



AARHUS UNIVERSITY



Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Laursen, M., Svejvig, P. & Rode, A. L. G. (2017). *Four approaches to project evaluation*. Paper presented at the 24th Nordic Academy of Management Conference, Bodø.

Publication metadata

Title: *Four approaches to project evaluation*. Paper presented at the 24th Nordic Academy of Management Conference, Bodø.

Author(s): Laursen, M., Svejvig, P. & Rode, A. L. G.

Document version: Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Four Approaches to Project Evaluation

Markus Laursen, Per Svejvig, Anna Le Gerstrøm Rode

Aarhus University, Aarhus BSS, Department of Management

Abstract

There are many theoretical and practical reasons for evaluating projects – including explorative arguments focusing on expanding descriptive knowledge on project work as well as normative arguments focusing on improving prescriptive models of project performance. Despite the need for project management methodologies that work and combat project failure, and research methods that can assess effective project management and methodologies, as well as empirical research on the actuality of projects as practice, evaluation research on projects including project management and methodologies is scarce.

Each of the framework's four approaches provides a distinct evaluation that sheds light on some issues while leaving others unattended. Following these lines, the paper calls for more multi-faceted project evaluations. Introducing a framework that can help analyze existing evaluations and structure upcoming evaluations by highlighting beneficial aspects and/or revealing hidden issues, the aim of this paper is to contribute to the theoretical and practical field of project management.

The paper contributes to project theory and practice by inspiring project researchers and aiding project workers in their efforts to open up the black box of projects and deliver relevant and valuable results.

1 Introduction

There are several reasons for conducting evaluations, but in general, evaluations can be done to exercise control and enable learning and enlightenment - for strategic, tactical, symbolic or constitutive reasons (Dahler-Larsen, 2013, pp. 208-212). More specifically, project evaluation is a relevant endeavor for a number of reasons (Svejvig & Hedegaard, 2016). One reason is explorative and aiming at developing a better understanding of projects and project management. Another reason is directive and aiming at improving or optimizing projects or project management. Project evaluation can be used ex-post in hindsight to document project work after a project is finished, interim to correct, adjust or align project work during a project, and ex-ante in advance to prioritize between alternative projects before one or several projects are started (Harri Laihonon, Linzalone, & Schiuma, 2015).

In general, evaluations can have a summative or formative purpose (Chen, 2015, pp. 7-9). Evaluations can be practically oriented and pertain to managers who wish to keep track of their projects and project performance or theoretically oriented and pertain to scholars who wish to nuance the understanding of projects and project performance. There has long been a lively debate concerning whether one project performs better than another project or is more or less successful: both in theoretical and professional communities such issues are important and enduring themes of discussion (Atkinson, 1999; Davis, 2014; Gemino, Sauer, & Reich, 2010; Pinto & Prescott, 1988; Zwikael & Smyrk, 2012). In short, project evaluation is an interesting and important subject both in theory and in practice.

Despite the large number and variety of arguments for evaluating projects – both for academic and pragmatic reasons - the literature within project management is scattered when it comes to project evaluation. Given the need to learn more about projects and improve upon project work, there is a need for addressing this topic specifically and not as a part of discussing project success that is often informed by some sort of evaluation. There is a gap of knowledge on project evaluation – knowledge that could help structure an evaluation process, and there is a need for research which supports the development of evaluation design.

The lack of advice on how to perform project evaluations makes it difficult to design an evaluation framework, which was the task the authors behind this paper was faced with, when engaging in a large practice-driven research program. As a consequence of the lack of systematic project evaluation methods, the team behind the study set out to develop an evaluation framework. This paper is an outcome of this endeavor and it outlines the resulting product of the designing process in a framework representing four approaches to project evaluation. By conceptualizing and presenting the framework in this paper, we seek to inspire project researchers and workers who wish to evaluate projects. Following these lines, the aim of

this paper is to present an evaluation framework and to illustrate its effectiveness using the action design research study from which it emerged to test its validity. The overall purpose is to aid project workers and scholars who wish to evaluate projects by presenting an artifact that can support project evaluation.

The paper follows the publication schema for a design science research study (Gregor & Hevner, 2013), in which the conceptual evaluation framework is treated as an artifact: a thing with a material existence - an artificially made object like a method and model (Gregor & Hevner, 2013, p. 341). The paper is structured as follows: After the current introduction in which the problem and relevance of project evaluation is presented and the purpose and scope of the developed artifact as a solution is specified, follows a theoretical section outlining research on evaluation in general and on project evaluation in specific, which serves as relevant or justificatory knowledge that informs the development of the artifact. The third section presents the research approach of the action design research study from which the artifact has emerged, and it includes the design process that led to the production of the artifact. The fourth section is the abstract domain, and it presents each of the four approaches to project evaluation that make up the artifact. The fifth section presents the instance domain, showing the application of the artifact, and at the end of the section, a brief evaluation is presented. We conclude the paper with a brief section of concluding remarks.

2 Theoretical Background

2.1 Evaluation Research

Evaluation is a vital word in everyday life often understood as *the action of appraising or valuing [something]* (Oxford English Dictionary). Evaluation research and evaluations are multi-faceted covering such diverse examples as community development projects, educational reform, public policy implementation (Stufflebeam & Shinkfield, 2007) or commercial and industrial corporations evaluating procedures for training and promoting employees (Rossi, Lipsey, & Freeman, 2004). The evaluand is very broadly described by Scriven as “something”, but needs to be defined (Scriven 1991 cited in Dahler-Larsen, 2013, pp. 55-60) and delimited in order to make the evaluation operational.

In this context and paper, we use evaluation more restricted as program evaluation interchangeable with evaluation research (Rossi et al., 2004, p. 2). Program evaluation is defined as follows (Chen, 2015, p. 6): *Program evaluation is the process of systematically gathering empirical data and contextual information about an intervention program – specifically answers what, who, how, whether, and why questions that will assist in assessing a program’s planning, implementation and/or effectiveness.*

Program evaluation has a long history with roots back in the 17th century; although systematic evaluation of programs started prior to World War I related to public health initiatives (Rossi et al., 2004). This was followed on by Lewin's pioneering "action research" studies about minority problems (white and black, Jew and non-Jew) (Lewin, 1946), commercial studies as the Hawthorne studies (Roethlisberger, Dickson, & Wright, 1961 (1939)) to mention a couple of historical studies.

Program evaluation could be classified in many ways where Chen (2015) states the following basic evaluation types: (1) constructive process evaluation, (2) conclusive process evaluation, (3) constructive outcome evaluation, (4) conclusive outcome evaluation and (5) hybrid evaluations derived from the first four basic types. Constructive or formative means providing information for improving a program while conclusive or summative means judging the overall merit or worth. Process is the stages in a program (e.g. program implementation) while outcome is the impact that the program has on its stakeholders (e.g. client, organization, society etc.). Dahler-Larsen (2013) provides another classification and mentions four types as objectives evaluation, outcome evaluation, process-based outcome evaluation and participatory evaluation where his categories partially overlap with Chen's (2015), although Dahler-Larsen emphasizes that process-based outcome evaluation differs from mainstream evaluation research by relying on constructivism epistemology (Dahler-Larsen, 2013). We will finally mention realistic evaluation based on a realism perspective, which is an epistemology between positivism and constructivism. Realistic evaluation has the formula Context (C) + Mechanism (M) => Outcome (O) where a program triggers a mechanism within a given context that gives a certain outcome. The key is the that the result is context-based (Befani, Ledermann, & Sager, 2007; Pawson & Tilley, 1997).

Evaluation research and program evaluation has progressed as its own discipline with evaluation societies, evaluation journals and beyond (Rossi et al., 2004), but apparently largely disconnected to project studies and project management research although recent research integrates the two disciplines and furthermore verbalizes it as evaluation of projects (Dahler-Larsen, 2013). In the following, we will not distinguish between program or project as this is a definitive question, so program evaluation and project evaluation are seen as part of the same entirety.

2.2 Project Evaluation Theory

Project evaluation is a central element in the literature on project studies and project management (Lenfle, 2012) despite the disconnectedness to evaluation research. We take a systems view for project evaluation as shown below in Figure 1 (Adapted from Andersen, 2010; Chen, 2015; Dahler-Larsen, 2013; Laursen & Svejvig, 2016):

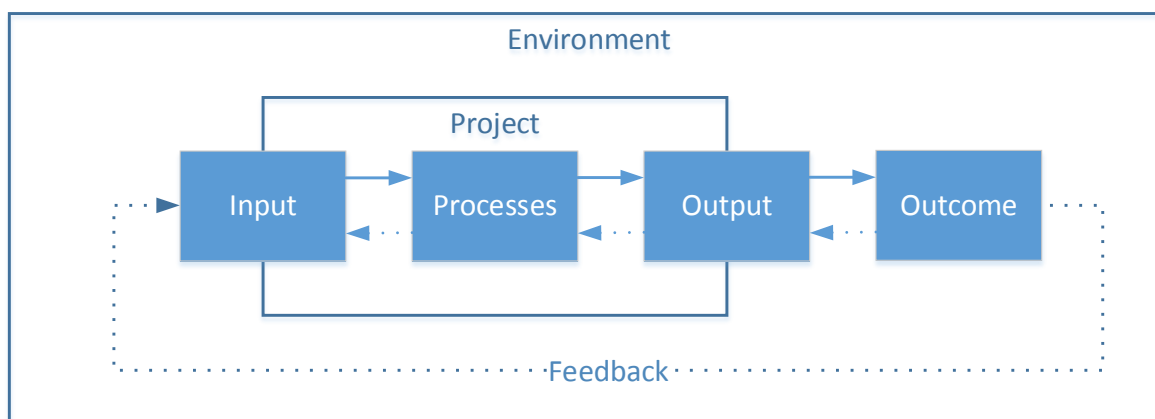


Figure 1: A systems view on project evaluation

Figure 1 shows a simplistic representation of a project as an open system relating to and depending on its environment (Bertalanffy, 1956). Inputs are resources from the environment such as money, technology, facilities and personnel, which are transformed to tangible and/or intangible outputs through project processes. Outcome is the resulting impact on its stakeholders derived from the project's output. The environment is interacting with the project either fostering and/or constraining the project processes influenced by social norms, organizational culture, political structures etc. Feedback mechanisms are shown with dashed lines and indicate how responses from the project and the environment can be used to regulate input, processes, output and outcome (Bertalanffy, 1968; Chen, 2015).

A common pattern in project evaluation is comparing projects where a comparison is the evaluation of two or more projects using the same evaluation criteria - as Swanson (1971, p. 145) puts it "thinking without comparison is unthinkable".

A classic project evaluation and comparison perspective is the iron triangle with the elements cost, time and quality (Atkinson, 1999). This is objectives evaluation (Dahler-Larsen, 2013) related to the output from the project and sometimes labeled success criteria for projects. Atkinson (1999) suggested a square route model to elaborate our understanding of success criteria in projects with dimensions such as benefits for organization and community but still including the iron triangle thereby focusing on both output and outcome in Figure 1 above. In the same vein Shenhar and Dvir (2007, pp. 23-36) propose a multidimensional strategic concept with five success dimensions and as a dynamic concept developing over time. The presentation shows the close relationship between project success criteria and project evaluation as well as the focus on output and/or outcome (objectives evaluation).

This brief presentation of evaluation theory and project evaluation theory forms the basis for the action design research described in next chapter where theories are involved when relevant.

3 Action Design Research Methodology

We frame our research approach as action design research (ADR) which is adapted from the information systems domain. “ADR is a research method for generating prescriptive design knowledge through building and evaluating...artifacts in an organizational setting” (Sein, Henfridsson, Purao, Rossi, & Lindgren, 2011, p. 40). The study has thus both elements of action research (interventions) and design research (building artifacts) (Goldkuhl, 2012).

The study is based on an initiative called Project Half Double, which has the purpose to build up a new and radical project paradigm to increase the competitiveness of the Danish industry. Project Half Double is a cooperation between Implement Consulting Group, manufacturing companies and universities (Svejvig et al., 2016; Svejvig & Grex, 2016). Project Half Double has created the Half Double Methodology (HDM) (methodology artifact). HDM has been tested in seven pilot projects (interventions), and the results have been evaluated using an evaluation model (evaluation artifact). This has taken place from June 2015 to December 2016, where a detailed account of the results are available elsewhere (Svejvig, Rode, & Frederiksen, 2017).

The study can be divided into two parallel cycles (Mathiassen, Chiasson, & Germonprez, 2012). First, a problem-solving cycle where the HDM is used in the seven pilot projects. Second, a research cycle with the purpose to evaluate the problem-solving cycle. The focus in this paper is on the design and evaluation of the evaluation model in the research cycle.

ADR consists of four interleaved stages: (1) problem formulation; (2) building, intervention, and evaluation; (3) reflection and learning; and (4) formalization of learning. ADR also describes seven principles which are shown together with the four stages in Table 1 below outlining the action design research process in this study (inspired by Gregor, Imran, & Turner, 2014):

Table 1: The action design research process divided into problem-solving cycle and research cycle

Stages and principles	Application of stages and principles in Project Half Double (Problem-solving cycle)	Application of stages and principles in the research part of Project Half Double (Research cycle)
<i>Stage 1 Problem formulation</i>		
Principle 1: Practice inspired research	Project Half Double is driven from practice with the overall objective to develop a new and radical project paradigm in order to increase the competitiveness of the Danish industry	The evaluation model is developed and used to evaluate the intervention process especially practices and impact in order to assess the degree to which the HDM is more successful than traditional approaches

Stages and principles	Application of stages and principles in Project Half Double (Problem-solving cycle)	Application of stages and principles in the research part of Project Half Double (Research cycle)
Principle 2: Theory-ingrained artifact	The artifact HDM is derived from lean and agile thinking (Axelos, 2015; Womack & Jones, 2003) and related to the rethinking project management research stream (Svejvig & Andersen, 2015; Winter, Smith, Morris, & Cicmil, 2006)	The evaluation model is based on open systems theory (Andersen, 2010; Chen, 2015), evaluation theory (Pawson & Tilley, 1997; Stufflebeam & Shinkfield, 2007), diamond model for project characteristics (Shenhar & Dvir, 2007) and beyond
Stage 2 Building, intervention, and evaluation		
Principle 3: Reciprocal shaping	The HDM is applied to the pilot projects and experiences from the pilot projects are used to revise and enhance the methodology	The evaluation model were initially developed as an abstract model and subsequently applied and re-shaped according to each pilot project and organization
Principle 4: Mutually influential roles	There is mutual learning between practitioners, consultants and researchers both within organizations and across organizations e.g. through knowledge sharing workshops – this learning process overlaps the problem-solving and research cycle	
Principle 5: Authentic and concurrent evaluation	The evaluation model is used to evaluate the pilot project and compare it with other projects called reference project	The evaluation model is discussed in interviews and workshops as part of the evaluation process
Stage 3: Reflection and learning		
Principle 6: Guided emergence	Guided emergence reflects that the initial design of the artifacts (the HDM and evaluation model) are shaped by their use and the participants who use the them (Sein et al., 2011, p. 44) – this adjusting process overlaps the problem-solving and research cycle	
Stage 4: Formalization of learning		
Principle 7: Generalized outcomes	The HDM as artifact is a generalized outcome which will (and has to) undergo more design cycles to reflect the learning that takes place in Project Half Double	The evaluation models are generalized outcomes that may be applied in other settings

The process outlined in Table 1 did not follow the linear fashion as described in the table above but was an iterative process moving back and forth between the stages as stipulated in the ADR method (Sein et al., 2011), and the two cycles are highly intertwined.

The theorizing related to the evaluation model has taken place in two domains: The abstract domain and the instance domain (Lee, Pries-Heje, & Baskerville, 2011). We started with an abstract problem about how

to evaluate projects using HDM and compare them with projects which have not used HDM. We developed an abstract solution based on open systems theory (Andersen, 2010; Chen, 2015), evaluation theory (Pawson & Tilley, 1997; Stufflebeam & Shinkfield, 2007), the diamond model for project characteristics (Shenhar & Dvir, 2007) and others. The abstract solution was then instantiated in the seven organizations each carrying out a pilot project (instance solution). The pilot projects were carried out at different times, so we were able to learn from organization to organization and thereby improve the evaluation model. Finally, the generalized abstract solution presented in this paper is a further refinement after the fieldwork in the seven organizations has taken place. Overall, this theorizing process could be described as abductive (Saunders, Lewis, & Thornhill, 2016).

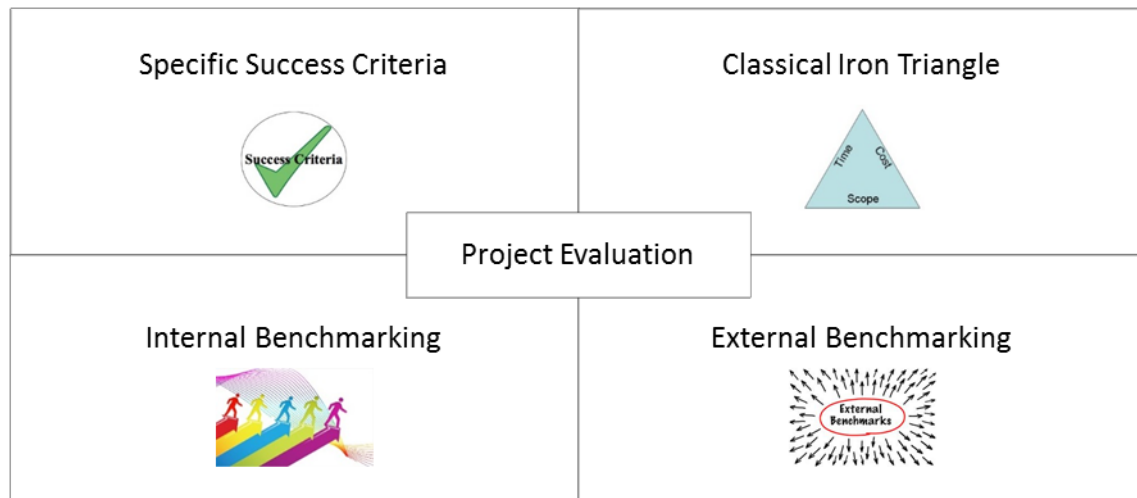
The presentation in this paper is first the generalized abstract solution (abstract domain) in the next chapter, and then the illustration of how the generalized solution is used in a specific example drawing on the Project Half Double (instance domain) in the chapter that follows.

4 Project Evaluation Framework

Based on the research methodology outlined above, the research team designed a conceptual framework to help solve the task of the research cycle and assess the working and value of the HDM – meaning an evaluation framework specifically for projects. In this section we present this conceptual model, meaning the result of the work taking place in the abstract domain, and we close the section by synthesizing what the approach contributes to project evaluation theory.

This project evaluation framework consists of four distinct approaches Specific Success Criteria, Classical Iron Triangle, Internal Benchmarking, and External Benchmarking as illustrated in Figure 2 below. The approaches are ordered according to their project specificity and we describe each of them in the following subsections.

Figure 2: Project evaluation framework



Across the framework, the four approaches share certain characteristics, e.g. the two approaches ‘Classical Iron Triangle’ and ‘Specific Success Criteria’ focus on the single project only. The two benchmarking approaches consider one project relative to other projects, and moreover, the benchmarking approaches do not entail predefined metrics, as with especially the ‘Classical Iron Triangle’ and to some extent the ‘Specific Success Criteria’. For benchmarking, internal and external is relative to the organization where a project is executed.

4.1 Specific Success Criteria

Projects have a degree of uniqueness, which is addressed by having specific success criteria for project evaluation. Thus, this approach primarily relates to the overall project objectives and the overall project success (McLeod, Doolin, & MacDonell, 2012). There can also be criteria relating to the success of the project management or project process. The distinction between project process and the project as a whole is reflected in many frameworks concerning success of projects and project management (Laursen & Svejvig, 2016). The overall objectives are likely to be the measures of effect and value that can only be measured after the project is finished and perhaps a product has been launched. In theory, there is no limit to the number of measures, but one of the challenges is to identify relevant measures for which a causal relationship to the project can be justified – even if causality is difficult to justify in practice. As we have labeled this approach specific, it means that it can and should be tailored to individual project objectives. This approach provides information about effect/impact over time of projects and their outputs, and the criteria take into account that the effect is for given stakeholders (McLeod et al., 2012).

This approach is based on what we may label objectives evaluation (Dahler-Larsen, 2013), where the objectives have been defined at the outset, and there is a natural desire to evaluate if the objectives were

fulfilled. There is an underlying assumption of the objectives being valid at the outset of the project, meaning that it makes sense to evaluate these even though the world has changed. Specific success criteria are often related to the perceived benefits for a stakeholder or group of stakeholders, such as the suggestion by Atkinson (1999) for benefits of information systems projects.

4.2 Classical Iron Triangle

The Classical Iron Triangle approach is a well-established concept to project evaluation for project management practitioners, as it has been applied for many years and still is today (McLeod et al., 2012). The Classical Iron Triangle approach is concerned with the project process or the success of the project management (Atkinson, 1999; McLeod et al., 2012). The measures of success are traditional; time (schedule), cost (budget), and quality (specification) that relate to the time from project initiation to the end of the project. The measure of project management success may also be expressed as the efficiency of the project management in relation to schedule and budget (Shenhar, Dvir, Levy, & Maltz, 2001). Thus, these measures provide an organization with indications of their ability for executing projects in relation to expectations.

The Classical Iron Triangle approach is based on objectives evaluation, and in a technical and rational way of thinking (Svejvig & Andersen, 2015) the project deliveries are defined in ways that the objectives are expected to be fulfilled by the end of the project (Dahler-Larsen, 2013). In this line of thinking, objectives are accepted as valid at the outset and by evaluating the dimensions in the iron triangle, the objectives are evaluated implicitly.

This approach has a limited scope which has been subject to much critique (e.g. Atkinson, 1999) as it has been applied without awareness of its limitations. However, it is not without reason that the approach is still in use; it is generic and simple, making it applicable across project types and methodologies. This approach provides easy-to-understand measures that are operational or they may be proxies for specific success criteria, e.g. a project schedule might be important as it represents the time to market of a product. Thus, an overall dimension is broken down to depend on meeting the scheduled goal.

4.3 Internal Benchmarking

Benchmarking is a concept that is used somewhat differently across sectors, and an academic journal (Benchmarking: An International Journal) is dedicated to studies on benchmarking, which is related to the research domain total quality programs. In this context, many definitions have been suggested over time (Nandi & Banwet, 2000), which have followed benchmarking theory through four evolutionary stages that we may sum up as going toward (1) priority to action, (2) evaluation of process, (3) satisfaction of

customer, (4) evaluation of strategies (Anand & Kodali, 2008; Maire, Bronet, & Pillet, 2005). Based on multiple definitions, Anand and Kodali (2008) suggest that benchmarking may be described as

a continuous analysis of strategies, functions, processes, products or services, performances, etc. compared within or between best-in-class organisations by obtaining information through appropriate data collection method, with the intention of assessing an organisation's current standards and thereby carry out self-improvement by implementing changes to scale or exceed those standards.(p. 259)

This description suggests a conceptualization of benchmarking as more than an approach to evaluation, rather it is described as an initiative to improve performance or quality of business operations. In this paper, we stay closer to the dictionary form where benchmarking is defined as “The action or practice of comparing something to a benchmark; evaluation against an established standard.” (OED, 2017). We can narrow this down even further through the definition of benchmarking within business: “A process in which a business evaluates its own operations (often specific procedures) by detailed comparison with those of another business (esp. a competitor), in order to establish best practice and improve performance; the examination and emulation of other organizations' strengths.” (OED, 2017). Yet, it is stated that the terms may also be applied internally in an organization. The business definition specifies that it is the operations, here projects, that are compared. We adopt an approach mainly related to systematic measurement and learning from projects (Kouzmin, Elke, Helmut, & Korac-Kakabadse, 1999).

The purpose of this approach is to provide an overall assessment of an invention or improvement initiative by benchmarking a project against several other projects. These other projects are labeled reference projects – a label inspired by reference class forecasting (Flyvbjerg, 2006). The project subject to an intervention may become a benchmark for further interventions to project management methodologies in an organization. Internal benchmarking may follow a somewhat similar approach as ex-post evaluations of projects in a project portfolio.

The benchmarks for projects may be derived from measures belonging to both prior outlined approaches; Specific Success Criteria and Classical Iron Triangle. Internal Benchmarking opens especially for Specific Success Criteria as projects within an organization share many traits and the approach is designed to isolate one parameter and measure the effect of changing it. However, benchmarking can be extended to other areas to expand the understanding of the specific context for each project. One way to inform the context systematically could be through the four dimensions of the diamond model (Shenhar & Dvir, 2007); Technology, Novelty, Pace, and Complexity. The complexity dimension may be supported by a framework mapping the project characteristics environment, tasks and processes, and resources and organization

(Fangel, 2010). The four dimensions are to be assessed for both the reference projects and the project subject to intervention. Another model for informing internal benchmarking in project evaluation is the project excellence model developed by the International Project Management Association (IPMA, 2016). In this way, our approach to benchmarking is not only related to performance and success itself, but also includes a desire to understand the basis for the measures.

The internal benchmarking approach suits a variance-based experiment, where the reference projects form the control group for the project subject to an intervention (Dahler-Larsen, 2013). Choosing the reference projects is a balancing act of comparability on especially the dimension similarity and proximity in terms of time.

4.4 External Benchmarking

The fourth approach to evaluation is also concerned with comparisons, but focuses on comparing the impacts of projects across organizations in order to learn from each other across organizations and to understand the influence of context on the outcome of the intervention. Here, our use of benchmarking is rather close to the dictionary's definition of benchmarking in a business context that we presented in the previous approach, as there is an implicit wish to improve current practices by benchmarking against the practices of other organizations. The benchmarking may be conducted by either one of the organizations or by an external team of evaluators such as a research team.

This approach to evaluation has a dual objective of both presenting conclusions on the outcomes of the evaluation and to learn from the evaluation which is labeled Hybrid outcome evaluation (Chen, 2015). Specifically, the approach follows real-world outcome evaluation and transferability evaluation. Real-world outcome evaluation is concerned with phenomena from the real world, covering both constructive and conclusive outcome assessments, and Transferability evaluation concerns the context, and it provides indications whether or not outcomes may be replicated in different contexts (Chen, 2015). Particularly Transferability evaluation is linked to realistic evaluation theory argued for by Pawson and Tilley (1997). Realistic evaluation proposes a model for causality focusing on mechanism and context explaining the outcome, also called the generative causation. In this way, this approach changes the perspective from the control group approach of internal benchmarking that focuses on treatment (mechanism) and control group.

Overall, the model by (Pawson & Tilley, 1997) is formulated as *Outcome=Mechanism+Context*. It adopts logics from the world of physics to the social world, also recognizing that a social program is a social system. As previously, we deem program evaluation suitable for projects too. Mechanism is to be understood as both tangible, such as a clockwork, and intangible in the meaning of ideas and opportunities.

Contexts are the social and cultural conditions that shape if the mechanism will be able to succeed. The authors use the example from the physical world of gunpowder only igniting provided the right conditions - e.g. it not being damp. The focus on mechanism and context is also expressed in the axioms of realism when dealing with change, here more specifically the questions which research has to answer.

Axiom 1: What are the mechanisms for change triggered by a program and how do they counteract the existing social processes? (Pawson & Tilley, 1997, p. 75).

Axiom 2: What are the social and cultural conditions necessary for change mechanisms to operate and how are they distributed within and between program contexts? (Pawson & Tilley, 1997, p. 77).

The external benchmarking approach focuses on the outcome of the same set of ideas across different contexts, thus the mechanism itself is similar, but how it counteracts existing social processes will not be the same across organizations. Thus, the focus of this evaluation approach is less on axiom 1 than axiom 2. It is not explicated in the formulation of axiom 2, yet we consider the nature and type of project to be reflected in the social and cultural conditions.

The means for informing this evaluation approach are overall more universal than the internal benchmarking, and apart from measures of performance, the approach applies assessments of characteristics of each organization. The comparison of outcome measures may be based on figures stemming from other evaluation approaches, in this way there is coherence across evaluation approaches. This evaluation approach does not only allow for cross-organizational comparison, but the universal measures also allow for comparing different types of projects for which the internal benchmarking approach is unsuitable. Moreover, comparing across organizations is likely to open up for more project types than found within one organization.

4.5 Knowledge Contribution of the Evaluation Framework

We have elaborated on each of the four approaches in terms of the purpose it has for an evaluation and the theoretical foundation in evaluation theory which is summarized in Table 2.

Table 2. Four approaches to evaluation

Approach	Purpose of evaluation	Evaluation theory
Specific Success Criteria	<ul style="list-style-type: none"> Project success, being measured against the overall objectives of the project (McLeod et al., 2012) 	<ul style="list-style-type: none"> Objectives evaluation (Dahler-Larsen, 2013)
Classical Iron Triangle	<ul style="list-style-type: none"> Project management success, being measured against the traditional gauges of performance (i.e., time, cost, and quality (Jugdev & Müller, 2005; McLeod et al., 2012) Process success – focus on project management (McLeod et al., 	<ul style="list-style-type: none"> Objectives evaluation (Dahler-Larsen, 2013)

	2012) <ul style="list-style-type: none"> • Project Efficiency meeting schedule and budget (Shenhar et al., 2001) 	
Internal Benchmarking	<ul style="list-style-type: none"> • To provide an overall judgement of an invention or improvement initiative by benchmarking a project against several other projects • Outcome / Impact measurement – impact on customer and business success (Shenhar et al., 2001) • Impact on stakeholders (e.g. client, organization, society) (Chen, 2015) 	<ul style="list-style-type: none"> • Conclusive outcome evaluation (Chen, 2015)
External Benchmarking	<ul style="list-style-type: none"> • Comparing projects in several organizations in order to learn from each other • Context discussion (Dahler-Larsen, 2013, pp.: 149-153) and how the context influences the results 	<ul style="list-style-type: none"> • Hybrid outcome evaluation (Chen, 2015) • Context discussion in realistic evaluation with context, mechanism and outcome) (Pawson & Tilley, 1997)

The contribution of the artifact is two-fold as there is varying maturity across the four approaches to evaluation in Table 2, but presenting an entire conceptual framework for project evaluation is a new perspective. We consider maturity in two dimensions: Solution Maturity and Application Domain Maturity suggested by Gregor and Hevner (2013) as the foundation for assessing a knowledge contribution of action design research. Solution Maturity concerns “current maturity of artifacts” (Gregor and Hevner (2013, p. 345), while the Application Domain Maturity concerns the context. These two dimension help determine the novelty of the artifact.

We deem the two approaches Classical Iron Triangle and Specific Success Criteria to be approaches that are Routine Design, meaning that both types of maturity is high. Our argument is that both approaches mainly provide “known solutions to known problems” (Gregor & Hevner, 2013, p. 345). on the contrary, the two benchmarking approaches are not high maturity for the Application Domain. Rigorous Internal Benchmarking is not commonly adopted for projects, even though it is portfolio management practiced in many organizations providing data on project performance. External Benchmarking is not well described in prior project evaluation literature, as the purpose and specific method is unlike general comparisons such as the Standish Group CHAOS report. We may claim a first application of this type of comparison for projects i.e. an exaptation (Gregor & Hevner, 2013). The learning perspective is known to the project management society, but it is also common knowledge that learning could be emphasized in project evaluation – our evaluation framework addresses this by applying benchmarking theory.

Having presented the artifact in the abstract domain, the next section will present instantiations through one example of each of the four approaches to evaluation. In doing so, we display the applicability of the artifact and provide a foundation for evaluating the evaluation framework.

5 Evaluating the Project Evaluation Framework

The following section outlines how the research team applied the framework to structure the project evaluation process, and in doing so, we illustrate the value of the artifact. Thus, this section concerns the instance domain. In accordance with the recommended structure of ADR (Gregor & Hevner, 2013), we present an evaluation of the artifact at the end of this section. This design evaluation is a product of an ongoing research project, meaning that it may be considered a proof-of-concept, and we focus here on the validity – meaning that the artifact works and does what it is supposed to do (Gregor & Hevner, 2013, p. 351). The evaluation strategy belongs to the naturalistic type, combining ex-ante and ex-post evaluation methods (Venable, Pries-Heje, & Baskerville, 2012). This means that the evaluation took place in a real-world setting and that it was evaluated during and after application.

As presented in the previous section, our framework consists of four approaches to evaluation that have been applied in the evaluation of the projects applying the Half Double Methodology. We present each approach separately and conclude with an overall section to conclude on the usefulness of the artifact. The examples are from what we called pilot projects, i.e. the projects subject to interventions and the reference projects.

5.1 Specific Success Criteria: First Approach to HDM Evaluation

In the first approach, we considered each pilot project in its own right – focusing on the particularities of the pilot project: its vision and mission – the *raison d'être*. From this consideration, specific success criteria pertaining to the pilot project were derived. Toward or upon completion of the pilot project, the research team engaged in a dialog with the project manager of the pilot project. The purpose was to evaluate the performance of the pilot project measured against the complete list of success criteria, which allowed the research team to investigate the success of the pilot project and to find out if the HDM was applicable in each case. Thus, the focus was on learnings. This evaluation approach gave a very deep understanding of one pilot project and a very project specific view on the workings of the HDM. In general, the number and variety of project specific success criteria is unlimited - in the Half Double study the number and variety varies.

A large manufacturing company provides an example of a specific evaluation based on the most comprehensive list of success criteria. The ten criteria range from sales progress and market share to phase duration and time to market as well as key stakeholder satisfaction in process and key stakeholder assessment of product. Evaluating the pilot project in terms of all these success criteria enables a very project relevant evaluation. The specific evaluation approach is shown in Table 3 below from Svejvig et al. (2016, p. 8):

Table 3: Specific success criteria and their fulfillment

SUCCESS CRITERIA		
	Target	Actual / Expected
#1	Obtaining and internal rate of return (Dynes & Aguirre) $\geq 14\%$	To be evaluated after launch of product
#2	Product should replace 90% of current pumps in the same series	To be evaluated after launch of product
#3	Standard unit cost below a certain number with specific technical data	To be evaluated after launch of product
#4	Reduce number of product variants by 50% without increasing number of platforms	To be evaluated after launch of product
#5	Sales doubled within five years and a market share of 20%	To be evaluated after launch of product and ultimately after five years
#6	Shorter time to market for pilot project where the frontloading phase from Gate 2 to Gate 3 is reduced from nine to six months	Current lead time is expected to be nine months although the project was able to finalize the phase in April 2016 – seven months after G2. It was, however, decided from a portfolio management perspective to postpone the project deadline to June 2016
#7	The first three phases of the product development project are done within six months (from development project gate DP0 to DP3 covering idea, pre-study and concept phases)	To be evaluated after gate DP3 is achieved in the product development project
#8	Pulse check shows satisfaction among key stakeholders on 4.4	Average rating differs between 3.5 and 4.0 from October 2015 (4.0) to January 2016 (3.5) to April 2016 (3.9)
#9	Key stakeholders assess that the product from the pilot project has a maturity level to be 4.5 on a scale from 1-5 (as an indicator of quality)	To be measured after completion of mature phase
#10	"Transition Readiness Assessment" (TRA) should reach a target of 90% after mature phase	The pilot project has gone from 63% in the beginning of the mature phase to 87% at the end of mature phase

In summary, the specific approach relies on evaluation criteria that are very near to the project, enabling an idiographic evaluation (Dahler-Larsen, 2013, pp. 63-64) which has the highest relevance to the specific project. It does, however, have its limits as it restricts the broadness of the evaluation, which often becomes narrow and limited to an intra-organizational and even intra-project comparison. This is a challenge when it comes to benchmarking the project in order to gain an understanding of it compared to

other projects – which is necessary for further evaluating the workings of the HDM. We address this challenge in the evaluation approach Classical Iron Triangle presented next.

5.2 Classical Iron Triangle: Second Approach to HDM Evaluation

Second, the scope of evaluation was expanded to consider both the pilot project and three comparable so-called reference projects. Consequently, we needed to expand the evaluation criteria to a range that was relevant across all four projects. The search for less specific evaluation criteria relevant across all four projects is guided by the task of documenting the benefits of the HDM and thereby by the overarching ambition of the methodology: to reduce time and increase impact. Thus, in order to investigate if the HDM can deliver on its promises, focus is on evaluating if the schedule was achieved and if the benefits were realized, but also criteria on budget and scope success were in use. In this way, the second step investigates the success of the pilot and reference projects individually by evaluating their performance – measured against the classical and more general success criteria.

A global pharmaceutical company provides an example of a more general evaluation which is based on two evaluation parameters pertaining to the classical iron triangle criteria and additional universal criteria. In this organization, the classical criteria were operationalized in a simple way; deeming whether the success criteria were fully or partially achieved for all but the quality criterion User satisfaction. This criterion was scored in a simple survey for project teams and optionally also steering committees and review teams. The more general evaluation approach based on classical criteria is shown in Table 4 below from Svejvig et al. (2017, p. 21).

Table 4: Selected criteria and their fulfillment

CRITERIA	PILOT PROJECT	REFERENCE PROJECT #1	REFERENCE PROJECT #2	REFERENCE PROJECT #3
Budget	Partially achieved (new estimate in execution phase)	Partially achieved (schedule delay increased cost)	Achieved	Achieved
Schedule	Achieved	Partially achieved (schedule was postponed twice)	Achieved	Achieved
Scope	Achieved	Achieved	Achieved	Achieved
Benefit realization	Partially achieved (two areas achieved and one partially achieved)	Achieved	Achieved	Partially achieved (super user training insufficient)
User satisfaction	Score 4.4 for core team and steering group and 4.5 for review team	Score 4.2 (max 5 and target was 4.0)	Score 4.4 (overall user satisfaction)	Score 3.6 (including user and super user)

In summary, the general approach relies on evaluation criteria that are less project specific and more nomothetic (Dahler-Larsen, 2013, pp. 63-64), which allows for comparisons between projects and

evaluations of projects' relative performance. According to Müller and colleagues, examples of universal parameters span areas such as project success characteristics of time, cost, scope, customer satisfaction, quality of deliverables and developed ideas (Joslin & Müller, 2016). Other relevant success criteria are pertaining to meeting self-defined success factors plus (Müller & Turner, 2007a, 2010) the project's overall performance and purpose as well as user requirements and satisfaction, in addition to reoccurring business with client, and satisfaction of client, supplier, project team and other stakeholders (Müller & Turner, 2007a, 2007b, 2010). While the list of universal criteria can be long, a rule of thumb might be suggested: the more general the approach, the fewer and the more classical the evaluation criteria. In the extreme instance, the general approach is restricted to one or a few aspects of the classical iron triangle of time, cost and quality (Atkinson, 1999) or scope (Müller & Turner, 2007a, p. 303). Such an evaluation is far from holistic. However, it does provide an opportunity for inter-project comparison that allows for a relative project understanding – which is necessary for further evaluating the workings of the HDM. This evaluation step is unfolded in the next sub-section.

5.3 Internal Benchmarking: Third Approach to HDM Evaluation

Third, we compared the pilot project and the three reference projects to evaluate their relative performance – measured on the parameters of time and impact selected in step 2. To assess whether the HDM makes a difference, we make an internal benchmarking and focus on relative performance to investigate if the pilot project is more successful than the reference projects within the same organization. While the Half Double study encompasses a limited and carefully selected number of projects within the same organization, the internal approach can span several programs and portfolios and in principle include all projects within an organization.

A FMCG company provides an example of an internal evaluation based on a comparison of the pilot and reference projects within the organization. The comparison allows for an evaluation of the projects' relative performance on selected parameters. Time is operationalized as time to market and measured in number of months from start until the first sales are generated. Impact is operationalized as sales and measured in indexed sales per month. The internal evaluation approach based on the FMCG's pilot and reference project comparison is shown in Figure 3 below (Svejvig et al., 2017, p. 15).

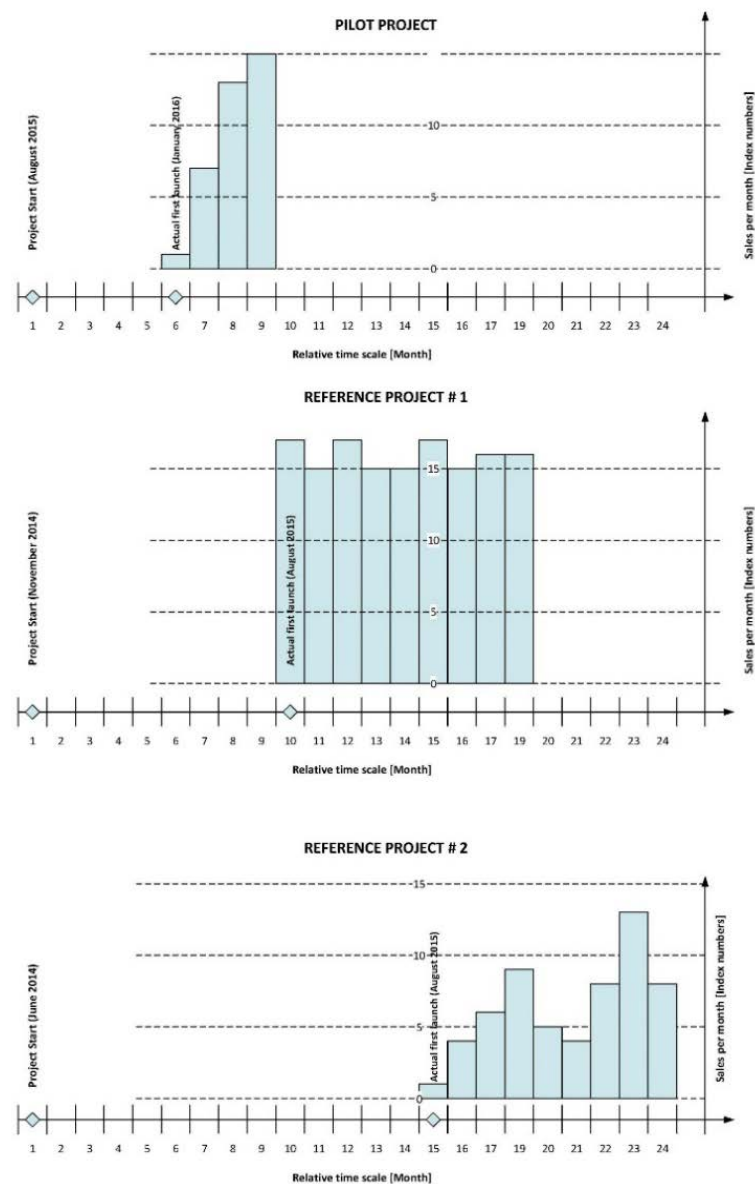









Figure 3: FMCG pilot and reference project comparison

In summary, the internal approach implies a comparison perspective – in which projects are contrasted with other projects inside the same organization. Applied in this way, the internal approach ranges from a comparison of two projects within the same organization to comparisons within and across both programs and portfolios that contrast projects within and across different programs and portfolios. The internal approach allows for a consideration of projects which are located within the same organization and conditioned by the same organizational conceptualities. Nevertheless, the internal approach does not provide knowledge on projects conditioned by other organizational circumstances – which is necessary for further evaluating the workings of the HDM. Therefore, we expanded the evaluation into the next step.

5.4 External Benchmarking: Fourth Approach to HDM Evaluation

Fourth, we compared the pilot projects in the different organizations in order to find out in which cases the HDM made a difference. In this last step, we began to evaluate the overall performance of all pilot projects – in terms of their relative performance compared to the reference projects. This part of the research investigated in which cases the pilot projects were more successful than the reference projects and focused on explanations to performance differences to gain knowledge on the boundaries of the HDM and to find its sweet spot. Compared to the earlier steps of the evaluation focusing on whether the HDM works, this last step directs attention to the question of where and when the HDM works.

The seven pilot organizations of the first phase of the PHD provide an example of a more general evaluation in which the relative performance of one pilot project in one organization is contrasted with the relative performance of another pilot project in another organization. In this way, the evaluation approach indicates that the HDM works better in some organizations than in others. In a structured search for the limits around the sweet spot of the HD methodology, both success and failure cases should be taken into account. The more external evaluation approach based on a comparison of the seven pilot organizations is shown in Figure 4 below from Svejvig et al. (2017, p. 6):

COMPANY	PROJECT TYPE	IMPACT FROM HALF DOUBLE METHODOLOGY
 GRUNDFOS	Product Development	○
 SIEMENS	Product Development	○
 Lantmännen Unibake	Market & Product Development	●
 Coloplast	Supply Chain Project	◐
 Novo Nordisk	IT Project	●
 GNI	E-commerce Project	●
 VELUX	Organizational Change	●

● Higher
○ Lower
◐ N/A

Figure 4: 7 pilot organization's project performance comparison

In summary, the external approach implies a comparison perspective – in which projects are contrasted with other projects from another organization. Applied in this way, the external approach ranges from a comparison of two projects from two different organizations to comparisons of multiple projects from multiple organizations. In this way, the external approach opens up for an understanding of projects on a broader level. Examples of an external approach, which is outside the scope of the HD task, is for instance professional associations like IPMA that compare hundreds of projects, and every year the publication of

the project comparison CHAOS report (Svejvig & Hedegaard, 2016). One of the most extreme examples of an external evaluation approach is a comparison of all projects registered in a very large research project database (Flyvbjerg, 2016). Extremely external comparisons can encompass different types of projects (construction versus innovation) with different contracts (fixed price versus flexible pay) within different sectors (public versus private) and industries (production versus service) across time (from early to late) and place (from local to global), just to mention a few of the dimensions on which projects can differ. In terms of the broadness of the benchmarking, a rule of thumb applies: the higher the number of projects and the more variety between them, the more difficult it is to gain relevant data. While big scale evaluations and evaluations using big data can be expensive (Olsson & Bull-Berg, 2015), they provide an opportunity to gain an understanding of projects on a higher level. An external evaluation on a larger scale would be beneficial to learn more about the HDM as well as the limits of its working and value. Following these lines, the conclusion of this sub-section is that the evaluation artifact has helped structure the evaluation process to collect data and find indicators that can generate answers to four questions pertaining to the overarching research task regarding the value of the HDM. Nevertheless, more research can still be done to investigate more precisely where and when and under which conditions the HDM works best.

5.5 Evaluation of the Artifact

Having illustrated an instantiation of the artifact, here we present an evaluation of the evaluation artifact, in a way taking our own medicine. The evaluation is guided by the seven guidelines suggested by Hevner, March, Park, and Ram (2004, p. 83) for design science research, which we adopt for our action design research approach.

Guideline 1: Design as an artifact – The research here has produced an evaluation artifact that we have also shown to be viable in section 5, and we may consider the research to live up to this guideline.

Guideline 2: Problem relevance – The research gap on project evaluation combined with the interest from practitioners to know if the Half Double methodology had an impact indicates that we have addressed an important and relevant business problem.

Guideline 3: Design evaluation – We have shown the qualities of the artifact in the previous section as part of the structured evaluation presented here. It is the outcome of an ongoing process of evaluating what worked.

Guideline 4: Research contribution – In line with authors arguing in favor of multi-aspect evaluation approaches (Olsson & Bull-Berg, 2015, p. 494), we designed a framework for project evaluation with different approaches and introduced benchmarking to the area of project evaluation.

Guideline 5: Research rigor – The rigorous approach to this study has been outlined in section 3, and here we emphasize the extensive use of templates and review processes in the research and development of the artifact.

Guideline 6: Design as a search process – The abductive approach of this study meant that the researchers continuously tested evaluation designs in a search for an artifact that would satisfy the needs for evaluation defined by the goals of Project Half Double.

Guideline 7: Communication of research – The research and artifact has been disseminated in reports along the process, and especially the research leading to the artifact is communicated in this paper.

In this way, the validity of the artifact has been demonstrated in the context for which it is designed.

Further application and publication need to show whether the artifact also has value outside the development environment – to document its usefulness in terms of its utility (Gregor & Hevner, 2013).

6 Concluding Remarks

In this paper, we have conceptualized and presented an evaluation framework that emerged during an action design research study which was carried out in order to track pilot projects applying a new project management methodology and contrast them with reference projects to find indicators of the effect of the methodology. The framework represents four distinct approaches to project evaluation from classical and specific success criteria to benchmarking against other projects.

The conceptual framework is evaluated as a design artifact. Its validity is assessed through an illustration of the applicability of the framework in an engaged action design research study with the goal of documenting the implications of the implementation of a project management methodology called Half Double in a number of projects and organizations.

The validity section illustrates some of the benefits and limitations of each approach and highlights the usefulness of combining all four approaches to avoid fragmented conclusions. Although a whole or holistic understanding can never be achieved, the framework helps direct attention to some of the evaluation dimensions that the evaluator needs to consider.

It is our hope that the framework can contribute by offering four approaches to project evaluation which can help analyze an existing evaluation and/or structure an upcoming evaluation by revealing the evaluation's insightful angles and/or hidden aspects.

Following these lines, the paper has implications for action design scholars and practitioners who wish to understand and/or improve projects - including professionals and policy makers interested in projects as well as their management, methodologies and performance.

Acknowledgement

The authors would like to thank the Danish Industry Foundation for funding this work and acknowledge contributions from Danish organizations involved in Project Half Double and Implement Consulting Group. The authors declare that they have no conflict of interest regarding the funding agency and other parties involved in Project Half Double.

References

- Anand, G., & Kodali, R. (2008). Benchmarking the benchmarking models. *Benchmarking: An International Journal*, 15(3), 257-291.
- Andersen, E. S. (2010). Are we getting any better? Comparing project management in the years 2000 and 2008. *Project Management Journal*, 41(4), 4-16.
- Atkinson, R. (1999). Project management: cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria. *International Journal of Project Management*, 17(6), 337-342.
- Axelos. (2015). *PRINCE2 Agile*. Norwich: TSO (The Stationery Office), part of Williams Lea.
- Befani, B., Ledermann, S., & Sager, F. (2007). Realistic Evaluation and QCA: Conceptual Parallels and an Empirical Application. *Evaluation*, 13(2), 171-192.
- Bertalanffy, L. v. (1956). General system theory. *General Systems*, 1, 1—10.
- Bertalanffy, L. v. (1968). *General system theory - Foundations, Development, Applications*. New York: George Braziller, Inc.
- Chen, H. T. (2015). *Practical Program Evaluation: Theory-Driven Evaluation and the Integrated Evaluation Perspective*. Thousand Oaks: SAGE Publications Inc.
- Dahler-Larsen, P. (2013). *Evaluering af projekter*. Odense: Syddansk Universitetsforlag.
- Davis, K. (2014). Different stakeholder groups and their perceptions of project success. *International Journal of Project Management*, 32(2), 189-201.
- Dynes, R. R., & Aguirre, B. E. (1979). Organizational Adaptation To Crises: Mechanisms Of Coordination And Structural Change. *Disasters*, 3(1), 71-74.
- Fangel, M. (2010). *Kompetencer i projektledelse. dansk grundlag for kompetencevurdering* (3. ed.). Hillerød: Foreningen for Dansk Projektledelse.
- Flyvbjerg, B. (2006). From Nobel prize to project management: getting risks right. *Project Management Journal*, 37(3), 5-15.
- Flyvbjerg, B. (2016). The Fallacy of Beneficial Ignorance: A Test of Hirschman's Hiding Hand. *World Development*, 84, 176-189.
- Gemino, A., Sauer, C., & Reich, B. H. (2010). Using Classification Trees to Predict Performance in Information Technology Projects. *Journal of Decision Systems*, 19(2), 201-223.

- Goldkuhl, G. (2012). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems*, 21(2), 135-146.
- Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(2), 337-355.
- Gregor, S., Imran, A., & Turner, T. (2014). A 'sweet spot' change strategy for a least developed country: leveraging e-Government in Bangladesh. *European Journal of Information Systems*, 23(6), 655-671.
- Harri Laihonon, P., Linzalone, R., & Schiuma, G. (2015). A review of program and project evaluation models. *Measuring Business Excellence*, 19(3), 90-99.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75-105.
- IPMA. (2016). The Project Excellence Model (PE model). Retrieved from <http://www.ipma.world/awards/project-excellence/the-pe-model/>
- Joslin, R., & Müller, R. (2016). Identifying interesting project phenomena using philosophical and methodological triangulation. *International Journal of Project Management*, 34(6), 1043-1056.
- Jugdev, K., & Müller, R. (2005). A RETROSPECTIVE LOOK AT OUR EVOLVING UNDERSTANDING OF PROJECT SUCCESS. *Project Management Journal*, 36(4), 19-31.
- Kouzes, A., Elme, L., Helmut, K., & Korac-Kakabadse, N. (1999). Benchmarking and performance measurement in public sectors Towards learning for agency effectiveness. *The International Journal of Public Sector Management*, 12(2), 121-144.
- Laursen, M., & Svejvig, P. (2016). Taking stock of project value creation: A structured literature review with future directions for research and practice. *International Journal of Project Management*, 34(4), 736-747.
- Lee, J. S., Pries-Heje, J., & Baskerville, R. (2011). Theorizing in Design Science Research. In H. Jain, A. P. Sinha, & P. Vitharana (Eds.), *Service-Oriented Perspectives in Design Science Research: 6th International Conference, DESRIST 2011, Milwaukee, WI, USA, May 5-6, 2011. Proceedings* (pp. 1-16). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Lenfle, S. (2012). Exploration, project evaluation and design theory: a rereading of the Manhattan case. *International Journal of Managing Projects in Business*, 5(3), 486-507.
- Lewin, K. (1946). Action research and minority problems. *Journal of social issues*, 2(4), 34-46.
- Maire, J. L., Bronet, V., & Pillet, M. (2005). A typology of "best practices" for a benchmarking process. *Benchmarking: An International Journal*, 12(1), 45-60.
- Mathiassen, L., Chiasson, M., & Germonprez, M. (2012). STYLE COMPOSITION IN ACTION RESEARCH PUBLICATION. *MIS Quarterly*, 36(2), 347-363.
- McLeod, L., Doolin, B., & MacDonell, S. G. (2012). A Perspective-Based Understanding of Project Success. *Project Management Journal*, 43(5), 68-86.
- Müller, R., & Turner, R. (2007a). The Influence of Project Managers on Project Success Criteria and Project Success by Type of Project. *European Management Journal*, 25(4), 298-309.
- Müller, R., & Turner, R. (2007b). Matching the project manager's leadership style to project type. *International Journal of Project Management*, 25(1), 21-32.
- Müller, R., & Turner, R. (2010). Leadership competency profiles of successful project managers. *International Journal of Project Management*, 28(5), 437-448.
- Nandi, S., & Banwet, D. (2000). Benchmarking for world class manufacturing—concept, framework and applications. *Productivity*, 41(2), 189-200.
- OED. (Ed.) (2017) Oxford English Dictionary. Oxford University Press.
- Olsson, N. O. E., & Bull-Berg, H. (2015). Use of big data in project evaluations. *International Journal of Managing Projects in Business*, 8(3), 491-512.
- Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. London: Sage Publications Ltd.
- Pinto, J. K., & Prescott, J. E. (1988). Variations in Critical Success Factors Over the Stages in the Project Life Cycle. *Journal of Management*, 14(1), 5-18.

- Roethlisberger, F. J., Dickson, W. J., & Wright, H. A. (1961 (1939)). *Management and the Worker*. Cambridge: Harvard University Press.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A Systematic Approach* (7. ed. ed.). Thousand Oaks, California: Sage Publications, Inc.
- Saunders, M., Lewis, P., & Thornhill, A. (2016). *Research methods for business students* (Seventh Edition ed.). Harlow: Pearson Education Limited.
- Scriven, M. (1991). Beyond formative and summative evaluation. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter century* (pp. 18-64). Chicago: University of Chicago Press.
- Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action Design Research. *MIS Quarterly*, 35(1), 37-56.
- Shenhar, A., & Dvir, D. (2007). *Reinventing project management: the diamond approach to successful growth and innovation*: Harvard Business Press.
- Shenhar, A. J., Dvir, D., Levy, O., & Maltz, A. C. (2001). Project success: A multidimensional strategic concept. *Long Range Planning*, 34(6), 699-725.
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation Theory, Models, & Applications*. San Francisco: Jossey-Bass.
- Svejvig, P., & Andersen, P. (2015). Rethinking project management: A structured literature review with a critical look at the brave new world. *International Journal of Project Management*, 33(2), 278-290.
- Svejvig, P., Ehlers, M., Adland, K. T., Grex, S., Frederiksen, S. H., Borch, M. M., . . . Pedersen, S. E. (2016). *Project Half Double, Preliminary Results for Phase 1, June 2016*. Retrieved from Aarhus:
- Svejvig, P., & Grex, S. (2016). The Danish agenda for rethinking project management. *International Journal of Managing Projects in Business*, 9(4), 822-844.
- Svejvig, P., & Hedegaard, F. (2016). The challenges of evaluating and comparing projects – An empirical study of designing a comparison framework. In J. Pries-Heje & P. Svejvig (Eds.), *Project Management for Achieving Change* (pp. 107-129). Frederiksberg: Roskilde University Press.
- Svejvig, P., Rode, A. L. G., & Frederiksen, S. H. (2017). *Project Half Double: Addendum: Current Results for Phase 1, January 2017*. Retrieved from Denmark:
- Swanson, G. (1971). Frameworks for comparative research: Structural anthropology and the theory of action *Comparative methods in sociology: Essays on trends and applications* (pp. 141-202). Berkeley and Los Angeles: University of California Press, Ltd.
- Venable, J., Pries-Heje, J., & Baskerville, R. (2012). *A comprehensive framework for evaluation in design science research*. Paper presented at the International Conference on Design Science Research in Information Systems.
- Winter, M., Smith, C., Morris, P., & Cicmil, S. (2006). Directions for future research in project management: The main findings of a UK government-funded research network. *International Journal of Project Management*, 24(8), 638-649.
- Womack, J. P., & Jones, D. T. (2003). *Lean thinking : banish waste and create wealth in your corporation*. New York: Touchstone Books.
- Zwikaël, O., & Smyrk, J. (2012). A general framework for gauging the performance of initiatives to enhance organizational value. *British Journal of Management*, 23, S6-S22.