 3: Example of an additional reflective indicator

How satisfied are you with the ERM implementation in your organization?		
Score	**Response option**	
5	○	Excellent
4	○	Good
3	○	Satisfactory
2	○	Sufficient
1	○	Unsatisfactory

Other research undertakings might aim to create a more detailed model to gain more information about the assessed entity and survey sample, respectively. For example, “partial ERM is in place” can be the response of firms that have completely different ERM implementations and the study of Beasley et al. (2005) doesn’t account for that. Additionally, as mentioned earlier, there has to be an incentive for organizations to participate in the survey. Participants do not learn anything new from the survey described above and their participation can only be attributed to goodwill.

7.2 Maturity Level as Index of Formed Attributes

Schumacher et al. (2016) designed an Industry 4.0 maturity model which should measure the readiness of manufacturing companies to comprehensively embed the new technologies in their production. For this, the domain was conceptualized as a 9-dimensional maturity model where each dimension represents another aspect of a manufacturing company.

The result of their assessment is a maturity level ranging from one to five for each of the nine dimensions, i.e. the assessed entities are described via a maturity profile composed of nine aspects. Figure 6 shows an exemplary assessment result.

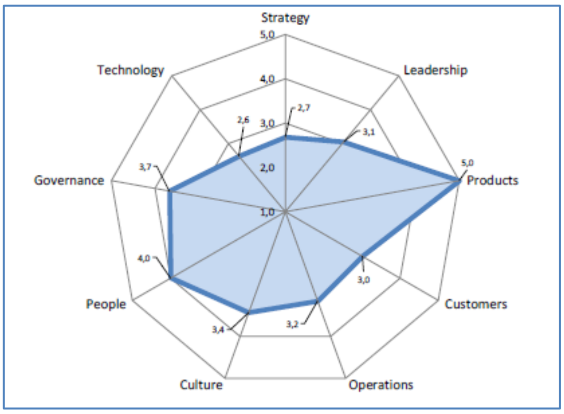

Figure 6: Example - Industry 4.0 maturity in 9 dimensions, Figure 2 in Schumacher et al. (2016)

The decision not to aggregate the dimensional maturity levels into an overall score obviously helps to avoid the interpretation fallacy on the top level (Edwards, 2011). However, a look at the conceptualization layer further down (that is the specification of items) shows that 62 *maturity items* were used as the measures of the assessment. An analysis of the maturity items immediately shows that they are very distinct aspects of their individual Industry 4.0 dimension and they are not designed to capture the respective dimension as a whole. Hence the maturity items have to be treated as forming attributes rather than reflective indicators. It is worth mentioning that the relationship between the items and the construct is, as so often the case, not explicitly discussed in the paper (Schumacher et al. 2016).

The 62 items were assigned to the nine dimensions and a logic was introduced of how to aggregate the items within the dimensions to derive the respective dimensional maturity level.

Prior to discussing the introduced aggregation logic and its consequences, the operationalization of the maturity items, i.e. the attributes of the maturity model needs some elaboration though.

Similar to the example discussed earlier, the items are operationalized as ordinal variables (see Figure 7). Again, the reliability of the measurement has to be questioned, since respondents have little guidance on how to classify their implemented system.

External Question	1	2	3	4	5
Do you use a road map for the planning of Industry 4.0 activities in your enterprise?	☐	☐	☐	☐	☐
1...Not implemented, 5...Fully implemented					

Figure 7: Exemplary question to measure a maturity item, Table 3 in (Schumacher et al. 2016)

From the wording of the question, one can clearly see that the question is intended to measure a factual phenomenon, as it is supposed to do in a maturity assessment, but as already discussed in chapter 5.4, these Likert-type questions where originally used to measure the wants, desires or conative dispositions of subjects (respondents) and not their opinions regarding matters of fact. Because it is such an important point, nothing is lost if Likert (1932, p. 13) is cited once more to point out what the intention of the question format was: *Without exception, the questions were presented in such a form as to permit a "judgment of value" rather than a "judgment of fact". […]. Since value judgments are required, it was conceived that every issue might be presented in such a way as to allow the subject to take sides between two clearly opposed alternatives.*

So naturally, one would say that the phenomenon which is subject to the question in Figure 7 can be fully captured by a binary indicator which is set to 1 if a road map for planning Industry 4.0 activities is used and set to 0 if it's not. Still, should there be any value in operationalizing the variable in a way to capture more information, e.g. the degree of implementation, than for factual measures the interim stages should be clearly specified in order to achieve a higher level of reliability (e.g. Table 2).

To conclude, there is no definition of the item's level that is present in response options 2-4, so the measurement is prone to reliability and interpretation problems from the outset. Also, regarding the prescriptive purpose for improvement activities that maturity models can serve thanks to their characteristic, it can already be said that this won't be the case for the model at hand. If the measures themselves are not unambiguous, the outcome of the assessment can be neither.

Now that the measurement is defined on the item level, the aggregation of the items to maturity levels in each of the nine dimensions can be discussed. In expert interviews it was found that not all maturity items are of equal importance, and so a weighting schema was introduced (Schumacher et al. 2016). With these weights the maturity is computed according to the following formula:

$$M_D = \frac{\sum_{i=1}^{n} M_{DAi} * g_{DAi}}{\sum_{i=1}^{n} g_{DAi}} \qquad (1)$$

with M being the maturity (level), D the dimension, A the item, g the weighting factor and n the total number of maturity items for the respective dimension. Though weights where assigned, thus the relative importance of the item is accounted for, the items have never been arranged in an ordered sequence and their relationship is not part of the conceptual model (as would be the case if they were assigned to different maturity levels). Through this aggregation, the resulting maturity level is finally impossible to interpret in absolute terms. In fact, a maturity model like this is the perfect example of an index and lost all necessary properties of a scale.

Diamantopoulos and Winklhofer (2001) proposed an alternative measurement perspective to reflective measurement, based on formative items which ends in the creation of an index rather than a scale. They give numerous examples where indexes where created based on a formative measurement perspective and admit that in many instances this perspective is not explicitly acknowledged as such (see Diamantopoulos and Winklhofer (2001, p. 270)). The example of Schumacher et al. (2016) shows that the inexplicit creation of indexes with formative items is not only limited to the economics literature, but also present in the literature of maturity modeling.

A look at the operationalization of the formative index in Diamantopoulos and Winklhofer (2001) shows, that it is no different to Schumacher et al. (2016). An example question is shown in Table 4. The items where again operationalized as ordinal variables, without clear specification of each item's level (thus no guidance for the respondent). Since the objective of the study was not to create a maturity model, the critique that relationships between the items where not accounted for does not apply. The aggregation to an overall index score was achieved via a MIMIC model (i.e. the model was statistically identified), which is neither common nor suggested for maturity models.

Table 4: Example question from Diamantopoulos and Winklhofer (2001, p. 276)

The costs of obtaining data useful for export sales forecasting purposes are often prohibitive.		
Score	**Response option**	
5	○	strongly agree
4	○	agree
3	○	neither agree nor disagree
2	○	disagree
1	○	strongly disagree

To summarize, the objective of the study of Diamantopoulos and Winklhofer (2001) was to present an alternative to scale development, since conventional scale development procedures assume reflective measurement, while they believe that several constructs would be better captured if approached from an formative perspective (see Diamantopoulos and Winklhofer (2001, p. 274)). The proposed alternative is to construct an index with formative items. By declaring it as an index, they concede that their model merely rank-orders the units of analysis and no response patterns stemming from logical relationships between the items are incorporated in the conceptualization. So the resulting score is only valid for relative comparisons and has little, if any, value when interpreted on its own, as is also argued in Edwards (2011).

When the dimensions are looked at individually, the maturity level in Schumacher et al. (2016) is derived quite similarly as an index score. The only difference is, that the item weights are derived from theoretical considerations rather than that they are statistically identified. But the knowledge about the item weights is not enough to make the formative construct unambiguously interpretable, since also the items' variances and covariance have to be taken into account (Edwards, 2011). Thus, their maturity model is rather an index and not capable of providing a roadmap for improvement for assessed entities via the identification and specification of appropriate improvement activities in an ordered sequence.

In this chapter, exemplary mainstream maturity models were presented and their design in terms of conceptualization and operationalization were discussed in the light of construct measurement research. It was demonstrated that they inadvertently may not represent empirical constructs that produce scale measures, which is the main characteristic that is needed for the aspired ERM maturity assessment. Thus, the evaluation in terms of the DSRM process triggers a red traffic light, which means the iteration loop has to go through once more and the brave maturity assessment designer has to go back to the design & development step to come up with another artefact, but this time going by the design of the CMMI.

8 Maturity Scales – the Capability Maturity Model Integration

This chapter will present the CMMI design principles for maturity models. This shall be done in the light of construct measurement research, which, to the knowledge of the author, hasn't been done before. Thus, this chapter in isolation should already represent a scientific contribution in its own right. This is because advocates of formative measurement in the management sciences (Petter et al., 2007; MacKenzie et al., 2011) are looking for successful implementations of formative/formed measurement models, but seem to overlook the CMMI. This could be because none of their usual, construct measurement research-based, key words are part of CMMI's documentation (SEI, 2010).

However, when discussing the CMMI-design principles, it is important to mention that the CMMI takes a clearly defined, process-based measurement perspective[11], which is not universally applicable. Therefor the subsequent discussions will identify the CMMI model characteristics, while stripping it from the process focus. The so derived design principles should be general enough that they are ready made for application in arbitrary measurement tasks in the management sciences, like for example an ERM maturity assessment. Nevertheless, *CMMI models provide guidance to use when developing processes* (SEI, 2010, p. 5), just like the ERM maturity assessment should provide guidance when developing an ERM system in businesses to fulfill its prescriptive purpose.

The analysis of the CMMI will be split into two parts. In the first part, the conceptualization of the CMMI will be analyzed, including the identification of the measurement perspective (formed vs. reflective), the specification of the model in terms of the relationship between maturity levels, dimensions and items, as well as the two conceptually different representations of the model.

The second part of the CMMI analysis will be concerned with the operationalization of the model, where especially the SCAMPI[12] A appraisal method will be elaborated. In this appraisal method a trained team of professionals audits the organization and uses all available channels (*The appraisal team observes, hears, and reads information…* (SCAMPI Upgrade Team, 2011, p.10)) to determine which model items are present and which are absent. The presentation of the operationalization will show the importance of specifying the model on fact-based items in the conceptualization, which can be determined as present or absent – based on objective evidence – in the operationalization.

While the conceptualization of the model should be discussed independently of the process focus of the CMMI, the operationalization should be discussed (partially) independent of the on-site audit basis of the SCAMPI A appraisal method. Keeping in

[11] *CMMs focus on improving processes in an organization. They contain the essential elements of effective processes for one or more disciplines and describe an evolutionary improvement path from ad hoc, immature processes to disciplined, mature processes with improved quality and effectiveness.* (SEI, 2010, p. 5)

[12] Standard CMMI Appraisal Method for Process Improvement (SCAMPI)

mind the requirements of the aspired ERM maturity assessment (Box 2), to which the CMMI-design principles should be applied, the operationalization must be done via a web-based self-assessment, as opposed to an on-site audit. Thus, the objective evidence behind each item must already be expressed in the wording of the questions. Still, there should be the possibility to assess the plausibility of the self-assessment result afterwards via on-site checks.

8.1 CMMI Conceptualization

8.1.1 Concept, Dimensionality and Measurement Perspective

The conceptual model of the CMMI is proposedly designed to contain all elements that are considered to be important for enterprise-wide process improvement in all kinds of organizations. The concept of process improvement is thereby discretized in capability maturity levels. The documentation stresses, that from their perspective, it is important not to focus improvement approaches on specific areas of the business, but to take a holistic view on the whole enterprise and coordinate process improvement activities in different areas.

Translated to construct measurement research, the outset of the conceptualization implies that to assess the overall stage of the holistic business development process, it was decided to:

1. Conceptualize the concept *process improvement* as an ordinal variable, i.e. a process improvement scale composed of ordinal capability maturity levels.
2. Conceptualize the concept *process improvement* as a multi-dimensional construct in order to capture the business as a whole and not to focus on one aspect of the business only.

Once the objective is set to create a measurement construct that produces a scale value as the measurement outcome, and the necessity is identified to conceptualize it as a multi-dimensional construct, the next step is to specify the constructs items. A second-order construct has formative relationships between the construct and dimensions (i.e. a formed object and it's components), whereas the dimensions themselves can be conceptualized as formed or reflective, depending primarily on the objective of the measurement (Petter et al., 2007) (i.e. the classification of the attribute that shall be rated (Rossiter, 2002)).

The CMMI documentation shows that, in essence the census of items of the model is thought of as a c*omprehensive integrated set of guidelines for developing products and services* and the itmes represent a *collection of best practices that help organizations to improve their processes.* (SEI, 2010, p. i)

It is immediately clear that these items are building blocks of mature (i.e. highly improved) processes and none of them is intended to capture the whole construct. Hence, within the CMMI, capable organizations are defined in terms of best practice activities to achieve high process maturity. The defining relationship between items and construct is assertive of a formed measurement model.

Similar to the Industry 4.0 maturity model, different aspects of organizations are grouped into separate dimensions of the CMMI. As a result of the process focus, dimensions are specified along different purposes of processes. Thus, so called *process areas* are representative of model dimensions. For each of these dimensions a development path from immature (i.e. *incomplete*) to mature (i.e. *defined*) processes is defined. The stages of the development process are not called maturity levels, but *capability levels*. Still, they possess the same conceptual meaning as maturity levels. This naming convention for the development stages within the dimensions (process areas) was probably introduced to avoid confusion with the overall maturity of the organization, which is labelled the organizations maturity level and derives from a clearly specified aggregation of the dimension's capability levels.

8.1.2 Continuous and Staged Representation of the CMMI

The differentiation of capability levels and the maturity levels rises the need to be more specific about two different representations of the CMMI, namely the *staged representation* and the *continuous representation*.

A clear objective of the CMMI is to be a prescriptive model, i.e. offer a clear path for process improvement. The different representations of the model enable organizations to follow two distinct improvement strategies:

- *Continuous Representation*: Results in a capability level profile for organizations that want to incrementally improve processes corresponding to an individual dimension, called process area (see Figure 9)
- *Staged Representation*: Results in an overall maturity level for organizations that want to improve processes corresponding to incrementally increasing sets of process areas

But regardless of their type, *Levels are used in CMMI-DEV to describe an evolutionary path recommended for an organization that wants to improve the processes it uses to develop products or services*.(SEI, 2010, p. 21)

Continuous Representation

Capability levels are the scale against which the process capability of a specific process area is measured. This is directly comparable to the dimensional maturity levels in Figure 6. Each capability level is defined in terms of a goal that has to be achieved in order to be awarded a capability level. To facilitate the assessment of these binary goals, sets of practices were defined for each goal (Figure 8).

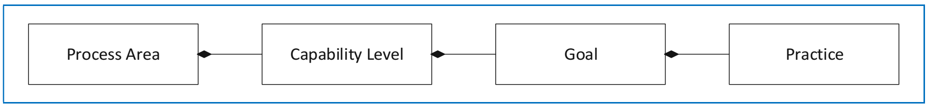

Figure 8: CMMI continuous representation – conceptual model

Since the mapping of goals to capability levels is one-to-one and in order to facilitate the comparison to other models already discussed earlier, the practices can be referred to as the attributes of the capability levels.

The result of a CMMI appraisal (the equivalent to a maturity assessment) for the continuous representation is presented in Figure 9.

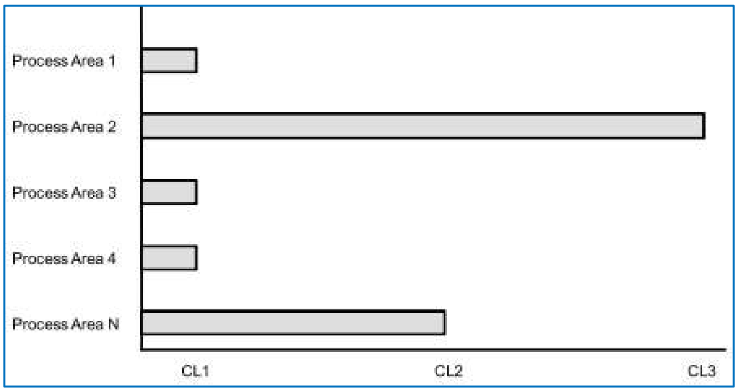

Figure 9: Continuous representation - capability levels form (capability) profile

The continuous representation is quite generic in the way that it does not prescribe a specific set of process areas based on which an organization should be assessed. In fact, the recommendation is, that organizations themselves decide which process areas are important to them, depending on the environment they are operation in and the specifics of the organization. Thus, an organization specifies a target profile for itself in terms of the relevant process areas and the appropriate process area capability level.

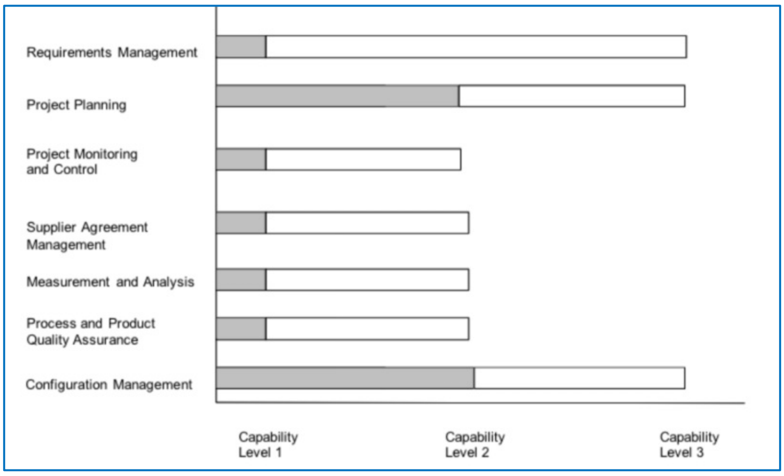

Figure 10: Target and achievement profile

The result of an CMMI appraisal via the continuous representation is thus a strength/weaknesses analysis where the as-is situation is benchmarked against a target profile (Figure 10).

Additive vs. Multiplicative Dimensions

The concept of maturity levels in the CMMI model is a more complex matter, but it is strongly related to capability levels. While for the continuous representation, dimensions, i.e. process areas are kept separate and a capability level is awarded in each of them, the process area's capability levels are combined in the staged representation and a maturity level is awarded for the whole organization.

There is a case for arguing that this aggregation of dimension scores into an overall score will result in a maturity index, since this happened for all other models that were discussed earlier. But the design of the CMMI staged representation allows for an aggregation of formative dimensions that preserves scale characteristics, which lets the resulting maturity level be well defined and interpretable.

As MacKenzie et al. (2011, p. 302) put it, *the focal construct is a function of the [sub-]dimensions that jointly define it. The question is, what type of function? Is it an additive or multiplicative one?*

They argue that additive dimensions are compensatory for each other, where the effect that one dimension has on the construct is independent from the effect of another dimension. Or, put differently, a change in an individual dimension is sufficient, but not necessary to produce a change in the construct (cf. MacKenzie et al., 2011, p. 302). If changes in dimensions go in opposite directions, they may even leave the construct unchanged. This is exactly how the dimensions in Schumacher et al. (2016) combine to form their maturity index. MacKenzie et al. (2011) describe this kind of construct as the union of its dimensions.

The alternative scenario is, that the construct represents the intersection of its dimensions, i.e. the dimension are multiplicative. There the dimensions are necessary and jointly sufficient for the meaning of the construct. I.e. the construct is defined by the exact state that dimension 1…N are in and it does make a difference if dimension one has score/maturity level 3 and dimension 2 has maturity level 1 or the other way around. MacKenzie et al. (2011, p. 302) conclude the discussion of constructs with multiplicative dimensions by stating that: *Although we are not aware of any specific examples of the use of this multiplicative structure for a measurement model, we do believe that this type of measurement model is appropriate for some constructs and should be explored in future research*.

However, as will be shown, the staged representation of the CMMI is such a specific example of a conceptual model with multiplicative dimensions. This is also manifested by the fact that each process area has related process areas assigned, where high-level relationships between the process areas are described in the model's conceptualization. Furthermore, it shall be seen that the incorporation of the multiplicative nature of the dimensions into the conceptual model is exactly what is needed to derive at a formed measurement model that produces scale values.

Staged Representation

To start going into the specifics of maturity levels, and thereby the staged representation of the CMMI, Figure 11 gives an overview of the available capability and maturity levels. As can be seen, a process area can be assessed as having a capability level (CL) from 0 to 3, whereas the organization as a whole may have a maturity level (ML) from 1 to 5. For a detailed elaboration of the concepts behind the different capability levels, see Table 5 in the appendix.

Level	*Continuous Representation Capability Levels*	*Staged Representation Maturity Levels*
Level 0	Incomplete	
Level 1	Performed	Initial
Level 2	Managed	Managed
Level 3	Defined	Defined
Level 4		Quantitatively Managed
Level 5		Optimizing

Figure 11: CMMI – Capability and maturity levels

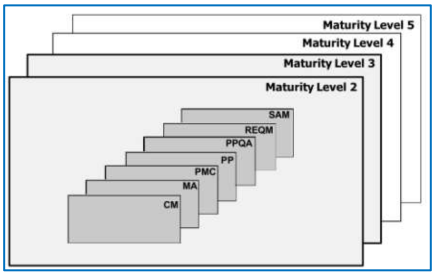

Figure 12: Assignment of process areas to maturity levels

For the continuous representation, it has been stated that organizations may want to specify their own target profile and thus determine themselves which process areas are important in the first place and further, how capable their processes are ought to be within those areas.

The staged representation on the other hand provides a standardized target profile against which all organizations are benchmarked. To be more specific, each maturity level of the staged representation in essence is a target profile. This is because the requirements for each maturity level are stated as capability levels in specific process areas. As can be seen in Figure 12, process areas, indicated by their abbreviations, are assigned to maturity levels. At the same time, the assigned process areas must be implemented to a certain capability level in organizations in order to be awarded the respective maturity level. Thus, as the SEI (2010, p. 26) puts it, a maturity level is a defined evolutionary plateau for organizational process improvement.

To obey the logic of ever-increasing properties, higher maturity levels have all process areas of lower maturity levels assigned, plus a higher capability level in each is demanded and/or additional process areas are assigned. Within the CMMI documentation, the process of defining target profiles (relevant process area + capability level) to satisfy maturity levels, i.e. choose the one continuous presentation of the model that matches the staged representation, is called equivalent staging. Since equivalent staging is the most intuitive way to understand the requirements for each maturity level, it should be used to present how capability levels in the different process areas are aggregated into maturity levels, thereby demonstrating how scores of a construct's multiplicative dimensions combine to produce an overall scale score. Figure 13 depicts the equivalent staging logic and also represents the names of the CMMI's process areas for the first time within this work, which helps to get a feeling for the model. In addition to Figure 13, the CMMI maturity level aggregation algorithm should be described textually in the following:

- ML 1: Is awarded if requirements for ML 2 are not met
- ML 2: Is awarded if all assigned process areas are at CL 2
- ML 3: Is awarded if all assigned process areas are at CL 3, including all process areas assigned to ML 2
- ML 4: Is awarded if all assigned process areas are at CL 3, including all process areas assigned to ML 3 and ML 2
- ML 5: Is awarded if all assigned process areas are at CL 3, including all process areas assigned to ML 4, ML 3 and ML 2

Name	*Abbr.*	*ML*	*CL1 CL2 CL3*
Configuration Management	CM	2	Target Profile 2
Measurement and Analysis	MA	2	
Project Monitoring and Control	PMC	2	
Project Planning	PP	2	
Process and Product Quality Assurance	PPQA	2	
Requirements Management	REQM	2	
Supplier Agreement Management	SAM	2	
Decision Analysis and Resolution	DAR	3	Target Profile 3
Integrated Project Management	IPM	3	
Organizational Process Definition	OPD	3	
Organizational Process Focus	OPF	3	
Organizational Training	OT	3	
Product Integration	PI	3	
Requirements Development	RD	3	
Risk Management	RSKM	3	
Technical Solution	TS	3	
Validation	VAL	3	
Verification	VER	3	
Organizational Process Performance	OPP	4	Target Profile 4
Quantitative Project Management	QPM	4	
Causal Analysis and Resolution	CAR	5	Target Profile 5
Organizational Performance Management	OPM	5	

Figure 13: Equivalent staging - Mapping the continuous representation to the staged representation of the CMMI

However, when reading through the CMMI documentation carefully, one might get the impression that the staged representation of the CMMI is a legacy of the CMM, where the assessment culminated in the assignment of a maturity level for the whole organization. The design that is truly promoted, however, is the continuous representation where an actual as-is profile is compared against a customized target profile. This approach helps of course to account for the idiosyncrasies of individual organizations and their business models. Whereas one figure, at such a high level of abstraction, attempts to compare what isn't to be compared…as intriguing as it might be.

This draws the attention of this work back to the continuous representation of the CMMI, where the specification of attributes for capability levels in the different process areas (Figure 14) needs some more elaboration before the discussion moves to the operational level.

8.1.3 Conceptualization of Process Areas

Descriptions about components of process areas are extensive in the CMMI documentation. Touching every component would contradict the objective of giving a parsimonious overview about the relevant aspects of the CMMI in terms of construct measurement research. Fittingly, the documentation distinguishes between required, expected and informative components of process areas whereby only the first two will be discussed further in this work[13]. But there are some more steps to do and characteristics to elaborate on in order to describe the CMMI precisely in construct measurement research terms. Although all that this sub-chapter aims to achieve is translating the conceptualization in the CMMI terminology depicted in Figure 8 to the construct measurement research terminology depicted in Figure 14.

Let's start from the top, i.e. from the left in Figure 8. A process area, as a dimension of the CMMI, is composed of four capability levels (CL0 – CL3). Since the capability levels themselves describe broad concepts (see Table 5), which are defined by multiple attributes, they need further specification to enable a precise measurement.
As can be seen in Figure 8, capability levels are compositions of goals. *Goal satisfaction is used in appraisals as the basis for deciding whether a [capability level of a] process area has been satisfied* (SEI, 2010, p. 9).

The complication herein, however, is that goals map one-to-one to capability levels and so do not help in the measurement's specification. Thus, for the construct measurement investigation, capability levels and goals can be collapsed into one entity. This is important to remember, because in the following elaborations the terms capability level and goal will be used simultaneously[14].

Before coming to the capability level's attributes, another characteristic that needs to be mentioned is, that the basic requirements for the capability levels remain consistent

[13] For an overview of all process area components see Figure 25 in the appendix

[14] This is necessary because it helps to compare the elaborations to the quite specific CMMI model terminology which holds more complications than the already generalized construct measurement research terminology.

over all process areas. This is achieved via implementing a universal progressive logic in the capability levels which requires at level:

1. that process area specific requirements are satisfied, where e.g. “Project Planning” may have completely different core/operative processes than “Configurations Management”
2. that general, common requirements over all process areas are satisfied, where in essence a process management is required for the operative processes of the respective process area (see Table 5)
3. that the process management principles, which are needed for CL2 and guide the operative processes defined in CL1, derive from a set of enterprise-wide standardized process principles and experiences are collected and used for continuous improvements (see Table 5)

These concepts, which are common amongst capability levels of all process areas are conceptualized as required practices, which exactly replicate what is defined as items in construct measurement research. *Before goals can be considered to be satisfied, either their practices as described, or acceptable alternatives to them, must be present in the planned and implemented processes of the organization* (SEI, 2010, p. 10).

This finally brings the CMMI conceptualization to Figure 14. As already discussed, the conceptualization included identifying the attributes based on which the construct will be measured, but the exact specification of the items in terms of their levels is part of the construct's operationalization. Nevertheless, conceptualization and operationalization are in no way independent processes and when it is stated ‘practices as described or acceptable alternatives must be present’ in the CMMI conceptualization, it gives already a hint that there should be some room for maneuver in the operationalization, as will be shown in the next sub-chapter.

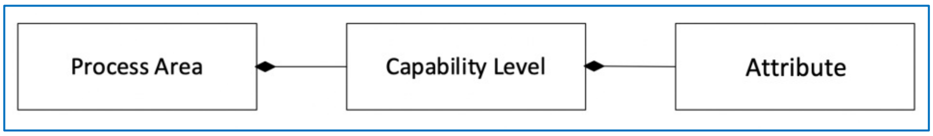

Figure 14: Process area – conceptual model

To give a more detailed presentation of the basic building blocks of the CMMI to the interested reader, its attributes, i.e. the practices required to satisfy goals which resemble capability levels are presented in Table 6 in the appendix.

To fully understand the specifics of the CMMI and at the same time the content of Table 6, a final characteristic has to be mentioned, which relates to elaborations earlier and leads to a final graphical presentation of the conceptual model.

With goals, as well as with practices, there is a distinction between specific and generic ones within the CMMI. The distinction can be traced to the characteristic that, whenever a goal or a practice is process area specific, it is labeled as such and whenever a goal or a practice is common across all process areas, it is labeled generic. Thus, requirements of capability level 1 are concerned with specific goals and practices and requirements of capability level 2 and 3 are concerned with generic goals and practices[15]. Additionally, generic goal 1 (generic practice 1), which represents capability level 1, is defined as the aggregation of all specific goals (specific practices) of the respective process area. This special characteristic can be seen in Table 6, and enables the presentation of the conceptual model of a process area as depicted in Figure 15, where capability levels can be inferred from the satisfaction of generic goals only.

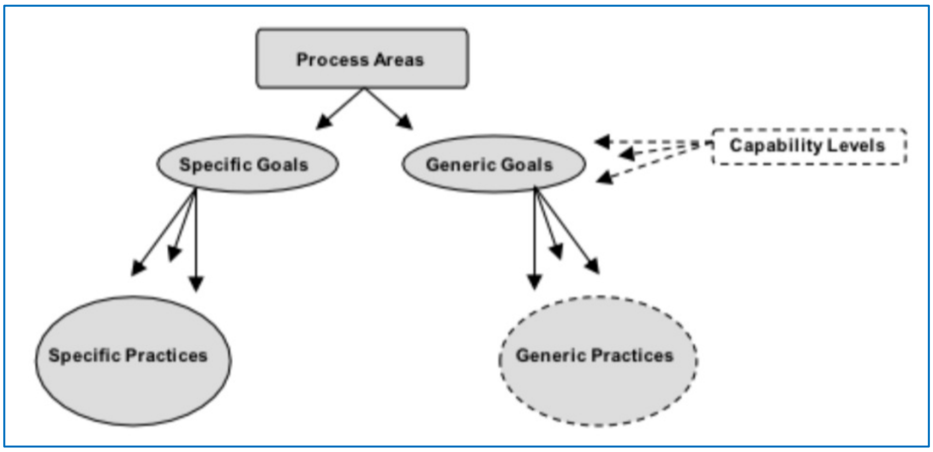

Figure 15: CMMI – process area, main components

[15] The fact that capability levels 2 and 3 use the same terms as generic goals 2 and 3 is intentional because each of these generic goals and practices reflects the meaning of the capability levels of the goals and practices. (SEI, 2010, p.24)

8.2 CMMI Operationalization

Chapter 8 serves two purposes, (1) in the overall framework of this thesis it is part of the ‘evaluation’ step of the DSRM process, where the objective is to find a suitable maturity assessment design for requirements outlined in Box 2. Furthermore (2), it presents the CMMI design principles in the light of construct measurement research and thus enriches the discussions in this field with a well-defined instance of a formed measurement model applied in the management domain. While purpose (2) was at the heart of the comprehensive elaborations of CMMI’s conceptual model in sub-chapter 8.1, purpose (1) should represent the focus of this sup-chapter. Therefore, the elaborations will center around important design principles that provide useful guidance in the operationalization of the ERM maturity assessment and not so much around other specifics discussed in the method definition document for CMMI operationalization (SCAMPI Upgrade Team, 2011).

The relevant method definition document for this work will be the one for the SCAMPI A appraisal method, which, as a Class A appraisal method, is the officially recognized and most rigorous method. Further, it is the only method that results in benchmarking quality ratings (i.e. comparative results), since it assures consistent ratings across different appraised organizations (SEI, 2010). Less formal Class B and C CMMI appraisals can’t guarantee this and thus are not able to give guidance for the aspired ERM maturity assessment’s operationalization, because consistent ratings across organizations are key for the intended purpose of being a decision mechanism to allocate government funds.

At this stage the conceptual model is already well defined in terms of the construct’s dimensions and items and a final step is missing to derive the empirical construct (Bisbe et al., 2007). This is the exact specification of the measurement including the specification of each item’s levels and the exact procedure to determine which level of the item is present (Babbie, 2006). Above all, however, is the premise that the measurement is based on objective evidence in SCAMPI A appraisals (SCAMPI Upgrade Team, 2011).

There are two types of objective evidence to determine to which degree a practice (= item) of the CMMI is satisfied (SCAMPI Upgrade Team, 2011), namely

- *artefacts* – a tangible form of objective evidence that a practice is implemented, e.g. organizational policies, meeting minutes, review results
- *affirmations* – an oral or written statement confirming (a lack of) implementation of a practice, which may be collected using interviews, demonstrations, questionnaires, …

While affirmations are the only form of objective evidence that can be used in the aspired web-based self-assessment, questions in the assessment may well be articulated to specifically ask for the existence of artefacts, which can later be fact-checked in an on-site assessment.

Affirmations in an interview, however, are more suited to represent objective evidence than answers to static questions in a web-based questionnaire. This is because ambiguities in the questions can be followed-up in an interview setting and if conducted properly, the only way to falsely collect objective evidence is if the interviewee is lying on purpose. While deliberately lying respondents are a risk that is not intended to be completely eliminated from the ERM self-assessment, the possibility that an auditor appears on-site and checks the answers afterwards (e.g. by demanding proof for the artifacts that where subject to the assessment questions) should be an acceptable mitigation measure.

Besides lying, however, the lack of a possibility to follow-up ambiguity is a severe risk for the validity and reliability of a web-based self-assessment priding itself of being based on objective evidence. This problem will be discussed in the remainder of this sub-chapter.

After highlighting differences in the audit-based SCAMPI A measurement procedure and the ERM model's self-assessment-based measurement procedure, the findings should be incorporated in the second part of the operationalization, namely the specification of item levels. Stepping back for a moment, it can be clearly seen at this stage that in construct measurement research every decision made regarding conceptualization and operationalization impacts other aspects of the measurement.

It was already stated that objective evidence is used to determine to which degree a practice of the CMMI is satisfied. The notion of a "degree of satisfaction" immediately brings the discussion to the specification of the model's items, i.e. the levels of the variable practice implementation. Rather than just assessing if a practice is implemented or not, in SCAMPI A appraisals five degrees of implementation are distinguished:

- Fully Implemented (FI)
- Largely Implemented (LI)
- Partially Implemented (PI)
- Not Implemented (NI)
- Not Yet (NY)

Definitions for each level of practice implementation can be found in Table 7, but to avoid confusion with maturity and capability levels, the stages of practice implementation shall be referred to as degrees of practice implementation hereafter.

The ordinal nature of the item evaluation is reminiscent of Likert-type question formats. The big difference to, e.g. Schumacher et al. (2016) however is, that each item level is very well defined and a team of experts is responsible to determine the true value of the ordinal variable, as opposed to a manager who participates in a survey and is not even aware of the conceptual model behind the questionnaire.

The operational choice to provide different degrees of practice implementation which can be determined in a SCAMPI A appraisal serves to give the experts some room for maneuver when determining if a capability level should be awarded or not. It avoids constellations where a capability level cannot be awarded in a process area when each practice is fully implemented and only one practice has slight insufficiencies (i.e. is partially implemented), which may play a very minor role in the specific context of the appraised organization.

For the aspired web-based ERM maturity assessment, similar to the situation of Schumacher et al. (2016), there is a manager of the assessed organization responsible to answer static questions. Furthermore, the assessment should be as parsimonious as possible in terms of the number of questions and necessary elaborations about the explicit concepts that are measured with the specific question. Thus, neither the questions can be as granular as is the SCAMPI A appraisal, nor can be the evaluations. In other words, there is no benefit in specifying very detailed response categories to fairly broad questions. For the ERM maturity assessment it makes therefor much more sense to operationalize the items as binary variables, more broadly capturing if e.g. a specific artefact in general is present or not, without going much into the details of the exact manifestation of the artifact. This recognition of limitations in the measurement's granularity should, before all, aid the realization of an objective evidence-based self-assessment.

9 ERM Maturity Assessment – Demonstration Case

The preceding discussions, from the introduction to the detailed analysis of CMMI-design principles, served to illustrate different steps of the DSRM process (Peffers *et al.*, 2008) (Figure 1) which was applied when the ERM maturity assessment was develop at the Institute of Management Science, TU Vienna. The first two steps, “Identify Problem & Motivate” and “Define Objectives of a Solution”, where covered at the start of the “Communication” step, i.e. in the introduction of this work. Subsequently, the Communication went on to the steps “Design & Development”, “Demonstration” and “Evaluation” quite dynamically, meaning that the

- presentation of design principles prominent in construct measurement research (chapter 5),
- the evaluation of mainstream maturity models found in the literature, based on fundamental characteristics of maturity models (chapter 6) as well as construct measurement research theory (chapter 7) and
- the presentation and evaluation of CMMI-design principles, based on construct measurement research theory (chapter 8) and compared against the objectives of the present study (Box 2),

represent aspects of all three steps at the same time. The deviation from the proposed DSRM sequence is not problematic since design science is an iterative process anyway. Furthermore, deviations were made consciously with the main objective to keep the presentation of fairly complex and interdisciplinary measurement approaches clear and interesting.

However, what is missing in the presentation of the design science approach to create a web-based ERM maturity self-assessment is a sound demonstration, i.e. a proof of concept of an empirical construct (Bisbe et al., 2007) that possesses all the characteristics deemed valuable in the preceding evaluations. At the time of writing (spring 2021), such an assessment tool is hosted on the servers of the TU Vienna (Figure 16), although not yet used by the Prime Minister.

In this chapter, the ERM maturity model, as the conceptual model of the maturity assessment will be presented as well as the operationalization of the maturity assessment as a web-based smart questionnaire.

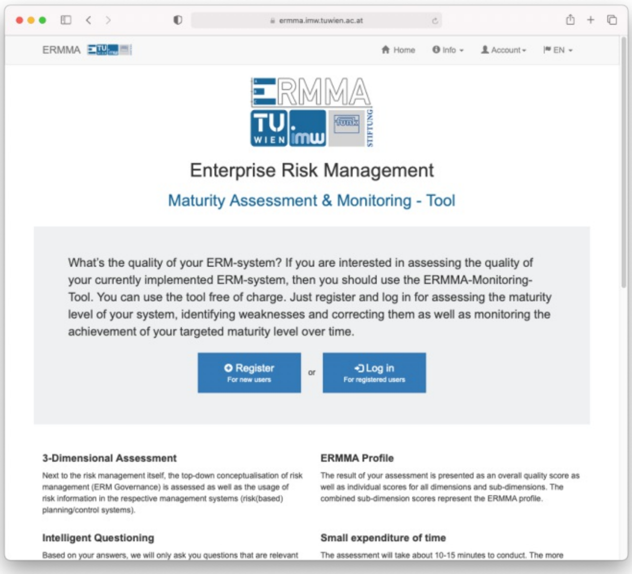

Figure 16: Web-based ERM maturity assessment tool

9.1 ERM Maturity Model

The intended purpose of the ERM maturity assessment (Box 2) is to rank-order all relevant businesses with respect to the maturity of their implemented ERM system, while keeping assessment results unambiguously interpretable. Thus, the utmost importance must be placed on the development of a true scale measure.

Consolidating the ERM literature with special attention to prominent ERM frameworks (ISO 31000, 2009; IIA-3LoD, 2013; COSO-ERM, 2017), it is clear that not all aspects of ERM can be attributed to one development path only. Hence different ERM dimensions/components need to be specified, where each component can be conceptualized as a scale via its own development path.

The formation of construct dimensions was based on a process focus in the CMMI model, since processes are supposed to be the backbone of businesses developing products and services (SEI, 2010).

In risk management, however, information is the live and blood of all activities. As can be seen, e.g. in Figure 3 where it is shown that information must be gathered, structured and categorized in order to conduct the risk management process[16] comprising risk identification, risk analysis and risk evaluation.

Thus, a three-dimensional conceptual construct for ERM was developed (Figure 17), based on an informational perspective where the information provider (dimension B) is separated from the information user (dimension C). Information provider and information user are separated, but by no means independent of one another. Referring back to the distinction between additive and multiplicative construct dimensions, it is immediately clear that a mature usage system of information can in no way compensate for an immature system that provides information. This finding has two implications:

1. If scores in dimension B and C should ever be aggregated to an overall ERM scale score, the aggregation has to deal with multiplicative dimensions and thus the aggregation algorithm in essence would have to rank possible combinations of B and C scores. As is the case in the staged CMMI representation.
2. The interplay between information provider and information user should not be left to chance, but should be well organized, derive from a well defined policy towards risk management and include a common risk understanding within the business. I.e. an ERM-governance is needed, which represents dimension A in the ERM maturity model.

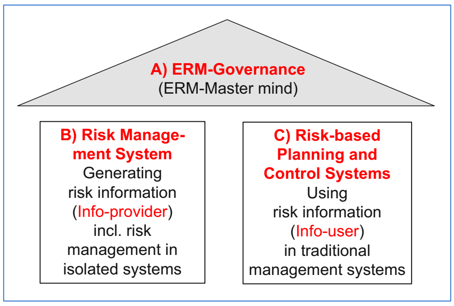

Figure 17: Information based perspective on three dimensional ERM construct

In the briefly mentioned design iterations leading up to the present ERM maturity assessment tool, it has been found that with the three theoretically well-defined dimensions, there are still too many aspects and too many loosely dependent development paths to specify three clean and consistent scale scores.

The solution found to keep the theoretically meaningful dimensions in the model, while getting an extra layer of granularity to set up the basic development path was to introduce three sub-dimensions for each dimension. Hence, in the end, ERM is

[16] To be precise, the elaboration focusses on the risk management processes' risk assessment part, according to ISO 31000 (2009)

constructed as a 9-(sub-)dimensional construct, where three sub-dimensions can always be attributed to one theoretically meaningful component of ERM. An example of how sub-dimensions naturally represent elements of a dimension was subtly provided in point 2 of the punctuation above.

The resulting nine (sub-)dimensions of the model where then conceptualized as a scale, composed of five ordinal maturity levels representing a natural development path from uncoordinated ad-hoc processes to systems interactively managed by top-management, i.e. best practice (Figure 18).

Dimensions	Maturity Levels: ML 1	ML 2	ML 3	ML 4	ML 5
A. ERM-Governance A1: Risk Strategy A2: Risk Understanding A3: Risk Organisation	Partial, i.e. silo-oriented process perspective	Process perspective incl. review and management (single loop)	Enterprise-wide (holistic) perspective	Corporate perspective (double loop)	Interactively managed by top management
B. Risk Mgt. System B1: RM Process B2: RM Training System B3: RM Information Sys.	Risk management process	Risk management system (single loop)	Enterprise-wide (holistic) risk management system	Corporate holistic risk management system (double loop)	
C. Risk-based Planning and Control Systems C1: Strategic MGT System C2: Fin. Perf. MGT System C3: Op. Process MGT Sys.	Risk limit systems	Key risk-based planning systems (incl. strategy and objective setting)	Key risk-based control systems (i.e. performance management)	Management systems with risk-adjusted performance measures	

Figure 18: Enterprise risk management maturity model

Each element of this maturity model, i.e. each maturity level of every sub-dimension was subsequently specified in terms of its necessary and sufficient requirements. The resulting binary attributes were designed to reflect objective evidence of the requirements being met. Hence the relationship between sub-dimensions, maturity levels and attributes in the ERM model (Figure 19) is no different to the relationship between process areas, capability levels and practices in the CMMI (cf. Figure 8 and Figure 14).

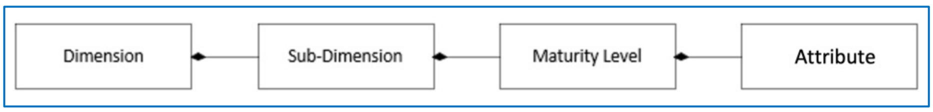

Figure 19: ERM maturity assessment - conceptual model

The consistent concepts behind the capability levels of all process areas of the CMMI (see Table 5 and Table 6) is replicated in the ERM model's maturity level specifications. The rational is that

- ML 1 represents processes that work on a silo basis, focused on the operational tasks of the respective business area only.
- In ML 2, the ML 1 processes have evolved to systems. These systems do still work in isolation between different business areas, but already include processes with are subject to monitoring activities. Such a system is comparable to the whole risk management construct that is shown in Figure 3.
- ML 3 represents harmonized ML 2 systems that are deployed enterprise-wide, i.e. in each relevant business area.
- ML 4 makes use of the harmonized ML 3 systems by aggregating the key figures over the whole enterprise. At ML 4, e.g. the risk management process as one ERM dimension produces holistic risk exposures representative for the whole business. Only at this high maturity level, the ERM systems becomes a truly comprehensive steering tool for Top-Management.
- ML 5 is conceptualized as a best practice system, which is the point when the Top-Management realizes what a mighty tool an ERM system at ML 4 is and actively uses it to manage the organization and conducts periodic reviews of ERM's effectiveness to drive a continuous improvement effort in ERM implementation.

As depicted in Figure 18, the model frame is ready to be populated by the attributes of the sub-dimension + maturity level concepts described in this sub-chapter.

The exact specification of the ERM maturity assessment attributes was already inspired by the predetermined operational choice that the assessment should be conducted questionnaire based, which necessitates binary attributes in order to achieve maximum validity and reliability for reasons already discussed (e.g. see chapter 8.2). Thus, attributes where specified in a way that they represent artefacts that can be either present or absent in organizations. Table 8 in the appendix presents the attributes for maturity levels 1-3 in the sub-dimension B1: Risk Management Process as an example.

This sub-chapter should be concluded with a final note to the ERM maturity model conceptualization, before the next sub-chapter briefly discusses its operationalization.

A system development, i.e. progression form one maturity model to the next, in the understanding of the ERM maturity model can be achieved in either of two ways:

1. A system is extended by an additional property.
2. A specific property of a system is replaced by a superior property.

These two ways of achieving progress for a system are also visible in Table 8, where red attributes represent completely new properties of the risk management process, and green attributes represent superior properties that replace already existing attributes in the model. Black attributes represent risk management process properties that already existed in lower maturity levels and remain unchanged.

9.2 ERM Maturity Assessment Operationalization

Summarizing the context in which the operationalization of the ERM maturity assessment is specified, it is known form the preceding discussions that static questions from a questionnaire are used in the measurement procedure. Furthermore, the questions must be articulated that they discover which one of the two attribute levels (present vs. absent) are true for the assessed organization.

In this quite predetermined scenario, there are still two important operational choices to make with the aim of achieving maximum validity and reliability as well as keeping the assessment as short as possible.

The first choice is to determine the sequence in which attributes of the models should be measured. Since the questionnaire in the end is not a chaotic assembly of questions with arbitrary relationships, the structure of the conceptual model from which it derives should be made use of in order to guide the participant through the assessment. This means that (1) each sub-dimension should be assessed on its own to keep the participant in the topic and (2) the ever-progressing nature of the system that is subject to the assessment within a sub-dimension should be made use of. Hence, in order to facilitate validity in the assessment, the measurement should start with the basic concepts in ML 1 and then go on to more complex systems to which the ML 1 system develops over the maturity levels.

The principle to start the measurement at maturity level 1 and then go on to higher maturity levels not only promises higher validity, it also offers the opportunity for a dynamic ending of the assessment in each sub-dimension, thus making the questionnaire smart.

Since the conceptual model is designed (in line with CMMI-design principles) in the way that a *maturity level forms a necessary foundation for the next level, trying to skip maturity levels is usually counterproductive* (SEI, 2010, p. 30). This model property can be used in the operationalization to end a dimension's assessment as soon as not all attributes of a maturity level are measured as present. This frees the participant from the burden of having to answer questions about artefacts that are not present in the organization.

To give an example of the ERM maturity assessment's operationalization, Figure 20 depicts an exemplary assessment screen as seen by participants. The figure holds two typical binary assessment questions regarding ML 3 attributes shown in Table 8. In addition to the question, participants are guided by context boxes at the top of the screen. These context boxes help the participant to grasp the perspective of the question as well as provide information about which aspect of the ERM system the questions relate to. For the specific questions shown, the ERM aspect is the risk management system (sub-dimension B1 in Figure 18).

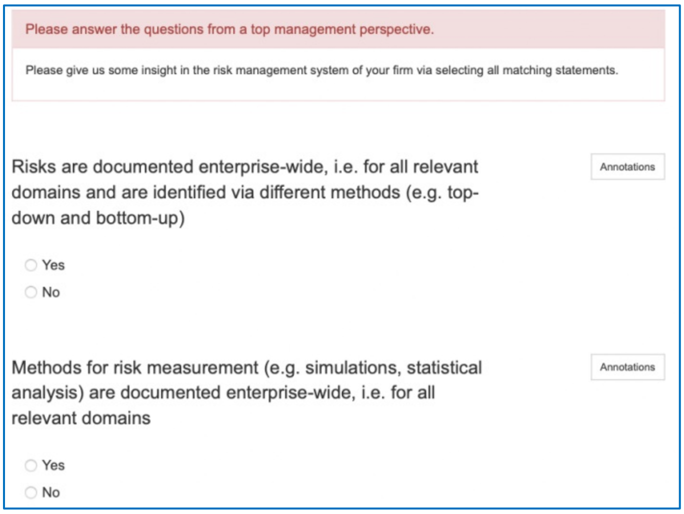

Figure 20: ERM maturity assessment - operationalization example

9.3 Exemplary ERM Assessment Result

The last sub-chapter shall serve to present assessment results as they are produced by the ERM maturity assessment tool (Figure 16). In the end assessment results, provided that the assessment is valid and reliable, are all that matters to top level executives, or the Prime Minister in the fictive scenario of the prologue.

As discussed, the measurement procedure reveals all necessary information, about which attributes are present in the assessed organization and which are missing, to produce the assessment results, i.e. score the entity on the maturity scale in each sub-dimension (Figure 21). Note that the left panel of Figure 21 shows the maturity level in each sub-dimension as a scale score that can be unambiguously interpreted in its own right, since the state of the development process that the organization is in is well defined. To fulfill the requirements (reach the attributes) of the next higher maturity level would represent the natural successive improvement activities.

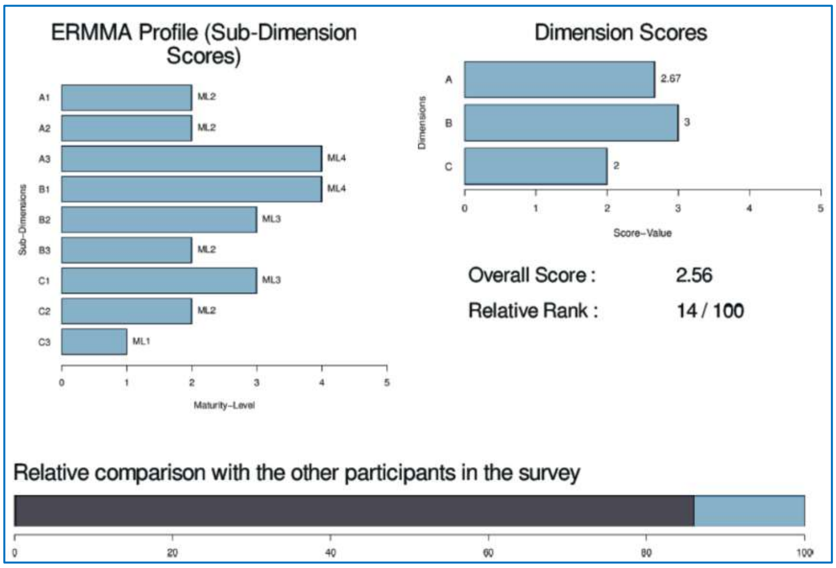

Figure 21: Maturity assessment results for participants (scores and rank)

Contrary, the information on the right panel of Figure 21 present index scores (arithmetic averages) that can be used to rank-order participants with respect to a specific dimension or the overall system, but provide no guidance on logical improvement strategies whatsoever. In the assessment results, the aggregation is even driven one step further to derive an overall index score for the participant as the result of an equally weighted aggregation of sub-dimension scale scores. With this overall score, it is possible to rank-order all participants of the assessment, as is shown in the graph at the bottom of Figure 21. Still, at this stage of the elaborations it should be clear that the overall score as an index has very little value in representing the as-is situation in the organization, controversial value in comparing organizations (i.e. ranking) and absolutely no value whatsoever in determining logical next improvement activities.

The last paragraph in this sub-chapter should be used to highlight the potential of the ERM maturity assessment tool not only to be used for one-off assessments, but to be a monitoring tool that can help to track developments over time.

For follow-up assessments, the scale characteristic of maturity levels can be used to parsimoniously monitor the development of the ERM systems over time. In follow-up assessments, participants may start with questions regarding attributes of the maturity level they were assigned in their last assessment. If answered positive, the presence of attributes in lower maturity levels is guaranteed and doesn't have to be assessed again. If answered negative, the respective maturity level can't be reached and higher ones can't be reached either, so doesn't have to be assessed.

10 Conclusion

This work demonstrated the development of a formed ERM maturity scale (Figure 18) which can be used as the basis for web-based ERM maturity self-assessments (Figure 16). The concept ERM maturity was conceptualized as a formed object, defined by its 9 main components. These components need to be rated separately because progress in terms of the attribute maturity can be achieved independently within them. The conceptual model of the respective ERM components was further specified in terms of its forming maturity attributes.

As a prescriptive maturity model, it not only helps to determine the as-is situation of businesses' ERM systems, but it also allows to derive consistent ratings (maturity levels) that can be compared across businesses and, most importantly, the underlying maturity scale serves as a roadmap for appropriate improvement activities.

The ERM maturity scale was developed in line with CMMI design principles (SEI, 2010) for maturity models, which implicitly provide all necessary conceptual and operational model characteristics needed to develop formed maturity scales of a prescriptive nature.

A major contribution of this work is to make the implicit conceptualization and operationalization of the CMMI explicit, i.e. analyze CMMI design principles in the light of construct measurement research and scale development. This should aid future maturity modeling endeavors to get a better understanding of the CMMI and apply its design principles in other domains. It was shown that maturity models in the literature fall way short of the full potential of prescriptive maturity models (de Bruin *et al.*, 2005; Enkel et al., 2011; Schumacher et al., 2016).

Further, the parsimonious elaborations of the CMMI design in the standardized terms of construct measurement research shall enrich the discussion in this scientific filed for a demonstration case of a formed measurement model. The discussions there are dominated by representatives of the social sciences, which take a realist perspective on construct measurement and start with the premises that the latent construct is something that exists and is reflected in certain observations (Likert, 1932; Nunnally and Bernstein, 1994; Babbie, 2006). However, especially in the field of MIS a constructivist measurement perspective might prove more usefull, where a concept is seen as a theoretical creation that is nothing else than the product of all its defining/forming attributes (Diamantopoulos and Winklhofer, 2001; Petter et. al., 2007; MacKenzie et al., 2011). The problem with traditional formative measurement, however, is that model identification, calibration and validation is not purely done based on theoretical reasoning (Rossiter, 2002), but instead the same empirical methods as in reflective measurement are used (e.g. see Diamantopoulos and Winklhofer, 2001). This leads to the loss of interpretability of the results of multi-dimensional constructs (Edwards, 2011) and prevents the creation of a prescriptive model.

The C-OAR-SE procedure (Rossiter, 2002) provides guidelines for certain types of concepts (e.g. formed objects, formed attributes) to be developed and validated based on expert knowledge only. Even though it has its critics (Diamantopoulos, 2005), its concepts in terms of construct measurement research are perfectly in line with the broadly accepted CMMI and where successfully applied to develop the ERM maturity model.

11 References

Altman, E. I. (1968) 'FINANCIAL RATIOS, DISCRIMINANT ANALYSIS AND THE PREDICTION OF CORPORATE BANKRUPTCY', *The Journal of Finance*, XXIII.

Babbie, E. (2006) *The Practice of Social Research*. 11th edn. Thomson Wadsworth. doi: 10.1017/CBO9781107415324.004.

Beasley, M. S., Clune, R. and Hermanson, D. R. (2005) 'Enterprise risk management: An empirical analysis of factors associated with the extent of implementation', *Journal of Accounting and Public Policy*, 24(6), pp. 521–531. doi: 10.1016/j.jaccpubpol.2005.10.001.

Bisbe, J., Batista-Foguet, J. M. and Chenhall, R. (2007) 'Defining management accounting constructs: A methodological note on the risks of conceptual misspecification', *Accounting, Organizations and Society*, 32(7–8), pp. 789–820. doi: 10.1016/j.aos.2006.09.010.

de Bruin, T. *et al.* (2005) 'Understanding the Main Phases of Developing a Maturity Assessment Model', in *16th Australasian Conference on Information Systems*. Sydney.

COSO-ERM (2017) 'Enterprise Risk Management Integrating with Strategy and Performance', *Committee of Sponsoring Organizations of the Treadway Commission*. Available at: www.coso.org.

Diamantopoulos, A. (2005) 'The C-OAR-SE procedure for scale development in marketing: A comment', *International Journal of Research in Marketing*, 22(1), pp. 1–9. doi: 10.1016/j.ijresmar.2003.08.002.

Diamantopoulos, A. and Winklhofer, H. M. (2001) 'Index Construction with Formative Indicators: An Alternative to Scale Development', *Journal of Marketing Research*, 38(2), pp. 269–277.

Edwards, J. R. (2011) 'The fallacy of formative measurement', *Organizational Research Methods*, 14(2), pp. 370–388. doi: 10.1177/1094428110378369.

Enkel, E., Bell, J. and Hogenkamp, H. (2011) 'Open innovation maturity framework', in *International Journal of Innovation Management*. doi: 10.1142/S1363919611003696.

Geerts, G. L. (2011) 'A design science research methodology and its application to accounting information systems research', *International Journal of Accounting Information Systems*. Elsevier Inc., 12(2), pp. 142–151. doi: 10.1016/j.accinf.2011.02.004.

Gordon, L. A., Loeb, M. P. and Tseng, C. Y. (2009) 'Enterprise risk management and firm performance: A contingency perspective', *Journal of Accounting and Public Policy*. Elsevier Inc., 28(4), pp. 301–327. doi: 10.1016/j.jaccpubpol.2009.06.006.

Göstl, P. (2017) *Risk Profile Contingent Analysis of Management Control Systems: Evidence from the Mechanical Engineering Industry*. Available at: https://books.google.co.uk/books?id=j0uzDwAAQBAJ.

Humphrey, W. S. (1988) 'Characterizing the Software Process: A Maturity Framework', *IEEE Software*, 5(2), pp. 73–79.

IIA-3LoD (2013) 'The three lines of defense in effective risk management and control. Position Paper, Institute of Internal Auditors'.

ISO 31000 (2009) 'ISO 31000:2009 Risk management - Principles and guidelines', *Risk Management*, 31000, p. 24. Available at: https://www.iso.org/standard.

Libby, R., Bloomfield, R. and Nelson, M. W. (2002) 'Experimental research in financial accounting', *Accounting, Organizations and Society*, 27, pp. 775–810. doi: 10.1016/S0361-3682(01)00011-3.

Likert, R. (1932) 'A Technique for the Measurement of Attitudes', *ARCHIVES OF PSYCHOLOGY*, 22, pp. 5–55.

Lundqvist, S. A. (2015) 'Why firms implement risk governance - Stepping beyond traditional risk management to enterprise risk management', *Journal of Accounting and Public Policy*. Elsevier Inc., 34(5), pp. 441–466. doi: 10.1016/j.jaccpubpol.2015.05.002.

MacKenzie, S., Podsakoff, P. and Podsakoff, N. (2011) 'Construct Measurement and Validation Procedures in MIS and Behavioral Research : Integrating New and Existing Techniques', *MIS Quarterly*, 35(2), pp. 293–334.

Monda, B. and Giorgino, M. (2013) 'An Enterprise Risk Management maturity model', in *Enterprise Risk Management Symposium*.

Nunnally, J. C. and Bernstein, I. H. (1994) *Psychometric theory*. 3rd edn. New York: McGraw-Hill.

Peffers, K. *et al.* (2008) 'A Design Science Research Methodology for Information Systems Research', *Journal of Management Information Systems*, 24(3), pp. 45–78.

Petter, S., Straub, D. W. and Rai, A. (2007) 'Specifying Formative Constructs in Information Systems Research', *MIS Quarterly*, 31(4), p. 623. doi: 10.2307/25148814.

Rossiter, J. (2002) 'The C-OAR-SE procedure for scale development in marketing: A comment', *International Journal of Research in Marketing*, 19(1), pp. 305–335.

Rossiter, J. R. (2008) 'Content validity of measures of abstract constructs in management and organizational research', *British Journal of Management*, 19(4), pp. 380–388. doi: 10.1111/j.1467-8551.2008.00587.x.

SCAMPI Upgrade Team (2011) *Standard CMMI Appraisal Method for Process Improvement (SCAMPI) A, Version 1.3: Method definition (CMU/SEI-2011-HB-001)*, *Software Engineering Institute, Carnegie Mello University*. Available at: http://www.sei.cmu.edu/library/abstracts/reports/11hb001.cfm.

Schumacher, A., Erol, S. and Sihn, W. (2016) 'A Maturity Model for Assessing Industry 4.0 Readiness and Maturity of Manufacturing Enterprises', in *Procedia CIRP*. doi: 10.1016/j.procir.2016.07.040.

SEI (1993) 'Capability Maturity Model'.

SEI (2010) 'CMMI for Development, Version 1.3', *Technical Report, CMU/SEI-2010-TR-033, Carnegie Mellon University*.

Simon, H. A. (1996) *The sciences of the artificial*, *Cambridge, MA*.

12 Appendix

12.1 Empirical Identification of Measurement Models

In reflective models (Figure 22, left panel) the loadings (λ_i) capturing the magnitude of the effects the construct (ξ) has on the indicators/measures (x_i) are empirically determined in a way to reduce the uniqueness in the measures (δ_i) as much as possible (Edwards, 2011). Similar as in factor analysis, the linear combination of measures is sought that best explains the variances in the measures. From a conceptual perspective, reflective models essentially represent a method to reduce measurement error. Thus, the technique to identify the model in a way that the error in the measures is reduced to their uniqueness is perfectly in line with the research objective.

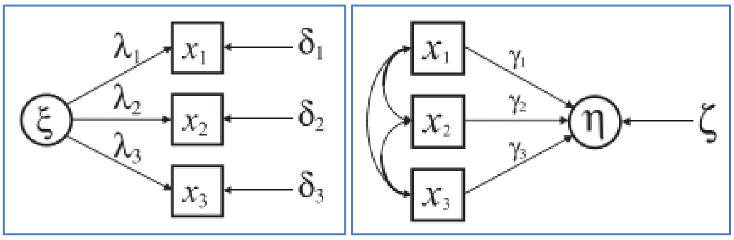

Figure 22: Reflective (left) and formative (right) measurement models (Figure 1&2 in Edwards (2011))

Formative models on the other hand, as depicted in the right panel of Figure 22, suffer from an identification problem. The indicators/measures (x_i) are unique aspects of the construct (η) and there is no assumption that they correlate in one way or another. The identification of the model should produce the loadings (γ_i) that specify how the indicators combine to define the construct. But different to reflective models where the objective of the identification was to minimize the error terms (δ_i), there is no such objective given for formative models.

To overcome this, the value for the construct (η) could, e.g. be derived from a measurement model with two reflective indicators (y_i) and then a combination of the formative indicators can be sought that minimizes the error term in the construct (ζ). This would produce a multiple indicator multiple cause (MIMIC) model (Figure 23). But as Edwards (2011) points out there are deficiencies in this method, not least that the model will change whenever other reflective indicators are chosen to determine the value of the construct.

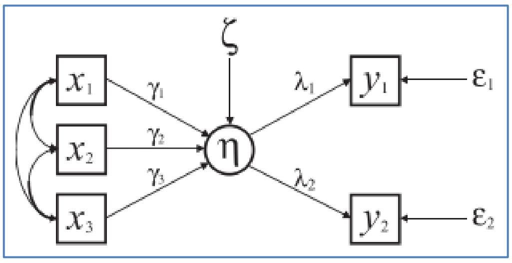

Figure 23: Formative measurement model incl. reflective indicators for identification (Figure 3 in Edwards (2011))

The need to implement formative models in bigger model structures in order to be able to identify them gives rise to the debate if there is any utility or interpretability whatsoever when such constructs are identified empirically.

12.2 Scale vs. Index, Additional Example

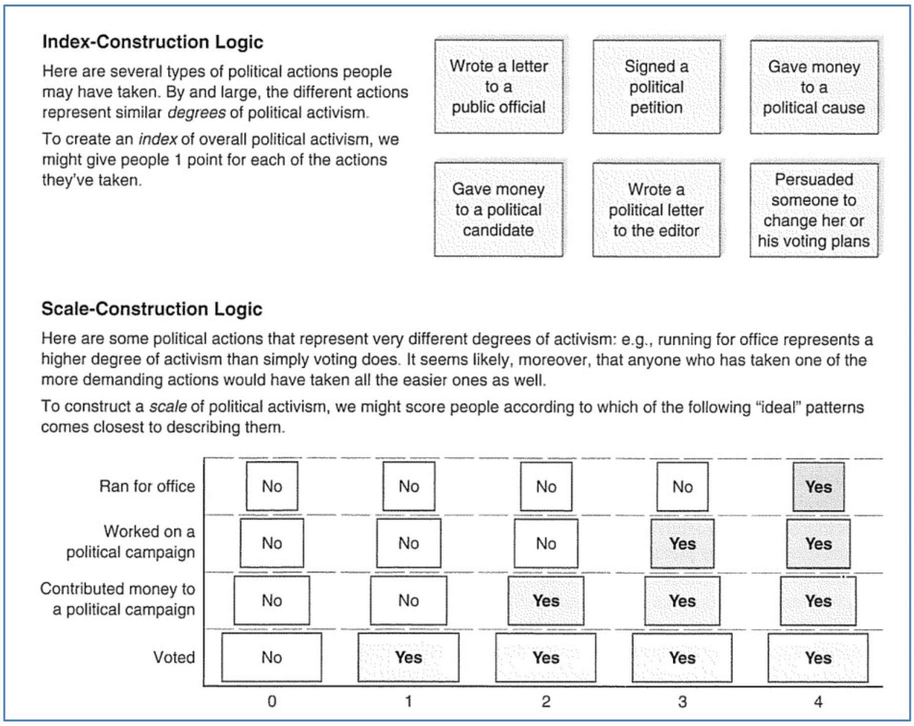

Figure 24: Scale vs. Index of political activism (Figure 6-1 in Babbie (2006, p. 155))

12.3 Additional CMMI Definitions

Table 5: Definition of CMMI capability levels

Capability Level	Definition
CL 0: Incomplete	A process that either is not performed or is partially performed. One or more of the specific goals of the process area are not satisfied.
CL 1: Performed	A process that accomplishes the needed work to produce work products; the specific goals of the process area are satisfied.
CL 2: Managed	A performed process that is planned and executed in accordance with policy; employs skilled people having adequate resources to produce controlled outputs; involves relevant stakeholders; is monitored, controlled, and reviewed; and is evaluated for adherence to its process description.
CL 3: Defined	A managed process that is tailored from the organization's set of standard processes according to the organization's tailoring guidelines; has a maintained process description; and contributes process related experiences to the organizational process assets.

Table 6: CMMI – Generic goals and generic practices

Generic Goal	Generic Practice
GG 1: Achieve Specific Goals	
	GP 1.1 Perform Specific Practices
GG 2: Institutionalize a Managed Process	
	GP 2.1 Establish an Organizational Policy
	GP 2.2 Plan the Process
	GP 2.3 Provide Resources
	GP 2.4 Assign Responsibility
	GP 2.5 Train People
	GP 2.6 Control Work Products
	GP 2.7 Identify and Involve Relevant Stakeholders
	GP 2.8 Monitor and Control the Process
	GP 2.9 Objectively Evaluate Adherence
	GP 2.10 Review Status with Higher Level Management
GG 3: Institutionalize a Defined Process	
	GP 3.1 Establish a Defined Process
	GP 3.2 Collect Process Related Experiences

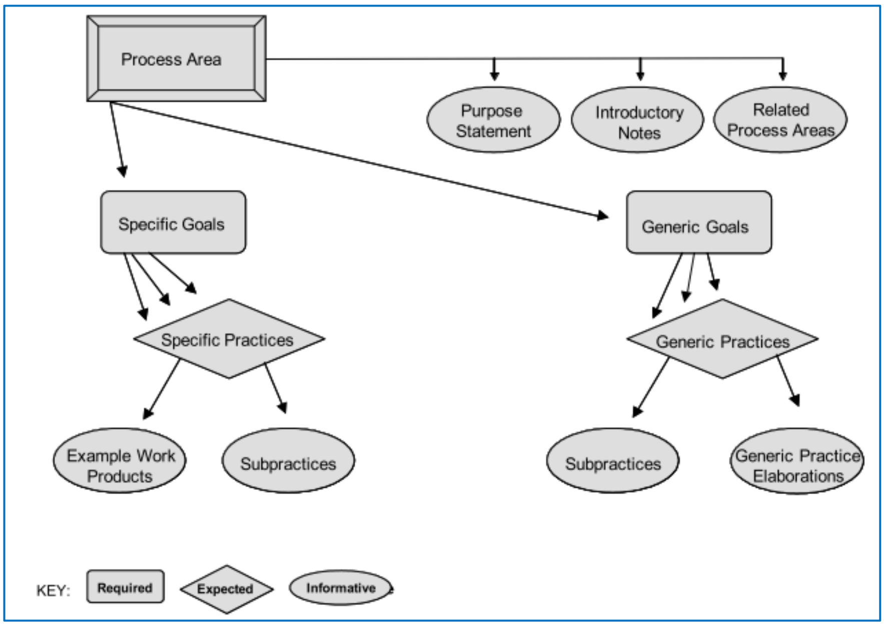

Figure 25: CMMI – Process Area, all components

Table 7: Characterization of practice implementations

Label	**Meaning**
Fully Implemented (FI)	Sufficient artifacts and/or affirmations are present (per 1.1.4, Determine Appraisal Scope and 2.4.1, Verify Objective Evidence) and judged to be adequate to demonstrate practice implementation, and no weaknesses are noted.
Largely Implemented (LI)	Sufficient artifacts and/or affirmations are present (per 1.1.4 and 2.4.1) and judged to be adequate to demonstrate practice implementation, and one or more weaknesses are noted.
Partially Implemented (PI)	Some or all data required (per 1.1.4, Determine Appraisal Scope and 2.4.1, Verify Objective Evidence) are absent or judged to be inadequate, some data are present to suggest some aspects of the practice are implemented, and one or more weaknesses are noted. OR Data supplied to the team (artifacts and/or affirmations) conflict –some data indicate the practice is implemented and some data indicate the practice is not implemented, and one or more weaknesses are noted.
Not Implemented (NI)	Some or all data required (per 1.1.4, Determine Appraisal Scope and 2.4.1, Verify Objective Evidence) are absent or judged to be inadequate, data supplied does not support the conclusion that the practice is implemented, and one or more weaknesses are noted.
Not Yet (NY)	The basic unit or support function has not yet reached the stage in the sequence of work, or point in time to have implemented the practice.

12.4 Additional ERM Maturity Assessment Definitions

Table 8: ERM maturity model - attributes

	ML 1	ML 2	ML 3
B1: RM Process	B1.ML1.1a: Specific risk identification B1.ML1.1b: Specific risk measurement B1.ML1.1c: Specific risk evaluation and risk response policies	B1.ML1.1a: Specific risk identification B1.ML1.1b: Specific risk measurement B1.ML1.1c: Specific risk evaluation and risk response policies B1.ML2.1d: Specific RM process monitoring B1.ML2.2: Specific review of RM process and monitoring	B1.ML3.1a: Enterprise-wide risk identification B1.ML3.1b: Enterprise-wide risk measurement B1.ML3.1c: Enterprise-wide risk evaluation and risk response policies B1.ML3.1d: Enterprise-wide RM process monitoring B1.ML3.1e: Enterprise-wide coordination of RM process B1.ML3.1h: Different concepts for monitoring and adjusting activities in different domains B1.ML3.2: Enterprise-wide review of RM process and monitoring

13 List of Figures

14 List of Tables