



A categorisation of testing in product design and development

Downloaded from: <https://research.chalmers.se>, 2026-10-13 10:03 UTC

Citation for the original published paper (version of record):

Tahera, K., Panarotto, M., Eckert, C. et al (2026). A categorisation of testing in product design and development. DESIGN SCIENCE, 12. <http://dx.doi.org/10.1017/dsj.2026.10075>

N.B. When citing this work, cite the original published paper.

A categorisation of testing in product design and development

Khadija Tahera ¹, Massimo Panarotto ², Claudia Eckert ³ and Ola Isaksson ⁴

¹*People and Organisation, The Open University, UK*

²*Department of Mechanical Engineering, Politecnico di Milano, Milano, Italy*

³*Engineering & Innovation, The Open University, UK*

⁴*Department of Mechanical Engineering, Chalmers University of Technology, Gothenburg, Sweden*

Abstract

Testing is often narrowly viewed in the context of verification and validation in product design and development. However, testing also plays a critical role in product design, especially in the early stages, for exploration, experimentation and evaluation. While agile and set-based design emphasise iterative testing, these approaches do not fully address testing purposes across stages or consider diverse stakeholder needs. Empirical studies examining the multifaceted roles of testing throughout the design process are scarce in the academic literature. This paper fills this gap by presenting two detailed case studies on engine design and space propulsion systems, covering both incremental and radical product development in large-scale manufacturing settings. The findings show how testing purposes evolve with design progress and increasing specificity. Drawing on insights from the case studies and a review of the existing literature, the paper develops a systematic categorisation of the purposes of testing across design and development stages. This provides a foundation for a shared understanding among diverse stakeholders to support the effective use of testing in practice and offers consistent terminology to connect the academic literature on testing.

Keywords: Categorisation of testing, Product design and development, Purpose of testing, Testing, Verification and validation

Received 28 February 2025 UTC

Revised 27 July 2026 UTC

Accepted 28 July 2026 UTC

Corresponding author

Khadija Tahera

khadija.tahera@open.ac.uk

© The Author(s), 2026. 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. 12, e31

journals.cambridge.org/dsj

DOI: [10.1017/dsj.2026.10075](https://doi.org/10.1017/dsj.2026.10075)



1. Introduction

Testing is often viewed as an activity carried out in the later stages of product design and development (Wynn & Clarkson 2018) and narrowly focused on verification and validation (V&V) (Shabi, Reich & Diamant 2017). However, in practice, testing is closely intertwined with design throughout the entire product development process, from early idea generation to final validation before production (Cooper & Sommer 2016; Olechowski *et al.* 2020). Testing has received limited systematic attention in design research despite its critical role in the design process.

Design is a co-evolutionary process in which problem and solution spaces develop together through iterative exploration (Dorst & Cross 2001; Dorst 2019). Testing can serve as a mechanism that links and informs both spaces (Dorst 2019) by generating feedback (Thomke 2003), revealing mismatches (Erat & Kavadias 2008) and guiding the refinement of design towards a solution (Camburn *et al.* 2017), for example. Early-stage testing is crucial in industry for learning and design



evaluation when developing new technology products or exploring new markets, where understanding of the products and their behaviour is limited (Thomke, 2008). However, testing consumes significant time and resources, accounting for approximately 30–50% of product development costs in the automotive industry (Tahera *et al.* 2019), with unplanned late-stage testing consuming up to 75% of the budget in aerospace (Sudol & Mavris 2018). Virtual testing through computer-aided engineering (CAE), such as Computational Fluid Dynamics (CFD) and Finite Element Analysis (FEA), has become more prevalent to reduce the costs of physical prototyping and tests. However, virtual testing models rely on calibration and validation using physical test data, and this paradoxically might increase the demand for physical testing for additional purposes, such as model validation (Eckert & Isaksson 2026). As a result, managing testing activities throughout the product design and development, both virtual and physical, has become increasingly complex in engineering practice (Tahera *et al.* 2023).

Another key challenge in industry is that tests are often designed to meet the specific needs of a particular team (Zammit *et al.* 2018). Similar observations were made in the diesel engine case study (Tahera 2014). For example, design teams conducted early component tests by modifying existing engines to explore emission performance. A different team carried out testing to evaluate reliability and durability at later stages. However, the later team was not closely involved in the early testing, and the design teams were typically informed only of failures, rather than broader insights into product behaviour. Learning from these tests often remains siloed within those teams, and the opportunity to learn systemically from each other's tests is missed, posing a critical managerial challenge. Empirical research to address these challenges and frameworks that support a more integrated and systematic understanding of testing across product design and development remain limited in the academic literature.

This paper addresses these challenges by examining two industrial case studies and proposing a categorisation framework that describes the purposes of testing in product design and development. The proposed framework maps testing across the design and development stages, clarifying its purposes and highlighting how tests can serve multiple stakeholders. This framework aims to bridge gaps in understanding among different stakeholders to support shared learning among practitioners. It also provides a clearer view of testing for the design research community.

2. Motivation

Industry practitioners understand that testing is challenging, and numerous testing activities occur, each serving specific purposes that meet the needs of a team. However, when it comes to project cost savings, there is a constant battle between teams over whether a test is needed. Due to a lack of shared understanding, a test that is highly valuable to one team may not be seen as valuable by another team. Teams may be unaware of a test and unable to gain value from other teams' tests.

The lack of shared terminology reinforces these issues. For instance, verification (the process of evaluating a design against engineering requirements) and validation (the process of assessing a product against stated or implicit customer requirements) are often used interchangeably, leading to confusion among teams regarding the expectations for these activities (Tahera *et al.* 2023). As a result, when

the role and purpose of testing are unclear among different stakeholders, the need for a test might be misunderstood, test data could be underutilised or the test could remain narrowly focused on meeting the team's needs. Valuable knowledge generated through testing can remain siloed and fail to inform subsequent design phases, leading to the repetition of avoidable mistakes and costly iterations.

To address this challenge, this research aims to “clarify the purposes of testing in design and development to bridge gaps of understanding among diverse stakeholders.”

The rest of the paper is organised as follows: [Section 3](#) reviews the relevant literature on the role of testing, how testing is discussed in design and development processes (DDP) and the involvement of different stakeholders. [Section 4](#) describes the research methodology, followed by insights from case studies in [Section 5](#). [Section 6](#) proposes the categorisation framework. The discussion and conclusion are in [Section 7](#).

3. Background and related literature in testing

Over the past few decades, the discussion on testing during product design and development has been dominated by the reduction of physical testing in favour of virtual testing, utilising advanced CAE, simulation and digital twins (Tao *et al.* 2018; Mauery *et al.* 2021; Popov *et al.* 2021; Bi *et al.* 2024). However, physical testing has gained renewed attention in some industries for purposes beyond V&V, especially for complex and innovative products. For example within the space industry, where there is an ongoing debate about the merits of incorporating early rapid prototyping and extensive physical testing of multiple conceptual solutions early in the design process (Öhrwall Rönnbäck & Isaksson 2018), with a focus to identify design flaws early, gather new data for analysis and enhance subsequent prototypes, thus minimising uncertainty, and accelerating development time (Peña Häufler *et al.* 2021; Shafqat *et al.* 2022). A similar trend has been observed in the automotive industry, with a strategy of “shift left” (see [Figure 1](#)), where V&V aspects are considered early in product design and development through virtual engineering (Bachorek & Jung 2023). Testing approaches are increasingly emphasising early prototype testing for learning and knowledge generation, balanced with qualification testing such as certification carried out throughout the development stages. Manufacturers are also involving customers or end users early and frequently, at every stage, to embed prototype demonstrations and tests for both technical assessments and validation (Tahera *et al.* 2019; Cooper 2024). This approach enables the rapid implementation of critical feedback loops and iterative design elements in agile development, particularly in high-uncertainty projects.

Testing is becoming increasingly challenging for complex products incorporating new technologies (Ferguson *et al.* 2022; Tahera *et al.* 2023). This complexity determines the testing approaches to demonstrate that products can meet the intended objectives (Valerdi & Sullivan 2023). Testing of incremental products with minor improvements or simple adjustments to the existing design can follow the procedures of the previous product generation, with additional testing required for any design changes. However, the design and development of complex innovative products need careful consideration of testing. For example, suppose a car manufacturer updates an existing model by redesigning the body for improved aerodynamics. Testing will primarily focus on functionality and may include

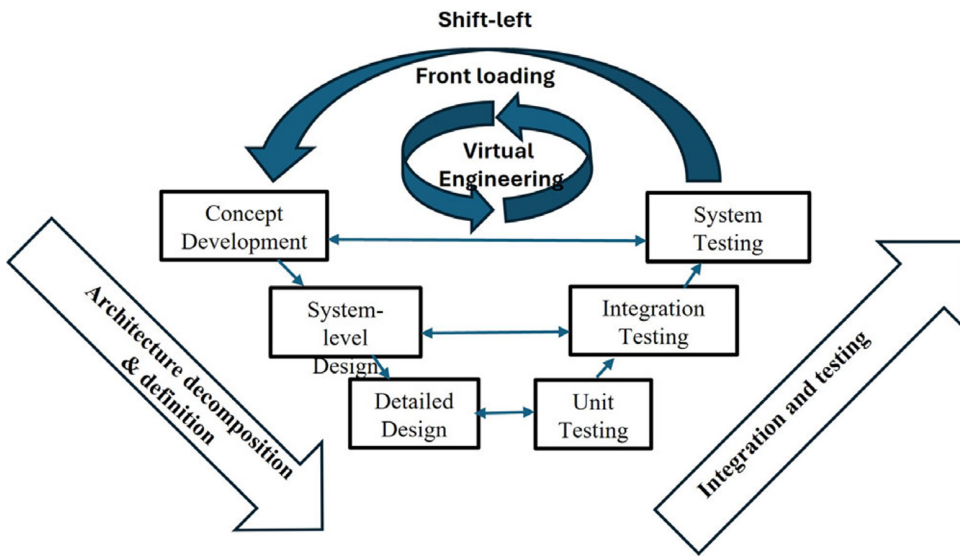


Figure 1. Simplified V model adopted from Bachorek & Jung (2023), redrawn to present the concept of “Shift-left,” where V&V and testing are considered through virtual engineering early in product design and development.

qualification testing to demonstrate compliance with relevant standards and regulations. However, a radically new car design, for example, with a lightweight carbon-fibre chassis integrated with floor-mounted electric batteries, introduces known and unknown uncertainties. Engineers must conduct additional experimental testing and low-fidelity prototype tests to build a sufficient understanding of the new design. For example, to learn how a new design behaves and performs when prior knowledge is limited and there is little similarity to existing designs. Testing during the early design phases may be conducted to identify feasible combinations of design parameters and to address any associated risks promptly, and can serve V&V purposes at later stages. This could include analysing battery–chassis interactions through thermal-stress analysis and testing, performing new structural integrity assessments for crash safety and evaluating how composite materials degrade under real-world vibrations and impacts. Product design with new technologies like this introduces far more physical testing, not just for functional performance, mechanical durability and reliability to verify the design and validate the product, but also early tests for design exploration, experimentation and uncertainty reduction (Thomke, 2008). These increase the overall development costs within the industry. Learning-focused experimental testing that incurs upfront costs but delivers value is often under-recognised (Arntzen *et al.* 2019) or challenged by different stakeholders (Canny *et al.* 2019), creating tension in testing practices and investment choices during product design and development.

The next sections examine the literature on the roles and purposes of testing, how testing is perceived in the models of product design and development and the involvement of different teams and stakeholders in testing.

3.1. The “roles” of testing

Testing is defined as “an activity in which a system or component is executed under specified conditions, the results are observed or recorded, and an evaluation is made of some aspect of the system or component” (IEEE 1990). According to this definition, tests are primarily conducted to evaluate and confirm that a design performs as intended and meets the specified technical and customer requirements before release (Engel 2010; Shabi *et al.*, 2017).

Testing has many other purposes identified in the literature (see Table 1). For example, Broyan *et al.* (2021) state that the main purpose of testing is to reduce uncertainties to acceptable levels by the end of the design process (Broyan *et al.* 2021). Testing can address two broad categories of uncertainty: “epistemic” and “aleatory” (Halfpenny *et al.* 2024). Epistemic uncertainty arises from a “lack of knowledge” or “lack of definition” (Mcmanus & Hastings 2006), which can be particularly relevant to new technology and innovative product development. In contrast, aleatory uncertainty, also known as objective or stochastic uncertainty, relates to inherent variations within a physical system or environment, such as dimensional variations in machined components (Wynn, Grebici & Clarkson 2011). For example, testing has been reported to be vital for evaluating design uncertainties of additive-manufactured components for space use (Dordlofva & Törlind 2020; Borgue, Panarotto & Isaksson 2022). New product design and development projects can face additional uncertainties related to estimation, project lifecycle stages and stakeholder involvement (Atkinson, Crawford & Ward 2006). Learning from testing can also mitigate these uncertainties and improve project outcomes (Thomke, 2008).

Testing plays an essential role in learning through early design experimentation and innovation (Thomke 2003). Erat & Kavadias (2008) describe experimentation as a four-step cycle: design, build, run and analyse, emphasising “run” and “analyse” as distinct steps. Testing occurs during the run phase, and learning happens during the analysis phase (Erat & Kavadias 2008). Similarly, Loch *et al.* (2001) highlight that iterative design-test cycles, in which learning from previous tests informs subsequent iterations, are crucial to a project. Software and service development in particular have a culture of repeated experimental user testing to generate insights or to understand the relationship between specific actions that deliver monetisable value to businesses (Lindgren & Münch 2016).

Other key roles of testing identified in product design are to generate information and obtain precise data on a proposed design’s properties, by revealing information such as material properties and dimensions (Song & Montoya-Weiss 1998; Camburn *et al.* 2017), identify discrepancies between expected and actual performance (Chen, Cheung & Yiu 2020), detect failure or degradation modes under intended conditions and reveal both direct and indirect system behaviours (Ktena, 2011).

Sim & Duffy (2003) describe testing as a process to assess feasibility and eliminate infeasible options, thereby narrowing the design space (Sim & Duffy 2003). Testing also facilitates the gradual refinement of the design (Camburn *et al.* 2017). Otto & Wood (2000) also suggest that testing aids in design feature optimisation and can uncover critical design issues (Otto & Wood 2000). Similarly, Luna *et al.* (2013) emphasise the importance of testing in evaluating uncertain scenarios in complex system-of-systems environments.

Table 1. Different purposes of testing identified in the literature with selected references

Purpose of testing	Selected publications (e.g.)
Testing for learning	(Erat & Kavadias 2008; Moogk 2012; Tahera <i>et al.</i> 2019)
Testing for experimentation	(Song & Montoya-Weiss 1998; Thomke 2001, 2003, 2008; Erat & Kavadias 2008; Moogk 2012; Luna <i>et al.</i> 2013; Lindgren & Münch 2016; Camburn <i>et al.</i> 2017; Peña Häufler <i>et al.</i> 2021)
Testing for generating information, revealing information, refinement, uncertainty reduction	(Song & Montoya-Weiss 1998; Thomke 2001); (Thomke & Bell 2001; Camburn <i>et al.</i> 2017)
Testing for simulation model building, simulation result validation	(Sargent 1991; Sargent 2013; Tahera, Earl & Eckert 2017)
Testing for design evaluation	(Sim & Duffy 2003; Luna <i>et al.</i> 2013; Peña Häufler <i>et al.</i> 2021)
Testing for demonstration	(Yadav, Singh & Goel 2006; Shabi <i>et al.</i> , 2017; Tahera <i>et al.</i> 2019)
Testing for verification and validation	(Hoppe, Engel & Shachar 2007; Engel 2010; Shabi <i>et al.</i> , 2017; Shankar, Summers & Phelan 2017; Tahera <i>et al.</i> 2019)
Testing for assurance, qualification, certification	(Bieda 1998; Tahera 2014)

Prior research indicates that testing plays many roles in product design, but these roles are often mentioned only briefly as passing concepts, without detailed context. This makes it hard to connect the roles of testing and the purpose within the overall design process, especially regarding when testing occurs and who the stakeholders are.

3.2. Testing in design and development process (DDP) models

A DDP model offers a structured view of activities, information flows, decision points and iterations throughout the DDP (Wynn & Clarkson 2018). Testing is placed towards the end of the development process in these widely known DDP models. For instance, the sequential stage-gate model (Cooper 1990) and Systems Engineering V-model (Figure 1) (Forsberg & Mooz 1994) position testing as a post-design phase, primarily for V&V purposes (Buede & Miller 2024). (Note: figures presenting these models have been avoided in this paper due to limited space, but can be found in Tahera *et al.* (2019)).

Design approaches, such as Hybrid and Agile (Cooper & Sommer 2016) and the set-based design approach (Kennedy 2008), iterative design and test cycles and frequent testing throughout the process are recognised as crucial elements. For example, the Hybrid and Agile approaches suggest that early testing helps explore the design space and make informed decisions. Testing throughout helps reduce uncertainties and avoid costly rework that can arise from discovering flaws late in

the process (Cooper & Sommer 2016). Similarly, the set-based design approach advocates conducting testing as early as possible in the design process, known as test-then-design (Kennedy 2008), to uncover trade-offs and limitations in technology before finalising design decisions. A fundamental principle of this approach is that a product or product parameter is tested not only for compliance with requirements but also to identify the available design margins (Del Rosario, Fenrich & Iaccarino 2019). Set-based design approaches advocate the use of testing to acquire knowledge not only for the specific design concept (for instance) but also for the broader design space. The ultimate objective of set-based design is to inform designers of the feasible solution space by finding (through testing) its limits.

3.3. Testing from a stakeholders' involvement view

Effective cross-functional teamwork in testing greatly benefits product design and development, leading to improved project outcomes across the industry (Cooper 2024). However, coordinating testing activities among multiple internal and external stakeholders can be challenging, particularly for large-scale, multifaceted projects (Kulkarni, Salado & Wernz 2022). The purpose of testing can vary depending on the stakeholders' perspectives and expectations involved in product development. For instance, internal stakeholders such as the design team may see testing as a way to assess the feasibility of proposed design concepts (Menold, Simpson & Jablow 2016). A CAE engineer typically sees physical testing as a means to learn whether their CAE models accurately predict real physical behaviour under specific conditions (model validation) or to assess the accuracy of the predictive models used (Moon *et al.* 2020). Test engineers, responsible for designing tests and defining the assessment criteria, aim to verify whether a product or its components meet specified requirements (Cornford 2001). Technicians mainly focus on whether tests indicate “pass” or “fail.”

External stakeholders have specific expectations for testing, as certification agencies and regulatory bodies establish protocols to guarantee correct procedures, product validation, quality assurance and certification. Government and regulatory bodies also monitor testing outcomes (Ciuffo *et al.* 2015). Many companies also involve suppliers in the testing process or engage beta-testing customers who evaluate the products under real-world conditions (Kirsilä, Hellström & Wikström 2007; Tahera 2022).

3.4. Gap in the literature

Surprisingly little empirical research examines testing in detail within product design and development despite the recognised importance of testing in industry. This view still supports Engel's (2010) observation that studies on testing in product design remain limited. Also aligns with Lévardy *et al.*'s (2004) finding that testing continues to receive significantly less attention than design and analysis tasks.

Although many roles and purposes of testing have been mentioned in the literature (see Table 1), most have not been contextualised within DPP models. Most studies focus on individual tests or specific purposes without relating them to other tests conducted for different objectives. This is also reflected in industry practices (discussed later in this article), where stakeholders often have varying

expectations and interpretations of testing activities and rarely coordinate with different teams.

Valerdi & Sullivan (2023) provide a valuable overview of the different types of testing undertaken during product development. Our case studies suggest testing in practice is often complex and does not exhibit a one-to-one relationship between test types and testing purposes. A single type of test may fulfil multiple purposes, while a specific purpose may require multiple types of tests throughout the product development process. For example, design verification may be addressed through several tests conducted at different stages of product development.

Systemic learning happens when knowledge from past tests is captured, shared and applied across project stages or even across projects (Verganti 1997) and is essential for deeper organisational learning (Thomke & Bell 2001; Erat & Kavadias 2008). Such learning is particularly critical in managing uncertainty during early development stages, refining testing strategies and fostering more adaptive approaches to design and innovation projects (Verganti 1997; Peña Häufler *et al.* 2021). The full value of testing can be underutilised without a clear and shared understanding among stakeholders. The literature review identifies a clear lack of a structured framework that helps understand testing activities and their purposes, and that clearly links to the design process.

4. Research methodology

As shown in Figure 2, the paper combines sources and authors' experiences. A 2023 workshop identified the need for systematic categorisation, leading to a literature review on testing purposes. Two case studies were revisited: one on diesel engine testing (Tahera 2014) and another on satellite propulsion system design (unpublished work). Figure 2 shows the overall steps followed in this research, although there were some iterations during the steps.

4.1. The workshop

The need for a categorisation framework was emphasised by an industry workshop held in 2022 (Tahera *et al.* 2023). The workshop titled “Physical testing in the era of digitalisation” aimed to gather a shared understanding from industrial experts and academics regarding challenges in testing. The purpose of this workshop was to gather insights into the significance of physical and virtual testing in product development and to identify factors influencing testing decisions during new product design. The online workshop was attended by 18 participants, including six from UK sectors such as automotive, aerospace, household appliances, oil and gas and software (with 5–30 years of experience), and 12 academics from the UK, USA and Europe. These industry professionals held roles, including product design, CAE analysis, modelling, simulation, verification, validation and testing.

The workshop began with three structured breakout groups, a mix of industrial and academic participants, each facilitated by an academic, discussing “When does testing happen in new product development?” and “What are the current challenges?”. All participants engaged in an open discussion in the second session addressing the question, “What are the challenges in the digitalisation of the testing process?”. The workshop was recorded, transcribed in Microsoft Teams and analysed using open coding (Miles 1994). The data were then clustered into themes

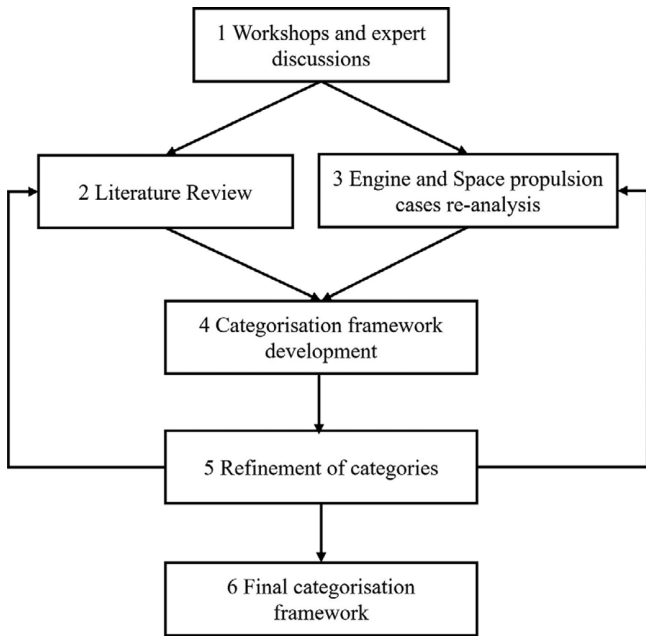


Figure 2. Research process demonstrating how expert input, literature review and case re-analysis contributed to the development and refinement of the final categorisation framework.

and subthemes and independently coded by three researchers to avoid bias. The results were refined collaboratively into agreed themes. More about themes can be found in Tahera *et al.* (2023).

4.2. The literature review method

This literature review employed a semi-structured process (Torraco 2005). Initially, a collection of publications was assembled, including influential papers identified from the reference list of Tahera *et al.* (2019). Selected papers were critically reviewed, and additional sources were obtained through citation analysis of these papers and a snowballing approach. The collection was also shaped by the authors’ 15 years of research in testing within engineering design and related fields.

Targeted searches were subsequently conducted in the databases of Web of Science, Scopus, IEEE Xplore, ScienceDirect and SpringerLink, focusing on the term “testing” and including the terms “new product design and development,” “product innovation,” “design processes,” “engineering design” and “engineering management.” We also searched in Google Scholar using specific terms such as “test,” “experiment,” “verification,” “validation,” together with “role of testing,” “purposes of testing” and “objectives of testing” to find articles that discussed specific purposes of testing. Many articles focused on test procedures and techniques, which were excluded from this study. Thirty-five articles were analysed. These key sources are shown in Table 1. Additional searches were conducted to identify examples of the subcategories of the proposed framework.

4.3. Case study method

This paper draws on two empirical case studies: (1) the design of space electric propulsion (EP) systems and (2) the design of diesel engines in a large manufacturing company.

The *Satellite EP systems* design case is based on an unpublished prior study focused on *highly innovative* products developed by a French manufacturer specialising in Hall Effect Thrusters (HETs). This case study was conducted by the second and fourth authors in 2021 and 2023 as part of a European research project on the design of Low-Energy EP systems. The interviews and data collection for this study focused on testing processes for three product types across three different Technology Readiness Levels (TRLs). The findings from this case study have not been published previously.

This study was conducted through interviews and workshops with eight industry participants. Approximately 80 current testing activities were identified and mapped using a Business Process Modelling Notation (BPMN) model in accordance with European Cooperation for Space Standardisation procedures. The second phase refined the BPMN model to address new business conditions, such as increased production demands from satellite mega-constellations, and proposed a “to-be” testing process. Finally, an academic expert on air-breathing EP was interviewed to explore whether new testing strategies might be required for radically different technologies, such as air-breathing propulsion, for new missions like low Earth orbit operations.

The representation of the “as-is” and “to-be” process models for three types of EP systems enabled discussions with company experts about the purpose and rationale behind the desired modifications to the current testing process. The priorities assigned to different types of tests were determined based on the time and cost data for 80 testing activities, as well as expert judgments during interviews. Data is presented in a qualitative form to protect sensitive company information.

The *Diesel engine design* case was revisited to investigate the various purposes of testing and the involvement of different stakeholders in an *incrementally developed product*. The case study was carried out from 2011 to 2014 during a PhD study (Tahera 2014). A total of 18 transcripts of semi-structured interviews and documents (diagrams, graphic representations, PowerPoint slides, procedures) collected during the original study were reanalysed. These interviews, lasting between 40 and 180 minutes, were conducted with eight engineers, which included senior business managers, product engineers, CAE engineers and V&V managers. The detailed methodology of the case study can be found in Tahera *et al.* (2019).

The original study focused on the testing process and its close link to design, identifying roles of testing (learning, demonstration, verification, validation and certification), but did not provide an exhaustive list. The focus of this study was on identifying the purposes of testing, specifically how different interviewees framed these purposes and linked them across various stages of product development, as well as the involvement of different stakeholders and teams in these stages.

First, the transcripts were systematically searched to identify all references to testing. Each reference was analysed based on (a) the stated role of testing (e.g., learning, validation, failure mode detection, demonstration), (b) the underlying purpose attributed to it (e.g., technological uncertainty reduction, regulatory compliance, customer validation), (c) the stakeholders involved (design engineer,

verification engineer, business manager, regulatory bodies, customer etc.) and (d) the stage of product development related to.

To ensure that the findings reflected current industrial practice, two additional interviews were conducted with a Business Transformation Manager in 2021 to discuss the implications of technological advancements and to identify any changes in their structured product development process. The discussion revealed that, despite advances in virtual testing and digital technologies, the structural testing process and procedures have largely remained unchanged within this company for the past 30 years. A similar observation within the automotive sector was reported by Yassine & Souweid (2021).

4.4. The framework development method

The list of purposes identified in the literature was first analysed. These purposes were often presented in a fragmented manner, with different terms used to describe similar yet distinct concepts. For instance, purposes such as “generating information” and “revealing information,” as well as “refinement” and “uncertainty reduction,” appeared together as similar concepts.

The case studies revealed four overall purposes of testing in the design process. Then we categorised these based on the problem and the solution space (Dorst 2011, 2019). The definitions of each category were derived from the literature (when available), and examples from case studies were used to explain the purposes. When a clear definition was not found in the literature, one was derived from the case studies and the combined understanding gained from the analysis of the literature. Some categories, such as exploration, exploitation and experimentation, were identified only in the literature, as these terms were used interchangeably in case studies.

The development process was iterative; as new terms emerged in the literature search, we examined their ontological differences and integrated them into the developing framework. Subcategories were therefore progressively developed through repeated cycles of analysis and refinement. This process continued until a comprehensive list of categories and subcategories was identified that could capture and represent the purposes of testing in the design process. The framework was also evaluated against the generic design process that could encompass both the problem and solution domains of design (Dorst 2019).

5. Insights from the case studies

This section presents insights from two empirical case studies. The satellite EP design case reports the challenges in aligning multistakeholder perspectives in complex projects, and the engine design case shows the complexity of testing in the design process, its purposes and stakeholders’ involvement.

5.1. Testing in Satellite Electric Propulsion (EP) design

This study examines the testing requirements and planning processes for three EP products in the space sector. Medium-Power HETs are designed for both orbit-raising and station-keeping missions. Low-power HETs are being developed for EP systems to meet the demands of the small satellite market. An air-breathing electric

thruster represents the most recent technology under development for this purpose.

5.1.1. *Medium-power EP*

Medium-power EP systems are the most mature and high-investment systems, with production targets ranging from tens to hundreds of units per year. The primary objective is to meet the stringent qualification requirements specific to these systems. For decades, government programmes have been the primary customers for medium-power EP systems, driving manufacturers to prioritise performance and reliability. This emphasis has resulted in rigorous verification, qualification and acceptance testing to ensure the systems meet specifications under intended environments with the required safety margins.

These thrusters are typically qualified for lifespan by testing a single thruster for the required duration, plus a safety margin. However, the qualification testing of this technology is still challenging, as a single thruster life test does not ensure statistically meaningful reliability. Two strategies are commonly employed:

1. Extended Lifetime Testing: Extending the test duration by 30%–50% to account for variations between units or inherent variability in wear-out processes.
2. Failure Mode Testing: Additional testing to address critical failure modes, often conducted at the component level using accelerated life testing techniques.

Despite advancements in the virtual simulation of EP, the complexity of the physics involved, such as plasma physics, makes it impossible to replace or reduce physical tests with digital simulations (Taccogna & Garrigues 2019). Additionally, no specific guidelines or standards for thruster life qualification for EP systems have been developed. Instead, these standards are often extrapolated from those used in chemical propulsion, as mission requirements can often be adapted.

5.1.2. *Low-power EP*

The rise of entrepreneurial actors has introduced new business scenarios in the space industry, such as mega-constellations for global internet coverage. These scenarios involve a dramatic increase in production rates, about 10–100 times higher than traditional satellite applications (e.g., telecommunication, science). Consequently, manufacturers face ambitious cost and lead time reduction targets. As this industry experiences a tenfold increase in production volume, conducting quality assurance tests for each thruster (as performed today) will be practically impossible (e.g., Brophy & Polk 2018).

Low-power EP systems often base their operational conditions and requirements on those of medium-power EP systems, as they share similar architecture and working principles. Consequently, there is a greater emphasis on qualification and acceptance testing, with less focus on demonstration testing. However, teams developing low-power EP systems (TRL 5–6, with target production volumes of hundreds of units per year) have a different view and want to focus on developing test strategies that minimise uncertainties regarding performance consistency across manufactured units.

Given time and schedule constraints, it is impractical to test every system individually (as is often done for medium-power EPs). Therefore, new testing approaches are required, emphasising batch testing and ensuring repeatability

through statistical analysis. While this approach is common in other industries, such as automotive, it represents a significant shift for the risk-averse space sector. A potential solution that was discussed was developing a statistical reliability model for thrusters based on tests conducted on multiple simulation models, currently one to six models (M1–M6). Quality assurance testing could then rely on statistical sampling rather than testing every unit.

A systems engineer highlighted the importance of seamless data flow between design-phase testing (uncertainty reduction) and quality assurance testing (compliance). However, these processes are currently decoupled due to differing stakeholder testing priorities. A new qualification approach would integrate early-phase test data with statistical models, increasing confidence in the design lifespan. However, applying this approach to the studied company would require increasing the number of engineering models to many (from M1–M6 to M1–MX), which contradicts objectives pursued by other departments. As one R&D manager noted, overcoming internal resistance to this testing mindset requires effort, but it is essential to align testing with evolving industry needs.

5.1.3. Air-breathing EP

Air-breathing EP systems are a radically new technology which operates on fundamentally new physical principles and targets innovative missions (e.g., orbits ~200 km above Earth). For teams working on air-breathing propulsion systems, the primary challenge is determining whether the technology is viable. If the technology fails, there will be no need to qualify or optimise it. The second challenge lies in understanding mission requirements, as these differ significantly from those of medium-power and low-power EP systems. This technology still requires extensive multiyear testing to establish feasibility and define requirements (Andreussi, Ferrato & Giannetti 2022). Researchers often demand increased investment in demonstration programmes to support novel mission scenarios.

5.1.4. Emphasis on testing on different product types

Insights from this case study provided a preliminary understanding of the different purposes of testing assigned to three product types within an EP manufacturing organisation. Based on expert discussions, the visual representation of the relative importance assigned to different testing types is drawn in Figure 3 to provide a visual summary of testing priorities rather than a precise quantitative comparison.

Two key factors found to drive the purposes of testing and priority in the EP business are: (1) the TRL and (2) the targeted size of production units. Medium-power HETs, as a more established technology, place a greater emphasis on product validation and compliance with government regulations. Low-power HETs, which are at a moderately advanced TRL but are produced in larger quantities, require extensive testing to address technological uncertainties and evaluate their suitability across various systems and environments. Whereas air-breathing thrusters, characterised by a low TRL, are primarily undergoing testing for proof-of-concept and concept demonstrations in laboratory and controlled environments.

Interviewees highlighted that the increased production demands for mega-constellations would require innovative testing strategies. Accelerated lifecycle testing of a single thruster may be sufficient when production volume is in the

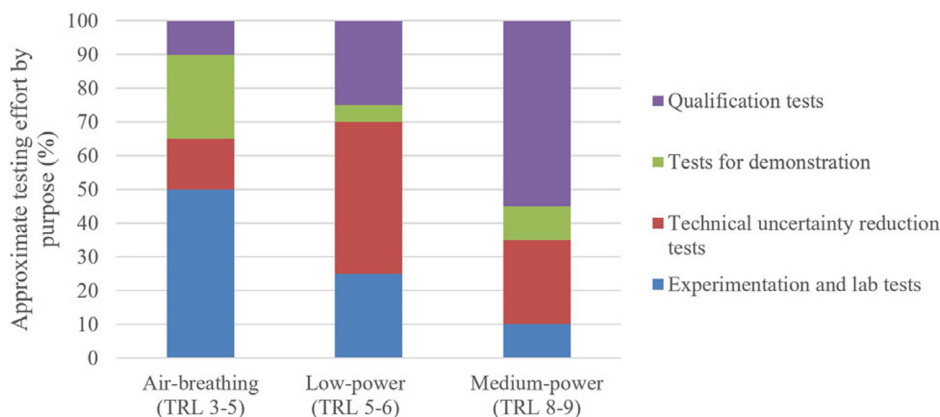


Figure 3. Distribution of testing effort across three EP types, highlighting distinct testing priorities for air-breathing, low-power and medium-power EP technologies at different TRL levels.

tens to hundreds, but it is not adequate to ensure quality for thousands of manufactured items. Strategic investments in testing, like early-stage failure mode testing, engineering modelling and probabilistic techniques, can minimise the need for extensive lifecycle physical testing. It is crucial to prioritise investment in laboratory testing (and even real-world demonstrations) to showcase the capabilities of the Air-Breathing EP and attract the investment needed to further develop these low TRL systems. However, the challenge in the sector is motivating governments and management to invest strategically in development.

5.2. Testing in the engine design and development process

The company’s new product development programme includes New Technology Introduction (NTI) and New Product Introduction (NPI) (outlined in Figure 4). Testing in NTI primarily focuses on understanding and evaluating the fundamental principles of emerging technologies. Only technologies that have been validated through rigorous laboratory testing in research and development (R&D) are considered. Technologies ideally reaching TRL 4 are considered in NPI. Formalised testing procedures and planning start in the NPI phase. It is important to highlight that the interviews primarily focused on NPI. However, interviewees mentioned the significance of experimental lab-based testing in several discussions and pointed out gaps in how these test results were not utilised during the NPI phase.

As the diesel engine is a mature product, design changes occur incrementally within the company. Engineers described their products as being incrementally improved, noting that the “functional capability” of the diesel engine does not change dramatically. Instead, regular small updates to the technology enhance performance or reduce the cost of ownership. When discussing the nature of testing for different types of product innovation, they stated:

Incremental design changes can be managed with standard testing or validated simulations, whereas step changes (radical product innovation) may require a new test plan.

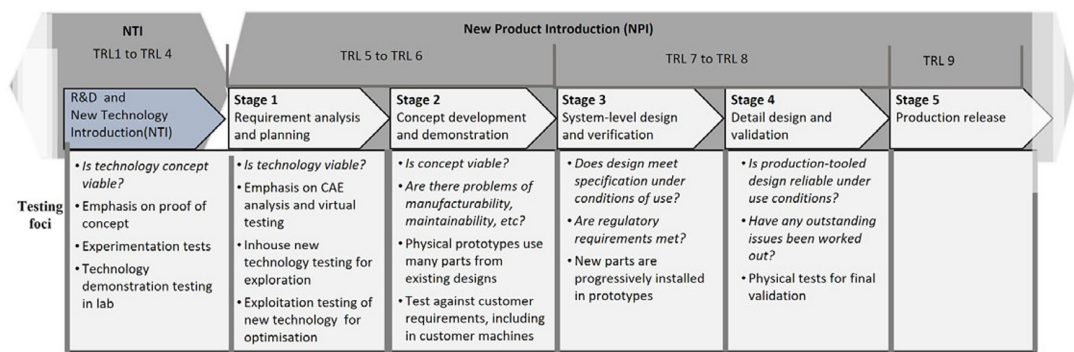


Figure 4. Focus of testing in each stage within the stage-gate process model used in the diesel engine company (adopted from Tahera *et al.* 2019).

For example, if the company is designing a cylinder block that is a scaled version of a previous product, the critically stressed areas would already be known, so it is possible to accurately assess risk through simulation, whereas a new cylinder block design would require physical testing for understanding stress behaviour and conditions.

Within an NPI programme, product objectives are reviewed against a Current Product Issues (CPI) database. The CPI database provides information about failure modes and effects observed in current products, which require special attention in the next product generation. This process is undertaken by key team members, including technical specialists, component owners, design owners and V&V managers.

5.2.1. Iterative design and testing in the stage-gate process

The company follows a formal NPI stage-gate process, mandated by the parent company to guide the development and testing of new engine designs. A detailed explanation of these stages is reported in Tahera *et al.* (2019). Figure 4 presents five key stages, from concept evaluation to production release, with defined objectives and decision points at each stage. The NPI process provides context for understanding how testing requirements evolve and testing focus throughout product development. Specifically, it illustrates how the purpose of testing shifts from early-stage technology viability assessment and concept demonstration to design verification, validation, qualification and production readiness in the later stages. Each stage described below shows how testing activities are prioritised according to design maturity and the development risks being addressed.

Stage 1 (Requirement analysis and planning): In-house technology or component-level testing or at the facilities of component suppliers to address any specific concerns. Primarily, the focus of testing is to explore the technology capabilities and potential adoption for new engine design and validate new technologies and components. There is also a particular focus on gathering data for building simulation models and CAE analysis.

From Stage 2 to Stage 4, testing goes on two levels: (1) engine-level testing involves evaluating standalone engines on test beds, and (2) machine-level testing

involves mounting engines in customer machines or vehicles to assess performance under anticipated use conditions. Performance and Emission tests are conducted first, followed by mechanical durability and reliability assessments (Tahera *et al.* 2019).

Stage 2 (Concept development and demonstration): demonstrates the viability and performance of new technologies integrated into new engine designs. Alternative designs are evaluated using various versions of a MULE engine, which is built by incorporating new technologies with existing components.

Stage 3 (System-level design and verification): Design verification testing happens in this stage to optimise performance and validate the hardware under optimised conditions. Prototype engines are assembled with a combination of production-intended and prototype parts that undergo both engine- and machine-level testing.

Stage 4 (Detailed Design and Validation): The effects of production and hardware variations on performance are evaluated, including comprehensive testing for reliability and durability, which will be conducted at this stage. The essential tests for compliance are typically planned to be concluded during product validation phases.

Stage 5 (Production release): continuation of mandatory tests (if necessary) and any remaining validation aspects.

5.2.2. Mapping testing purposes and stakeholders

Internal stakeholders frequently mentioned during the interviews include product development engineers (often called component owners or product owners within the company), design engineers, the CAE team, V&V managers, the business manager and testing teams. External stakeholders comprise suppliers, customers and regulatory and governmental bodies. First, it was pinpointed when various stakeholders, whether official or unofficial, engage at each stage, and the reasons for their involvement were clarified. Then, stakeholder engagement throughout the NPI process was mapped. The main purposes of testing broadly were to reduce technical and design uncertainty, demonstrate concepts/systems and perform V&V. Figure 5, which is built on Figure 4, shows how “testing purposes,” “internal stakeholders” and “external stakeholders” are mapped with NPI stages in this engine company.

The main purpose of testing in *Stage 1 and Stage 2* is “defining design requirements” and the viability of “concept demonstration”. The *Product development team*, *Design team* and *CAE team* are primarily involved in testing to explore alternative technologies and establish the design requirements for the engine. These teams closely engage in iterative processes to identify design issues, mitigate design and technological *uncertainties*, refine design requirements and concurrently develop and validate CAE models using measurement data from physical tests. The *V&V team* only oversees the testing activities and gets informed of the outcome as necessary. In *Stages 2–4*, the primary purpose of testing is to *demonstrate* the engine’s performance, verify the design and validate the product to meet the requirements of both internal and external stakeholders. The *V&V team* and *Testing team* play a central role in this stage, focusing on identifying actual failure modes through tests and overseeing the planning, execution and analysis of tests. The leader of the *V&V team* is responsible for coordinating testing activities

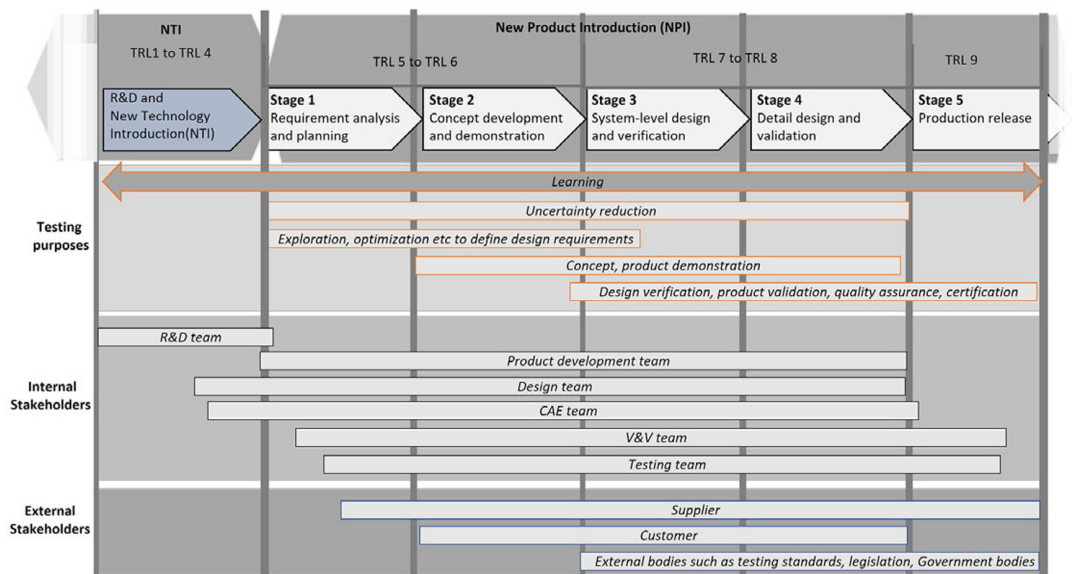


Figure 5. Mapping of testing purposes and stakeholder involvement across five stages of engine development, highlighting the increasing complexity and multistakeholder involvement in testing activities throughout the process.

with the *Testing team and other teams* in alignment with the requirements and overseeing the verification plan through tools like the design verification plan and report. The *Product development team*, *Design team* and *CAE team* engage actively in interpreting test results and addressing any design issues uncovered during testing.

External stakeholders, such as suppliers and customers, get closely involved from *Stage 2 through Stage 5*. MULE and production-intended engines are tested within *customer* machines for demonstration, verification and validation purposes. This process facilitates early feedback and overall assessment of the engine's performance. *Suppliers* conduct the component and sub-system validation tests according to the specifications provided by this manufacturer. Regulatory bodies and guidelines are consulted throughout the process, overseeing and guiding the qualification and compliance process. This ensures that all qualification requirements are met at each stage, and the product complies with relevant standards and regulations for its intended applications and market use. They may become closely involved in certification tests or be informed of the outcome.

The current process highlighted the challenges of disconnect and silos between teams, as discussed in this case study. For example, many early tests, especially during R&D and Stage 1, are team-driven and focus on addressing specific issues related to technology readiness or design, often operating within teams and disconnected from later stages of V&V. For instance, it was noted that the V&V manager had minimal involvement or visibility in earlier in-house experimental testing and component-level testing carried out with suppliers, restricting their ability to incorporate these insights into the final verification plan. However, they emphasised that early engagement of the validation team with experimental testing

activities could greatly enhance the overall V&V programme. The V&V teams could gain valuable insights by incorporating additional requirements for tests that identify real failure modes or by focusing on specific aspects or results that are crucial in later stages, such as gathering useful measurement data and validating simulation models for subsequent stages. Similarly, tests may be performed, particularly at Stages 4 and 5, in the event of test failure or to address manufacturability issues. These lead to tests happening outside the planned V&V activities without full awareness or involvement of the CAE and design teams.

To summarise, observations from the diesel engine case study highlight the inherent complexity of the current testing process, driven by the multifaceted roles of testing activities. For instances,

1. A test can serve multiple purposes: A single testing activity can fulfil multiple purposes for different stakeholders simultaneously. For example, a prototype test may serve to “verify” a design concept for the design engineers while also “demonstrating” the concept to customers. Therefore, it is crucial to identify the purposes of multiple stakeholders that can be aggregated into a single test.
2. Multiple tests serve the same purpose: Multiple tests are conducted concurrently or at various stages to fulfil a purpose. For instance, verification of a product design may require several testing activities. FEA analysis can sufficiently verify a design parameter for CAE and design engineers, but the V&V manager would also need to perform physical tests to verify the failure mode. Often, the V&V team will have to follow the company’s standard test plan even though realising a failure mode is overly assessed by multiple teams and at different stages.
3. A test serves different purposes across stages: A test may have varying purposes depending on the stage of product development. For instance, a gross thermal test of a MULE engine might be used for customer demonstration during the concept phase, while the same test on a new engine would serve design verification and product validation purposes in later stages. However, such a test may not meet certification requirements if conducted at an earlier stage.

5.3. Key findings from both case studies

Both case studies show that in practice, testing can have many purposes. For innovative technologies like air-breathing EP, which is at early TRL stages, the purpose of testing was primarily design exploration, experimentation, uncertainty reduction and model validation. However, testing mainly focused on compliance activities such as qualification, verification and validation for more mature products manufactured in large quantities, such as the engine and medium-power EP. This suggests that the emphasis on testing during development can vary not only with the product’s maturity level but also with production scale (as indicated in the space propulsion case study).

Testing activities across different stakeholders are fragmented and not fully coordinated in both case studies, and often one stakeholder’s testing needs are not fully understood by other stakeholders in large-scale projects (particularly evident in the space case). As a result, there is a misalignment between stakeholders’ priorities on testing (as per the space propulsion case study), and opportunities for systemic learning and value-added insights from testing for multiple stakeholders are often lost. The case study indicated that, while tests are typically

designed for a specific purpose at a given stage of the DDP, planned aggregated tests can evaluate multiple failure modes in a single test, benefiting other stakeholders. Therefore, it is suggested that coordination among stakeholders can maximise the value of testing. These observations highlight the need for creating a structured framework that systematically maps tests to their purposes, stakeholders and stages. This approach would improve coordination and better align costly testing efforts across the product development process. At present, however, no such framework has been found in practice in either case that systematically supports this coordination or, at the very least, makes all the purposes of testing visible among different stakeholders (also absent in the current literature).

6. Categorisation framework of purposes of testing

This section discusses the proposed categorisation framework for testing in design, drawing insights from both the case study analysis and the literature review. The first category addresses the discovery, definition and framing of the problem space of design. The remaining three categories are aligned with the solution space, focusing on the development and delivery of the solution or product. [Figure 6](#) illustrates the categories and sub-categories.

- (1) *Testing for defining design requirements* – for defining and framing the problem space,
- (2) *Testing for uncertainty reduction* – related to the set of potential solutions within the solution space, and for the purpose of evaluation and refinement,
- (3) *Testing for demonstration* – to choose the best solution from a set of alternative solutions, and
- (4) *Testing for compliance* – for the final V&V of the production-intended design and tests for regulations.

The following sections discuss each category in detail, concluding with a brief definition, the primary purposes of the tests and potential benefits. (A summarised table of all categories is in [Appendix A1](#). To note, categories are named according to the specific purposes of tests; for instance, a test conducted for the purpose of experimentation is termed an “experimentation test.”

6.1. Testing for defining design requirements

Testing activities in this category have the purpose of acquiring information related to the design or problem space. In the early stages of product design and development, design problems and requirements are often ill-defined, leading to epistemic uncertainty arising from inadequate knowledge or insufficient definition. Significant challenges frequently stem from the “newness” inherent in innovative designs, which may involve internal factors such as technological advancements or the adoption of new technologies in products, as well as external factors like evolving customer demands and regulatory requirements, as suggested by the case studies. Literature also suggests that testing plays a pivotal role in defining design specifications that align with both customer and technical requirements (Zwolinski & Gaur 2003; Cooper 2024).

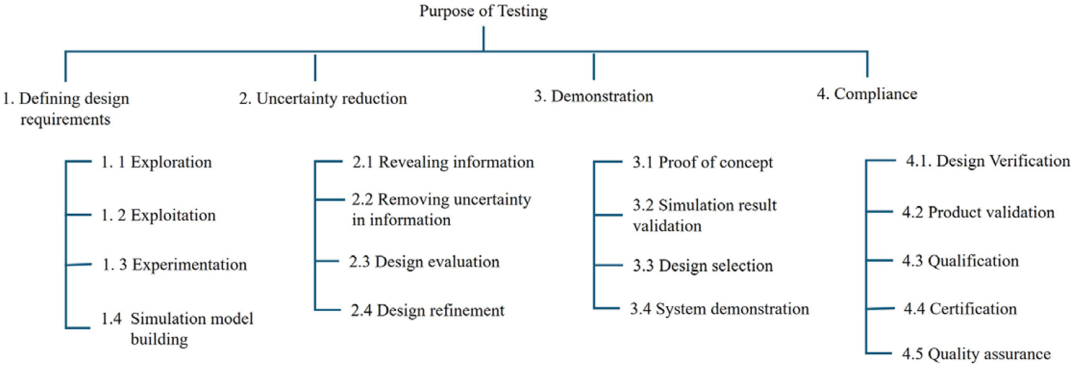


Figure 6. Top categories and corresponding subcategories of different purposes of testing that may happen in product design and development.

6.1.1. Exploration

Exploration testing is primarily conducted to acquire new knowledge, typically addressing the foundational question: “If I do this, what will happen?” This type of testing aims to uncover information about the design space, including its boundaries, parameters and the complex interactions between these parameters. According to (Steinle 1997), exploration could be a specific type of experimentation “– an experimentation which is not, as in the ‘standard view,’ driven by specific theories. It is typically practised in periods in which no theory or – even more fundamentally – no conceptual framework is readily available. I call it exploratory experimentation” (Steinle 1997). Unlike theory-driven experimentation (discussed later), exploratory testing can reveal critical design characteristics, such as in the context of design changes (Zwolinski & Gaur 2003). For instance, exploration testing might be used to determine the aerodynamic properties of a vehicle undergoing structural modification.

Exploration testing in the engine case study often involved computational analyses, such as FEA), integrated with low-fidelity prototype testing and measurement data to gain insights into design performance, rather than assessing reliability or durability. Such exploratory tests are essential, as they frequently uncover unexpected phenomena, which can significantly impact the design process and highlight areas that require further investigation. Thus, exploration testing not only provides a deeper understanding of the design space but also challenges design assumptions. Table 2

Table 2. T Testing for Exploration: Definition, Purpose, and Benefits	
Definition	Test to pursue and acquire new knowledge in periods when no theory or, even more fundamentally, no conceptual framework is readily available
Purpose	Acquire new knowledge; reveal new characteristics and phenomena
Potential benefit	Identification of design boundaries and areas of best performance; Generation of new questions, hypotheses and problems

summarises the definition, purposes, and potential benefits of conducting testing for exploration.

6.1.2. *Exploitation*

Exploitation testing is primarily focused on identifying “good enough” configurations within the existing design space, often aimed at optimising current designs rather than exploring new concepts. According to Balachandran *et al.* (2016), these tests strive to achieve a “better” configuration by maximising the utilisation of specific design parameters, thereby enhancing performance or efficiency within known boundaries.

Exploitation testing is valuable in identifying the potential for incremental improvements within the constraints of existing knowledge, technologies and products, offering more certain and immediate benefits compared to exploratory approaches. For industries, exploitation tests can be used to uncover opportunities for structural weight reduction, for instance, by leveraging the existing margins within the structures of a previous design. This process often involves accurate and reliable simulations coupled with prototype testing to confirm the feasibility of these improvements (Degenhardt *et al.* 2006).

Exploration and exploitation learning processes are often depicted as paradoxical due to their divergent objectives and the fundamentally different mindsets and routines they entail (March 1991). Exploration involves searching for novel design space regions characterised by high uncertainty in parameter values, while exploitation focuses on optimising existing configurations to identify the best values within the defined search space (Balachandran *et al.* 2016). Exploration is a learning process that prioritises search, experimentation, playfulness and discovery, fostering innovation and the uncovering of new possibilities, whereas exploitation centres on refining existing ideas, executing them efficiently and maintaining consistency (March 1991; Nemanich, Keller & Vera 2007). Table 3 summarises the definition, purposes, and potential benefits of conducting testing for exploitation.

6.1.3. *Experimentation*

Experimental testing, which is generally guided by theoretical frameworks, systematically assesses whether a concept or experimental prototype meets predefined expectations. Typically, this form of testing is conducted during the R&D stage, within laboratory settings and controlled environments, employing a hypothesis-driven approach.

Table 3. Testing for Exploitation: Definition, Purpose, and Benefits	
Definition	Test to identify additional or improved solutions within a known problem space
Purpose	Obtain better design configuration; maximise utilisation of parameters
Potential benefit	Identifying design performance, narrowing down the design space and identifying optimal or suitable solutions potential for optimisation

Experimentation is a widely recognised method in both science and engineering for testing theories and establishing scientific knowledge. Hall (2011) distinguishes between experimentation in scientific contexts and in design. In scientific experimentation, validity is achieved through peer scrutiny and reproducibility, adhering to a convergent process. Conversely, experimentation in design is characterised by both divergent and convergent phases, typically involving iterative cycles. Experimental design in this context ensures that a trajectory is established along which verification metrics can be defined (Hall 2011).

Erat & Kavadias (2008) argue that the experimentation process encompasses both exploratory and exploitation learning aspects when a series of tests are integrated. During the exploration phase, the emphasis is on acquiring knowledge about the design space, while the subsequent exploitation phase focuses on identifying optimal configurations within that space. This approach aligns with the exploration and exploitation discussion outlined previously. Initially, the focus is on generating knowledge about feasible design alternatives, and later it shifts to refining knowledge about specific candidate solutions. Experimentation is often associated with start-ups. The results of experiments can either confirm the validity of an original concept, allowing continued development in the same direction, or indicate the need for a strategic redesign. However, even for a very mature product like a diesel engine, the company’s R&D department carries out experiments, for example, when looking for new materials that can achieve higher combustion temperatures. Table 4 Summarises the definition, purposes, and potential benefits of conducting testing for Experimentation purposes.

6.1.4. Simulation model building

The purpose of these testing activities is to gather empirical data from physical tests to develop computational simulation models. To construct sophisticated models using techniques such as FEA and CFD for virtual testing, a thorough understanding of the system, along with empirical data from physical experiments, is essential. Testing serves as the primary method for verifying numerical algorithms and ensuring the accuracy of these computational models (Oberkampf & Trucano 2008).

Both case studies indicate that discrepancies between virtual and physical testing measurements can arise from several factors: (a) inaccuracies in the virtual model’s construction, (b) incorrect theoretical assumptions and (c) insufficient calibration or validation of the model using empirical data. Consequently, experimental or physical test data are deemed the most reliable for model development and validation. Incorporating historical test data can help the model-building process both for building the models and for validating these (Sargent 2013). Table 5 summarises

Table 4. Testing for Experimentation: Definition, Purpose, and Benefits	
Definition	Test to establish the true value of a hypothesis
Purpose	Systematically evaluate concepts and prototypes; find optimal designs
Potential benefit	Comparison of performance; understanding of targeted or potential designs; design strategies

Table 5. Testing for Simulation model building: Definition, Purpose, and Benefits

Definition	Test to build and calibrate a computational model
Purpose	Collect experimental test data to build simulation models
Potential benefit	Information of the behaviour of products and components; measured values of model parameters

the definition, purposes, and potential benefits of conducting testing for Simulation model building.

6.2. Testing for uncertainty reduction

Most product development processes, where significant cost commitments are made, commence after the R&D phase is completed, once the company has reasonable confidence in the feasibility of developing a successful product. During this phase, development teams focus on designing the product or service, which involves addressing and resolving numerous uncertainties. The various types of testing conducted during this stage are specifically aimed at reducing these uncertainties.

6.2.1. Revealing information

The purpose of these tests is to reveal information about a proposed design and support the development of a concrete product specification. These tests may assess an object’s properties (e.g., materials or dimensions), identify discrepancies between expected and measured design parameters (Chen *et al.*, 2020), detect known failure or degradation modes or reveal both direct system behaviours and indirect knock-on effects (Ktena 2011). For example, while a diesel engine designer may anticipate improved energy consumption following a design change, testing is needed to quantify the actual gains. Prototype tests can measure key performance characteristics, such as fuel flow, airflow and blow-by, to better understand their interaction within a new engine design. These tests ultimately generate concrete values for specific parameters. Table 6 summarises the definition, purposes, and potential benefits of conducting testing for revealing information.

Table 6. Testing for Revealing information: Definition, Purpose, and Benefits

Definition	Test to ascertain specific properties of an object/value of a parameter
Purpose	Reveal errors and mismatches; detect failure or degradation modes; reveal design behaviour
Potential benefit	Measured values of parameters, understanding of the nature of failure and its consequences

Table 7. Testing for Removing Uncertainty in information: Definition, Purpose, and Benefits

Definition	Test to remove or reduce a specifically identified uncertainty
Purpose	Recognise information adequacy; improve model definition; improve estimates
Potential benefit	Knowledge about the variations in product specification, performance and properties (e.g. materials, dimensions) and models' data accuracy

6.2.2. Removing uncertainty in information

The purpose of these tests is to assess the adequacy of information by addressing epistemic uncertainty. Sources of uncertainty include incomplete or inconsistent information, poor-quality measurements (Mcmanus & Hastings 2006), ambiguities in information description or modelling, omitted details, unclear model scope (Wynn *et al.*, 2011) and the use of approximate or predicted data. These tests improve estimates and verify the precision of model data by reducing uncertainty. Table 7 summarises the definition, purposes, and potential benefits of conducting testing for removing uncertainty in information.

6.2.3. Design evaluation

Testing for design evaluation assesses how well a design meets its intended specifications at a particular stage of the design process. Sim & Duffy (2003) describe design evaluation testing as a process that analyses the feasibility of potential design solutions and reduces the design solution space by eliminating infeasible options. It measures the quality of design solutions against predefined criteria and evaluates prototype performance for specific design configurations. For example, an engine's structural integrity and component functionality may be assessed to determine whether actual power absorption matches expected performance. In complex systems, these tests can also assess performance under unknown or uncertain operational scenarios within system-of-systems environments (Luna *et al.* 2013). Table 8 summarises the definition, purposes, and potential benefits of conducting testing for design evaluation.

6.2.4. Design refinement

The purpose of testing for design refinement is to evaluate “the progress of design performance towards the desired solution and to optimise design features”

Table 8. Testing for Design evaluation: Definition, Purpose, and Benefits

Definition	Test to assess design solution(s) against given sets of criteria
Purpose	Measure the quality of proposed design solutions
Potential benefit	Quality of design solutions, comparison and rationale of design change

(Derrick & Boiten 1999). Camburn *et al.* (2017) describe this testing as instrumental in refining models and simulation results and adjusting specifications to meet the product’s specific use conditions. These tests support the improvement of design performance by guiding design modifications. Design refinement can be achieved through individual or multiple testing efforts (Otto & Wood 2000). For instance, in diesel engine design, tests conducted throughout product development can track improvements in specific parameters, such as successive evaluations to determine whether NOx exhaust emissions have been reduced. Table 9 summarises the definition, purposes, and potential benefits of conducting testing for design refinement.

6.3. Testing for demonstration

The purpose of testing for demonstration is to show that a design or set of design solutions works and fulfils its intended design objectives. These tests provide evidence that a proposed solution can meet its intended use and satisfy stakeholder requirements. As stakeholders may hold different views regarding a technology, its maturity and the extent of its demonstration, concept and system demonstration testing play an important role in supporting the selection of a design among alternative solutions. Following selection, further testing may be required to demonstrate performance and reliability at specific stages of the product development process.

6.3.1. Testing for proof-of-concept

The purpose of testing for proof of concept is to establish the feasibility of a novel concept and assess its potential for further development. These tests provide evidence that the concept is practical and capable of progressing to subsequent stages of design. Proof-of-concept testing is typically conducted using a low-fidelity prototype to evaluate the viability of the proposed idea. For example, a small-scale demonstration system may be tested under different conditions to validate feasibility before progressing to more complex system designs. Table 10 summarises the definition, purposes, and potential benefits of conducting testing for proof-of-concept.

6.3.2. Simulation result validation

The purpose of testing for simulation result validation is to assess the accuracy of model predictions by validating simulation results. This purpose differs from

Table 9. Testing for Design refinement: Definition, Purpose, and Benefits	
Definition	Test to investigate the progress of design performance towards the desired solution
Purpose	Assess progress of design towards the desired solution; optimise design features
Potential benefit	Understanding of refined models, specifications and designs; modified design space; trajectory of design change

Table 10. Testing for proof-of-concept: Definition, Purpose, and Benefits

Definition	Test of an instance of a product concept for a specific purpose
Purpose	Prove feasibility of a new concept; establish viability for development and production
Potential benefit	Operating conditions and lab settings for assessment; performance for a chosen concept, market acceptance, functional requirements fulfilment potential

model building, discussed in [Section 6.1.4](#). Sargent (2013) and Moon *et al.* (2020) describe these tests as a means of validating models by comparing model input–output relationships with those of the real system. These tests establish confidence in simulation results and verify the extent to which model predictions represent real-world behaviour. Physical test results provide a reliable source of data for model validation (Mauery *et al.* 2021). [Table 11](#) summarises the definition, purposes, and potential benefits of conducting testing for simulation result validation.

6.3.3. Design selection

The purpose of testing for design selection is to identify the most suitable design solution from a set of available alternatives. These tests support decision-making by evaluating the performance of competing options. Multiple prototyping and testing activities are often required before committing to a solution (Toche, Pellerin & Fortin 2020). As observed in the engine case study, this approach enables learning from alternative concepts and supports the selection of the most viable option. For example, the impact of material choices on performance can be tested to inform material selection for a specific application. [Table 12](#) summarises the definition, purposes, and potential benefits of conducting testing for design selection.

6.3.4. System demonstration

The purpose of system demonstration testing is to show that a chosen concept or technology can operate successfully under specified operating conditions or environments. These tests provide evidence that the technology can achieve the required performance metrics and support stakeholder confidence in its feasibility. They are

Table 11. Testing for Simulation result validation: Definition, Purpose, and Benefits

Definition	Test to generate data to compare the results of a numerical simulation
Purpose	Assess the accuracy of model prediction; validate simulation results
Potential benefit	Knowledge and understanding of assumptions enabling further model improvement

Table 12. Testing for Design selection: Definition, Purpose, and Benefits

Definition	Test to establish which of a given set of design alternatives performs best against a set of criteria
Purpose	Choose a feasible design from alternatives
Potential benefit	Knowledge and information of performance of alternatives; criteria used in selection knowledge of object and design selected

particularly valuable when organisations have limited knowledge of a new technology that offers potential advantages over established alternatives. The engine case study highlights that system and concept demonstration tests focus on proving that the technology can deliver the required level of performance. Other demonstration testing, such as reliability demonstration testing, supports the achievement of acceptable system reliability, the estimation of lifecycle costs and the planning of operations and maintenance strategies Yadav *et al.*, 2006. These tests can also identify design weaknesses and inform system improvements (Johanning *et al.* 2018). Table 13 summarises the definition, purposes, and potential benefits of conducting testing for system demonstration.

6.4. Testing for compliance

The purposes of compliance testing are to support verification, validation, quality assurance and certification by confirming that the product meets technical specifications, fulfils customer requirements and complies with certification standards. As the design progresses through manufacturing and towards service introduction, these tests provide evidence of its readiness for deployment and market introduction.

6.4.1. Design verification

The purpose of verification testing is to confirm that a design’s actual performance aligns with its expected behaviour and that design outputs meet specified requirements. Verification testing is a critical component of systems verification (Hoppe *et al.*, 2007; Engel 2010; Shabi *et al.*, 2017). Engel (2010) describes testing as a method for verifying specifications and requirements through established test

Table 13. Testing for System Demonstration: Definition, Purpose, and Benefits

Definition	Test to demonstrate the performance characteristics (functionality, reliability and/or reliability) of a system
Purpose	Demonstrate that the chosen concept/technology will deliver required performance
Potential benefit	Understanding of the extent to which technology delivers required performance reliably

Table 14. Testing for Design Verification: Definition, Purpose, and Benefits	
Definition	Test if specified requirements have been fulfilled
Purpose	Verify actual behaviour against expected design requirements
Potential benefit	Assessment of quality of performance; design produced according to specifications; knowledge of deviations from requirements and their rationale

procedures conducted under specified environmental conditions, with compliance confirmed through analysis of the resulting test data. These tests provide evidence that a product has been constructed in accordance with its design specifications and requirements. Wheatcraft (2012) further argues that verification through testing provides the highest level of confidence that a system meets its requirements. Table 14 summarises the definition, purposes, and potential benefits of conducting testing for design verification.

6.4.2. Product validation

The purpose of validation testing is to demonstrate that a product meets the specified application and intended-use requirements of the customer (Hoppe *et al.*, 2007; Engel 2010; Shabi *et al.*, 2017). These tests provide evidence that the product performs as expected within its intended operational context before entering production. Validation testing also supports the assessment of product performance under real or representative operating conditions. In addition, these tests help address stochastic uncertainties arising from inherent variations in physical systems or environments, such as dimensional deviations in machined components. Table 15 summarises the definition, purposes, and potential benefits of conducting testing for product validation.

6.4.3. Qualification

The purpose of qualification testing is to assess that the design complies with established standards and is suitable for its intended application. They also confirm adherence to standardised evaluation procedures prior to full-scale production. For example, lubricant qualification testing uses standardised engine dynamometer tests to evaluate crankcase lubricant performance and verify compliance with

Table 15. Testing for Product Validation: Definition, Purpose, and Benefits	
Definition	Test to find out if a product can fulfil the requirements for the specified application or intended use
Purpose	Check that the product fulfils customer requirements, operating conditions and specifications
Potential benefit	Knowledge of compliance with customer and stakeholder specifications; knowledge that appropriate design processes have been followed

Table 16. Testing for Qualification: Definition, Purpose, and Benefits	
Definition	Test to demonstrate that qualification standards are met
Purpose	To qualify an advanced prototype for functionality and safety
Potential benefit	Understanding of compliance of design/technology/product/process with standards

specifications established by organisations such as the American Petroleum Institute. Testing procedures must also comply with American Society for Testing and Materials guidelines in this case to ensure accuracy and reliability. Table 16 summarises the definition, purposes, and potential benefits of conducting testing for qualification.

6.4.4. Certification

The purpose of certification testing is to assess whether a final product complies with regulatory requirements for specific use conditions. These tests provide evidence that the product meets the criteria necessary for accreditation and market deployment. Certification testing also confirms that products have been developed and validated in accordance with prescribed regulations. Certification tests follow legislatively established test cycles that specify the characteristics of the measurement system, including the required configuration, devices and procedures for accurate assessment (Arregle *et al.* 2006). For example, to demonstrate in-use compliance, a manufacturer may be required to show that an engine meets specified particulate emission limits throughout its operational life. Similarly, heavy-duty diesel engines intended for use in the United States must undergo certification testing to confirm compliance with Environmental Protection Agency exhaust emission standards. Table 17 summarises the definition, purposes, and potential benefits of conducting testing for certification.

6.4.5. Quality assurance

The purpose of quality assurance testing is to establish that a product in production meets quality standards under specific conditions at various phases of the development cycle. This process is integral to the product DDP. Assurance testing helps identify and address emerging issues by assessing risks through both simulation

Table 17. Testing for Certification: Definition, Purpose, and Benefits	
Definition	Test to investigate that certification criteria are met within a predefined parameter range
Purpose	Demonstrate that the final product complies with regulations for its use conditions
Potential benefit	Knowledge of product compliance and regulatory acceptance

Table 18. Testing for Quality assurance: Definition, Purpose, and Benefits	
Definition	Test to assure that a set of given quality criteria are met
Purpose	Ensuring product quality; reducing risks of failure in design and production
Potential benefit	Understanding of product performance and process alignment with quality standards and regulations

and testing, thereby aiming to minimise the potential for failure in both design and production (Bieda 1998). For instance, samples of diesel engines are subjected to dynamometer testing as part of a quality audit process to evaluate the consistency and reliability of the production process. Table 18 summarises the definition, purposes, and potential benefits of conducting testing for quality assurance.

6.5. Summary

The purpose of testing changes depends on the specificity required and the stage of the design process. Figure 7 illustrates how the purposes of testing evolve in relation to two key elements: progress and specificity in design. The x-axis represents the progression of the design process over time, and the y-axis indicates increasing specificity in design, ranging from general problem exploration to a final product instance. However, depending on the product type, whether incremental or radical, testing for the same purposes can occur at different stages due to the varying levels

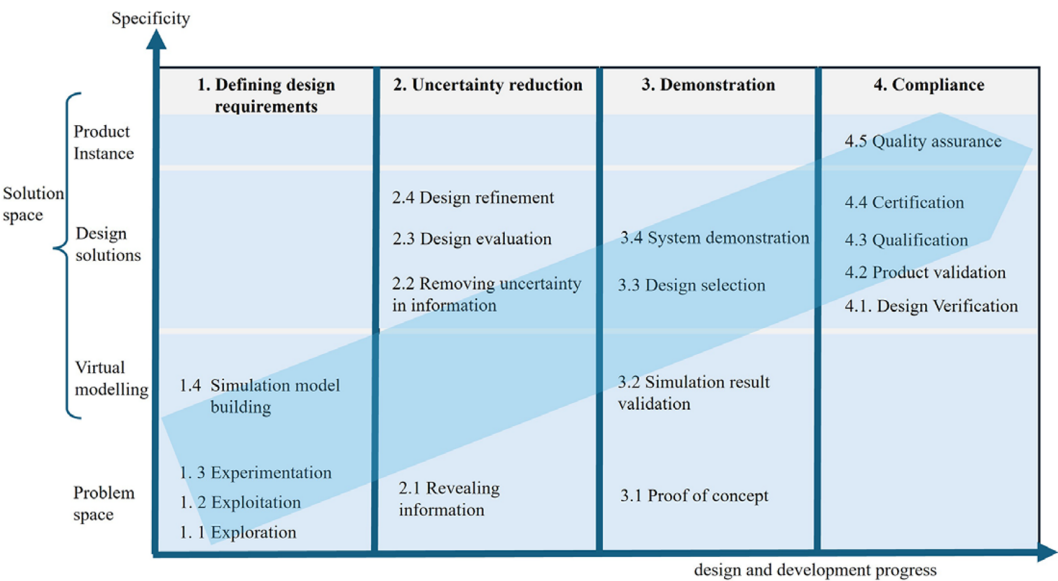


Figure 7. Illustration of the various testing purposes in product design and development, showing increased specificity from the bottom-left to the top-right as the design progresses.

of specificity required for the test subject, be this a design concept or a principal solution. For instance, “proof of concept” testing of novel technologies may be conducted during the R&D phase in a lab setting to establish the feasibility of the concept, but subsequent proof-of-concept tests may also be required during the system development stages in a more mature prototype level to integrate and validate the technology within the broader product context. In Figure 7, the arrow pointing from the bottom left to the top right emphasises the increasing specificity and evolving testing purposes in design and development.

The purpose of testing would evolve as the design process shifts between exploring, framing and defining the problem space, transitioning to the virtual modelling space and then developing alternative design solutions, ultimately leading to more solution-based product instances, with a constant shift between problem and solution spaces as the design evolves.

7. Discussion and conclusion

This research aimed to “clarify the purposes of testing in design and development to bridge gaps of understanding among diverse stakeholders.” To address this, first present insights from two empirical case studies and literature reviews to better understand the evolving nature of testing in product design and development, taking into account the different stakeholders involved (in Section 5). Second, a proposed categorisation framework for systematically recognising the different purposes of testing that can happen in the DDP (in Section 6). Although earlier design literature has explored testing for experimentation (Thomke 2003), learning (Erat & Kavadias 2008) and V&V (Shabi *et al.*, 2017), these discussions often do not clearly connect to the wider context of the design process. This framework fills that gap by offering consistent terminology that links and integrates various strands of academic literature on testing.

Industry focus remains heavily on compliance and validation tests. Although there is an increasing emphasis on integrating virtual testing (e.g., simulation models) and early, low-fidelity prototype testing, the focus remains on reducing the cost of validation and conformance testing (as observed in both case studies). However, it is increasingly realised that strategically designed early-stage tests linked with later test design can substantially accelerate product development (Peña Häufler *et al.* 2021), particularly those aimed at identifying new failure modes. Learning from early experimental testing activities can decrease the overall effort required for acceptance and qualification tests before commercial deployment. The findings suggest engineers recognise that a test can fulfil multiple purposes for different stakeholders; therefore, organising testing activities around core purposes, understanding what is learnt, what can be learned from tests and identifying the stakeholders who can benefit from this knowledge provides systemic learning (Verganti 1997) and can maximise the value of testing in product design. Although managers and engineers understand the needs, convincing regulators and government stakeholders to invest in early testing and adapt to the evolving testing requirements of advanced technologies, such as those in the space sector, remains a major challenge.

This research aimed to clarify the purposes of testing in product design and development, rather than exhaustively categorising test types or analysing test outcomes. While testing is used to evaluate product characteristics (e.g., durability,

usability, reliability, safety), the focus here was on why testing occurs at each stage of the design process. As mentioned earlier, a particular test, depending on the context, can serve different purposes such as detecting design flaws early, verifying functional performance or fulfilling regulatory requirements. The value of focusing exclusively on purposes lies in uncovering the underlying intentions of different stakeholders that influence testing decisions. Clarifying purposes with a categorisation framework helps identify gaps in situations where testing is conducted as a legacy process or overly focuses on technical risk mitigation and compliance, and where critical purposes, such as learning, experimentation or long-term usability, may be overlooked. Examining the purposes of testing through this framework will provide systemic insights into how testing activities are carried out in PDD, supporting systemic decision-making.

While not every stakeholder or team needs to understand every aspect of testing, a shared framework helps ensure testing is integrated, purposeful and cumulative rather than siloed and fragmented. Our proposed categorisation framework is intended for those responsible for managing, coordinating and integrating testing activities throughout the product design and development, including senior business managers, product engineers, CAE engineers and V&V managers. It also benefits individual teams by making their testing activities more visible to other stakeholders and aligning these with the overall project testing plan.

We acknowledge some limitations in our study. One challenge was conducting a structured search on the topic, as the term “test” is used in many different contexts, such as hypothesis testing and statistical tests, which led to irrelevant articles. We manually sorted and excluded papers, which may have led to the exclusion of key papers. While we conducted a thorough literature review and consulted with experts, because of the evolving nature of testing activities, the categorisation of testing purposes presented in this paper may not be complete. New categories or purposes may emerge in future research. Additionally, our case studies primarily focus on test-intensive industries, such as automotive and space, which may not be fully representative of other sectors, such as household appliances, oil and gas or nuclear power, where iterative prototyping and testing may not be as necessary or feasible. Stakeholder involvement in testing also may vary across industries, depending on product size, type and complexity. Future research could explore these variations and assess the effectiveness of different test strategies, including the balance between virtual and physical testing. Further research is also needed to explore the management of multistakeholder involvement in testing to optimise learning outcomes. Key questions for future inquiry include: Who is responsible for recognising and managing multistakeholder contributions to testing? How should testing information be planned, retrieved and transferred to meet various stakeholders’ needs? What are the cost, time and complexity implications of managing these multistakeholder needs in testing? Addressing these questions will be essential for advancing our understanding of how to maximise the value of testing in the development of radical new products and for innovation.

To conclude, this research demonstrates that testing is no longer merely a compliance-driven or cost-management activity but a critical activity that runs alongside product design to support learning, experimentation and evaluation, enabling informed decision-making. Looking forward, as industries transition towards a circular economy model, the importance of testing would be amplified, highlighting its essential role across all stages of the product lifecycle, including development, production, reuse, repurposing and remanufacturing. This

transition necessitates a coherent, shared understanding of testing among supply chain stakeholders to ensure effective management and strategic deployment. Without a clear understanding of the value of testing, there is a risk that testing activities may be perceived merely as a cost factor rather than as a critical enabler of systemic learning and value-adding activity. This misconception could lead to a reduction in testing efforts across the supply chain. We aim that our research findings, along with the proposed categorisation framework, help identify the purposes of testing and contribute to an improved understanding of this area.

Acknowledgements

The authors acknowledge the use of Grammarly [v1.2.208.1781] and Copilot for grammar, spelling and punctuation checks and for polishing the English of the entire manuscript.

References

- Andreussi, T., Ferrato, E. & Giannetti, V. 2022 A review of air-breathing electric propulsion: from mission studies to technology verification. *Journal of Electric Propulsion* 1 (1), 31; doi:[10.1007/s44205-022-00024-9](https://doi.org/10.1007/s44205-022-00024-9).
- Arntzen, S., Wilcox, Z., Lee, N., Hadfield, C. & Rae, J. 2019 *Testing Innovation in the Real World*. Nesta.
- Arregle, J., Bermudez, V., Serrano, J. & Fuentes, E. 2006 Procedure for engine transient cycle emissions testing in real time. *Experimental Thermal and Fluid Science* 30 (5), 485–496; doi:[10.1016/j.expthermflusci.2005.10.002](https://doi.org/10.1016/j.expthermflusci.2005.10.002).
- Atkinson, R., Crawford, L. & Ward, S. 2006 Fundamental uncertainties in projects and the scope of project management. *International Journal of Project Management* 24 (8), 687–698; doi:[10.1016/j.ijproman.2006.09.011](https://doi.org/10.1016/j.ijproman.2006.09.011).
- Bachorek, A. & Jung, J. 2023 Establishing virtual test-driven development environments in the automotive domain: a continuous engineering approach. In *2023 IEEE/ACM 11th International Workshop on Software Engineering for Systems-of-Systems and Software Ecosystems (SESOS)*, pp. 54–57. IEEE.
- Balachandran, P. V., Xue, D., Theiler, J., Hogden, J. & Lookman, T. 2016 Adaptive strategies for materials design using uncertainties. *Scientific Reports* 6, 1–9.
- Bi, Z. M., Mikkola, A., Ip, A. W. H., Yung, K. L. & Luo, C. M. 2024 Virtual verification and validation to enhance sustainability of manufacturing systems. *IEEE Transactions on Automation Science and Engineering*. <https://doi.org/10.1109/tase.2024.3370053>
- Bieda, J. 1998 Product assurance structure and management. *Quality Progress* 31, 71.
- Borgue, O., Panarotto, M. & Isaksson, O. 2022 Fuzzy model-based design for testing and qualification of additive manufacturing components. *Design Science* 8, e11; doi:[10.1017/dsj.2022.6](https://doi.org/10.1017/dsj.2022.6).
- Brophy, J. R. & Polk, J. E. 2018 *Life Qualification of Hall Thrusters by Analysis and Test*. Jet Propulsion Laboratory, National Aeronautics and Space Administration.
- Broyan, J., Gatens, R., Schneider, W., Shaw, L., Mckinley, M., Ewert, M., Meyer, M., Ruff, G., Owens, A. & Meyer, C. 2021 NASA environmental control and life support technology development for exploration: 2020 to 2021 overview. In *50th International Conference on Environmental Systems*.
- Buede, D. M., & Miller, W. D. 2024 *The Engineering Design of Systems: Models and Methods* (4th ed.). John Wiley & Sons. <https://doi.org/10.1002/9781119984030>

- Camburn, B., Viswanathan, V., Linsey, J., Anderson, D., Jensen, D., Crawford, R., Otto, K. & Wood, K. 2017 Design prototyping methods: state of the art in strategies, techniques, and guidelines. *Design Science* 3; doi:[10.1017/dsj.2017.10](https://doi.org/10.1017/dsj.2017.10).
- Canny, A., Bouzekri, E., Martinie, C. & Palanque, P. 2019 On the importance of supporting multiple stakeholders points of view for the testing of interactive systems. In *1st Workshop on Research and Practice Challenges for Engineering Interactive Systems while Integrating Multiple Stakeholders Viewpoints (EISM 2019)*. CEUR-WS: Workshop Proceedings, pp. 113–121.
- Chen, T. Y., Cheung, S. C. & Yiu, S. M. 2020 Metamorphic testing: a new approach for generating next test cases. *arXiv preprint arXiv:2002.12543*.
- Ciuffo, B., Maratta, A., Tutuianu, M., Anagnostopoulos, K., Fontaras, G., Pavlovic, J., Serra, S., Tsiakmakis, S. & Zacharof, N. 2015 Development of the worldwide harmonized test procedure for light-duty vehicles: Pathway for implementation in European Union legislation. *Transportation Research Record* 2503 (1), 110–118; doi:[10.3141/2503-12](https://doi.org/10.3141/2503-12).
- Cooper, R. G. 1990 Stage-gate systems: a new tool for managing new products. *Business Horizons* 33 (3), 44–54; doi:[10.1016/0007-6813\(90\)90040-I](https://doi.org/10.1016/0007-6813(90)90040-I).
- Cooper, R. G. 2024 Deploying agile for physical-product development: big challenges and effective solutions. *IEEE Engineering Management Review* 52, 15.
- Cooper, R. G. & Sommer, A. F. 2016 The agile–stage-gate hybrid model: a promising new approach and a new research opportunity. *Journal of Product Innovation Management* 33 (5), 513–526; doi:[10.1111/jpim.12314](https://doi.org/10.1111/jpim.12314).
- Cornford, S. L. 2001 Introduction to testing and test effectiveness for reliability assessment. In *Product Integrity and Reliability in Design*, pp. 253–287. Springer.
- Degenhardt, R., Rolfes, R., Zimmermann, R. & Rohwer, K. 2006 COCOMAT – improved material exploitation of composite airframe structures by accurate simulation of post-buckling and collapse. *Composite Structures* 73 (2), 175–178; doi:[10.1016/j.comp-struct.2005.11.042](https://doi.org/10.1016/j.comp-struct.2005.11.042).
- Del Rosario, Z., Fenrich, R.W. & Iaccarino, G. 2019 Margin as model: some answers to “How Many Tests Should I Perform?”. In *AIAA Aviation 2019 Forum*, p. 3554.
- Derrick, J. & Boiten, E. 1999 Testing refinements of state-based formal specifications. *Software Testing, Verification and Reliability* 9 (1), 27–50; doi:[10.1002/\(SICI\)1099-1689\(199903\)9:1<27::AID-STVR172>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1099-1689(199903)9:1<27::AID-STVR172>3.0.CO;2-B).
- Dordolova, C. & Törlind, P. 2020 Evaluating design uncertainties in additive manufacturing using design artefacts: examples from space industry. *Design Science* 6, e12; doi:[10.1017/dsj.2020.11](https://doi.org/10.1017/dsj.2020.11).
- Dorst, K. 2011 The core of ‘design thinking’ and its application. *Design Studies* 32 (6), 521–532; doi:[10.1016/j.destud.2011.07.006](https://doi.org/10.1016/j.destud.2011.07.006).
- Dorst, K. 2019 Co-evolution and emergence in design. *Design Studies* 65, 60–77; doi:[10.1016/j.destud.2019.10.005](https://doi.org/10.1016/j.destud.2019.10.005).
- Dorst, K. & Cross, N. 2001 Creativity in the design process: co-evolution of problem-solution. *Design Studies* 22 (5), 425–437; doi:[10.1016/S0142-694X\(01\)00009-6](https://doi.org/10.1016/S0142-694X(01)00009-6).
- Eckert, C. & Isaksson, O. 2026 Coevolution of design research and industrial development: empirical insights from European manufacturing companies. *Journal of Engineering Design* 37 (2), 428–457.
- Engel, A. 2010 *Verification, Validation, and Testing of Engineered Systems*. Wiley; doi:[10.1002/9780470618851](https://doi.org/10.1002/9780470618851).
- Erat, S. & Kavadias, S. 2008 Sequential testing of product designs: implications for learning. *Management Science* 54 (5), 956–968; doi:[10.1287/mnsc.1070.0784](https://doi.org/10.1287/mnsc.1070.0784).

- Ferguson, S., Lai, K., Chen, J., Faidi, S., Leonardo, K. & Olechowski, A. 2022 Why couldn't we do this more often?: Exploring the feasibility of virtual and distributed work in product design engineering. *Research in Engineering Design* 33 (4), 413–436; doi:[10.1007/s00163-022-00391-2](https://doi.org/10.1007/s00163-022-00391-2).
- Forsberg, K., Mooz, K. 1994 *The Relationship of System Engineering to the Project Cycle*, Vol. 9. NCOSE.
- Halfpenny, A., Chabod, A., Thumati, B. & Miu, T. 2024 Achieving high confidence in fatigue reliability by quantifying the effects of uncertainty. *Procedia Structural Integrity* 57, 718–730; doi:[10.1016/j.prostr.2024.03.078](https://doi.org/10.1016/j.prostr.2024.03.078).
- Hall, A. 2011 Experimental design: design experimentation. *Design Issues* 27 (2), 17–26; doi:[10.1162/DESI_a_00074-Hall](https://doi.org/10.1162/DESI_a_00074-Hall).
- Hoppe, M., Engel, A. & Shachar, S. 2007 SysTest: improving the verification, validation, and testing process – assessing six industrial pilot projects. *Systems Engineering* 10 (4), 323–347; doi:[10.1002/sys.20082](https://doi.org/10.1002/sys.20082).
- IEEE 1990 *IEEE Standard Glossary of Software Engineering Terminology*, p. 121990. IEEE.
- Johanning, L., Weller, S. D., Thies, P. R., Holmes, B. & Griffiths, J. 2018 Device design. In Wave and Tidal Energy (eds D. Greaves and G. Iglesias). doi:[10.1002/9781119014492.ch5](https://doi.org/10.1002/9781119014492.ch5).
- Kennedy, M. 2008 *Ready, Set, Dominate: Implement Toyota's Set-based Learning for Developing Products and Nobody Can Catch You*. Oaklea Press.
- Kirsilä, J., Hellström, M. & Wikström, K. 2007 Integration as a project management concept: a study of the commissioning process in industrial deliveries. *International Journal of Project Management* 25 (7), 714–721; doi:[10.1016/j.ijproman.2007.02.005](https://doi.org/10.1016/j.ijproman.2007.02.005).
- Ktena, A. 2011 Physics and modeling of magnetic non destructive testing techniques. In *Key Engineering Materials* (Vol. 495), pp. 265–268. Trans Tech Publ; doi:[10.4028/www.scientific.net/KEM.495.265](https://doi.org/10.4028/www.scientific.net/KEM.495.265).
- Kulkarni, A. U., Salado, A. & Wernz, C. 2022 Optimal verification strategies in multi-firm projects. *Systems Engineering* 25 (3), 254–270; doi:[10.1002/sys.21615](https://doi.org/10.1002/sys.21615).
- Leckenby, E., Dawoud, D., Bouvy, J., & Jónsson, P. 2021 The Sandbox Approach and its Potential for Use in Health Technology Assessment: A Literature Review. *Applied health economics and health policy* 19 (6), 857–869; doi:[10.1007/s40258-021-00665-1](https://doi.org/10.1007/s40258-021-00665-1).
- Lindgren, E. & Münch, J. 2016 Raising the odds of success: the current state of experimentation in product development. *Information and Software Technology* 77, 80–91; doi:[10.1016/j.infsof.2016.04.008](https://doi.org/10.1016/j.infsof.2016.04.008).
- Loch, C. H., Terwiesch, C. & Thomke, S. 2001 Parallel and sequential testing of design alternatives. *Management Science* 47 (5), 663–678.
- Luna, S., Lopes, A., Tao, H. Y. S., Zapata, F. & Pineda, R. 2013 Integration, verification, validation, test, and evaluation (IVVT&E) framework for system of systems (SoS). *Procedia Computer Science* 20, 298–305; doi:[10.1016/j.procs.2013.09.276](https://doi.org/10.1016/j.procs.2013.09.276).
- March, J. G. 1991 Exploration and exploitation in organizational learning. *Organization Science* 2 (1), 71–87; doi:[10.1287/orsc.2.1.71](https://doi.org/10.1287/orsc.2.1.71).
- Mauery, T., Alonso, J., Cary, A., Lee, V., Malecki, R., Mavriplis, D., Medic, G., Schaefer, J. & Slotnick, J. 2021 A guide for aircraft certification by analysis. NASA STI Program Report Series.
- Mcmanus, H. & Hastings, D. 2006 A framework for understanding uncertainty and its mitigation and exploitation in complex systems. *IEEE Engineering Management Review* 34 (3), 81–94; doi:[10.1109/EMR.2006.261384](https://doi.org/10.1109/EMR.2006.261384).
- Menold, J., Simpson, T. W. & Jablow, K. W. 2016 The Prototype for X (PFX) framework: assessing the impact of PFX on desirability, feasibility, and viability of end designs. In *International Design Engineering Technical Conferences and Computers and*

- Information in Engineering Conference*, p. V007T06A040. American Society of Mechanical Engineers.
- Miles, M. B.** 1994 *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks.
- Moogk, D. R.** 2012 Minimum viable product and the importance of experimentation in technology startups. *Technology Innovation Management Review* 2, 23–26.
- Moon, M.-Y., Kim, H.-S., Lee, K., Park, B. & Choi, K.** 2020 Uncertainty quantification and statistical model validation for an offshore jacket structure panel given limited test data and simulation model. *Structural and Multidisciplinary Optimization* 61 (6), 2305–2318; doi:[10.1007/s00158-020-02520-8](https://doi.org/10.1007/s00158-020-02520-8).
- Nemanich, L. A., Keller, R. T. & Vera, D.** 2007 Managing the exploration/exploitation paradox in new product development: How top executives define their firm's innovation trajectory. *International Journal of Innovation and Technology Management* 04 (03), 351–374; doi:[10.1142/S0219877007001132](https://doi.org/10.1142/S0219877007001132).
- Oberkampff, W. L. & Trucano, T. G.** 2008 Verification and validation benchmarks. *Nuclear Engineering and Design* 238 (3), 716–743; doi:[10.1016/j.nucengdes.2007.02.032](https://doi.org/10.1016/j.nucengdes.2007.02.032).
- Öhrwall-Rönneback, A. & Isaksson, O.** 2018 Product development challenges for space subsystem manufacturers. In *Proceedings of International Design Conference, DESIGN* (Vol. 4, pp. 1937–1944).
- Olechowski, A. L., Eppinger, S. D., Joglekar, N. & Tomaszek, K.** 2020 Technology readiness levels: Shortcomings and improvement opportunities. *Systems Engineering* 23 (4), 395–408; doi:[10.1002/sys.21533](https://doi.org/10.1002/sys.21533).
- Otto, K. & Wood, K.** 2000 *Product Design: Techniques in Reverse Engineering and New Product Development*. Pearson.
- Peña Häufler, B., Globocnik, D., Landaeta Saldías, P. & Salomo, S.** 2021 Rapid validity testing at the front end of innovation. *Journal of Product Innovation Management* 38 (4), 447–472; doi:[10.1111/jpim.12585](https://doi.org/10.1111/jpim.12585).
- Popov, O., Iatsyshyn, A., Sokolov, D., Dement, M., Neklonskyi, I. & Yelizarov, A.** 2021 Application of virtual and augmented reality at nuclear power plants. In: Zaporozhets, A., Artemchuk, V. (eds) *Systems, Decision and Control in Energy II. Studies in Systems, Decision and Control in Energy II, Studies in Systems, Decision and Control*, vol 346. pp. 243–260. Springer, Cham. doi:[10.1007/978-3-030-69189-9_14](https://doi.org/10.1007/978-3-030-69189-9_14).
- Sargent, R. G.** 1991 *Simulation Model Verification and Validation*. Institute of Electrical and Electronics Engineers (IEEE).
- Sargent, R. G.** 2013 Verification and validation of simulation models. *Journal of Simulation* 7 (1), 12–24; doi:[10.1057/jos.2012.20](https://doi.org/10.1057/jos.2012.20).
- Shabi, J., Reich, Y. & Diamant, R.** 2017 Planning the verification, validation, and testing process: a case study demonstrating a decision support model. *Journal of Engineering Design* 28 (3), 171–204; doi:[10.1080/09544828.2016.1274964](https://doi.org/10.1080/09544828.2016.1274964).
- Shafqat, A., Oehmen, J., Welo, T. & Ringen, G.** 2022 The role of risk mitigation actions in engineering projects: An empirical investigation. *Systems Engineering* 25 (6), 584–608.
- Shankar, P., Summers, J. D. & Phelan, K.** 2017 A verification and validation planning method to address change propagation effects in engineering design and manufacturing. *Concurrent Engineering* 25 (2), 151–162; doi:[10.1177/1063293X16671771](https://doi.org/10.1177/1063293X16671771).
- Sim, S. K. & Duffy, A. H.** 2003 Towards an ontology of generic engineering design activities. *Research in Engineering Design* 14 (4), 200–223; doi:[10.1007/s00163-003-0037-1](https://doi.org/10.1007/s00163-003-0037-1).
- Song, X. M. & Montoya-Weiss, M. M.** 1998 Critical development activities for really new versus incremental products. *Journal of Product Innovation Management* 15 (2), 124–135; doi:[10.1111/1540-5885.1520124](https://doi.org/10.1111/1540-5885.1520124).
- Steinle, F.** 1997 Entering new fields: exploratory uses of experimentation. *Philosophy of Science* 64 (S4), S65–S74; doi:[10.1086/392587](https://doi.org/10.1086/392587).

- Sudol, A. & Mavris, D. N. 2018 Framework for reliability growth and rework projections for launch vehicles during testing. In *2018 IEEE Aerospace Conference*, pp. 1–15. IEEE.
- Taccogna, F. & Garrigues, L. 2019 Latest progress in Hall thrusters plasma modelling. *Reviews of Modern Plasma Physics* 3 (1), 12; doi:[10.1007/s41614-019-0033-1](https://doi.org/10.1007/s41614-019-0033-1).
- Tahera, K. 2014 *The Role of Testing in Engineering Product Development Processes*. Ph.D., The Open University.
- Tahera, K. 2022 Procurement in new product development for physical testing and collaboration with suppliers. In *Annual Logistics Research Network (LRN) Conference Proceedings*, pp. 146–151.
- Tahera, K., Earl, C. & Eckert, C. 2017 A method for improving overlapping of testing and design. *IEEE Transactions on Engineering Management* 64 (2), 179–192; doi:[10.1109/TEM.2017.2654223](https://doi.org/10.1109/TEM.2017.2654223).
- Tahera, K., Jowers, I., Loureiro-Koechlin, C., Eckert, C. & Lockett, H. 2023 Industrial perspectives on the adoption of virtual testing. *Proceedings of the Design Society* 3, 2305–2314; doi:[10.1017/pds.2023.231](https://doi.org/10.1017/pds.2023.231).
- Tahera, K., Wynn, D. C., Earl, C. C. & Eckert, C. M. 2019 Testing in the incremental design and development of complex products. *Research in Engineering Design*, 30 (2), 291–316; doi:[10.1007/s00163-018-0295-6](https://doi.org/10.1007/s00163-018-0295-6).
- Tao, F., Zhang, H., Liu, A. & Nee, A. Y. 2018 Digital twin in industry: state-of-the-art. *IEEE Transactions on Industrial Informatics* 15 (4), 2405–2415.
- Thomke, S. 2001 Enlightened experimentation. The new imperative for innovation. *Harvard Business Review* 79 (2), 66–75.
- Thomke, S. H. 2003 *Experimentation Matters: Unlocking the Potential of New Technologies for Innovation*. Harvard Business Press.
- Thomke, S. 2008 Learning by experimentation: prototyping and testing. In *Learning by Experimentation: Prototyping and Testing. Handbook of New Product Development Management* (ed. C. H. Loch & S. Kavadias), Elsevier; doi:[10.1016/B978-0-7506-8552-8.50018-9](https://doi.org/10.1016/B978-0-7506-8552-8.50018-9).
- Thomke, S. & Bell, D. E. 2001 *Sequential Testing in Product Development*, pp. 308–323. Management Science.
- Toche, B., Pellerin, R. & Fortin, C. 2020 Set-based design: a review and new directions. *Design Science* 6; doi:[10.1017/dsj.2020.16](https://doi.org/10.1017/dsj.2020.16).
- Torraco, R. J. 2005 Writing integrative literature reviews: guidelines and examples. *Human Resource Development Review* 4 (3), 356–367; doi:[10.1177/1534484305278283](https://doi.org/10.1177/1534484305278283).
- Valerdi, R. & Sullivan, B. P. 2023 Engineering systems integration, testing, and validation. In *Handbook of Engineering Systems Design* (ed. A. Maier, J. Oehmen & P. E. Vermaas), pp. 1–31. Springer International Publishing; doi:[10.1007/978-3-030-46054-9_20-2](https://doi.org/10.1007/978-3-030-46054-9_20-2).
- Verganti, R. 1997 Leveraging on systemic learning to manage the early phases of product innovation projects. *R&D Management* 27 (4), 377–392; doi:[10.1111/1467-9310.00072](https://doi.org/10.1111/1467-9310.00072).
- Wheatcraft, L. S. 2012 *Thinking Ahead to Verification and Validation*. Requirements Experts.
- Wynn, D. C. & Clarkson, P. J. 2018 Process models in design and development. *Research in Engineering Design* 29 (2), 161–202; doi:[10.1007/s00163-017-0262-7](https://doi.org/10.1007/s00163-017-0262-7).
- Wynn, D. C., Grebici, K. & Clarkson, P. J. 2011 Modelling the evolution of uncertainty levels during design. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 5 (3), 187–202; doi:[10.1007/s12008-011-0131-y](https://doi.org/10.1007/s12008-011-0131-y).
- Yadav, O. P., Singh, N. & Goel, P. S. 2006 Reliability demonstration test planning: a three dimensional consideration. *Reliability Engineering & System Safety* 91 (8), 882–893; doi:[10.1016/j.res.2005.09.001](https://doi.org/10.1016/j.res.2005.09.001).

- Yassine, A. & Souweid, S.** 2021 Time-to-market and product performance tradeoff revisited. *IEEE Transactions on Engineering Management* **70** (9), 3244–3259.
- Zammit, J., Gao, J., Evans, R. & Maropoulos, P.** 2018 A knowledge capturing and sharing framework for improving the testing processes in global product development using storytelling and video sharing. *Proceedings of the Institution of Mechanical Engineers Part B-Journal of Engineering Manufacture* **232** (13), 2286–2296; doi: [10.1177/0954405417694062](https://doi.org/10.1177/0954405417694062).
- Zwolinski, M. & Gaur, M. S.** 2003 Integrating testability with design space exploration. *Microelectronics Reliability* **43** (5), 685–693; doi: [10.1016/S0026-2714\(03\)00034-9](https://doi.org/10.1016/S0026-2714(03)00034-9).

Appendix

Appendix A1.			
	Definition	Purpose	Potential benefit
Testing for defining design requirements	Exploration	Test to pursue and acquisition of new knowledge in periods when no theory or even more fundamentally no conceptual framework is readily available	Identification of design boundaries and areas of best performance; Generation of new questions, hypotheses and problems
	Exploitation	Test to identify additional or improved solutions within a known problem space	Identifying design performance; narrowing down the design space and identifying optimal or suitable solutions potential for optimisation;
	Experimentation	Test to identify additional or improved solutions within a known problem space	Comparison of performance, understanding of targeted or potential designs; design strategies
	Simulation	Test to build and calibrate a computational model	Information of the behaviour of products and components; measured values of model parameters
Testing for uncertainty reduction	Revealing information	Test to ascertain specific properties of an object/value of a parameter	Measured values of parameters, understanding of the nature of failure and its consequences;
	Removing uncertainty in information	Test to remove or reduce a specifically identified uncertainty	Knowledge about the variations in product specification, performance and properties (e.g. materials, dimensions), and models' data accuracy
Continued			

Appendix A1. Continued			
Definition		Purpose	Potential benefit
Design evaluation	Test to assess design solution(s) against given sets of criteria	Measure the quality of proposed design solutions	Quality of design solutions, comparison and rationale of design change
	Test to investigate the progress of design performance towards the desired solution	Assess progress of design towards the desired solution; optimise design features	Understanding of refined models, specifications and designs; modified design space; trajectory of design change
Testing for demonstration	Testing for proof-of-concept	Test of an instance of a product concept for a specific purpose	Operating conditions and lab settings for assessment; performance for a chosen concept, market acceptance, functional requirements fulfilment potential
	Simulation result validation	Test to generate data to compare the results of a numerical simulation	Knowledge and understanding of assumptions enabling further model improvement
	Design selection	Test to establish which of a given set of design alternatives that performs best against a set of criteria	Knowledge and information of performance of alternatives; criteria used in selection knowledge of object and design selected
	System demonstration	Test to demonstrate the performance characteristics (functionality, reliability and/or reliability) of a system	Understanding of the extent to which technology delivers required performance reliably
Testing for compliance	Design verification	Test if specified requirements have been fulfilled	Assessment of quality of performance; design produced according to specifications; knowledge of deviations from requirements and their rationale

Continued

Appendix A1. Continued			
	Definition	Purpose	Potential benefit
Product validation	Test to find out if a product can fulfil the requirements for the specified application or intended use	Check the product fulfils customer requirements, operating conditions and specifications	Knowledge of compliance with customer and stakeholder specifications; knowledge that appropriate design processes have been followed
Qualification	Test to demonstrate that qualification standards are met	To qualify an advanced prototype for functionality and safety	Understanding of compliance of design/technology/product/process with standards
Certification	Test to investigate that certification criteria are met within a predefined parameter range	Demonstrate that the final product complies with regulations for its use conditions	Knowledge of product compliance and regulatory acceptance
Quality assurance	Test to assure that a set of given quality criteria are met	Ensuring product quality; reducing risks of failure in design and production	Understanding of product performance and process alignment with quality standards and regulations