

On the factors influencing confidence in models and simulations for decision-making: a survey

Johannes Schwarzburg^{1,2}, Jakob Trauer^{1,3} and Eric Rebentisch²

¹Laboratory for Product Development and Lightweight Design, TUM School of Engineering and Design, Technical University of Munich, Garching, Germany

²Sociotechnical Systems Research Center, Massachusetts Institute of Technology, Cambridge, MA, USA

³em engineering methods AG, Darmstadt, Germany

Abstract

Over the last decades, modeling and simulation have become central methods in engineering design. Today's technologies enable previously unachievable levels of sophistication and accuracy. However, if decision-makers are unaware of the confidence they can place in models and simulations (M&S), they either fail to leverage their potential by not involving them in processes or make judgments based on unreliable results. Assessments to evaluate M&S exist, but factors that enable decision-makers to have confidence and improve acceptance of using M&S need to be researched in more detail. Therefore, a literature review analyzing design requirements and an online survey to measure factors associated with confidence were conducted. As a result, the survey identified nine predictors of confidence: (1) capability, (2) history, (3) validity, (4) reliability and (5) accessibility of the model. Further, (6) integrity and (7) competence of the modeler, as well as (8) trusting nature and (9) risk awareness of the stakeholder were identified. Having confidence in M&S results significantly increases the reliance on them and leads to better-informed decision-making. Therefore, based on the findings, a framework and an initial application model were developed. The results were initially evaluated and are described.

Keywords: Models and simulations, Decision-making, Confidence, Advanced engineering, MBSE, VVUQ

Received 31 July 2023
Revised 23 May 2024
Accepted 24 May 2024

Corresponding author
Eric Rebentisch
erebenti@mit.edu

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Des. Sci., vol. 10, e15
journals.cambridge.org/dsj
DOI: 10.1017/dsj.2024.14

the **Design Society**
a worldwide community

 **CAMBRIDGE**
UNIVERSITY PRESS

1. Introduction

Driven by increasing global competition, ongoing digital transformation, fast-paced trends and constantly changing market demands, engineering companies need to radically adapt to remain successful (Dumitrescu *et al.* 2021). To cope with these challenges and increasing uncertainties, evermore modeling and simulation (M&S) is used in situations that are too costly, complex, dangerous or time-consuming to be analyzed empirically – especially for decision-making activities, M&S becomes increasingly important (Isaksson & Eckert 2020). Applications range from model-based systems engineering (e.g., Madni *et al.* 2018), to data-driven engineering approaches (e.g., Trauer *et al.* 2020), to digital twins (e.g., Trauer *et al.* 2022b). However, as Box (1979) rightly stated, “All models are wrong, but some are useful”. Therefore, decision-makers carefully need to assess whether they can trust a model

or not. Particularly in the collaboration of design and simulation engineers, this trust may not exist and hinders collaboration (Maier *et al.* 2008; Schweigert-Recksiek 2021). Consequently, it is essential to offer possibilities to assess the credibility and validity of models and simulations (Isaksson & Eckert 2020).

Determining confidence in M&S results to better manage the mentioned trade-offs and advance toward increasing the reliance on M&S results has been known for more than 40 years. At that time, confidence was already seen as an attribute of the users, which involved engineers, modelers, managers and other decision-makers (Gass & Joel 1981). In the field of concurrent engineering, Maier *et al.* (2008) investigated several factors hindering the collaboration and communication of design and simulation engineers. However, they did not focus on confidence in M&S but on trust between the engineers. Further, they did not derive guidelines that could be adopted by practitioners. There are a few guidelines, assessments and frameworks for assessing confidence in M&S results, but primarily, theoretical perspectives have been explored for now. Chaudhari (2022) introduced a framework based on a literature review and expert interviews, which suggests that confidence in M&S emerges through a combination of model-, modeler- and stakeholder-related constructs. The framework is not ready for practical implementation because the relationships between these constructs were not validated.

To understand whether decision-makers could rely on M&S results, the focus has been on the evaluation of credibility, which predominantly includes technical aspects and is closely related to verification and validation (V&V) (Steele 2008). Factors restricting effective model-centric decisions are known and can be categorized into model-related (e.g., lack of transparency) and human-related (e.g., communication barriers) components (Rhodes 2018; Schweigert-Recksiek 2021). In comparison, understanding factors that predict confidence and influence relationships between them to build confidence in M&S is limited (Chaudhari, Rebentisch & Rhodes 2022). Trust, which incorporates many social aspects, is another crucial prerequisite before integrating M&S results into decision-making processes (German & Rhodes 2017). The main challenges for research are to objectively determine confidence and trust through sociotechnical assessments and quantitative measures to be able to evaluate the suitability of M&S outcomes for their integration into decision-making processes (Chaudhari *et al.* 2022; Trauer *et al.* 2022a,2022b).

Within the future directions for assessing confidence in M&S, Chaudhari *et al.* (2022) state that an objective assessment requires an empirically validated set of factors influencing confidence as well as their relationships. Initial hypotheses based on confidence-inspiring factors were formulated as a precursor for the framework that need to be tested and consolidated using tools such as surveys and observational studies. Practical implementation of the framework requires measuring the characteristics of M&S and decisions based on them, stakeholder preferences, and the expertise of the modelers (Chaudhari *et al.* 2022).

The overarching objective for this study is to design and develop an assessment for M&S, which decision-makers could look at to evaluate the confidence in M&S results to improve the existing framework. This article aims to provide factors that predict confidence in M&S to extend the framework. Consequently, this framework offers the possibility for modelers to manage the trustworthiness of their models as well as for decision-makers to be able to comprehensively assess whether they can have confidence in a model to use it for their decisions.

1.1. Structure of the article

[Section 2](#), state of the art, begins with current practices in M&S-based decision-making. It includes a literature review on confidence, credibility and trust for evaluating M&S results and the relationships between them. Within the approaches to determine and evaluate the application readiness of M&S, the existing framework to assess confidence in M&S is described. Based on the state of the art, the detailed research objective ([Section 3](#)) is derived. The analysis of the current situation and improvement needs of the framework ([Section 4](#)), and a survey as a research methodology to measure factors associated with confidence ([Section 5](#)) lead to the results in [Section 6](#). Core insights from the survey are presented, and an application model that complements the updated framework is introduced, focusing on a confidence assessment and decision support. The discussion ([Section 7](#)) interprets the results, addresses limitations and constraints and explores the implications and relevance of the findings for academia and industry. The conclusion ([Section 8](#)) and the outlook for future research directions ([Section 9](#)) finalize the article.

2. State of the art

2.1. Model- and simulation-based decision-making

The demand to base decisions on M&S results is increasing continuously (German & Rhodes 2017). The major concern of decision-makers, which could include developers and users, but also individuals who are dependent on the decisions, is whether the results are correct for each problem to be solved (Sargent 2015).

According to Oberkampf (2021), six main factors influence the decision process to achieve a decision result. These include the familiarity and reliability of information sources, risk tolerance and potential reward, experience with available options, personal goals and value systems, organizational goals and competition, and the return on investment/ profit margin. Organizations need to consider the risks and benefits of a decision and therefore should only make an investment in M&S if the benefits (e.g., reducing the time to market, increasing the profit margin) exceed the perceived risks. Factors that contribute to risk, in that case, are the probability of occurrence and the negative outcome. Especially in regulatory decision-making, an incorrect decision would have higher consequences for public and environmental safety and is therefore very impactful. While surprises are hard to manage for high-consequence decision-making, uncertainty is omnipresent (Oberkampf 2021).

As specified by German & Rhodes (2017), M&S-based decision-making involves several key components, which include a representative model, human actors and a decision that needs to be made. The model is seen as an important asset for decision-making, as it not only generates information facilitating the decision during its use but also throughout its development. Model information flows from one actor to another until it reaches a final decision-maker. Actors can be modelers, analysts, engineers, managers and other people in the organization. In the context of different decisions, individuals could find themselves in various roles during the process of making a final decision. The decisions to be made can either be discrete or even a set of decisions that are influenced by several smaller decisions during the information flow (German & Rhodes 2017). Open feedback and transparent

communication are among the key attributes for effective decision-making, especially as senior decision-makers are not necessarily technical experts in the models and need to develop trust in individuals who have the required expertise and capability (German & Rhodes 2017; Schweigert-Recksiek 2021).

A model is a “[...] physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process” (Sokolowski & Banks 2010). A further definition describes a model as a “simplified reproduction of a planned or existing system with its processes in a different conceptual or physical system” (VDI 2018). A model that differs from the original system within a certain tolerance, serves to perform a task that could not be done in physical operation or only with a great expenditure (VDI 2018).

Models are used for simulation, architecture definitions, requirements, behaviors and verification and validation (V&V). Additional applications are to gain a general understanding and insight, performance analysis, documentation support, communication, visualization and model-based decision support (Madni *et al.* 2018).

A simulation is seen as the “application of a model to produce a result [...]” (Roy & Oberkampff 2011). The Association of German Engineers (VDI) defines simulation as “the representation of a system with its dynamics in an experimental model to reach findings which are transferable to reality” (VDI 2018). Moreover, Balci (2010) adds to this definition of simulation as he explains it as “[...] experimenting with or exercising a model or several models under diverse objectives such as problem-solving, training, acquisition, entertainment, research, and education” (Balci 2010). Figure 1 presents an overview of how M&S are used to receive insights based on their results. Models are implemented in simulations that are executed and subsequently create results. Analyzing the results leads to insights that can be used to make decisions or refine models (Sokolowski & Banks 2010).

M&S helps to understand the behavior of complex systems, especially for those not yet created or where modifications are expensive (Olsen & Raunak 2019). The abstraction of reality, one of the core M&S functionalities, is enabled through assumption-based predictions that try to make sense of and explain the real world

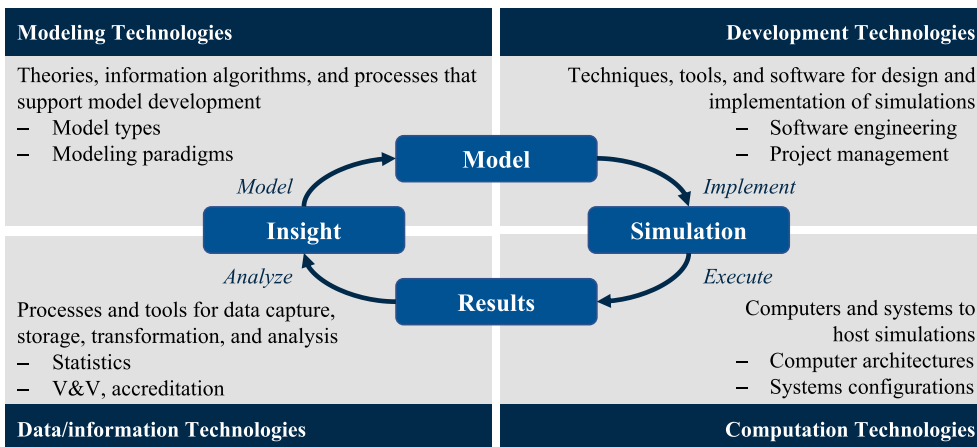


Figure 1. Connection of implemented models, executed simulations, analyzed results and gained insights supported by relevant technologies (adapted from Sokolowski & Banks 2010).

and support decision-makers (German & Rhodes 2017). When these future directions for systems are ambiguous, a crucial functionality of M&S is to provide guidance in uncertain environments. In that way, M&S helps to understand system and component relations by representing a real-world system (Dunke & Nickel 2021).

M&S are utilized in multiple domains to understand system behavior and abstract reality and predict future directions for product development (German & Rhodes 2017; Olsen & Raunak 2019). Marshall *et al.* (2017) define model-based engineering as the application of digital tools, artifacts and environments in executing engineering activities intended to offer improved efficiency, flexibility and further benefits. Figure 1 presents an overview of how M&S are used to receive valuable insights based on their results supported by different technologies for each step.

The judgment or choice to use a model or simulation for critical decisions depends on “(i) the consequences from selecting a model, (ii) stakeholder’s beliefs about model and modeler characteristics, (iii) individual preferences, for example, willingness to take the risk and (iv) contextual factors, for example, organizational boundaries [...]” (Chaudhari *et al.* 2022). If the trustworthiness of the M&S and the modeler surpasses the perceived decision risk, it is expected that stakeholders have a higher level of confidence to rely on the results to make a decision (Chaudhari *et al.* 2022).

The significance of M&S results for the decision process mainly depends on the decision-maker’s understanding and assessment of the overall M&S process. As decision-makers are among the primary users of the M&S results, it is their responsibility to determine the amount of confidence that should be given to the results. The degree to which results influence a decision specifies the amount of confidence a decision-maker has in M&S. In this instance, it is possible to distinguish between confidence in the usefulness and outputs of M&S. The amount or level of confidence mentioned is application- and user-specific due to differing requirements and judgmental preferences (Gass & Joel 1981).

However, Harper, Mustafee & Yearworth (2021) acknowledge that decision-makers need to utilize insights beyond the M&S results to identify opportunities for action. They also highlight that the overall M&S process should already incorporate elements that positively impact trust that are important for the stakeholder or decision-maker (Harper *et al.* 2021). Regarding perception, it is important to mention that individual perspectives can differ from organizational ones. Perception is relevant for stakeholders regarding the M&S results, but also concerning the modeler (Oberkampff 2021; Chaudhari *et al.* 2022).

2.2. Literature on confidence, credibility and trust measures to evaluate model and simulation results

2.2.1. Confidence

In general, confidence is defined as the certain expectation that something is going to happen without the consideration of failing (Blomqvist 1997). Regarding model predictions, confidence is specified as estimating “how likely the prediction would be correct” (Rechkemmer & Yin 2022).

In the context of M&S, confidence is also described as model confidence, which is explained as a complex sociotechnical concept that “determines a stakeholder’s decision to select a model for critical engineering work” (Chaudhari *et al.* 2022). In this article, we stick to this definition of confidence in M&S. Model confidence

should be seen as an attribute of the user rather than the model itself. To achieve confidence in a model, a close collaboration between the developer and the user is crucial (Gass & Joel 1981). Achieved model confidence enables efficient transdisciplinary engineering as organizations often rely on cross-functional teams where a knowledge exchange across disciplines is required (Chaudhari *et al.* 2022).

According to Gass & Joel (1981), confidence results from summarized information that defines and leads to a decision-maker's judgment. The formation of this decision is affected by the beliefs and preferences of the decision-maker (Chaudhari *et al.* 2022). Sargent (2015) states that confidence in a model increases its value for a user up to a certain degree. In the context of making critical decisions, model confidence represents that using the model "results in expected outcomes under given contextual circumstances" (Chaudhari *et al.* 2022).

Specifically, for utilizing M&S results for decision-making, model confidence includes the willingness to base decisions on them and expresses the user's attitude regarding the model (Gass & Joel 1981). Harper *et al.* (2021) complement that the nature of the supported decision-making task also contributes to the confidence in M&S results. Determining whether a model suits a task is critical before relying on its results for decision-making. Attributes such as accuracy, validity, representativeness and coherence should be part of an assessment that aims to identify the degree of confidence that a stakeholder can associate with a model (Chaudhari *et al.* 2022).

2.2.2. Credibility

Similar to confidence, the importance of credibility for M&S in various application fields increased during the last decades and has its origin in the 1980s (Mehta *et al.* 2016). The concept has significance for scientists, engineers and decision-makers as it combines qualitative and quantitative aspects of evaluating M&S results (Steele 2008). Credibility is generally defined as an "actor's perceived ability to perform something he claims he can do on request" (Blomqvist 1997). In an M&S context, credibility is also referred to as simulation credibility and specified as "the principal omnipresent measure in the utilization of simulated quantities" (Mehta *et al.* 2016). Another description of credibility takes the "quality to elicit belief or trust in M&S results" (NASA 2019) into account. Credibility refers to building the needed confidence in the use of M&S but also deriving insights that could influence decisions. Components that influence the overall credibility of M&S results can be distinguished into data-, model- and M&S use-related credibility (Sargent 2015; Vin 2015).

Credibility is considered a forward-looking concept, which is evaluated and determined over time. As outlined above, credibility depends on the correctness and accuracy of results. If the models do not demonstrate these, overall credibility declines. Further aspects that contribute to credibility are the context of the model use and supporting features that have high credibility themselves (Yilmaz & Liu 2020; Chaudhari *et al.* 2022).

Furthermore, the credibility of the M&S results indicates that they are believable and worthy of confidence. Underlying elements for credibility, in this case, are the quality of modeling and, thus, the analysts conducting the work, as well as verification, validation and uncertainty quantification (VVUQ) activities, including sensitivity analysis. The credibility of results is essential when M&S are utilized in decision-making for engineering projects. Chaudhari *et al.* (2022) mention that

evaluating credibility should go beyond numerical ratings and highlight that communication is one key aspect in a multi-stakeholder environment. A manager must judge the credibility of M&S results based on previous and own judgment, which includes knowledge about the actors involved in the process (Oberkampff & Roy 2010).

Model credibility is a fundamental prerequisite for model curation (Rhodes 2022). A suggested set of heuristics highlights the relationship between credibility and curation by specifying factors and attributes that contribute to the overall M&S credibility within a model curation context. These include aspects such as communication, trustworthiness, expertise and acceptance that influence credibility (Rhodes 2022).

2.2.3. Trust

Trust can be defined as “the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party” (Mayer, Davis & Schoorman 1995). Lee & See (2004) explain trust as “the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability,” which is used as the definition of trust within this article.

Both definitions include at least two stakeholders, which Thielsch, Meeßen & Hertel (2018) describe as the trustor and the trustee. The relationship toward trusting each other is built incrementally and can therefore also be recognized as a process outcome (Blomqvist 1997). Specifically in the context of M&S, trust is “a stakeholder’s belief that the model performs its functions accurately and efficiently” (Chaudhari *et al.* 2022).

Trustworthiness mainly evaluates the degree of trust while addressing the reliability of model- and modeler-related attributes, which largely impacts the overall success of an M&S study (Chaudhari *et al.* 2022). High levels of trustworthiness are needed for M&S-based decisions, which could have high-risk outcomes (Yilmaz & Liu 2020). Next to credibility, trustworthiness is highlighted in the dimensions of confidence in a model and the provider’s credibility when determining the successful delivery of a simulation project (Robinson & Pidd 1998). The process of developing trust spans and evolves through the whole life cycle of an M&S study and starts with the very first step of creating the model or simulation (Harper *et al.* 2021). A decision-maker can gain trust in the model if its results match the requirements of the user regarding reliability and accuracy (Yilmaz & Liu 2020).

2.2.4. Relationships between confidence, credibility and trust

Rhodes (2022) specifies model confidence, trust and value as linked constructs to model credibility, which changes with time and context. Gass (1993) utilizes confidence and credibility interchangeably when describing critical prerequisites for decision-makers seeking value from models. Differentiating trust from confidence is characterized by attribution and perception, not probability. In a situation of trust, specific action is preferred to alternatives, although there is a potential for disappointment. Risk is acknowledged in a situation of trust, whereas confidence accepts it. The relationship between both is not a “zero-sum game in which the

more confidence is given, the less trust is required and vice versa” (Luhmann 2000). Instead, confidence, credibility and trust are interrelated and need to be assessed context-specifically. Credibility and trust are rooted in stakeholders and their perception (Luhmann 2000; Chaudhari *et al.* 2022).

Gass & Joel (1981) conclude that confidence indicates credibility and reliability, which are measurable after using a model. Yilmaz & Liu (2020) even identify credibility as trust in the context of building confidence in M&S. Rhodes (2022) characterizes model confidence and model trust as concepts associated with model credibility. Blomqvist (1997) explores the connection between confidence and credibility toward trust. He claims that trust involves considering alternatives, while confidence does not. Credibility is seen as a “passive concept referring to the actor’s claimed ability” (Blomqvist 1997), which does not necessarily include the intention or action of the stakeholder. In comparison, trust would include these aspects of performing the requested action. In summary, confidence in M&S results is identified as a precursor for trustworthy action and behavior. To build confidence initially, a decision-maker should have a certain amount of trust in the model and modeler, but also the process which leads to deriving insights from the model (Blomqvist 1997; Harper *et al.* 2021). Further, according to Maier *et al.* (2008), “mutual trust” among engineers is a core factor influencing communication in product development.

2.3. Determining application readiness of models and simulations

2.3.1. Approaches to assess and evaluate model and simulation results

The need for comprehensive M&S assessments in science and industry to ensure that M&S produces credible results that can be trusted and utilized with confidence in decision-making is widely established. For example, Trauer *et al.* (2022b) introduced a Digital Twin Trust Framework for industrial application, which includes specific recommendations and concrete measures to increase trust in the concept itself and among stakeholders. As digital twins are based on M&S, these measures might also be applicable to M&S for decision-making. However, the framework has not yet been applied in the industry and lacks quantitative assessments of trust. Also, in other areas, M&S assessments are far from being implemented for various reasons, such as the lack of expertise, especially in VVUQ methods (Wright *et al.* 2020). VVUQ are referred to as processes to assess prerequisites to employ the results of M&S. Moreover, VVUQ is essential for developing trust and confidence in M&S to utilize results for critical decisions. VVUQ includes confidence-building assessments and evaluations to genuinely empower decision-makers (Oberkampf 2021). For M&S results to be broadly adopted and thus incorporated into decision-making processes, it is important that these sociotechnical concepts, such as trust in the model, can be quantified (Wright *et al.* 2020).

To effectively utilize M&S and thus make critical decisions based on them, decision-makers and other involved stakeholders need to know qualitative and quantitative aspects contributing to credibility and confidence of the M&S and their results (Balci 2012; Mehta *et al.* 2016). Not only M&S independently but also a combination of involved and related products, processes, people and projects leads to a significantly higher degree of certainty and decision-making confidence.

The objective is to strive for the highest level of completeness and detail to provide the highest achievable quality for the M&S application assessment (Balci 2012).

Further aspects that indicate that M&S are ready for use are associated with the decision criticality and M&S users. Users should be familiar with and prepared for the specific M&S results. This includes developing awareness of practices and training with the specific M&S types used, as well as comprehending a user guide (NASA 2019; Schweigert-Recksiek 2021; Trauer *et al.* 2022b). Aspects for the M&S that form the guide's basis and should be included in any case are assumptions, abstractions and their rationales, basic structure and mathematics, operational limits and permissible uses (NASA 2019).

2.3.2. Existing framework to assess confidence in models and simulations

The existing confidence framework of Chaudhari *et al.* (2022) consists of three main parts and associated types of constructs that should contribute to confidence in M&S. These constructs are model-, modeler- and stakeholder-related. Stakeholders are understood as individuals who assess or use the M&S. Each type is further separated into three constructs, and each construct is further divided into at least two attributes. The constructs can be seen as suggested factors contributing to confidence in M&S. A description provided for each attribute can contain multiple technical or social aspects. The framework aims to increase the acceptance and effectiveness of M&S and their utilization for critical decisions. The constructs are designed to facilitate measuring confidence in M&S (Chaudhari *et al.* 2022).

The scope of the framework indicates that measuring and evaluating confidence in M&S and their results requires a comprehensive approach. Many technical and social factors need to be taken into consideration to assess confidence. Especially the descriptions of the attributes offer a variety of options for the operationalization and improvement of the framework for its practical implementation. All constructs and attributes of the framework, as well as hypothesized relationships to predict confidence in using M&S, are visualized in Figure 2 (Chaudhari *et al.* 2022).

The described framework constructs are connected to each other through relationships that are conceptualized through four main hypotheses. Chaudhari *et al.* (2022) describe these hypotheses as follows:

H1. Modeler-related competence, benevolence and integrity influence a model's accuracy, capability and usability.

H2. A stakeholder's perception of a modeler's trustworthiness depends on the modeler's competence, benevolence and integrity and is mediated by individual and contextual factors.

H3. A stakeholder's perception of a model's trustworthiness is a function of the model's accuracy, capability and usability; and is mediated by individual and contextual factors.

H4. A stakeholder's confidence in model usage depends on the model and modeler's perceived risk and trustworthiness.

The first hypothesis H1 connects the modeler- and model-related constructs. H2 establishes a link between modeler- and stakeholder-related constructs and targets the perceived trustworthiness of the modeler. Individual and contextual

Type	Construct	Attribute	Description
Model-related constructs	Capability	Intended use	Explicit requirements, expectations, assumptions, limitations; Consistency of runtime environment, experiments, and scenario
		Fidelity	Representativeness between model specification and the real world setting
	Accuracy	Verification	Accuracy of model implementation and input data with respect to the conceptual specification
		Pedigree	Input data quality and quantity; the credibility of data source
		Validation	Level of expected uncertainty in model predictions
	Usability	Interactivity	Simulatability, textual explanations and graphic visualization
Transparency		Precise technical information including inputs, outputs, and internal code.	
Reusability		Availability of reusable solutions to address recurrent problems	
Modeler-related constructs	Competence	Domain expertise	Accumulated process knowledge from extended practice
		Task-specific ability	Influential skills, competencies, and characteristics within context
	Benevolence	Cooperativeness	Behavior demonstrating individual support
		Communication	Exchange of information and knowledge across boundaries
	Integrity	Adherence	Personal commitment to a set of principles
		Ethics	Person's moral habits, ability to decide right or wrong, and free will
Stakeholder-related constructs	Perceived trustworthiness	Modeler-specific	Perceived reliability of modeler's actions
		Model-specific	Perceived reliability of model implementation, results and insights
	Perceived risk	Criticality	Consequences to safety and the degree of decision influence
		Vulnerability	Expectation of ill-effects from actions of the modeler
	Individual and contextual factors	Trust propensity	Generalized disposition to trust others irrespective of the situation
		Organizational culture	Responsibility for acting to the results, participation policies, implementation criteria

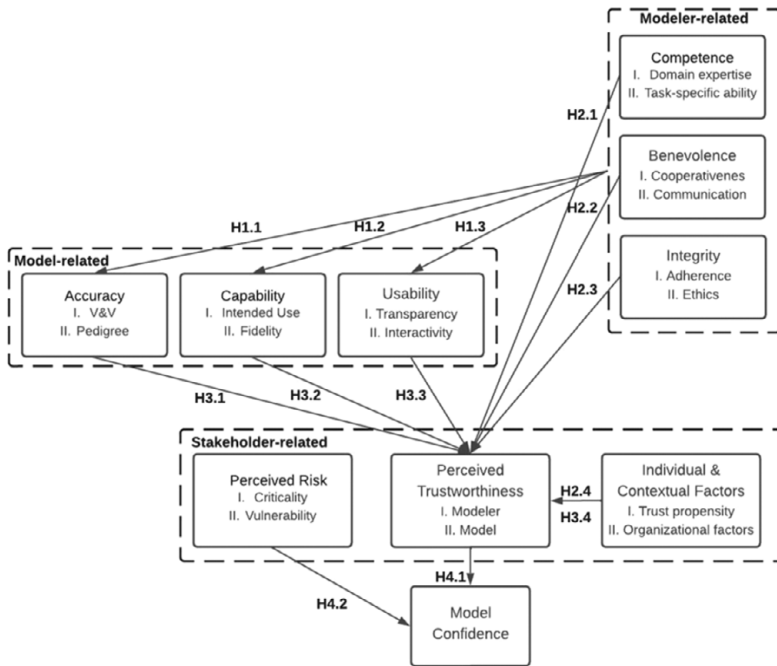


Figure 2. Top: Model confidence constructs and examples of attributes (Chaudhari *et al.* 2022). Bottom: Hypothesized construct relationships and their connection to model confidence (Chaudhari *et al.* 2022).

factors are essential components of H2 and H3. In comparison to H2, H3 associates model- and stakeholder-related constructs but focuses on the perceived trustworthiness of the model. Hypothesis H4 incorporates the perceived risk and trustworthiness of the stakeholder and influences the overall confidence (Chaudhari *et al.* 2022).

3. Research objective

As described in the introduction, the overarching goal of this research was to propose an assessment model for decision-makers to determine confidence in M&S results with the support of factors that aim at predicting confidence in M&S. To achieve this goal, we have formulated specific objectives that will be accomplished through the support of the survey:

- a. Refine and measure predictors (constructs) and indicators (attributes) that establish confidence in M&S and their results.
- b. Test and empirically validate previously hypothesized relationships between confidence predictors (see [Section 2.3](#)).
- c. Improve the usability of the updated framework through an application model as a guideline to develop M&S.
- d. Aggregate and summarize principles, practices and measures that contribute to increasing confidence in M&S.

The target state is an extended and improved framework that contributes to the acceptance of M&S for key engineering decisions and adds value to all stakeholders across the organization. It is desired that the framework provides some systematic inspirations to improve M&S-based decision-making and to support the understanding of decisive factors that contribute to building confidence. The improved framework to assess confidence in M&S should serve as a basis for digital transformation by providing a structured way of measuring confidence and trust to improve an organization's ability to use M&S for decisions (Chaudhari *et al.* 2022).

The main use case for the developed solution is to improve product feature design performance. The use case's goal is to reduce the time to develop the features and integrate them into products to save costs from the avoidance of prototyping. M&S results supported by the framework and application model should assist stakeholders during relevant milestone decisions. The scope of the use case is cross-functional, and the precondition could be that a new technology enables a new product feature. The success end condition is determined by including the framework and application model in the decision-making process to evaluate confidence in M&S results. For the use case to improve product feature design, the stakeholders' feature owner and feature manager are most applicable and provide connected user stories. An identified secondary actor next to them is an executive who makes a project or milestone decision. The trigger of the use case is that a new product feature is identified by research. Other considerable use cases are related to integrating the framework and application model into decision-making processes for the virtual testing of entire products or components.

4. Analysis of the current situation and improvement needs of the framework

Previously conducted semi-structured interviews with 20 modelers, engineers and managers within product development at an automotive company revealed the challenges of creating and using M&S as well as evaluating the suitability of their results to make decisions based on them. The interviews supported the identification of relevant stakeholders, their challenges, potential user stories and desired benefits (Chaudhari 2022).

Relevant stakeholders contributing to the creation, management and utilization of M&S are manifold and can contain users and operators, sponsors and beneficiaries, as well as simulation project managers. System analysts, subject matter experts and modeling and programming experts are also essential within the M&S life cycle. One individual can have several roles, while several individuals or teams can also share roles. Teams for M&S projects should include design and system engineers, system operators and users and simulation experts. They could involve other roles such as planners, controllers and representatives of further departments (Brade 2003; VDI 2021).

The outlined roles could be stakeholders utilizing the framework to assess confidence in M&S and its application model to evaluate M&S. They were identified in consultation with the project sponsor and are also desirable respondents for the survey: (1) Concept/Research Analysts: utilize M&S in early product development phases; (2) Software Developers/Modelers: directly involved in building/creating M&S; (3) Engineers: involved in M&S development and design or utilizing them for further engineering activities such as testing; (4) Manager: makes design and engineering decisions supported by M&S; (5) Director/Chief Engineer: makes key engineering and strategic decisions supported by M&S results.

Among other aspects, the interviews revealed that stakeholders were most skeptical about the unclear maturity of M&S and didn't have clarity about the usefulness of M&S. Overall, their environment for using M&S for engineering decisions was characterized by skepticism, complexity, lack of transparency, uncertainty and ambiguity. M&S-related, person-related and organizational as well as cultural aspects raise skepticism among stakeholders. If these aspects are not addressed, and the skepticism remains, it could lead to not accepting the M&S and making inefficient decisions based on unreliable results. Among others, organizational and cultural aspects contain inefficient collaboration with partners or suppliers, lack of a principled approach, inability to utilize industry standards and inefficiently allocated resources.

User stories of modelers and related roles include the desire to enable product development to improve the capability of product features through the development of initial concepts supported by the evaluation of M&S and the assessment of confidence with the framework. Further, the design and validation of M&S of product features could be done more efficiently for production implementation.

User stories of engineers in the context of using M&S for product development and design reflect the desire to improve component testing with reliable M&S results. The framework and application model should provide a rating of confidence predictors and suggest metrics to accelerate decisions. To improve the overall design and reduce the engineering time through integrated and optimized M&S of product features are further demands.

The user stories of executives, especially managers, who make decisions based on M&S and their results include the aim to deliver applicable product features faster to customers through developed and validated requirements. The framework and application model should support the M&S development to evaluate the application readiness of features.

Analyzing the number of responses of the interviewees revealed that most user stories are connected to the requirements specification and concept

readiness of product feature development processes and further cluster around the preliminary and critical design review in the overall product development process. User stories and their assignment to the processes indicate that the framework and application model can be of particular importance for multiple M&S stakeholders within feature and product development processes.

The desired benefits of the study to develop the framework and application model and using M&S can be structured in the three phases of improving M&S comprehensively, accelerating M&S-based decision-making and building confidence in using M&S. The main desired benefit for decision-making supported by the framework is to establish confidence within leadership teams and communities. M&S impact decision-making in beneficial ways by determining and solving critical issues early, and discovering potential failures with minimized effort. The framework should also improve decision-making by establishing collaboration to increase the transparency of the M&S life cycle and provide a rationale that requirements are being met.

As described previously, the required improvement needs include measuring M&S characteristics, testing hypotheses of the framework constructs with a survey, validating relationships between confidence, and assessing further topics such as stakeholder preferences and decision relevance (Chaudhari *et al.* 2022). Further improvement needs are to complement the framework by an application model (e.g., summary sheet) for decision-makers and to develop an assessment to evaluate confidence in M&S. Therefore, confidence attributes should be extended and operationalized to refine the framework. To evaluate M&S from different sociotechnical perspectives, the application model should be complemented by measures of confidence, credibility and trust. To assess confidence in detail, a quantitative approach and thus quantified confidence is required, along with qualitative evaluations to determine M&S application readiness.

The framework to assess confidence in M&S presented here is currently mainly based on theoretical perspectives, which were accompanied by previously conducted expert interviews. The included constructs can be extended or reduced by measuring the included and further suggested attributes. A subsequent task will be to categorize them into updated validated constructs with potentially different attributes. For the operationalization of attributes, aspects included in their descriptions will be part of the survey questions. Based on the individually measured attributes and derived updated constructs, the four initial hypotheses of the framework can be tested. The relationships between attributes, constructs and confidence types (model, modeler and stakeholder) will be evaluated, refined and potentially empirically validated with the support of the designed survey, observational studies and previous research. These steps will ultimately lead to establishing the framework to assess confidence in M&S, measuring outcomes such as reliance and confidence, and consolidating relationships between constructs, which are then seen as confidence predictors. The practical implementation of the framework needs to be addressed by measuring M&S characteristics, decision risk, preferences of stakeholders and expertise of the modelers. A subsequent desired outcome as part of the framework validation could be a ranking and objective assessment of the confidence predictors (Chaudhari *et al.* 2022).

5. Survey to measure factors associated with confidence

5.1. Design and structure of the survey

In alignment with the described objectives in the previous section, the survey should deliver insights related to M&S development, when M&S are used to make critical decisions, and what decision-makers look to establish confidence in M&S. The research methods selected for survey development were literature analysis, an analysis of previous research, and previous expert interviews.

The structure of the survey is divided into characteristics related to the participants, the framework to assess confidence and outcome measures. The framework with all constructs and attributes and outcome measures, which include measuring confidence and aspects of decision-making, are part of the survey (see [Figure 3](#)). The included questions aim at operationalizing the attributes and measuring the relevance of framework constructs toward building confidence. Therefore, aspects of explaining the attributes included in the general structure of the framework (see [Section 2.3](#)) are part of the survey questions, often without mentioning the actual attributes. Most questions to measure framework attributes are model-related and include VVUQ aspects. Additionally, the questions cover modeler-related and stakeholder-related attributes. Aspects to include in the survey were further identified and developed through discussions regarding the requirements of the project sponsor.

A general introductory text to the survey respondents included the purpose of the study, information on data collection, a broad overview of the survey and contact information. Specific instructions to follow throughout the survey were explained afterward. The instructions aimed to ensure that the current state of the adaption and utilization of M&S in organizations is captured. Therefore, participants were asked to consider a recent instance involving developing or using a model/simulation that contributed to product-related decisions, ensuring that it reflects their general experience and refer to that specific instance while answering subsequent questions.

The survey questions are consistent in their terminology and mainly close-ended to capture quantitative responses regarding the attributes. This is realized with ranking, multiple choice and mainly Likert scale questions.

Characteristics and background include designed questions regarding the use instance, the respondent's position, a differentiation between industry and research, specific industries, size of the organization, milestones where the M&S were used, the M&S type and experience with it (see [Figure 3](#)).

Survey questions related to the framework constructs and attributes are structured according to the main sections of the framework (model, modeler and stakeholder). The survey questions were operationalized, considering explanations and important aspects for each attribute. As the survey design proceeded, the rigid assignment of specific survey questions to framework attributes was partially dissolved to introduce suggestions for further attributes, measure several aspects of particular attributes and introduce VVUQ-related questions. Response options for measuring the attributes are outlined on a similar five-point Likert scale in the same direction to ensure comparability during the analysis and evaluation.

Characteristics

Use instance	Open question during which the specific use instance should be described briefly with less than 20 words.
Position	Response options are aligned with the previously described stakeholders (concept/research analyst, software developer/modeler, engineering, manager, director/chief engineer)
Industry/ research	The industry is requested with twelve response options (e.g., automotive, aerospace) of which one is an open response option to specify the industry if not included. A question is included to ask if the respondent works in industry or research.
Company size	Response options for the number of the employees who work at the organization include specifications for small and medium-sized companies and range from up to nine employees to above 500.
Milestones	Included response options for milestones around which M&S results were primarily used to make decisions are based on feature development and generic product development processes.
M&S type	The response options include types such as engineering, system, dynamic, and agent-based M&S.
Experience	Years of experience with the specific type of M&S are requested with an open response question.

Initial framework with attributes to measure

Type	Construct	Attribute
Model-related constructs	Capability	Intended Use
		Fidelity
	Accuracy	Verification
		Pedigree
		Validation
	Usability	Interactivity
Transparency		
Reusability		
Modeler-related constructs	Competence	Domain expertise
		Task-specific ability
	Benevolence	Cooperativeness
		Communication
	Integrity	Adherence
Ethics		
Stakeholder-related constructs	Perceived trustworthiness	Modeler-specific
		Model-specific
	Perceived risk	Criticality
		Vulnerability
	Individual and contextual factors	Trust propensity
Organizational culture		

Further suggested attributes

<i>Use of M&S standards</i>	Assessed by determining the degree to which the process of building the specific M&S adheres to defined standards or guidelines.
<i>Curation</i>	Measured regarding its use and adherence. The use determines how well curation is established in the organization, and adherence how individuals adhere to curation processes.
<i>Testing</i>	Measured by the degree with which the likelihood of potential edge cases (extreme operating parameters) produced by the M&S, which could be caused by uncertainties, was assessed.

Outcome measures

Overall confidence in results (internally/ externally developed)	Measured by determining how confident the respondents were overall with respect to M&S that were developed internally or provided by third parties.
Confidence inspiration (general M&S/ VVUQ activities)	Includes response options that are based on the findings of the literature review and insights from the interview analysis.
Relevance of/ reliance on M&S for decision making	Measured by asking whether M&S was a fundamental part of the decision making process and the outcomes.
Developed using cross-functional teams	Response options range from developed with no teams or virtual teams up to cross-functional teams that may include members from other organizations (e.g., partners).
Organizational benefits	Response options include a variety of feature-related (e.g., assessing feasibility), process-related (e.g., reducing cycle time), and decision-related (e.g., reducing committed resources) benefits.

Figure 3. The structure of the survey instrument with stakeholder characteristics, attributes of the initial confidence framework by Chaudhari *et al.* (2022), further suggested attributes and outcome measures for decision-making with descriptions. All attributes were measured on a comparable 5-point Likert scale. Not included attributes of the framework are highlighted in gray.

The Likert scale is a method for assessing attitudes regarding a particular topic. Typically, it comprises a set of statements associated with the issue in question, and respondents are tasked with expressing their level of agreement for each statement on a five-point scale (Singh 2006). As recommended by Krosnick, Narayan & Smith (1996) and Menold *et al.* (2014), verbal scales were used in this study, and each scale was labeled to avoid distortion of results. Tailored response options were designed for each Likert-scale item. The process of developing the Likert scale response options involved testing, integrating feedback and piloting to ensure validity and adequacy of the questions. The Likert scales include generic (e.g., level of agreement) and individually tailored scales to the content of the questions (Vagias 2006). The survey questions related to the attributes can be found in the updated framework as questions for measurement in Figure A9.

Model attributes to be measured with the survey include all current attributes of the framework. Additionally suggested attributes based on the literature findings and interview insights focus on VVUQ-related aspects (e.g., correlation analysis,

testing) and M&S development practices (e.g., curation). Due to the requested focus on VVUQ, validation is measured with multiple attributes that include correlation analysis and uncertainty management.

The modeler attributes to be measured with the survey include all six attributes currently incorporated in the framework to assess confidence in M&S. The attributes cooperativeness, communication, adherence and ethics are all evaluated on a scale from strong disagreement to strong agreement. Adherence is determined by requesting if modelers demonstrated adherence to professional standards and willingness to reconcile differences. Ethics are measured by evaluating whether the modelers were fully transparent with the strengths and limitations of M&S within their reports.

The stakeholder attributes measured in the survey are most of the included attributes of the framework section. Decision criticality is measured with three attributes by evaluating the level of comfort with using M&S to make key decisions related to human safety, the business case of the program and regulatory authorities. The response options range from not at all comfortable to very comfortable as well as not applicable for the use instance, which is also provided as an aspect for the three criticality attributes. Trust propensity and vulnerability are not included as they are complex to capture with a few attributes.

The outcome measures included in the survey address the use of M&S and their results for engineering decision-making (see [Figure 3](#)). The focus of the questions related to this section is on confidence measures (e.g., confidence-inspiring activities), M&S practices in use (e.g., cross-functional development) and improvements due to M&S (e.g., benefits of using M&S). The confidence in M&S used in the instance is measured by determining how confident the respondents were overall with respect to M&S that were developed internally or provided by third parties, such as suppliers. The response options are that the results were rejected (no confidence), they could not rely on the results (fall back on empirical test data), results were alright (only a small part of a decision), results were useful (needed more supporting evidence) and fully confident in the results (willing to commit). General and VVUQ-related confidence-inspiring activities and practices are measured with two separate questions with clear instructions. Both questions include response options that are based on the findings of the literature review and insights from the interview analysis. Response options for the general M&S activities (ranking question) are trust in the modeler, credibility of results, explainability of results, V&V, UQ, accredited by a third party, accepted by a third party, and robustness across a range of test case conditions.

5.2. Participants of the survey

Ideally, most participants are identified as stakeholders of the framework and the application model. The targeted audience of stakeholders and profiles for the survey are modeling and simulation engineers, modelers, developers and other experts involved in the M&S life cycle, such as virtual testing engineers. Suppliers of M&S, decision-makers such as managers and other stakeholders who rely on M&S results are also part of the target audience. Experience-wise, it is important that potential respondents were at some point involved as stakeholders in model- and/or simulation-based decision-making processes, which should be the case for most individuals involved in the product development process at industrial companies.

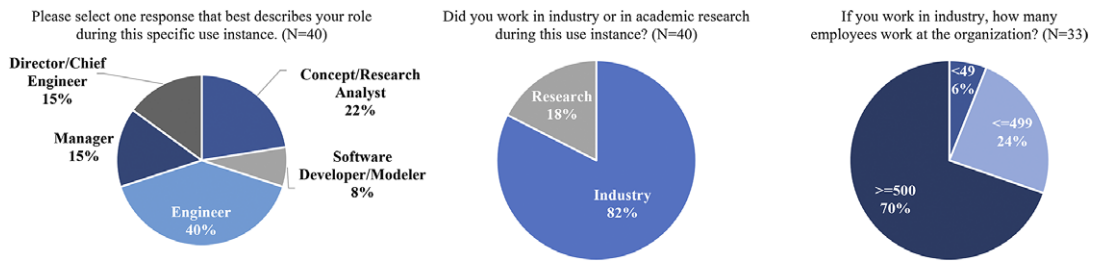


Figure 4. Characteristics of the survey participants related to their position during the M&S use (left), a differentiation in industry and research (middle), and the size of the organization (right).

A European sample and a sample of alumni of the System and Design Management (SDM) program at the Massachusetts Institute of Technology (MIT) were utilized for the survey distribution. The European sample mainly comprises experts from a network of automotive manufacturers, suppliers and startups. Individuals working in the aerospace industry are also part of the sample, and the survey was later shared with research institutes from several German universities and authors of the most relevant papers of the literature review. The MIT SDM sample comprises mid- to later-career engineering management professionals from a variety of industries. Therefore, the sample design type can be seen as a non-probability convenience sample. The survey was pilot-tested with detailed reviews and resulting adaptations in cooperation with the project sponsor. Forty responses from different sources were collected.

Most respondents were engineers during the specific use instance, followed by concept/research analysts (see Figure 4). A similar number of managers and directors/chief engineers participated in the survey. Both included samples are complementary, as the European sample included many more modelers and the SDM sample more executives. Most survey respondents had 2 years of experience with the M&S type used in the instance, followed by participants with more than 13 years of experience. Generally, the experience level of survey participants is balanced among the remaining categories. Most survey respondents worked in the industry during the M&S use instance. Most of the participants in the industry work in large organizations with 500 employees or more, while around one-third work in organizations with less than 500. The respondents worked in many different industries, with automotive being the most frequently selected, with slightly above one-third. Aerospace is the second most often represented industry, with energy close behind. Further industries where participants worked during the M&S use instance are health and medical technology, agriculture, manufacturing, military/marine, technology and retail.

5.3. Data analysis and interpretation

5.3.1. Utilized statistical analysis methods

Data were collected and acquired with the experience management software Qualtrics, through which the survey was hosted online and administered. The statistical methods used are the chi-squared test, independent *t*-test, correlation analysis, principal component analysis (PCA) and exploratory factor analysis (EFA), analysis of variance (ANOVA), regression analysis and factor scores.

The survey analysis¹ starts with correlation analysis, PCA, and EFA for each section (model, modeler, stakeholder) of the framework and the associated attributes. Pearson's correlation is utilized to identify significant relationships between each of the model, modeler and stakeholder attributes and, in Section 6.1, also with outcome measures such as confidence.

PCA and EFA are used within the survey analysis to test, validate and extend the confidence framework constructs and all previously included and additional suggested attributes. The constructs thereby represent the derived components/factors from the analysis. Factor rotation specifies the components from the PCA. It provides more details about the association of attributes, which need to be refined with suitable construct descriptions dependent on the updated included attributes. Conducted rotation methods are orthogonal rotation, which results in the rotated component matrix, and oblique (non-orthogonal) rotation, which creates the pattern matrix. The number of derived confidence predictors is compared with the initial number of framework constructs for each section. It is analyzed if the framework attributes can be combined in summarizing updated constructs that can be seen as confidence predictors with attributes as indicators.

Correlation analysis of derived constructs (predictors) and outcome measures as well as regression analysis are performed to test the framework hypotheses and connections between confidence constructs. Subsequently, additional analysis is performed for multiple-response questions. Outcome measures are investigated with independent *t*-tests and ANOVAs based on groups identified through the characteristics. ANOVAs were useful for determining differences regarding the outcome measures between groups such as different stakeholders (e.g., modelers, engineers, executives). It continues with an exploration of characteristics of the respondents (e.g., position, industry) which are investigated using chi-squared and *t*-tests to identify significant differences between groups. Pearson's chi-squared tests are used to further identify significant associations between derived categorical variables (e.g., clusters of decision-making milestones and binary outcome measures).

The results and their interpretation generate insights to refine the framework and specify the application model. Interpreting the data focuses on evaluating the results regarding the hypotheses associated with the framework. Implications from findings should generate insights to improve the overall framework and aim to derive metrics or a confidence score.

5.3.2. Model-related attributes

Primarily intended use, fidelity, verification, validation, transparency, use of M&S standards and curation show significant and strong correlations.² The results

¹All explanations regarding the statistical analysis methods chi-squared test, correlation analysis, *t*-test, PCA and ANOVA are reported as suggested by Field (2009).

²The highest and most significant correlations at the 0.01 level are between intended use (e.g., with fidelity, $r = 0.57$), fidelity (e.g., with curation adherence, $r = 0.59$), use of standards (e.g., intended use, $r = 0.5$), curation use (e.g., transparency, $r = 0.47$) and curation adherence (e.g., use of standards, $r = 0.52$). There was a significant relationship between fidelity and pedigree that also had the highest correlation among all the attributes of the model, $r = 0.73$, p (two-tailed) < 0.01 . Verification and validation have a significant relationship with each other with a moderate correlation, $r = 0.34$, p (two-tailed) < 0.05 . The highest number of correlations at the 0.05 and 0.01 levels are associated with curation adherence and use of standards. The lowest and smallest correlations are associated with interactivity and uncertainty management. $N = 40$.

indicate that there are frequent and strong relationships among model attributes, but some might be prioritized and summarized. Pearson correlations for model-related attributes are displayed in [Figure A1](#).

The PCA³ revealed that the model attributes are summarized with five components, with all existing and suggested attributes included. The first component stands out as nearly all attributes have high loadings on them. Five derived components from the initial PCA are more than the three constructs included previously in the framework. The first component has an eigenvalue that is 2.6 times higher than the eigenvalue of the second component, which indicates a higher relevance. The EFA carried out on the 15 attributes is performed with varimax (orthogonal) and oblique rotation (direct oblimin). The results of the PCA and EFA of modeler-related attributes are summarized in [Figure A2](#).

The first component is associated with fidelity, pedigree and curation as attributes and is also related to the performance of the M&S according to the EFA results. A possible description to summarize the included attributes could be capability, which requires efforts and investments to ensure reliable empirical foundations and data sources of the M&S. The second component contains intended use, fidelity, reusability and curation as attributes and relates to the robustness of the M&S and its utilization for further purposes. History might be a suitable description to summarize them as this could be used for M&S that have existed for a while, are application-specific, and individuals worked with it and documented the use of M&S. The third component, which could be described as validity, incorporates the verification, validation and testing attributes as they relate to the quality of M&S and establishing credible results. The fourth component is described as reliability, as it includes the attributes of correlation analysis and uncertainty management and focuses on techniques to assess uncertainties through the M&S life cycle. The attributes interactivity and transparency deal with the presentation of results, relate to usability aspects and can be summarized as accessibility.

5.3.3. Modeler-related attributes

Correlation analysis reveals that all included modeler attributes show a number of significant correlations, especially between expertise and ability, as well as among adherence and ethics.⁴ Pearson correlations for model-related attributes are displayed in [Figure A3](#).

³A PCA was performed on the 15 items (model attributes). The Kaiser–Meyer–Olkin (KMO) measure verified the sampling adequacy for the analysis, KMO = 0.706 ('good' according to Field 2009) and all KMO values for individual items were > 0.504, which is above the acceptable limit of 0.5. Bartlett's test of sphericity (105) = 214.82, $p < 0.001$, indicated that correlations between items were sufficiently large for PCA. An initial analysis was performed to obtain eigenvalues for each component in the data. Five components had eigenvalues over Kaiser's criterion of 1 and, in combination, explained 71.842% of the variance; this is the number of components that are included in the further analysis. The model attribute subscales of the survey demonstrate a good level of reliability and consistency between attributes of the constructs, Cronbach's $\alpha = 0.844$.

⁴There was a significant relationship between domain expertise and task-specific ability, $r = 0.83$, p (two-tailed) < 0.01. There was a significant relationship between cooperativeness and all other attributes. Significant at the 0.05 level (two-tailed) with domain expertise ($r = 0.38$) and task-specific ability ($r = 0.36$). Significant at the 0.01 level (two-tailed) with communication ($r = 0.46$), adherence ($r = 0.59$) and ethics ($r = 0.52$). Communication was significantly correlated with adherence, $r = 0.53$, and ethics, $r = 0.57$; adherence was also correlated with ethics, $r = 0.66$ (all p (two-tailed) < 0.01). $N = 38$.

The PCA reveals that modeler attributes can be summarized with two components with varying but also negative influences. The first component stands out due to the high loadings of all modeler attributes on it. The two derived components from the initial PCA are less than the three constructs included in the framework previously. The first component is influenced by all the modeler attributes with generally high loadings, although domain expertise has a relatively low loading. The second component is positively influenced by domain expertise and task-specific ability, but negatively by communication and ethics.⁵

The first component has an eigenvalue that is 1.7 times higher than the eigenvalue of the second component, indicating greater relevance. For the EFA it is expected that modeler attributes will be clearly divided and suitable component descriptions that fit the framework constructs can be identified. The conducted EFA on the six attributes is performed with varimax (orthogonal) and oblique rotation (direct oblimin). The results of the PCA and EFA of modeler-related attributes are summarized in [Figure A4](#).

The first component derived includes cooperativeness, communication, adherence and ethics as attributes and is characterized by people working cooperatively and ethically. This indicates that individuals who possess these attributes are seen as trustworthy, collaborative and generally pleasant to work with. A possible description to summarize the attributes could be integrity. The second component incorporates domain expertise and task-specific ability as attributes. The previous construct competence summarizes the included attributes well.

5.3.4. Stakeholder-related attributes

The correlation analysis revealed that all included stakeholder attributes show a number of significant correlations with a focus on trustworthiness⁶ and criticality.⁷ Pearson correlations for model-related attributes are displayed in [Figure A5](#).

⁵A PCA was conducted on the six items (modeler attributes). The KMO measure verified the sampling adequacy for the analysis, KMO = 0.669 ('mediocre' according to Field 2009), and all KMO values for individual items were > 0.539, which is above the acceptable limit of 0.5. Only domain expertise is slightly below the limit with 0.487, but because the KMO value for the overall analysis is above the acceptable limit of 0.5, it is possible to proceed with the analysis with caution when interpreting the results. Bartlett's test of sphericity $\chi^2(15) = 104.36$, $p < 0.001$, indicated that correlations between the items were sufficiently large for PCA. An initial analysis was performed to obtain the eigenvalues for each component in the data. Two components had eigenvalues above Kaiser's criterion of 1 and in combination explained 76.765% of the variance, this is the number of components that were included in the further analysis. The model attribute subscales of the survey demonstrate a good level of reliability and consistency between attributes of the constructs, Cronbach's $\alpha = 0.766$.

⁶There was a significant relationship between model-specific trustworthiness and all other attributes at the 0.01 level (two-tailed). The strongest correlations are with modeler-specific trustworthiness ($r = 0.63$) and business case criticality ($r = 0.64$). Modeler-specific trustworthiness was further significantly correlated with human safety criticality, $r = 0.40$, and organizational culture, $r = 0.34$ (both at the 0.05 level (two-tailed)); business case criticality was also significantly correlated with modeler-specific trustworthiness, $r = 0.56$ (p (two-tailed) < 0.01). $N = 37$.

⁷There were significant relationships between all attributes related to criticality. Namely, human safety and business case, $r = 0.60$, human safety and regulation, $r = 0.51$, as well as a business case and regulation, $r = 0.47$ (all p (two-tailed) < 0.01). Organizational culture is significantly correlated with modeler-specific trustworthiness ($r = 0.34$) and human safety criticality ($r = 0.40$), both at the 0.05 level (two-tailed). Strong correlations of organizational culture exist with business case criticality ($r = 0.55$) and regulation criticality ($r = 0.57$), both at the 0.01 level (two-tailed). The only non-significant

Analysis revealed that two components are derived from the initial PCA,⁸ which is less than the three constructs previously included in the framework. The first component is influenced by all stakeholder attributes with generally high loadings. The second component is positively influenced by regulation criticality and organizational culture, but negatively influenced by modeler-specific trustworthiness. The first component has an eigenvalue that is 3.5 times higher than the eigenvalue of the second component, indicating a much stronger relevance of aspects related to trustworthiness. In summary, the included stakeholder attributes that were part of the PCA conducted all have a high impact on stakeholder confidence.

The EFA performed included varimax (orthogonal) and oblique rotation (direct oblimin). The results of the PCA and EFA of modeler-related attributes are summarized in [Figure A6](#).

The first component includes modeler- and model-specific trustworthiness and business case criticality according to the results. Modeler-specific trustworthiness is a crucial aspect of building trust in M&S results. If M&S results are critical to the business case, then there is a higher expectation that M&S are reliable and accurate. A suitable description of this component as a characteristic of a stakeholder could be a trusting nature. The second component derived incorporates human safety criticality, regulation criticality and organizational culture. The component could be described as risk awareness, which includes external factors that influence M&S results and highlights that a collaborative and responsible work environment with the highest standards for conduct and behavior is important to ensure reliable M&S-based decisions.

The complete version of the updated framework with tested and extended confidence predictors and indicators (previously represented by constructs and attributes) based on the statistical analysis is displayed in [Figure A9](#). Questions for measurement and aspects related to the attributes are part of the visualization as well.

6. Results

6.1. Core insights of the survey on the use of M&S related to the outcome measures

One of the first investigations with respect to the outcome measures is to identify relationships between the confidence predictors derived from the framework constructs (through PCA and EFA) and the outcome measures. The derived components show strong and significant relationships with the measured overall

correlation among stakeholder attributes is between modeler-specific trustworthiness and regulation criticality ($r = 0.26$). $N = 37$.

⁸A PCA was conducted on the six items (stakeholder attributes). The KMO measure verified the sampling adequacy for the analysis, $KMO = 0.811$ ('great' according to Field 2009), and all KMO values for individual items were > 0.767 , which is well above the acceptable limit of 0.5 (Field 2009). Bartlett's test of sphericity $\chi^2(15) = 64.88$, $p < 0.001$, indicated that the correlations between items were sufficiently large for PCA. An initial analysis was performed to obtain eigenvalues for each component of the data. Two components had eigenvalues above Jolliffe's criterion of 0.7 and in combination explained 71.53% of the variance; this is the number of components that were retained in the further analysis. The model attribute subscales of the survey demonstrate a good level of reliability and consistency between the attributes of the constructs, Cronbach's $\alpha = 0.819$.

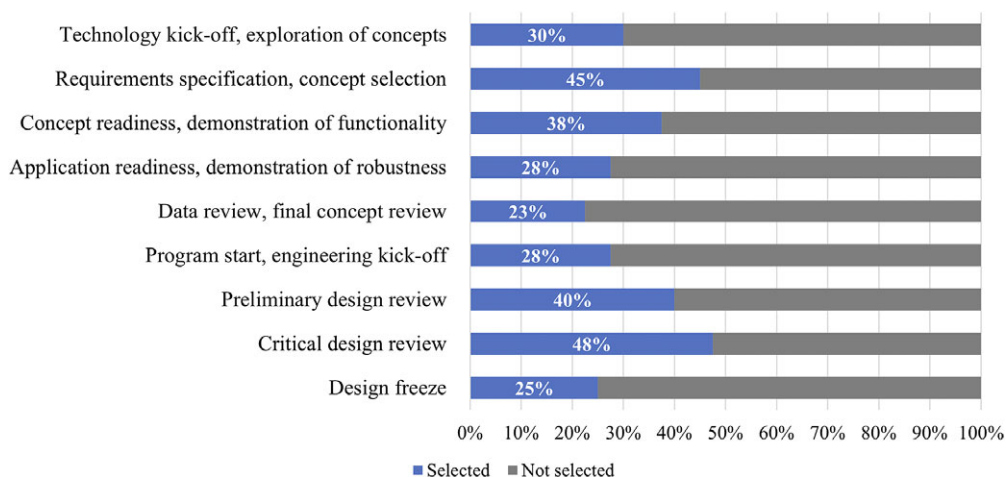


Figure 5. Milestones at which model/simulation results were primarily used in decision-making by the survey participants ($N = 40$). Multiple responses were possible.

confidence in M&S, and confidence has a strong correlation with reliance on M&S for decisions.⁹

Stakeholder-related confidence predictors are influenced by many model- and modeler-related constructs. Especially the predictor’s capability and accessibility show strong relationships with them. The measured overall confidence is mainly related to validity, integrity, trusting nature and risk awareness. It is also directly related to the reliance on M&S for the decision. The correlation matrix of the predictors and outcome measures is shown in Figure A7.

The visualization of responses for milestones around which M&S results were primarily used for decision-making is displayed in Figure 5. During the analysis, a binary variable of the milestones was created, which separates and clusters milestones around requirements specification, concept selection (response options 1–5: early milestones) and critical design review (response options 6–9: late milestones). The included industries are automotive and aerospace, while a differentiation is made between modelers, engineers and executives. Statistical analysis revealed that there are no significant differences between which industries¹⁰ or positions¹¹ prioritize specific milestones.

Binary variables for the outcome measures of reliance on M&S decision-making and overall confidence in M&S results were created (e.g., no confidence, fully confident). The previously introduced milestone clusters were also used.

⁹There were significant relationships between capability and risk awareness, which also had the highest correlation among all predictors, $r = 0.60$, p (two-tailed) < 0.01 . Trustworthiness shows a significant correlation at the 0.01 level with history, $r = 0.56$. The outcome measures decision-making reliance on M&S and overall confidence also show a significant relationship, $r = 0.52$, p (two-tailed) < 0.05 . The highest number of correlations at the 0.05 and 0.01 levels are associated with both stakeholder-related predictors and measured overall confidence. The lowest and smallest correlations are associated with the model- and modeler-related predictors history, reliability and competence. $N = 38$.

¹⁰There was no significant association between milestone clusters and different industries (automotive, aerospace) $\chi^2(2) = 3.455$, $p = 0.178$. $N = 33$.

¹¹There was no significant association between milestone clusters and different roles (modelers, engineers and executives) $\chi^2(2) = 1.024$, $p = 0.599$. $N = 40$.

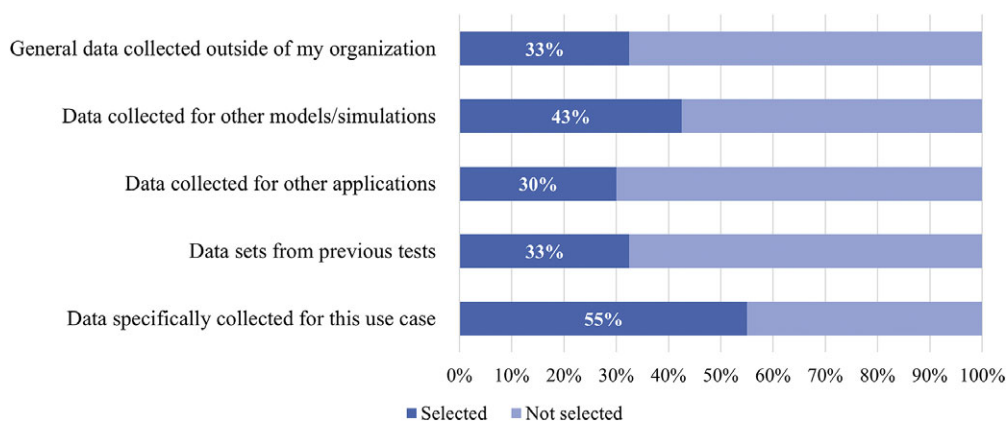


Figure 6. Nature of the input data for the model/simulation used by the survey participants ($N = 40$). Multiple responses were possible.

These binary variables were used for crosstabulations and chi-square tests. There were no significant differences between the milestones and the reliance on M&S for decision-making,¹² as well as the confidence¹³ in the results. Decisions, where the results of using M&S are relevant, are slightly more often used by the respondents in early milestones related to requirements specification and concept selection. There was no specific milestone that was more prominent than others for the stakeholders which indicates that they use M&S for decision-making throughout the life cycle.

ANOVAs regarding the reliance on M&S for decision-making suggest that there are significant differences between years of experience with a specific M&S type and the use of it for making critical decisions.¹⁴ The reliance on M&S is the lowest for individuals with less than 2 years of experience and the highest for those with 13 years or more.

The nature of the input data (see Figure 6) was investigated to determine whether data specifically collected for a use case increases confidence and reliance on M&S for decision-making. Data specifically collected for a use case were most often selected by respondents. A binary variable was created to separate the responses in use case-specific selected data and data collected elsewhere. The results reveal that using case-specific data increases the reliance on M&S for

¹²There was no significant association between the milestone cluster and the reliance on M&S for decision-making $\chi^2(1) = 0.009, p = 0.923$. On average, participants utilized M&S for decisions at late milestones ($M = 3.54, SE = 0.215$) than if they did at early milestones ($M = 3.48, SE = 0.176$). This difference was not significant $t(34) = 0.230, p = 0.635, N = 36$.

¹³There was no significant association between milestone clusters and confidence that stakeholders have in the results $\chi^2(2) = 0.376, p = 0.540$. On average, participants expressed greater confidence if they use M&S at late milestones ($M = 4.23, SE = 0.257$) than if they used it at early milestones ($M = 4.17, SE = 0.120$). This difference was not significant $t(34) = 3.241, p = 0.081, N = 36$.

¹⁴There was a significant effect of years of experience with the M&S type on the reliance on M&S for decisions, $F(3, 30) = 3.031, p < 0.05, \omega^2 = 0.152, N = 34$. There was no significant effect of the use instances on the reliance on M&S for decisions, $F(2, 29) = 0.432, p = 0.653, \omega^2 = -0.037, N = 32$. There was no significant effect of positions on the reliance on M&S for decisions, $F(4, 31) = 0.375, p = 0.825, \omega^2 = -0.075, N = 36$. There was no significant effect of industries on the reliance on M&S for decisions, $F(2, 26) = 0.078, p = 0.926, \omega^2 = -0.068, N = 29$.

making critical decisions significantly,¹⁵ which might be caused by the credibility of the data. The confidence in the results is also slightly higher.¹⁶

Analysis indicates that if data are not specifically collected for a use case, the number of uncertainty sources may increase.¹⁷ This can be caused by not knowing the conditions and parameters under which the data from other sources were collected. Uncertainties caused by input data were most often selected and impacted the decision-making relevance and overall confidence in M&S results. No significant differences were determined between uncertainties caused by input data and outcome measures. The negative relationship between the number of uncertainties and the overall confidence could indicate that the confidence declines if the number of uncertainty sources increases.¹⁸

Further analysis reveals differences between epistemic and aleatory uncertainties on outcome measures. Respondents had slightly, but not significantly higher levels of overall confidence,¹⁹ and the reliance of M&S for decisions was higher²⁰ if the uncertainties were based on stochastic effects and variability (aleatory uncertainties) instead of a lack of knowledge and model bias (epistemic uncertainties) or other sources.

The results indicate that the development method of M&S has a strong and significant impact on the outcome measures of reliance on M&S for decision-making and measured overall confidence. The highest values were received if people worked in cross-functional teams, especially if members of other organizations (e.g., partners) were involved.²¹

Confidence-inspiring activities (see Figure 7) were investigated to identify significant relationships between measures and to determine which are most relevant to different stakeholders. According to the percentages in each rank, trust in the modeler, credibility of results and their explainability are in the top three

¹⁵On average, data specifically collected for a use case had a higher relevance and therefore reliance on M&S for decision-making ($M = 3.53$, $SE = 0.140$) than data collected elsewhere ($M = 3.47$, $SE = 0.244$). This difference was significant $t(34) = 5.661$, $p < 0.05$.

¹⁶On average, data specifically collected for a use case resulted in higher confidence in M&S results ($M = 4.26$, $SE = 0.129$) than data collected elsewhere ($M = 4.12$, $SE = 0.208$). This difference was not significant $t(34) = 1.122$, $p = 0.297$. $N = 36$.

¹⁷On average, data collected somewhere else increased the number of uncertainties in M&S ($M = 2.62$, $SE = 0.474$) than data specifically collected for a use case ($M = 2.33$, $SE = 0.211$). This difference was not significant $t(32) = 1.978$, $p = 0.169$. $N = 34$.

¹⁸There are no significant correlations between the four most frequently selected uncertainties (noise factors, epistemic uncertainty, aleatory uncertainty, uncertainties caused by input data) and overall confidence. There was a significant negative relationship between the selected number of uncertainties and overall confidence in M&S results at the 0.05 level (two-tailed), $r = -0.45$. $N = 36$.

¹⁹On average, focusing on aleatory uncertainties increased overall confidence in M&S ($M = 4.33$, $SE = 0.333$) than epistemic uncertainties ($M = 4.20$, $SE = 0.223$). This difference was not significant $t(34) = 2.850$, $p = 0.101$. $N = 36$.

²⁰On average, focusing on the aleatory uncertainties increased the reliance on M&S for decision-making ($M = 4.00$, $SE = 0.236$) than the epistemic uncertainties ($M = 3.53$, $SE = 0.215$). This difference was not significant $t(34) = 1.821$, $p = 0.186$. $N = 36$.

²¹There was a significant effect of the development method on the relevance for using M&S for critical decisions, $F(5, 30) = 4.088$, $p < 0.01$, $\omega^2 = 0.300$, $N = 36$. There was a significant effect of the development method on the overall confidence measured, $F(5, 30) = 2.546$, $p < 0.05$, $\omega^2 = 0.177$, $N = 36$. There were no significant associations between the development method and positions ($\chi^2(10) = 10.606$, $p = 0.389$, $N = 36$) as well as industries ($\chi^2(10) = 9.262$, $p = 0.507$, $N = 29$).

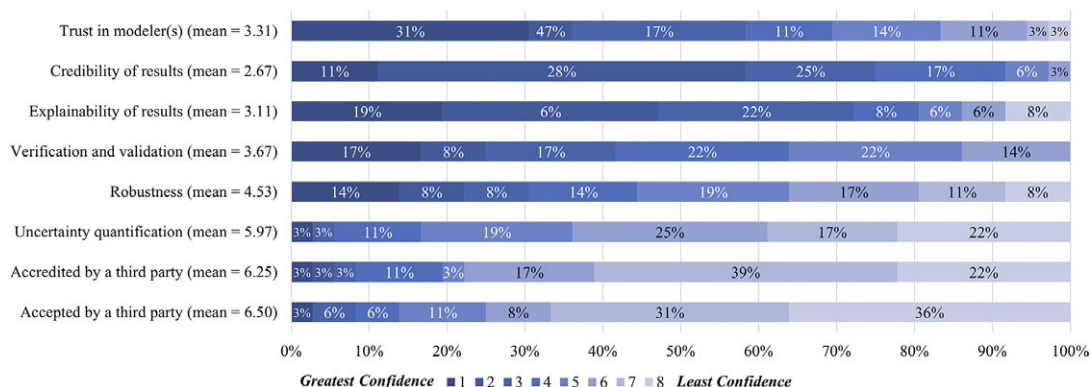


Figure 7. Responses ranked by the degree to which they inspired confidence in the model/simulation results for the participants ($N = 36$) (1 – inspired the greatest confidence; 8 – inspired the least confidence).

priorities with V&V, robustness and UQ afterward. If means are considered, credibility and explainability of results rank above trust in the modeler.

Trust in the modeler and the credibility of results are among the most important aspects of confidence inspiration. This might be caused by the close connection between the three sociotechnical concepts of confidence, credibility and trust. It should be highlighted that third-party acceptance and accreditation do not seem to be very relevant for inspiring confidence in M&S results. Multiple significant negative relationships occur between responses, especially between third-party accreditation/acceptance and VVUQ activities.

Further analysis between stakeholders revealed that the importance of trust is stakeholder-specific, and UQ might not be as important in practice as indicated by the literature. For modelers, V&V, as well as the credibility and explainability of results, are most important; trust in the modeler ranks last. For engineers, trust in the modeler, and the credibility of results are equally important. V&V, explainability and robustness follow. Executives almost had the same ranking as engineers. UQ is slightly more important for them. There are no differences in the ranking according to the respondents' confidence. It can be summarized that trust in the modeler and the credibility of results inspire the greatest confidence in M&S, especially for engineers and executives, and explainability and VVUQ are prioritized by modelers.

For confidence-inspiring VVUQ practices that contribute to confidence in results, the sensitivity analysis of the M&S results and the verification of the data were most often selected by the respondents. Solution verification was equally often selected as application assessment, while uncertainty visualization was slightly more often selected than validation of computational model and validation metrics. Code verification and uncertainty characterization were selected more often than reporting procedures of VVUQ results, which were selected least often. The results indicate that there are differences in the importance of VVUQ practices among stakeholders. Modelers care more about code verification, while engineers prioritize solution verification and data verification. Executives focus on the application assessment, which includes aspects such as the limits of scope. The sensitivity analysis of M&S results is ranked among the first three practices for each stakeholder.

6.2. Application model to complement the framework

Based on the literature findings, expert interview insights, and survey results, the application readiness of M&S to build confidence may be determined through an application model of the framework. This guideline is structured in the form of a confidence assessment and decision support. The approach is of qualitative and quantitative nature as it provides guiding principles, measures and assessment methods to evaluate decisions that include the reliance on M&S results. This application model is an initial suggestion based on the insights of the survey. To date, it is of a purely theoretical nature and has not been tested in a practical environment.

The confidence assessment is intended to guide confidence measures and confidence-inspiring activities. Presenting confidence predictors and indicators based on the updated framework constructs is a fundamental part of the application model. Decision support summarizes the refined and validated relationships between confidence predictors and their association with the outcome measures of decision-making. Based on the explained predictors and confidence scores, it also includes an example approach with interpretative weightings and achieved confidence levels to illustrate quantifying overall confidence in M&S.

6.2.1. Confidence assessment

The first section of the application model, the confidence assessment, is based on the survey analysis results of the measured framework constructs and attributes (see Section 5.3) combined with outcome measures and differentiated by stakeholders (see Section 6.1). The included confidence measures present findings related to the decision-making outcomes, such as the reliance on M&S. They also describe strong relationships and introduce the idea of a calculated confidence score. Confidence-inspiring activities highlight the importance of developing M&S in cross-functional teams, present actions that resulted in high confidence levels, and characterize the VVUQ practices most frequently used for stakeholder groups as displayed in Figure 8.

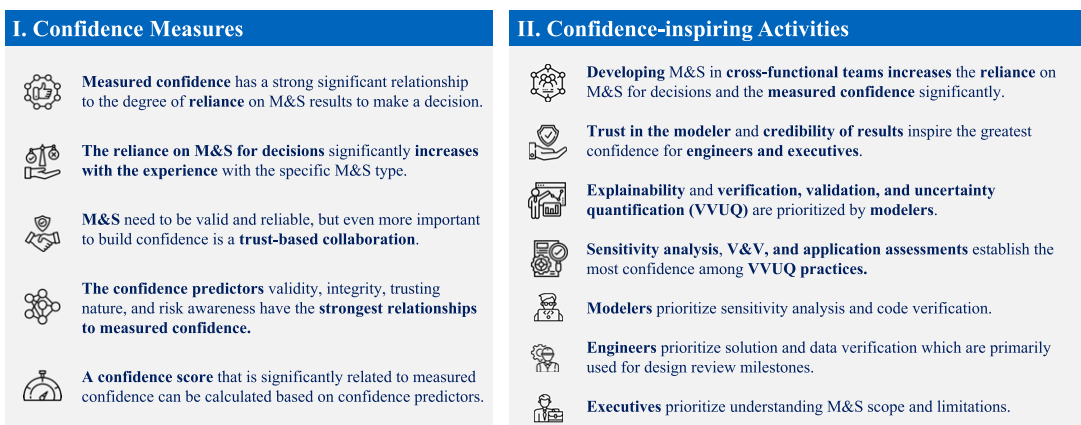
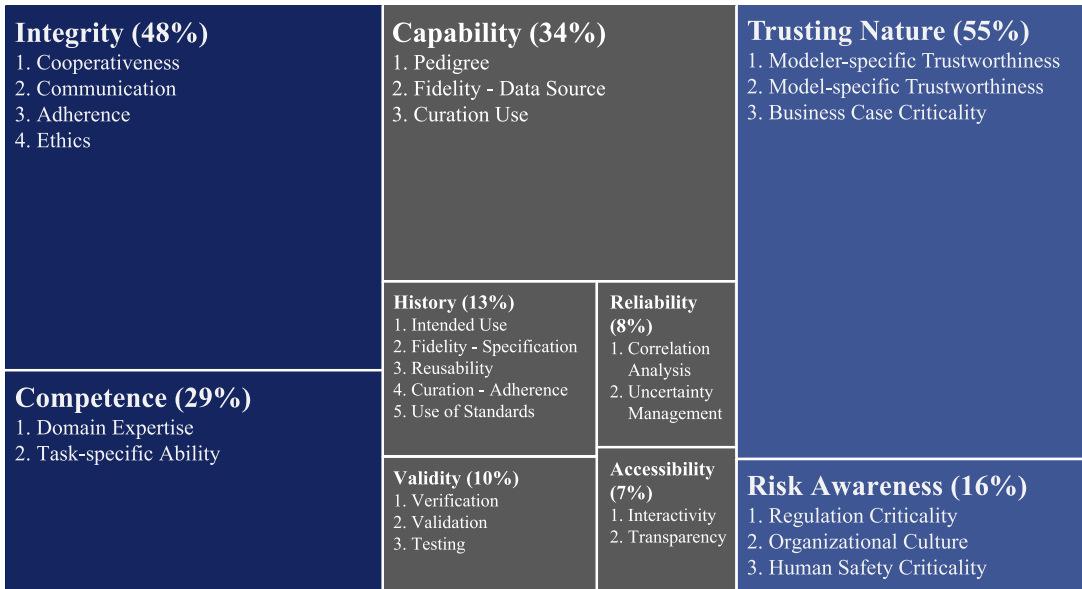


Figure 8. Confidence measures and confidence-inspiring activities. The included aspects are based on the outlined survey analysis results and presented core insights.



■ Modeler-related ■ Model-related ■ Stakeholder-related

Figure 9. Derived confidence predictors and associated indicators. The size of the boxes, as well as the percentages behind the predictors, represent the eigenvalues in % of variance derived from the principal component analysis and indicate the importance within the specific section. The three columns are separate from each other and do not necessarily add up to the same fraction based on differential analysis.

The confidence measures are ordered according to their perceived importance for decision-makers. Confidence in M&S results strongly impacts the degree of reliance on them to make a decision. The reliance thereby also depends on the experience with the specific M&S types. Although valid and reliable results are crucial, trust-based collaboration between involved stakeholders is even more relevant to establishing confidence. Regarding confidence-inspiring activities, trust in the modeler and the credibility of results inspired the most confidence for engineers and executives, while modelers prioritize the explainability of results and VVUQ. Among the VVUQ practices, the sensitivity analysis, V&V, and application assessments established the most confidence. Modelers preferably use sensitivity analysis and code verification, engineers prioritize solution and data verification and executives focus on understanding the scope and limitations of M&S through application assessments.

The confidence predictors and indicators that represent the constructs and attributes of the updated framework are included in the second part of the confidence assessment. For each confidence predictor, a set of indicators that should be assessed to develop model-, modeler- and stakeholder-related confidence is outlined (see Figure 9).

Within model-related confidence predictors, capability represents the largest variance, followed by history, validity, reliability and accessibility. As outlined in the survey analysis, there are strong significant relationships between the indicators related to the predictor capability. Moderate significant relationships between included indicators occur for the confidence predictor validity. Validity also has a strong significant relationship with measured confidence in M&S. Among the

predictors related to modelers, integrity possesses a moderately significant relationship to measured confidence. Strong significant relationships occur between the included indicators of cooperativeness, communication, adherence and ethics. The confidence predictor competence, in comparison, contains a significant relationship between its indicators, domain expertise and task-specific ability. Both stakeholder-related confidence predictors, trusting nature and risk awareness, have a moderately significant relationship to measured confidence. There are strong relationships between the indicators associated with trusting nature and moderate relationships between the indicators of risk awareness (see Appendix).

Decision-makers benefit from the results of the confidence assessment section through an overview of confidence measures, confidence-inspiring activities and confidence predictors. Especially the relation between confidence in M&S results and reliance on them for making decisions highlights the importance of establishing confidence. The stakeholder-specific insights of general and VVUQ-related activities clarify prioritized activities. The main part of the confidence assessment presents predictors and indicators that need to be addressed by decision-makers to build and establish confidence.

6.2.2. Decision support

The first part of the decision support visualizes the validated relationships between confidence predictors (see Figure 10) and presents further analysis results. In addition to refining the framework, the survey results are used to test and update the initial hypothesized relationships between confidence predictors. The results of EFA and correlation studies (see Figure A7) are valuable to validate the relationships. Confirmatory regression analysis was performed to validate the relationships between constructs. Especially between stakeholder-related confidence predictors and the measured overall confidence in M&S (see Figure A8), significant relationships at the 0.05 level (two-tailed) exist (e.g., between modeler- and model-specific trustworthiness and confidence). There are highly significant relationships at the 0.001 level (two-tailed) between business case criticality and confidence and between measured confidence in M&S and the reliance on M&S results for decision-making.

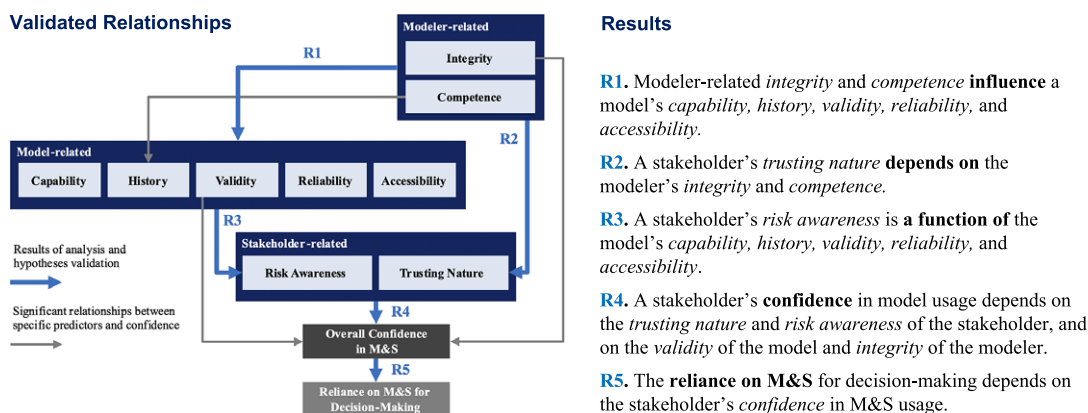


Figure 10. Relationships between confidence predictors and outcome measures. The statistical evidence of the relationships is described within this and previous sections and supported by analysis results in the Appendix.

The summarized findings reveal that mainly the modeler's integrity and competence influence a model's capability, history, validity, reliability and accessibility. The model constructs were extended, while the modeler constructs were reduced by benevolence which is now part of integrity. Competence has a strong relation to the history of a model and contributes to perceiving M&S as accurate and usable. Based on the results, the integrity of the modeler has a strong impact on the overall confidence in M&S.

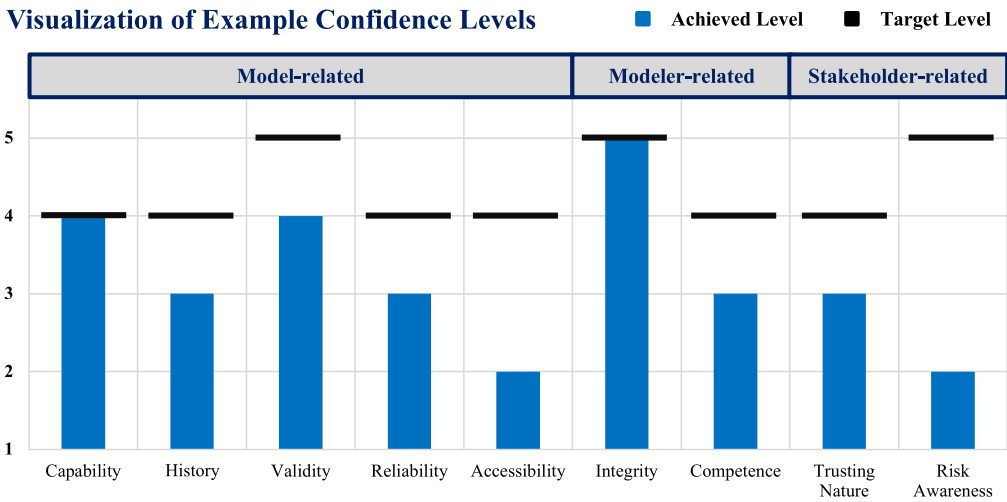
The trusting nature of a stakeholder that includes business case criticality, as well as modeler and model-specific trustworthiness, is strongly related to the integrity and competence of the modeler. It is also impacted by model constructs, especially the capability and accessibility. The risk awareness of the stakeholder is related to most of the model constructs, namely, the capability, validity, reliability and accessibility. The regression analysis indicates that the more trustworthy the modeler and model are perceived, the more likely it is that individuals will have the confidence to make decisions based on them. The confidence of a stakeholder in model usage or, as measured, the overall confidence in M&S, depends on the stakeholder characteristics of trusting nature and risk awareness, but also significantly relates to the validity of the model and integrity of the modeler.

Results of the analysis and validation of the hypotheses are highlighted and formulated to characterize dependencies between the confidence predictors. In summary, modeler-related integrity and competence influence the confidence predictors related to the model and the stakeholder. Modeler competence additionally has an impact on the M&S history. At the same time, integrity directly affects the overall confidence in M&S. Model-related predictors influence stakeholder-related risk awareness, while modeler-related predictors impact their trusting nature. The validity of M&S additionally impacts overall confidence. Stakeholder-related predictors ultimately affect overall confidence in M&S. Therefore, the reliance on M&S for decision-making depends on the overall confidence of the stakeholder in M&S (see [Figure 10](#)).

The second part of the decision support section presents a quantitative approach to determine confidence scores through interpretative weighting of predictors. The approach (see [Figure 11](#)) contains recommendations to quantify the predictors through confidence levels, a visualization of example confidence levels based on Blattnig *et al.* (2008), and formulas including example weighting factors to calculate the confidence scores.

The predictors can be evaluated with the achieved and target confidence levels, while achieved levels can be determined with survey questions to measure each predictor or the included indicators individually (e.g., with a five-point Likert scale). Another possibility is to utilize the NASA credibility assessment levels introduced by Babula *et al.* (2008). The target levels depend on the organization and are subjective. The levels achieved for each predictor are combined with weighting factors to calculate the confidence related to the model, the modeler and the stakeholder individually. In total, the weighting factors for each confidence type should be 1.

The next step combines the three confidence types to calculate the overall confidence in the M&S score. The previous statistical analysis revealed that a high calculated confidence score is a good indicator of actual confidence in M&S. To improve the score, it is recommended to focus on predictors that have high weighting factors. The weighting factors can be based on the strength of



Formulas for Confidence Scores with Achieved Levels and Weighting Factors

Model-related confidence = $a_1 * \text{Capability} + a_2 * \text{History} + a_3 * \text{Validity} + a_4 * \text{Reliability} + a_5 * \text{Accessibility}$ (1)

Modeler-related confidence = $b_1 * \text{Integrity} + b_2 * \text{Competence}$ (2)

Stakeholder-related confidence = $c_1 * \text{Risk Awareness} + c_2 * \text{Trusting Nature}$ (3)

Confidence in M&S = $a * \text{Model Confidence} + b * \text{Modeler Confidence} + c * \text{Stakeholder Confidence}$ (4)

Figure 11. Illustrative visualization of confidence levels and formulas to calculate confidence scores with interpretative predictor weightings.

relationships between measured predictors, outcomes and calculated scores. Organizations should aim to identify their own suitable weighting factors.

The decision support section of the application model enables decision-makers to gain insight into validated relationships between confidence predictors and outcome measures. The presented approach to calculating confidence scores quantifies the predictors and combines them with interpretative weightings to determine a comparable measure. The resulting confidence score can be used among involved stakeholders to indicate the actual confidence they can place in the M&S results to rely on them for decision-making.

Formats for communicating the overall approach can be knowledge transition sessions that aim to implement the framework and the application model by explaining the main ideas, content and approaches to enable potential users. Workshop-based implementations that require more time are also suitable communication formats. If a M&S project has already started and the framework wants to be used to assess the results, it is recommended to introduce the concepts at a cross-functional meeting.

7. Discussion

7.1. Interpretation of the results

The application model and the updated framework to assess confidence in M&S were designed and developed according to the objectives formulated by previous

research studies, and the improvement needs were identified based on an analysis of an existing framework. Many of the elements of the model confidence framework presented here are similar to elements found in existing M&S practice guides (see, e.g., NASA 2019), and indeed, many of those publications were sources used in the development of the model confidence framework. The objective in developing the model confidence framework used in this study was not to reproduce a collection of M&S best practices but to identify which practices correlate with decision-maker confidence in M&S results. This includes not only the technical attributes of the model but also the social context in which it was created and the inclinations of the decision-makers. Standards such as the NASA Std. 7009a (NASA 2019) infer that adherence to M&S best practices will result in decision-maker confidence rather than attempting to demonstrate empirical correlation. Other treatments of decision-maker trust in models such as Hazelrigg (2023) are primarily theoretical and not specific to characteristics that can be documented and presented to a practicing engineering decision-maker. The sociotechnical system perspective presented here attempts to address the factors in a practical engineering decision-making context that predict confidence in M&S results.

The challenges and pain points of using M&S within the assessment of key stakeholders were addressed during the design of the application model. The intended applications were explored with the user stories, and the developed use case was used to improve the design of product features. The survey instrument to measure factors associated with confidence and to refine the solution was developed with the support of feedback sessions and pilot tested with the project sponsor to ensure applicability. Pilot testing confirmed and refined the suggested characteristics, response options and outcome measures. The formulation and testing of further use cases could obtain further feedback regarding the applicability of the application model and the framework.

Indicators of usability of the application model resulted from the assessment of existing approaches to evaluate M&S with the success criteria that identified specific improvement needs. An outcome of the assessments was that the application model could be a guideline or summary sheet with adaptive and interactive tools to improve the usability of previous approaches. Insights from expert interviews, such as the demand for improved documentation and traceability to improve the usability of M&S assessments, were used and incorporated into the design of the application model. Another aspect that indicates the usability of the developed solution is that the process of designing the application model and the comprehensibility of the improved framework was regularly reviewed with the project sponsor, especially at the beginning of the project. Testing and demonstrating the functionality of the application model with case studies could support improving the solution's usability.

The solution's usefulness is indicated by developing the application model concerning the desired benefits of supporting the creation of M&S, building confidence in their use and improving subsequent M&S-based decision-making. The previous challenge of a lack of a principled approach is addressed with the suggested structure of the application model divided into a confidence assessment and decision support section. The validated and visualized relationships between the confidence predictors that impact the overall confidence in M&S and the reliance on M&S further contribute to the usefulness of the improved framework. Another aspect that indicates the usefulness of the application model is its

structure, which allows iterations to refine the solution through acquisition of further empirical evidence (e.g., weighting factors for confidence scores). In general, the usefulness of the application model should be improved and evaluated by presenting the solution to key stakeholders and implementing feedback iteratively.

7.2. Limitations and constraints

The roles of confidence, credibility and trust were evaluated in an M&S context and the results were used to illustrate a structured guideline for evaluating M&S results with a focus on assessing confidence. Additional strengths are that the framework was improved based on empirical evidence with refined and validated relationships and that the application model introduced an approach to determine confidence qualitatively and quantitatively.

A primary concern about the results of the survey to measure factors associated with confidence that were presented here is that they are based on a relatively small convenience sample ($N = 40$). Other significant relationships could potentially be identified using a different sample. While the survey respondents all came from industrial sectors with high levels of M&S use, a larger, more systematic sampling of M&S practitioners and decision-makers is needed before these findings can be interpreted in any generalized way.

The potential impact of excluding the vulnerability and trust propensity of stakeholder-related confidence attributes due to the complexity in measuring them should be investigated. This directly relates to survey participant comments suggesting that the measured factors might be too model-centric and should include more factors associated with modelers and stakeholders.

The suggested application model is still only of a theoretical nature. It is based on the insights of the survey and previous work. However, it has never been applied in practice. Therefore, claims on the applicability and usefulness of the model cannot be made. Nevertheless, the application model is a possible starting point for future research and a potential inspiration for practitioners, in an area currently lacking this kind of systematic support.

7.3. Implications and relevance

7.3.1. Academia

Attributes and constructs of the framework were extended and operationalized to measure M&S characteristics, modeler expertise and stakeholder preferences, which were used to refine the framework and generate confidence predictors and indicators. For academics, validated relationships between confidence predictors and indicators are especially relevant, as they connect research on socio-technical aspects of evaluating M&S. The hypothesized relationships of the framework were tested and empirically validated with a survey to measure factors that contribute to confidence. Therefore, the results contribute to future research suggesting a multi-stakeholder approach to determine the confidence of a decision-maker in M&S with a framework by Chaudhari *et al.* (2022). The presented approaches to determine the application readiness of M&S are a further contribution to research in addition to the empirical data collected with the survey and subsequently analyzed.

7.3.2. Industry

The main contribution of this research is a designed and developed assessment in the form of an application model for decision-makers to evaluate confidence in M&S as a complement to the framework. For practitioners, the updated framework and application model as a guideline for M&S evaluation and documentation are the main contributions. The application model provides an approach for practically implementing the updated framework. Both combined represent an approach to determine confidence qualitatively and quantitatively to evaluate the application readiness of M&S. The application model can be used cross-departmentally as part of formal documentation for people using M&S. An important contribution for them are the confidence predictors with specific suggestions that should be emphasized when presenting M&S results to decision-makers. The developed survey instrument is another contribution, as it can be adapted and used by organizations to measure factors associated with confidence and decision-making outcomes within the specific M&S context.

8. Conclusion

The validation and refinement of the relationships between confidence predictors and the improvement of an existing framework to assess confidence in M&S with an emphasis on practical usability were the main objectives of this research. Without the awareness of how much confidence to place in M&S and their results, decision-makers may neglect to leverage their potential by not integrating them into operations or by making subsequent judgments based on questionable outcomes. To overcome these challenges, predictors and indicators that inspire greater levels of confidence in stakeholders and drive greater acceptance of M&S use were researched with the support of a survey and presented with the contribution of an application model of the framework.

The design and development of the application model were initiated by a literature review that contained fundamentals in the model-based engineering context and existing approaches to determining the application readiness of M&S with a focus on confidence, credibility and trust.

To design the application model, the existing framework was analyzed to identify improvement needs, focusing on its empirical validation and practical usability. Context and requirements were explored based on modelers, engineers and executives as stakeholders for which user stories were formulated that were connected to a use case for the framework and application model to improve the design of the product features through a guideline and summary sheet to establish confidence in M&S. The identified challenges and pain points of people who are skeptical about using M&S are addressed by summarizing the desired benefits of improving M&S comprehensively, accelerating M&S-based decision-making and building confidence in using M&S.

The model-related confidence predictors, capability, history, validity, reliability and accessibility confirm and extend the previous framework constructs. Modeler-related predictors are integrity and competence, which reduce the previous constructs by including benevolence. Trusting nature and risk awareness are stakeholder-related confidence predictors that directly influence the confidence in M&S. Integrity of the modeler and validity of the model also demonstrated strong significant relationships to the measured confidence. Based on the analysis,

the hypothesized relationships between the predictors were validated and refined. Greater confidence in M&S significantly increases the reliance on M&S for decision-making, which is also directly related to the years of experience with the M&S type used. Data specifically collected for a use case significantly improves the reliance on M&S results. Developing M&S in cross-functional teams significantly increases reliance on M&S and measured confidence in the results. The trust in the modeler and the credibility of the results inspired the greatest confidence among engineers and executives, while modelers predominantly focus on explainability and VVUQ. Among VVUQ practices, sensitivity analysis, data verification and application assessments establish the greatest confidence.

With the support of the findings, the framework was updated, and the results were used to realize the application model that includes a confidence assessment and decision support. The improved framework complemented by the application model – although not validated yet – aims to enable the widespread use of M&S and its results through an assessment of confidence for the involved stakeholders. Communicating the framework and the application model should empower the M&S stakeholders to use the solution. Therefore, knowledge transition sessions or workshop-based implementations should contain an introduction of the framework structure and use case, an overview of the application model with an explanation of the sections and specific objectives, and a summary of confidence predictors.

9. Outlook

Based on the findings of an initial evaluation and the feedback from the survey respondents, it is recommended to focus additional research on organizational and interpersonal trust factors, as the survey focused mainly on model-related attributes, as indicated by previous research. A refined focus on modeler and stakeholder-related attributes to extend confidence predictors could result in separate studies on these aspects to expand beyond the existing literature and create bridges between different domains (e.g., technology, management, psychology). For these studies, it could help to involve multiple companies from various industries to collect data from different perspectives.

Industrial case studies could be a starting point for continuing the practical implementation of the framework to assess confidence in M&S. Collecting more data with the designed and potentially shortened survey would be another recommendation for future research. Factors contributing to confidence and relationships between them could be further visualized with structural equation modeling, allowing an improved understanding of the strength of relationships between indicators, predictors and outcomes.

For future research, it would also be important to analyze how decision-makers evaluate M&S that they do not directly use operationally but on whose results they depend because of the nature of business processes. Finally, it is recommended to focus on appropriately representing confidence predictors as an objective assessment and reliable source for decision-makers. Therefore, it is necessary to investigate what information they want to see and what stakeholders are looking for when determining the application readiness of M&S. Conducting knowledge transition sessions or workshops with key decision-makers segmented by use cases

could support defining the required information for different situations and stakeholders.

References

- Babula, M., Bertch, W. J., Green, L. L., Hale, J. P., Mosier, G. E., Steele, M. J. & Woods, J.** 2008 Nasa standard for models and simulations: Credibility assessment scale. 47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition. <https://doi.org/10.2514/6.2009-1011>.
- Balci, O.** 2010 “Golden Rules of Verification, Validation, Testing, and Certification of Modeling and Simulation Applications.”
- Balci, O.** 2012 A life cycle for modeling and simulation. *Simulation* **88**, 870–883.
- Blattnig, S., Green, L., Luckring, J., Morrison, J., Tripathi, R. & Zang, T.** 2008 Towards a credibility assessment of models and simulations. 49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials, Schaumburg, IL. <https://doi.org/10.2514/6.2008-2156>.
- Blomqvist, K.** 1997 The many faces of trust. *Scandinavian Journal of Management* **13**(3), 271–286. [https://doi.org/10.1016/S0956-5221\(97\)84644-1](https://doi.org/10.1016/S0956-5221(97)84644-1).
- Box, G.** 1979 *Robustness in the Strategy of Scientific Model Building*. Elsevier.
- Brade, D.** 2003 A Generalized Process for the Verification and Validation of Models and Simulation Results.
- Chaudhari, A. M.** 2022 Common themes in exploratory interviews (conducted by the research team of a co-author). Unpublished Presentation. MIT, Cambridge.
- Chaudhari, A. M., Rebentisch, E. & Rhodes, D. H.** 2022 Confidence in models and simulations: A multi-stakeholder analysis. In *29th ISTE International Conference on Transdisciplinary Engineering, TE 2022*. <https://doi.org/10.3233/ATDE220636>.
- Dumitrescu, R., Albers, A., Riedel, O. & Stark, R.,** Eds. 2021 *Engineering in Germany - The status quo in business and science, a contribution to Advanced Systems Engineering*. Fraunhofer IEM, Paderborn.
- Dunke, F. & Nickel, S.** 2021 Simulation-based multi-criteria decision making: An interactive method with a case study on infectious disease epidemics. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-021-04321-8>.
- Field, A.** 2009 *Discovering Statistics Using SPSS*. SAGE Publications Ltd.
- Gass, S. I.** 1993 Model accreditation: A rationale and process for determining a numerical rating. *European Journal of Operational Research*. Elsevier **66**(2), 250–258.
- Gass, S. I. & Joel, L. S.** 1981 Concepts of model confidence. *Computers & Operations Research* **8**, 341–346.
- German, E. & Rhodes, D. H.** 2017 *Model-Centric Decision-Making: Exploring Decision-Maker Trust and Perception of Models*. Springer International Publishing.
- Harper, A., Mustafee, N. & Yearworth, M.** 2021 Facets of trust in simulation studies. *European Journal of Operational Research* **289**, 197–213.
- Hazelrigg, G. A.** 2023 Model validation based on value-of-information theory. In Vol 6: *35th International Conference on Design Theory and Methodology (DTM)*. American Society of Mechanical Engineers.
- Isaksson, O. & Eckert, C.** 2020 “Product development 2040.” 1–56.
- Krosnick, J. A., Narayan, S. & Smith, W. R.** 1996 Satisficing in surveys: Initial evidence. *New Directions for Evaluation* **1996**, 29–44.
- Lee, J. D. & See, K. A.** 2004 Trust in automation: Designing for appropriate reliance. *Human Factors* **46**, 50–80.

- Luhmann, N.** 2000 *Familiarity, Confidence, Trust: Problems and Alternatives*. Department of Sociology, University of Oxford.
- Madni, A. M., Ghanem, R. G., Wheaton, M. J., Boehm, B. & Erwin, D.** 2018 *Model-Based Systems Engineering: Motivation, Current Status, and Needed Advances*. Springer.
- Maier, A. M., Kreimeyer, M., Hepperle, C., Eckert, C. M., Lindemann, U. & Clarkson, P. J.** 2008 *Exploration of Correlations between Factors Influencing Communication in Complex Product Development*, Vol. 16. Sage Publications, Inc.
- Marshall, G. C., Hale, J. P., Zimmerman, P., Kukkala, G., Kobryn, P., Puchek, B., Bisconti, M., Baldwin, C. & Mulpuri, M.** 2017 Digital model-based engineering: Expectations, prerequisites, and challenges of infusion. NASA/TM—2017–219633.
- Mayer, R. C., Davis, J. H. & Schoorman, F. D.** An Integrative Model of Organizational Trust. *The Academy of Management Review* 20(3), 709–34. JSTOR, <https://doi.org/10.2307/258792>.
- Mehta, U. B., Eklund, D. R., Romero, V. J., Pearce, J. A. & Keim, N. S.** 2016 Simulation Credibility: Advances in Verification, Validation, and Uncertainty Quantification. NASA/TP-2016-219422.
- Menold, N., Kaczmirek, L., Lenzner, T. & Neusar, A.** 2014 How do respondents attend to verbal labels in rating scales? *Field Methods* 26, 21–39.
- NASA** 2019 NASA handbook for models and simulations: An implementation guide for nasa-std-7009a. NASA technical handbook NASA-HDBK-7009.
- Oberkampff** 2021 Simulation-informed decision making.
- Oberkampff, W. L. & Roy, C. J.** 2010 *Verification and Validation in Scientific Computing*. Cambridge University Press.
- Olsen, M. & Raunak, M.** 2019 *Quantitative Measurements of Model Credibility*. Elsevier.
- Rechkemmer, A. & Yin, M.** 2022 When Confidence Meets Accuracy: Exploring the Effects of Multiple Performance Indicators on Trust in Machine Learning Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI'22)*. Association for Computing Machinery, New York, NY, USA. 1–14. <https://doi.org/10.1145/3491102.3501967>.
- Rhodes, D. H.** 2018 Using human-model interaction heuristics to enable model-centric enterprise transformation. 2018 Annual IEEE International Systems Conference (SysCon). IEEE.
- Rhodes, D. H.** 2022 *Investigating model credibility within a model curation context*. Recent Trends and Advances in Model Based Systems Engineering. Cham: Springer International Publishing, 67–77.
- Rhodes, D. H.** 2022 *Investigating Model Credibility within a Model Curation Context*. Springer International Publishing.
- Robinson, S. & Pidd, M.** 1998 Provider and customer expectations of successful simulation projects. *Journal of the Operational Research Society* 49, 200–209.
- Roy, C. J. & Oberkampff, W. L.** 2011 A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. *Computer Methods in Applied Mechanics and Engineering* 200, 2131–2144.
- Sargent, R. G.** (2015). An introductory tutorial on verification and validation of simulation models. In *2015 Winter Simulation Conference (WSC)* (pp. 1729–1740) IEEE.
- Schweigert-Recksiek, S.** 2021 Enhancing the Collaboration of Design and Simulation – Bridging Barriers in Technical Product Development. <https://doi.org/10.13140/RG.2.2.13110.27209>.
- Singh, Y. K.** 2006 *Fundamental of Research Methodology and Statistics*. New Age International.

- Sokolowski, J. A. & Banks, C. M.** 2010 *Modeling and Simulation Fundamentals: Theoretical Underpinnings and Practical Domains*. Wiley
- Steele, M. J.** 2008 Dimensions of credibility in models and simulations. In *Proceedings of the 2008 Summer Computer Simulation Conference (SCSC '08)*. Society for Modeling & Simulation International, Vista, CA, Article 57, 1–9.
- Thielsch, M. T., Meeßen, S. M. & Hertel, G.** 2018 Trust and distrust in information systems at the workplace. *PeerJ* 6, e5483.
- Trauer, J., Mutschler, M., Mörtl, M. & Zimmermann, M.** 2022a Challenges in Implementing Digital Twins – a Survey. *Proceedings of the ASME 2022 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. Volume 2: 42nd Computers and Information in Engineering Conference (CIE). St. Louis, Missouri, USA. August 14–17, 2022. V002T02A055. ASME. <https://doi.org/10.1115/DETC2022-88786>.
- Trauer, J., Schweigert-Recksiek, S., Schenk, T., Baudisch, T., Mörtl, M. & Zimmermann, M.** 2022b A digital twin trust framework for industrial application. *Proceedings of the Design Society* 2, 293–302.
- Trauer, J., Schweigert-Recksiek, S., Okamoto, L., Spreitzer, K., Mörtl, M. & Zimmermann, M.** 2020 Data-driven engineering definitions and insights from an industrial case study for a new approach in technical product development. In *Proceedings of the NordDesign 2020 Conference, NordDesign 2020*. <https://doi.org/10.35199/NORDDESIGN2020.46>.
- Vagias, W. M.** 2006. Likert-type scale response anchors. <https://media.clemson.edu/cbshs/prtm/research/resources-for-research-page-2/Vagias-Likert-Type-Scale-Response-Anchors.pdf>.
- VDI** 2018 Vdi-richtlinie: Simulation of systems in logistics, materials handling, and production - simulation and visualization (vdi 3633).
- VDI** 2021. Vdi-richtlinie: Modelling and simulation - building the model (vdi 4456).
- Vin, L. J. D.** 2015 Simulation, models, and results: Reflections on their nature and credibility. *Proceedings of FAIM2015*. 148–155.
- Wright, D. W., Richardson, R. A., Edeling, W., Lakhili, J., Sinclair, R. C., Jancauskas, V., Suleimenova, D., Bosak, B., Kulczewski, M., Piontek, T., Kopta, P., Chirca, I., Arabnejad, H., Luk, O. O., Hoenen, O., Weglarz, J., Crommelin, D., Groen, D. & Coveney, P. V.** 2020 Building confidence in simulation: Applications of easyvuuq. *Advanced Theory and Simulations* 3, 1900246.
- Yilmaz, L. & Liu, B.** 2020 Model credibility revisited: Concepts and considerations for appropriate trust. *Journal of Simulation* 16, 312–325.

A. Appendix

Pearson Correlations

	Intended Use	Fidelity – Degree	Fidelity – Representativeness	Verification	Pedigree	Validation – Degree	Validation – Correlation Analysis	Validation – Uncertainty Management	Interactivity	Transparency	Reusability	Use of Standards	Curation – Use	Curation – Adherence	Testing
Intended Use	1	.219	.566**	.577**	.045	.298	.221	.278	.187	.334*	.351*	.499**	.326*	.546**	.375*
Fidelity – Degree	.219	1	.169	.275	.727**	.341*	.296	.297	.224	.460**	-.097	.409**	.489**	.587**	.341*
Fidelity – Representativeness	.566**	.169	1	.516**	-.053	.286	.111	.094	-.093	.230	.415**	.337*	.356*	.326*	.140
Verification	.577**	.275	.516**	1	.201	.336*	.137	.055	.197	.223	.189	.411**	.266	.354*	.376*
Pedigree	.045	.727**	-.053	.201	1	.253	.194	.171	.076	.279	-.123	.432**	.364*	.400*	.233
Validation – Degree	.298	.341*	.286	.336*	.253	1	.439**	.253	.167	.353*	-.016	.348*	.158	.203	.380*
Validation – Correlation Analysis	.221	.296	.111	.137	.194	.439**	1	.503**	.256	.343*	.000	.287	.148	.354*	.298
Validation – Uncertainty Management	.278	.297	.094	.055	.171	.253	.503**	1	.018	.282	-.013	.252	.291	.195	.298
Interactivity	.187	.224	-.093	.197	.076	.167	.256	.018	1	.387*	-.034	.164	.002	.186	.326*
Transparency	.334*	.460**	.230	.223	.279	.353*	.343*	.282	.387*	1	.079	.386*	.465**	.564**	.276
Reusability	.351*	-.097	.415**	.189	-.123	-.016	.000	-.013	-.034	.079	1	.355*	.309	.394**	.057
Use of Standards	.499**	.409**	.337*	.411**	.432**	.348*	.287	.252	.164	.386*	.355*	1	.489**	.516**	.383*
Curation – Use	.326*	.489**	.356*	.266	.364*	.158	.148	.291	.002	.465**	.309	.489**	1	.635**	.107
Curation – Adherence	.546**	.587**	.326*	.354*	.400*	.203	.354*	.195	.186	.564**	.394*	.516**	.635**	1	.294
Testing	.375*	.341*	.140	.376*	.233	.380*	.298	.298	.326*	.276	.057	.383*	.107	.294	1

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Figure A1. Pearson correlations for model-related attributes (N = 40).

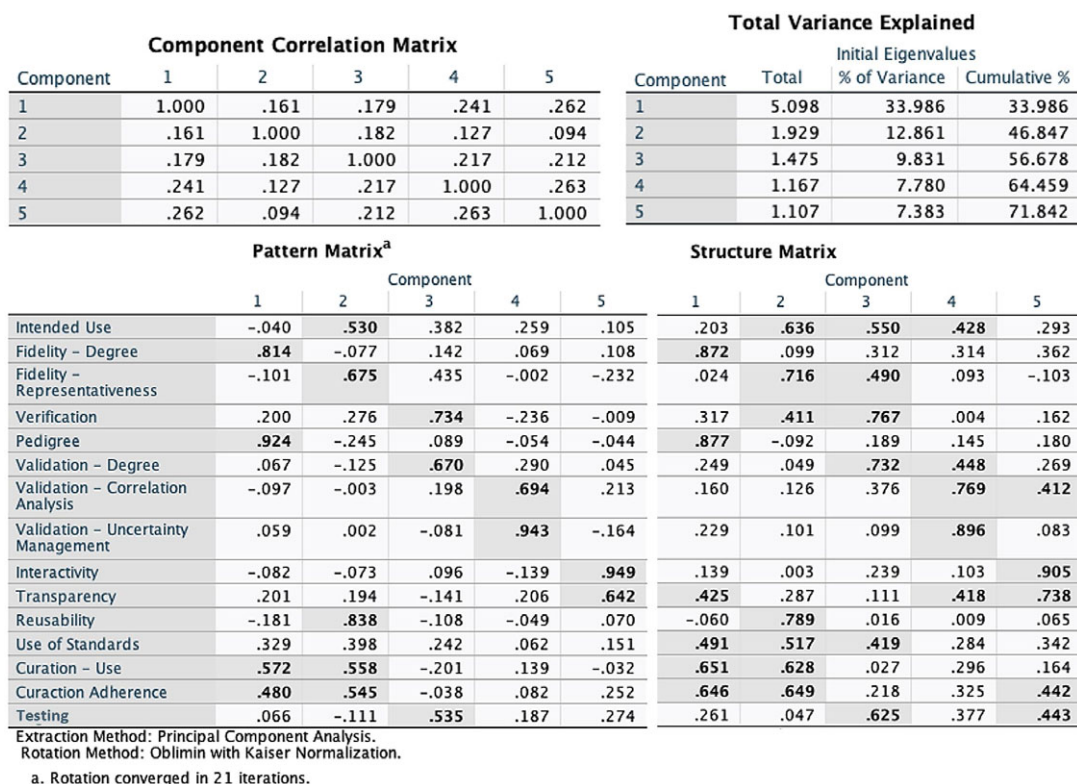


Figure A2. Summarized results of the PCA and EFA of model-related attributes (N = 40).

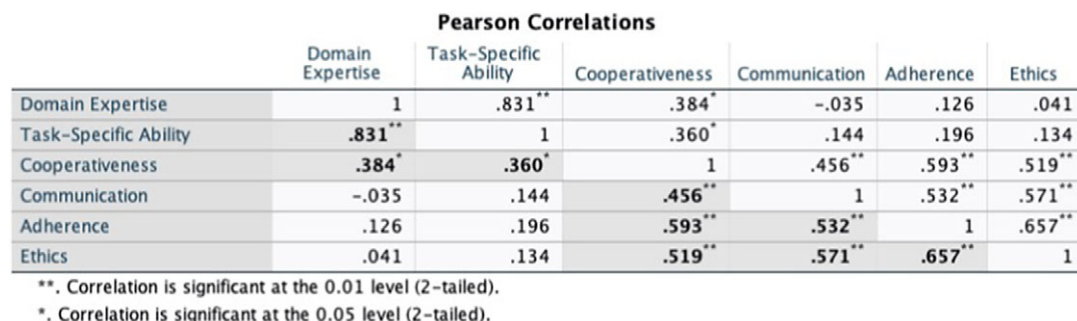


Figure A3. Pearson correlations for modeler-related attributes (N = 38).

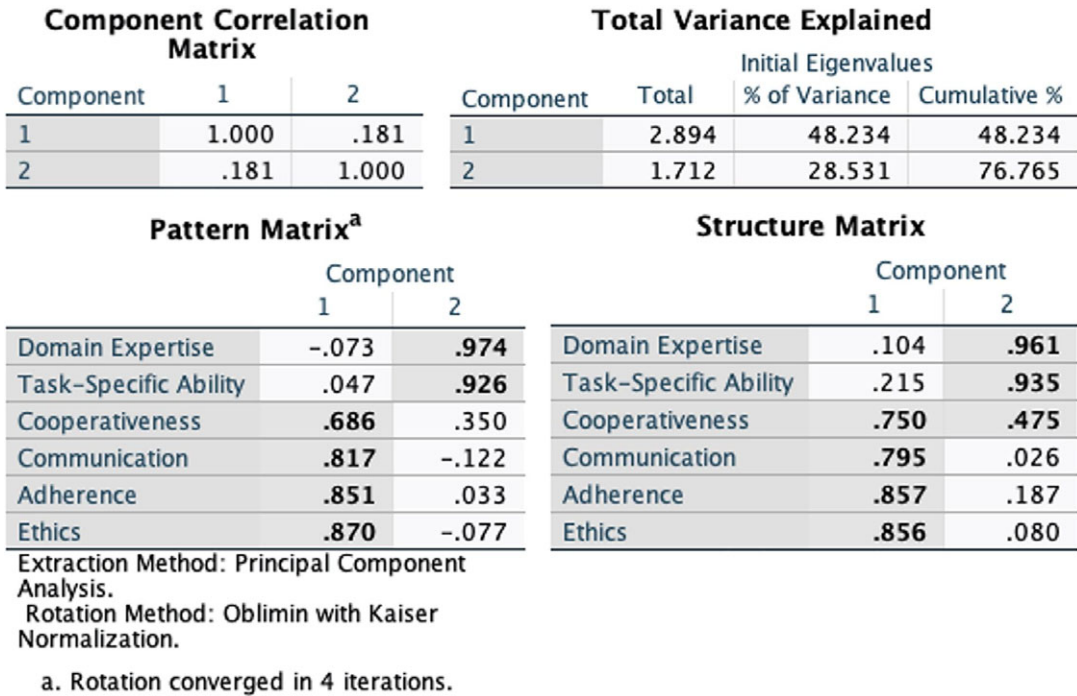


Figure A4. Summarized results of the PCA and EFA of modeler-related attributes (N = 38).

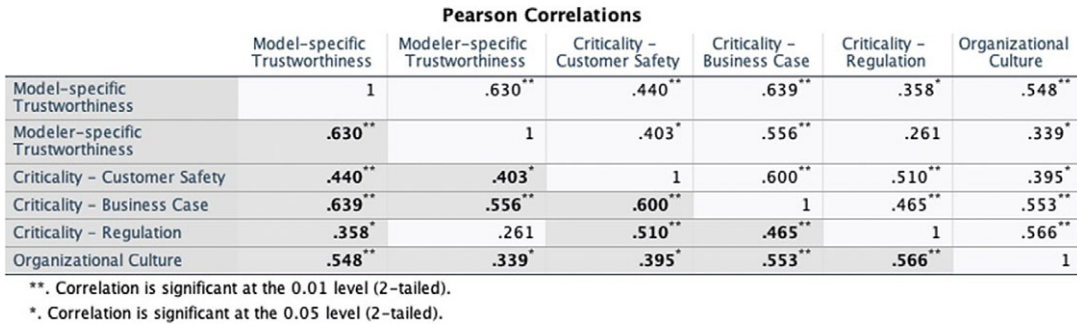


Figure A5. Pearson correlations for stakeholder-related attributes (N = 37).

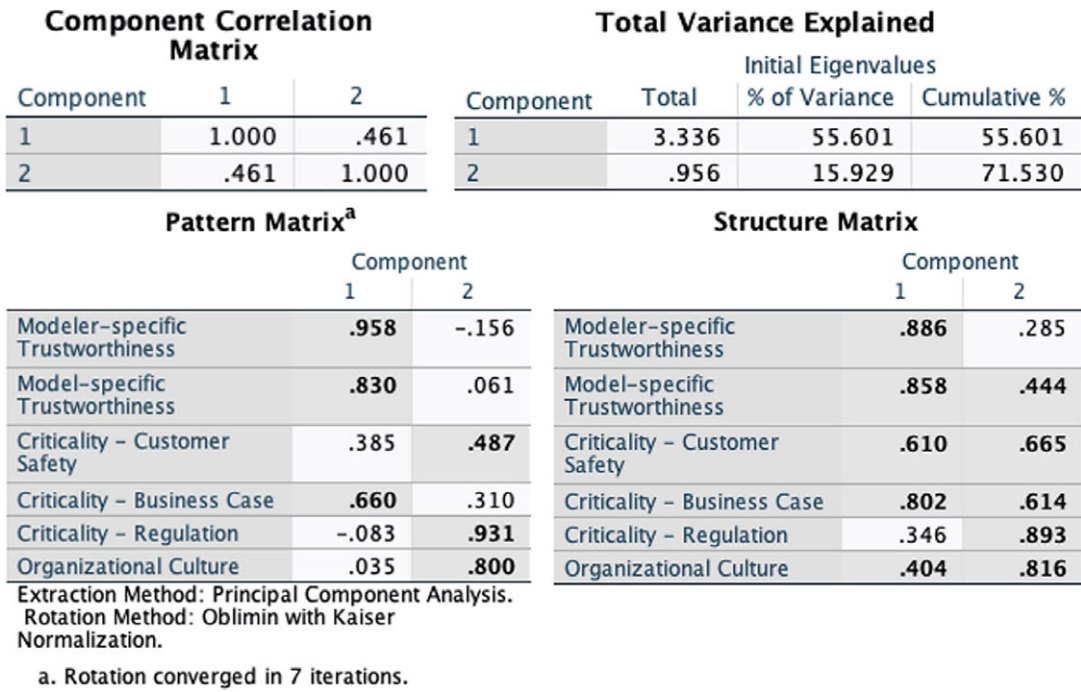


Figure A6. Summarized results of the PCA and EFA of stakeholder-related attributes (N = 37).

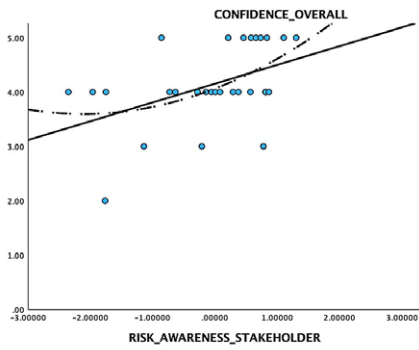
Pearson Correlations

	CAPABILITY_ MODEL	HISTORY_ MODEL	VALIDITY_ MODEL	RELIABILITY_ MODEL	ACCESSIBILITY_ MODEL	INTEGRITY_ MODELER	COMPETENCE_ MODELER	TRUSTING_NATURE_ STAKEHOLDER	RISK_AWARENESS_ STAKEHOLDER	DECISIONMAKING_ RELEVANCE	CONFIDENCE_ OVERALL
CAPABILITY_MODEL	1	.161	.179	.241	.262	.102	.184	.555**	.595**	.112	.268
HISTORY_MODEL	.161	1	.182	.127	.094	.052	.343*	.323	.134	.117	.106
VALIDITY_MODEL	.179	.182	1	.217	.212	.262	.305	.342	.543**	.047	.512**
RELIABILITY_MODEL	.241	.127	.217	1	.263	.003	.176	.170	.404*	-.111	-.079
ACCESSIBILITY_MODEL	.262	.094	.212	.263	1	.264	-.177	.399*	.560**	.061	.247
INTEGRITY_MODELER	.102	.052	.262	.003	.264	1	.181	.467**	.297	.147	.459**
COMPETENCE_MODELER	.184	.343*	.305	.176	-.177	.181	1	.315	-.058	.328	.258
TRUSTING_NATURE_STAKEHOLDER	.555**	.323	.342	.170	.399*	.467**	.315	1	.461**	.241	.471**
RISK_AWARENESS_STAKEHOLDER	.595**	.134	.543**	.404*	.560**	.297	-.058	.461**	1	.139	.457*
DECISIONMAKING_RELEVANCE	.112	.117	.047	-.111	.061	.147	.328	.241	.139	1	.521**
CONFIDENCE_OVERALL	.268	.106	.512**	-.079	.247	.459**	.258	.471**	.457*	.521**	1

** . Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

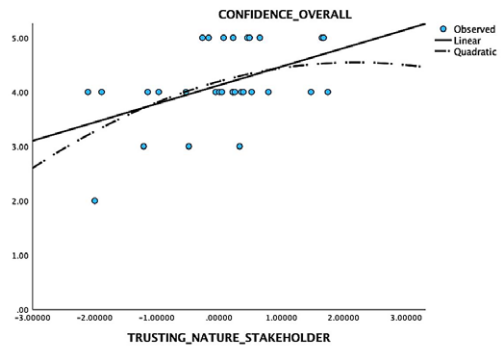
Figure A7. Correlation matrix of confidence predictors and outcome measures ($N = 38$).



Model Summary and Parameter Estimates
Dependent Variable: CONFIDENCE_OVERALL

Equation	Model Summary					Parameter Estimates		
	R Square	F	df1	df2	Sig.	Constant	b1	b2
Linear	.209	7.388	1	28	.011	4.153	.345	
Quadratic	.228	3.980	2	27	.031	4.063	.443	-.105

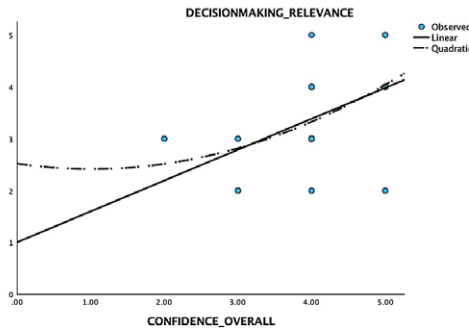
The independent variable is RISK_AWARENESS_STAKEHOLDER.



Model Summary and Parameter Estimates
Dependent Variable: CONFIDENCE_OVERALL

Equation	Model Summary					Parameter Estimates		
	R Square	F	df1	df2	Sig.	Constant	b1	b2
Linear	.222	7.980	1	28	.009	4.125	.341	
Quadratic	.239	4.237	2	27	.025	4.197	.316	-.072

The independent variable is TRUSTING_NATURE_STAKEHOLDER.



Model Summary and Parameter Estimates
Dependent Variable: DECISIONMAKING_RELEVANCE

Equation	Model Summary					Parameter Estimates		
	R Square	F	df1	df2	Sig.	Constant	b1	b2
Linear	.272	12.688	1	34	.001	1.003	.595	
Quadratic	.281	6.452	2	33	.004	2.521	-.208	.102

The independent variable is CONFIDENCE_OVERALL.

Figure A8. Regression analysis of the stakeholder-related confidence predictors and outcome measures.

Confidence Types	Construct (Predictor)	Attribute (Indicator)	Question for Measurement (Utilize five-point Likert scale with generic or with customized response options)	Related Aspects (mostly adapted from initial framework)
Model-related	Capability	Pedigree	<i>How confident are you that the data sources you used for the model/simulation sufficiently represent the full application space for which the model/simulation was intended?</i>	Input data quality and quantity; credibility of data source
		Fidelity – Data Source	<i>How confident are you that the data sources you used for the model/simulation include the fidelity required to accurately predict the system behavior?</i>	Fidelity of the data source, predict system behavior accurately
		Curation Use	<i>The curation of models/simulations you use for decision-making is well established in your organization.</i>	Lifecycle management and preservation
	History	Intended Use	<i>The original functionality and design of the model/simulation was consistent with the application for which it was used in this instance.</i>	Explicit requirements, expectations, assumptions, limitations; Consistency of runtime environment, experiments, and scenario
		Fidelity - Specification	<i>The specification of the model/simulation was consistent with the physical conditions of the application scenario.</i>	Representativeness between model specification and the real-world setting
		Reusability	<i>To what degree did this model/simulation reuse existing model and simulation elements?</i>	Availability of reusable solutions to address recurrent problems
		Curation - Adherence	<i>Individuals adhere to the curation process for models/simulations in your organization.</i>	Active enhancement to ensure value and future use
	Validity	Use of Standards	<i>To what degree did the process of building this specific model/simulation adhere to defined standards or specified guidelines?</i>	Internally developed guidelines, industry specific standards
		Verification	<i>The features and/or attributes implemented in the model/simulation were aligned with the model specifications, especially for its verification.</i>	Accuracy of model implementation and input data with respect to the conceptual specification
		Validation	<i>To what degree did you validate the model/simulation with data from testing of representative hardware and/or software in a representative environment?</i>	Level of expected uncertainty in model predictions
	Reliability	Testing	<i>To what degree did you assess the likelihood of potential edge cases (extreme operating parameters) produced by the model/simulation which might be caused by uncertainties?</i>	Test extreme operating parameters with representative scenarios
		Correlation Analysis	<i>The validation and especially the correlation analysis using empirical test data were included in the presentation of results of the model/simulation.</i>	Identify relationships between outcomes to explain results
	Accessibility	Uncertainty Management	<i>The model/simulation we used included a defined uncertainty measurement system to guide the development and to assess the robustness of the model/simulation results.</i>	Uncertainty characterization, quantification, propagation
		Interactivity	<i>What is your perception of how intuitively understandable the output of the model was for the key stakeholders who were not the creators of the model/simulation?</i>	Simulatability, textual explanations and graphic visualization
		Transparency	<i>Stakeholders had a clear understanding of the process, steps, and operations that the model/simulation uses to produce the output.</i>	Precise technical information including inputs, outputs, and internal code.
Modeler-related	Integrity	Cooperativeness	<i>Modelers were always available and willing to provide support to address problems with or modifications to the model/simulation.</i>	Behavior demonstrating individual support
		Communication	<i>Modelers openly shared information and knowledge with other teams/departments about the specific details of the model/simulation.</i>	Exchange of information and knowledge across boundaries
		Adherence	<i>Modelers demonstrated adherence to professional standards and willingness to reconcile differences.</i>	Personal commitment to a set of principles
		Ethics	<i>Modelers were fully transparent with the strengths and limitations of the model/simulation within their reports.</i>	Person's moral habits, ability to decide right or wrong, and free will
	Competence	Domain Expertise	<i>Please select the response that best represents the modeler's domain expertise in creating these types of models/simulations.</i>	Accumulated process knowledge from extended practice
		Task-specific Ability	<i>Please select the response that best represents the task-specific ability of the person who created the model/simulation.</i>	Influential skills, competencies, characteristics in context
Stakeholder-related	Trusting Nature	Modeler-specific Trustworthiness	<i>To what degree did you trust that the modelers produced a model/simulation with the quality that is needed to make a critical decision?</i>	Perceived reliability of modeler's actions
		Model-specific Trustworthiness	<i>To what degree did you trust that the model/simulation produced the quality of work that is needed to make a critical decision?</i>	Perceived reliability of model implementation, results and insights
		Business Case Criticality	<i>Describe your level of comfort using the model/simulation to make a key decision related to the business case for the program.</i>	Consequences for business case and the degree of decision influence
	Risk Awareness	Regulation Criticality	<i>Describe your level of comfort using the model/simulation to make a key decision related to regulatory authorities.</i>	Consequences to regulatory authorities and the degree of decision influence
		Organizational Culture	<i>People in your organization take responsibility to ensure that the models/simulations they produced are accurate and trustworthy.</i>	Responsibility for acting to the results, participation policies, implementation criteria
	Human Safety Criticality	<i>Describe your level of comfort using the results from the model/simulation to make a key decision related to human safety.</i>	Consequences to human safety and the degree of decision influence	

Figure A9. Updated framework to assess confidence in M&S with confidence types, predictors, indicators, questions for measurement and related aspects (partly based on Chaudhari et al. 2022).