Measuring Agile Capability

-

Survey Tool Development and Empirical Pre-Test

By
Sebastian Engelmann (115692)
MSc. E-Business

A thesis submitted for the degree of
Master of Science
Copenhagen Business School

Under supervision of Till J. Winkler

September 16th, 2019

STUs: 144,331
Pages: 69

Acknowledgements

First and foremost, I would like to thank Till Winkler, my supervisor at CBS, for making me aware of the topic. He provided me with valuable guidance during the thesis process by sharing his experience of measurement instrument development.

In this vein, I would also like to thank Jacob Nørbjerg, who took over the thesis process for the oral defence.

I would especially like to thank the experts who generously took the time to share their experience with me. All the interviews were very insightful to me, even outside the specific topic of this thesis. I truly enjoyed some of the discussions that developed during the interviews.

Furthermore, I would like to thank those anonymous respondents to my pre-test that made the effort to assess their organisation. I especially appreciate the people that went a step further and commented on the survey with the intention to increase everybody's knowledge about the topic.

Lastly, I would like to thank my friends for their endless support be it in Copenhagen or from the distance. Special thanks to

Finally, special thanks to Viet Thai for supplying me with the best local Thai dishes in Nordvest.

To everyone not mentioned, I still want to say thank you for your support!

Abstract

Many organisations are adopting agile software development practices as they promise to improve their ability to manage a changing environment. This transformation is often guided by an agile framework such as the Scaled Agile Framework (SAFe). Measurement instruments are an established tool to support transformations and thereby, are beneficial for practice. At the same time, measurement instruments support research by enabling quantitative assessments of the characteristics of organisations. Consequently, this thesis investigates through the lens of the resource-based view *how can organizations reliably, comprehensively, and parsimoniously measure their agile capability?* In a first step, several agile frameworks are analysed to identify the common practices that form agile capabilities in an organisation. In the second part of this thesis, results from a small sample pre-test are analysed to identify potential issues with the measurement instrument as well as to obtain an indication for the tool's validity and reliability. The results show that agile practices should be measured on team, program, and portfolio level. Furthermore, measuring agile capabilities based on practices provides a lidded assessment of organisations' true agile capability. Therefore, the measurement instrument is complemented by an assessment of organisational culture. Future research should implement the findings of this thesis to refine the measurement instrument.

Keywords: scaled agile; organisational ambidexterity; measurement instrument; IT capability; digital transformation

Table of Contents

Table of Tables

Table of Figures

List of Abbreviations

CBS	Copenhagen Business School
CMM	Capability Maturity Model
CMMI	Capability Maturity Model Integration
CMMI-DEV	Capability Maturity Model Integration - Development
COBIT	Control Objectives for Information and Related Technology
DA	Disciplined Agile
ISD	Information System Development
KMO	Kaiser-Meyer-Olkin
LeSS	Large Scale Scrum
RAGE	Recipes for Agile Governance in the Enterprise
RBV	Resource Based View
S@S	Scrum@Scale
SAFe	Scaled Agile Framework
SPICE	Software Process Improvement and Capability Determination
VIF	Variance Inflation Factor

1 Introduction

1.1 Background

Product development capabilities provide a central criterion for the responsiveness of an organisation in a volatile environment (Brown & Eisenhardt, 1995) making it a crucial aspect of competitive advantage (Wind & Mahajan, 1997). The different strategic approaches to product development of innovator and imitator (Kerin, Varadarajan, & Peterson, 1992) and the associated discussion of first-mover advantage and disadvantage (Lieberman & Montgomery, 1988) have in common that for an organisation's success it is key to be responsive to a constantly changing environment (Christopher, 2000). One of the most popular options to enable this responsiveness in established organisation is to implement *agile methods* (Lee & Xia, 2010).

The concept of *organisational ambidexterity* explains that organisations have to fulfil today's business demands while at the same time prepare for the future to sustain success (He & Wong, 2004). Concurrently, increased unpredictability through digitalisation and globalisation stresses the importance of organisational responsiveness (Hitt, Keats, & DeMarie, 1998). One way to improve the responsiveness, or *exploration* as it is called in the ambidextrous view of the organisation (March, 1991), is through agility (Vinekar, Slinkman, & Nerur, 2006). As innovative IT systems offer a great opportunity for an organisation's competitiveness (Henderson & Venkatraman, 1993), the development of these becomes important.

Agile software development methods were developed to achieve better product-market-fit through short feedback cycles (Boehm & Turner, 2005) and consequently higher responsiveness to change. Traditionally, the agile approach to software development was applied at smaller projects on a team level only (Boehm & Turner, 2005), but the associated benefits made it interesting for larger projects, teams, and organisations alike (Dikert, Paasivaara, & Lassenius, 2016). However, the application of agile software development beyond a single team, *scaling agile*, is related to certain difficulties (Dybå & Dingsøyr, 2009). In contrast to small projects and single teams, larger organisations require additional coordination, which some see as opposing agile values (Dikert et al., 2016). More concrete, coordination between the teams and with the organisation outside the IT department is required. To avoid slowing down the development process due to the need of coordination, the teams need to have the capacity to make independent decisions based on an understanding of the overall vision of the organisation (Gibson & Birkinshaw, 2004).

In response, several frameworks were developed by practitioners that provide guidance on scaling agility beyond a small single team through ensuring contextual ambidexterity. Even though large scale agile is associated with downsides such as increased distance between the development team and other stakeholders such as users compared to agility in a single team (Dikert et al., 2016), an increasing number of organisations adopt these frameworks (Torgeir Dingsøyr & Moe, 2014; VersionOne Inc., 2015, 2016, 2017, 2018).

1.2 Motivation

To support the implementation of scaled agile practices in organisations and achieve the associated benefits, it is necessary to understand where in the process of adapting agility an organisation is. One way to make the status-quo explicit is through measurement instruments such as maturity models. Measurement provides several benefits: it allows comparison to earlier states, to the competition or industry, as well as best practices (Becker, Knackstedt, & Pöppelbuß, 2009). Thereby, measurement can provide orientation and direction on the next steps to improvement by providing a tool for reflection. In general, organisations benefit from measuring agility before, during, and after the process if they want to make a transition from traditional approach (Gren, Torkar, & Feldt, 2015).

From a scientific point of view, a self-assessment measurement instrument provides additional research opportunities, because it simplifies assessment of organisations compared to qualitative approaches like interviewing. This could result for example in classification studies or studies where correlation of agile capability and other organisational characteristics are investigated. Despite the need for a quantitative measurement of agile capability, its research is still in its infancy and no valid model has been developed yet (Gren et al., 2015; Leppänen, 2013).

1.3 Objective and Research Gap

While many assessments on team agility exist (e.g., Little, 2007), and some focus only on one framework (e.g., Turetken, Stojanov, & Trienekens, 2017), only limited research has been conducted on how to parsimoniously measure the capability of an organisation to manage several agile teams (Javdani Gandomani & Ziaei Nafchi, 2015). Nowadays, commercial agile assessments exist, however, they lack scientific documentation (e.g., Tousignant, 2018) or again focus on one framework only such as the Lean Enterprise Assessment (Scaled Agile, 2019). Sidky, Arthur, and Bohner (2007) focus on assessing the agile potential, but do not provide guidance on implementing agility in software development.

Thus far, no measurement instrument for organisations employing more than one team has been developed and empirically evaluated that provides guidance on how to achieve agile capability. Therefore, the objective of this thesis is in a first step to develop a survey instrument to measure the agile capability of organisations that follows scientific standards in its development and consolidates several frameworks in order to allow a more general assessment. To provide the most value, the goal is a comprehensive tool that can be applied without extensive effort or guidance when using it. Instead, it should be applicable for organisational self-assessment conducted by for example management or consultants. Thereby, the tool would be generally accessible including smaller companies with limited resources, which again, is important for both research and practice. In a second step, this study empirically tests the developed tool through a pre-test to assess its validity and reliability as well as to identify potential issues that should be addressed in subsequent studies.

Consequently, considering the practical relevance as well as the theoretical blind spot of agile capabilities for organisations and research, this paper investigates the following research question: *How can organisations reliably, comprehensively, and parsimoniously measure their agile capability?*

1.4 Scope and Delimitations

The key aim of this study is to develop a measurement instrument that is grounded in scientific methods. The overall process of investigating the research question is guided by established measurement methodology (DeVellis, 2012) and enhanced by methods for maturity model development (Becker et al., 2009). Since no established models that guide large-scale agile transformation exist in literature (Paasivaara, 2017), the measurement instrument is based on commercial frameworks developed by practitioners that provide practices to implement agile software development into the organisation's IT departments such as the *Scaled Agile Framework* (SAFe).

Even though agile thinking can be applied in non-engineering departments or on chief-officer level (Greening, 2010), the research presented in this thesis focuses on agile software development in IT departments. In fact, the focus is on practices that enable the usage of agile methods for larger projects that require a larger team that does not fit agile practices due to team size and the coordination of several teams.

The research focuses on how the implementation of agile practices in formerly non-agile organisations can be assessed. Thereby, mainly "organisations considering, planning, or in the midst

of agile transformation" (Conboy & Carroll, 2019, p. 8) are addressed. This implies a limited applicability for 'natural' or 'born' agile organisations such as grown start-ups.

The research objective is investigated from the perspective of the resource-based view (Barney, 1991) on organisations, which states that skills and processes lead to competitive advantage for an organisation. Despite the importance to measure agility in general, the focus lays on observable practices and thereby this research only briefly covers organisational culture and how it relates to agile capabilities. However, it is important to acknowledge that the right mindset is an important aspect of the success of agile practices. Consequently, measuring mindset or culture in an organisation could provide an alternative valuable way of measuring agility in organisations. The rationale behind choosing the capabilities perspective is that capabilities seem to provide a more objective, accessible, and quantifiable way of measuring agile capabilities. This study aims to contribute to the knowledge base of agility by:

1) identifying common practices from different frameworks that form agile capability.
2) presenting arguments for a specified scale to assess the maturity of these practices.
3) conducting a pre-test of the model with a small-sample size[1] to obtain an empirical indication for the model's strengths and weaknesses.
4) suggest changes to the model based on a quantitative and qualitative evaluation of the pre-test.

1.5 Advanced Organiser

This thesis is organised as follows: chapter 2 presents the existing knowledge on the topic by presenting a literature review as well as the theoretical concepts that this work draws on. In chapter 3 the choice of methods is argued for and its implementation presented. Chapter 4 contains the results of the scale development process and the empirical pre-test. The discussion of the results is presented in chapter 5. Finally, chapter 6 contains the conclusion of this thesis.

[1] Due to the time constraints related to student projects, it is expected that only a limited number of responses can be collected. Saunders et al. (2016) indicate that ten responses would be the minimum number for a student project.

2 Theoretical Underpinnings

The following chapter provides the theoretical foundation for this thesis by indicating how this research is related to existing research on organisational resources and ambidexterity. Furthermore, in this chapter a conceptualisation of the focal object of this work (i.e., agility) is provided. It is explained how agility relates to organisations in order to propose a way to make it measurable. Next, the most important frameworks that provide guidance on scaling agile are introduced. Finally, established ways of measuring capabilities are presented and compared.

2.1 Resource-Based View and Organisational Capabilities

The resource-based view (RBV) focuses on organisations' internal resources as a source of competitive advantage (Barney, 1991). Thereby, it detaches the explanation of an organisation's success from a product focus and environmental factors (Wernerfelt, 1984) as in industry analysis frameworks (e.g., Porter, 1980; Schmalensee, 1985). Instead, according to the RBV, a competitive advantage is gained when developing the "rare, imitable, and non-substitutable resources" (Barney, 1991, p. 117) that exist in an organisation.

Resources are defined as "available factors that are owned or controlled by the firm" (Schoemaker & Amit, 1993, p. 35) and can be classified as either assets or capabilities (Wade & Hulland, 2004). Assets are "anything tangible or intangible the firm can use in its processes for creating, producing, and/or offering its products (goods or services) to a market" (Wade & Hulland, 2004, p. 109). Capabilities, in contrast, are skills or processes that increase the value of the input (Wade & Hulland, 2004), through "repeatable patterns of actions in the use of assets to create, produce, and/or offer products to a market" (Wade & Hulland, 2004, p. 109). Similarly, Amit and Schoemaker (1993) describe capabilities as "a firm's capacity to deploy resources, usually in combination, using organisational processes, to effect a desired end". Winter (2003, p. 991) sees capabilities as "a high-level routine (or collection of routines) that, together with its implementing input flows, confers upon an organisation's management a set of decision options for producing significant outputs of a particular type". Consequently, capabilities can be defined as processes or high-level routines representing "a firm's ability to use its other resources to produce outputs for meeting desired objectives" (Kishore, Swinarski, Jackson, & Rao, 2012, p. 459). Feldman and Pentland (2003) show that routines have an ostensive (i.e., structural) aspect that guides people in an organisation as well as a performative aspect that "creates, maintains, and modifies" (p. 94) the

routine through reflection. Both aspects need to be considered when looking at organisational routines (Feldman & Pentland, 2003).

In the context of this thesis, the definition of Wade and Hulland (2004) for capabilities will be followed. It provides the most suitable explanation related to software development methods through its focus on creation and production of products that is enabled by unique skills or processes.

2.2 Organisational Ambidexterity

The long-term success of an organisation depends on its ability to efficiently manage and exploit their existing resources while simultaneously adapting to changing market conditions and exploring opportunities to create value in the future (Gibson & Birkinshaw, 2004; He & Wong, 2004; Tushman & O'Reilly III, 1996). This ability of an organisation to be successful at both exploration and exploitation at the same time is called *organisational ambidexterity* (March, 1991).

The term ambidexterity is derived from the Latin words for both ('ambo') and right hand ('dexter'). It can be translated as both-handedness meaning the ability to use both hands like the right (strong) hand. Duncan (1976) introduced the term to organisational theory by presenting the paradox of stability and change at the same time, which can be managed by introducing dual-structures. March (1991) introduced the terms *exploitation*, referring to the "refinement and extension of existing competences, technologies, and paradigms" (p. 71), and *exploration*, which focuses on "experimentation with new alternatives" (p. 71). Tushman & O'Reilly III (1996) look at the subject from an innovation perspective characterising an organisation's development by alternating evolutionary and revolutionary change. Consequently, they define the challenge for an organisation as being able to increase "alignment of strategy, structure, and culture" (p. 24) in the short term and simultaneously shifting them in the long term. Gibson and Birkinshaw (2004) investigate how organisations can achieve alignment (i.e., everyone working toward the same goal) and adaptability (i.e., "the capacity to reconfigure activities (…) quickly to meet changing demands" (p. 209)).

What all of the aforementioned authors and the pairs they define have in common is that the pairs of terms are opposing each other. Thereby, in an organisational context, the term ambidexterity refers to an organisation's ability to manage two opposing modes of operation at the same time. For an organisation it is important to balance both sides, as the focus on only exploitation will harm the long-term success and a focus on exploration will lead to a lack of benefiting from the new (March, 1991). Thus, organisational ambidexterity can in summary be defined as "an organisations' ability to

be aligned and efficient in its management of today's business demands while simultaneously being adaptive to changes in the environment" (Raisch & Birkinshaw, 2008, p. 375).

The result of engaging in exploration is characterised as "uncertain, distant, and often negative" by March (1991, p. 85). Furthermore, unforeseeable opportunities might emerge at a later stage or the probability of success of a project might depend on the environment, which increases the difficulty to assess the expected return. In contrast, a focus on exploitation increases predictability for an organisation. As exploitation and exploration compete for resources, it becomes necessary for organisations to balance the two despite the difficulty to measure and compare their outcomes, making it complicated to describe the trade-offs and consequently to decide on the path to follow.

Successfully integrating the organisation's exploitation and exploration activities can lead to a long-term competitive advantage by creating a dynamic capability (O'Reilly III & Tushman, 2008). Raisch and Birkinshaw (2008) state that in order to achieve integration "a common set of values, a shared vision, and an overarching governance process" is required. Traditionally, researchers (e.g., Duncan, 1976; March, 1991) saw introducing independent structures in the organisation, where one group focuses on exploitation and another one on exploration, as the best way to achieve ambidexterity. This is referred to as structural ambidexterity (Gibson & Birkinshaw, 2004). However, externalising exploitation or exploration make it difficult to achieve this integration (Benner & Tushman, 2003). In contrast, *contextual ambidexterity* means "the behavioural capacity to simultaneously demonstrate" (Gibson & Birkinshaw, 2004, p. 209) both manners in the same organisational group of people (Gibson & Birkinshaw, 2004), which is more complex than managing one consistent strategy according to Gupta, Smith, and Shalley (2006). Gibson and Birkinshaw (2004) suggest that the best way to resolve the tension between the two disparate poles is to create an organisational context where individuals can decide how to deal with the different demands instead of separating the disperse modes by structure, task, or time.

Summarizing the above stated, the ambidextrous view on organisation suggests that for the long-term success of an organisation it is necessary to both exploit existing resources and simultaneously develop new ones. This can lead to an organisational capability and consequently to a long-term competitive advantage. To achieve this, organisations should develop an environment that enables individuals to react to different challenges, which is especially important in demanding volatile environment (Teece, Pisano, & Shuen, 1997).

2.3 Agile Software Development

Since the 1990s agile software development approaches have evolved as an alternative to traditional solutions with the goal of addressing issues such as budget and time exceedance as well as a lack of responsiveness to changing requirements (Boehm & Turner, 2004). In contrast to the traditional, heavyweight, and plan-driven methods such as waterfall (Boehm & Turner, 2005), agile software development is characterised as lightweight, iterative, and incremental (Dikert et al., 2016). The Oxford Dictionary of English states that agility is the ability "to move quickly and easily" ("Agile," 2010). Early exemplary methods exhibiting agile characteristics are Extreme Programming (K. Beck, 1999) and Scrum (Schwaber & Beedle, 2001).

The focus of agile software development methods is "on adaptation and innovation rather than prediction and control" (Vinekar et al., 2006, p. 32). The characteristics found in the new methods were the foundation for the agile manifesto (Kent Beck et al., 2001), even though the methods only reflect certain aspects of the agile manifesto while they disregard others (Conboy & Fitzgerald, 2004). The Manifesto for Agile Software Development is an often-cited definition of agile (Laanti, Similä, & Abrahamsson, 2013) and even described as the official definition by some (e.g., Chow & Cao, 2008). It presents the common understanding of agile software development of a group of influential software engineering professionals:

> Individuals and interactions over processes and tools
> Working software over comprehensive documentation
> Customer collaboration over contract negotiation
> Responding to change over following a plan
> (Kent Beck et al., 2001)

As the manifesto stems from 2001 and has not been updated since, several proposals to change have been made (e.g., Ambler, 2011) but not implemented. This led to several different groups presenting their own updated agile manifestos (e.g., Kolt, 2014). Despite its importance for providing evidence when a method can be described as agile, neither the agile manifesto nor the agile principles presented together with the agile manifesto define the term agile (Keplinger, 2007). Instead they are "guidelines for delivering high-quality software in an agile manner" (Torgeir Dingsøyr, Nerur, Balijepally, & Moe, 2012, p. 1214). Finally, the manifesto lacks grounding in theory and philosophy according to Conboy and Fitzgerald (2004), which leads to a differing understanding of the term

(Keplinger, 2007). Hence, the agile manifesto and the accompanying principles must be reflected in a definition of agile, but they do not provide this definition themselves.

However, in literature several definitions of agility can be found. A representative sample is presented in the following. Conboy and Fitzgerald (2004) provide an attempt to conceptualise agile by comparing it to lean manufacturing and systems thinking, defining agility as "the continual readiness of an entity to rapidly or inherently, proactively or reactively, embrace change, through high quality, simplistic, economical components and relationships with its environment" (p. 40). Boehm and Turner (2004) define agility as being innovative, flexible, and adaptive to new environments. In a similar way, van Oosterhout, Waarts, and van Hillegersberg (2006) define agility through the concept of flexibility as the ability to quickly react to change. Sambamurthy, Bharadwaj, and Grover (2003) define agility as the "ability to detect and seize market opportunities with speed and surprise". While stressing the ability to identify and cope with change, these definitions miss the people focus, simplicity of processes, and incremental delivery mentioned in the agile manifesto (Laanti et al., 2013). Gallagher and Worrell (2008) become more concrete by characterising agility as "the ability to sense and respond to changes in an organisation's internal and external environment by quickly assembling resources, relationships and capabilities" (p. 73). Kochikar and Ravindra (2007) developed a definition of agile capability as „the ability of an individual or organisation to extend or reconfigure existing competencies, or acquire new competencies, so as to deliver continued high performance in the face of rapid environmental change."

Lee and Xia (2010) provide an overview of existing definitions and identify *response extensiveness* (i.e., "the extent, range, scope, or variety of software team responses") and *response efficiency* (i.e., "the time, cost, resources, or effort associated with software team responses") as two prevailing elements in literature. Further, they provide their own definition for agile software development on a team level: "the software team's capability to efficiently and effectively respond to and incorporate user requirement changes during the project life cycle." (2010, p. 90).

Conboy (2009) conceptualises agility as going beyond flexibility (i.e., the ability of a systems development method to "create change or proactively, reactively, or inherently embrace change in a timely manner, through its internal components and its relationships with its environment" (2009, p. 336)) and leanness (i.e., "contribution to perceived customer value through economy, quality, and simplicity" (2009, p. 339)). More general, according to Agarwal, Shankar, and Tiwari (2006), the difference between leanness and agility is that leanness focuses on eliminating waste and

inefficiencies to reduce costs, while agile uses leanness to emphasise the effort on creating effective responses and valuable outcomes.

Table 1: Concept Matrix for Agile Definitions

	respond to change	speed	flexible	adaptive	innovative	external change	internal change	extend competencies	cost
Conboy and Fitzgerald (2004)	x	x	x						
Boehm and Turner (2004)			x	x	x				
van Oosterhout et al. (2006)	x	x	x						
Sambamurthy et al. (2003)	x	x		x		x			
Gallagher and Worrell (2008)	x	x				x	x		
Kochikar and Ravindra (2007)	x	x				x		x	
Lee and Xia (2010)	x	x							x
Conboy (2009)	x	x	x	x		x	x	x	x

Summarizing, Conboy (2009) defines agility as being continuously ready "to rapidly or inherently create change, proactively or reactively embrace change, and learn from change while contributing to perceived customer value (economy, quality, and simplicity), through its collective components and relationships with its environment."

Furthermore, he provides an information system development (ISD) taxonomy to assess if parts of an ISD method can be seen as agile based on his definition of agility. There, he states three requirements:

1. To be agile, an ISD method component must contribute to one or more of the following:
   i. creation of change
   ii. proaction in advance of change
   iii. reaction to change
   iv. learning from change

2. To be agile, an ISD method component must contribute to one or more of the following, and must not detract from any:
   i. perceived economy
   ii. perceived quality
   iii. perceived simplicity
3. To be agile, an ISD method component must be continually ready i.e. minimal time and cost to prepare the component for use

(Conboy, 2009, p. 341)

As the definition of Conboy (2009) in connection with the taxonomy provides the most comprehensive definition of agility (Torgeir Dingsøyr et al., 2012) and directly or indirectly covers the attributes of the other definitions (Table 1), it will be used for the course of this thesis.

2.4 Scaled Agile Frameworks

Initially, agile software development methods were tailored towards the use of an individual small team with collocated members (Paasivaara & Lassenius, 2011; Vallon, da Silva Estácio, Prikladnicki, & Grechenig, 2018). For projects that require more than a single small team, applying agile would not work (Pernstål, Feldt, & Gorschek, 2013).

This makes it more difficult to apply agile software development methods (Dybå & Dingsøyr, 2009) when an organisation requires software that can only be developed by larger teams or when software development takes place globally (Hossain, Babar, & Paik, 2009). To address these challenges (Reifer, Maurer, & Erdogmus, 2003), agile software development methods must be able to scale[2]. There are mainly two situations where scaled agile can be found: born agile organisations (e.g., Spotify) that have grown beyond a single team and large enterprises that make a transition from traditional methods to agile (Rigby, Darrell K.; Sutherland, Jeff; Noble, 2018).

Some of the founders of agile proposed how agile can scale (Sutherland, 2001) and be used in globally distributed teams (Sutherland, Viktorov, Blount, & Puntikov, 2007). Researchers have investigated how agile is applied in a large software development project through a case study (e.g.,

[2] Despite some disagreement about the definition of scalability (Hill, 1990), scaling can be defined as the ability of a system to respond „to a variation over a range of environmental or design qualities“ (Duboc, Rosenblum, & Wicks, 2007, p. 376).

Torgeir Dingsøyr, Moe, Fægri, & Seim, 2018; Paasivaara, Behm, Lassenius, & Hallikainen, 2018; Paasivaara & Lassenius, 2011). Others focused on how well agile performs in larger, distributed projects in comparison to traditional methods (e.g., Papadopoulos, 2015). One simple method to scale agile beyond a single team where an additional meeting called Scrum of Scrum is used to coordinate several teams was found to be inefficient (Dingsøyr et al., 2018). Hence, several frameworks from practitioners addressing the issue of scaling agile by providing guidance to organisations have been developed and gained significant utilization. Analysing a representative set of the literature about scaling agile as well as annual industry reports (VersionOne Inc., 2015, 2016, 2017, 2018) allowed to identify the most important of these frameworks (Table 2): Scaled Agile Framework (Scaled Agile Inc., 2019) (SAFe), Large-Scale Scrum (The LeSS Company B.V., 2019) (LeSS), Disciplined Agile (Disciplined Agile Inc., 2019) (DA), Nexus (Schwaber & Scrum.org, 2018), Scrum@Scale (Sutherland & Scrum Inc., 2019) (S@S), Recipes for Agile Governance in the Enterprise (Thompson & CPrime, 2013) (RAGE), and Spotify[3] (Kniberg, 2014a, 2014b; Kniberg & Ivarsson, 2012). The frameworks[4] provide practices as well as values and principles to guide implementation of agile in the organisation. Most of these frameworks are based on Scrum (LeSS, Nexus, S@S, and Spotify). While SAFe provides the most comprehensive set of practices for scaling, DA is focused on guidance for adopting scaled agile, and RAGE is mostly concerned with providing recommendations for governance mechanisms.

The practices for implementing agile in an organisation can be grouped into three domains: team, project, and program as found in SAFe and RAGE. In the team domain, the practices ensuring that an individual team delivers value are collected. These cross-functional and self-organising teams are the foundation for agile software development in the different frameworks. As scaling agile requires coordination of several teams, the program domain contains practices to ensure alignment among the different teams. Additionally, this domain provides practices to centralise support for the team. Practices in the portfolio domain ensure IT and business alignment by providing strategic

[3] Spotify does not publish a framework but is an often-cited example of implementing agile at scale and is therefore included in this research.

[4] All frameworks are presented online with additional material available offline mostly in the form of books. Information about the frameworks were collected from their websites unless otherwise cited. Therefore, and for reasons of simplicity, they will be quoted with their abbreviation referring to their website in the remainder of this thesis.

business objectives and an adequate budgeting. This structure is similar to traditional project and development organisation (PMI, 2013).

Table 2: Frameworks mentioned in Research and Industry Report

	Source	Alqudah and Razali (2016)	Laanti (2014)	Vaidya (2014)	Bass (2016)	Dikert et al. (2016)	Turetken et al. (2017)	Dingsøyr et al. (2018)	Kalenda, Hyna, and Rossi (2018)	Paasivaara et al. (2018)	Canna et al. (2018)	Hobbs and Petit (2017)	Putta (2018)	Buchalcevova (2018)	Paasivaara (2017)	VersionOne Inc. (2015)	VersionOne Inc. (2016)	VersionOne Inc. (2017)	VersionOne Inc. (2018)	Horlach et al. (2018)
Frameworks	SAFe	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	LeSS	x		x		x	x	x	x	x		x	x	x	x	x	x	x	x	x
	DA	x	x	x			x		x	x		x	x	x	x	x	x	x	x	x
	Nexus	x							x			x		x				x	x	x
	Spotify	x																		x
	RAGE	x							x							x	x	x	x	x
	Scrum@Scale													x						x
	DSDM																			x
	Crystal																			x
	Agility Path		x																	
	Lean Management															x	x	x	x	
	APM															x	x	x	x	
	Enterprise Agile															x	x			x
	Enterprise Scrum															x	x		x	x
	FAST Agile																			x
	Goal Driven Agile																			x
	Prince 2 Agile																			x
	Scrum PloP																			x
	Matrix of Services																			x
	SLIM																			x
	EUP																			x
	laCoCaModel																			x
	XScale																			x

Apart from the practices, the different frameworks provide a set of principles that are based on the agile manifesto and support the implementation of agile in organisations. Strode, Huff, and

Tretiakov (2009) show in a multi-case study that organisational culture correlates with the effectiveness of agile methods and another study showed that software development agility is indicated by organisational culture (Sheffield & Lemétayer, 2013). Thereby it is indicated that only if both practices and culture are implemented, agile is a resource for an organisation that provides competitive advantage. The incompatibility of parts of agile practices with certain aspects of a (traditional) organisational culture have been described extensively in literature (e.g., Boehm & Turner, 2005; Cao, Mohan, Xu, & Ramesh, 2009; Chan & Thong, 2009; Cockburn & Highsmith, 2001; Nerur, Mahapatra, & Mangalaraj, 2005; Vijayasarathy & Turk, 2008). Further, following the practices without having an agile mindset in the organisation may result in a cargo-cult, where an organisation follows the practices but lacks the return of value because the mindset is missing (McConnell, 2000). Thereby, cargo-cult stands for an issue that has been described as ostensive and performative aspects of routines by Feldman and Pentland (2003). In contrast to Feldman and Pentland (2003), that see the interplay of ostensive and performative as a source for change, in the context of cargo-cult the two aspects provide an explanation for failing agile implementation on the basis of organisational retention.

2.5 Agile Maturity Assessment

Maturity models describe how something develops (Fontana, Meyer, Reinehr, & Malucelli, 2015). This could be a range of elements such as a person, an object, or a system, whose situation can be assessed and thereby the model provides guidance (Kohlegger, Maier, & Thalmann, 2009). In general, maturity can mean two things: (1) „the state of being complete, perfect, or ready“, (2) “to bring to maturity or full growth; to ripen” (Oxford Eglish Dictionary (OED), “Mature,” 2009; “Maturity,” 2009). Both definitions mention the aspect of completeness or perfectionist, but the second stresses the process aspect of achieving maturity. In an organisational context maturity models are used to describe organisational capabilities, and hence support managing and improving these capabilities (Maier, Moultrie, & Clarkson, 2012). The basic idea behind a maturity model is that by describing typical behaviour for an organisation at a certain level, organisations can be placed on a certain maturity stage (Fraser, Moultrie, & Gregory, 2002). Attributes further describe what goals need to be achieved to reach a maturity level. Thereby, the purpose of maturity models is two folded. First, it supports to establish the capability of an organisation in a specific area and second the results provide orientation and direction for improvement in accordance with best practices in the area (Essmann & Du Preez, 2009). There are two kinds of assessments: staged and continuous. The first

requires a set of characteristics that needs be completed to achieve the next level, while for the latter maturity is reflected by each characteristic's own rating (Wulf, Winkler, & Brenner, 2015).

Table 3: Comparison of Maturity Models

	CMMI-DEV (V1.3)	**SPICE (ISO)**	**COBIT 4.1**	**COBIT 5**
Maturity Levels	Initial; Managed; Defined; Quantitatively Managed; Optimizing	Incomplete Process; Performed Process; Managed Process; Established Process; Predictable Process Optimizing Process	Non-existent; Initial/ ad hoc; Repeatable but intuitive; Defined; Managed and measurable; Optimised	Incomplete Process; Performed Process; Managed Process; Established Process; Predictable Process; Optimising Process
Attributes	Achieve Specific Goals; Institutionalize a Managed Process; Institutionalize a Defined Process	Process Performance; Performance Management; Work Product Management; Process Definition; Process Deployment; Process Measurement; Process Control; Process Innovation; Process Optimisation	Awareness and Communication; Policies, Plans, Procedures; Tools and Automation; Skills and Expertise; Responsibility and Accountability; Goals Setting and Measurement	1.1 Process Performance; 2.1 Performance Management; 2.2 Work Product Management; 3.1 Process Definition; 3.2 Process Deployment; 4.1 Process Management; 4.2 Process Control; 5.1 Process Innovation; 5.2 Process Optimisation

The most widespread of maturity assessment models is the Capability Maturity Model (CMM) that was originally designed for software development process maturity and defines software process maturity as "the extent to which a specific process is explicitly defined, managed, measured, controlled, and effective." (Paulk, Weber, Garcia, Chrissis, & Bush, 1993, p. 408). The model provides characteristics for five maturity levels: *initial*, *managed*, *defined*, *quantitatively managed*, and *optimising*. A maturity level covers several process areas that require a certain capability level in the area to achieve the next maturity level. The Capability Maturity Model Integration for Development (CMMI-DEV) (CMMI Product Team, 2010) is an extension of the CMM that focuses on processes required for product and service development. Another assessment model that originated in software development is ISO/IEC 15504 (ISO/IEC, 2006), which is also called Software Process Improvement and Capability Determination (SPICE). The model defines capability levels for processes as *incomplete*, *performed*, *managed*, *established*, *predictable*, or *optimising*. The Control Objectives for Information and Related Technology (COBIT) framework provides good practices for IT governance. It presents processes and control objectives that can be assessed on a generic scale provided. In version 4.1 (IT Governance Institute, 2009) these are: *non-existent*, *initial/ad hoc*, *repeatable but intuitive*, *defined process*, *managed and measurable*, and *optimised*. For a more fine-

grained assessment the framework also provides maturity attributes with specific characteristics for each attribute on each of the levels. In version 5 (ISACA, 2012) the levels are: *incomplete*, *performed*, *managed*, *established*, *predictable*, and *optimising*. Thereby, the levels are closer to SPICE. This has been criticised, because of the bigger gap between incomplete and performed than the other levels as achieving the first level (i.e., performed) means that the goal of the process are already largely achieved (Pasquini & Galiè, 2013).

The above presented models provide a good basis for developing an agile maturity assessment, but their underlying assumption of strictly defined processes is somewhat contradictory to agile development processes. In fact, case studies show how CMMI and agile can work together (e.g., (Jakobsen & Johnson, 2008; Jakobsen & Sutherland, 2009). However, as it remains questionable whether it is possible to maintain agility at the higher maturity levels (Lukasiewicz & Miler, 2012; Paulk, 2001; Turner & Jain, 2002), the maturity models cannot be applied directly to assess agile maturity (Fontana, Fontana, Da Rosa Garbuio, Reinehr, & Malucelli, 2014).

3 Methodology

The purpose of this study is to identify a reliable and parsimonies way to measure the agile capability of an organisation. This research objective was addressed by a critical realist research philosophy. The following chapter explains how the research presented in this thesis was conducted. Starting with the underlying assumptions of the study, followed by presenting the overall research design and logic, and concluded with the concrete implementation of the different methods used to gather, process, and analyse information. Further this chapter contains an explanation of the decisions taken during the process and thereby makes the research process more transparent.

3.1 Paradigmatic Assumptions

The following section presents my research philosophy based on my view on ontology, epistemology, and axiology to clarify how the research problem should be understood and addressed (Kuhn, 1962). This helps clarifying the researcher's role in this project and influences the research design by providing foundation for the research design as well as data and evidence collection and interpretation (Easterby-Smith, Thorpe, & Jackson, 2015) As these aspects are fundamental for the research procedure, it is important to make them transparent and consequently allow for a coherent research project (Saunders, Lewis, & Thornhill, 2016).

3.1.1 Ontological Assumptions

Ontological assumptions describe how one sees reality (Saunders et al., 2016). The research conducted is based on the assumption that objective structures and causal mechanisms determine reality and thereby follows a realist understanding of ontology. The researcher believes that reality exists independently of our perception or knowledge of it (Archer, Bhaskar, Collier, Lawson, & Norrie, 1998) resulting in an objective understanding of reality. However, reality is not directly accessible for us. Instead, we experience the manifestation of the things. Thereby, we can describe the reality as consisting of layers (Archer et al., 1998) and consequently the researcher needs to find ways to access the reality (Zachariadis, Scott, & Barrett, 2013). Summarizing, the underlying ontological assumption for this research project is that of ontological realism.

3.1.2 Epistemological Assumptions

Epistemological assumptions describe what is considered "valid and legitimate knowledge" (Saunders et al., 2016, p. 127). From the researchers perspective, knowledge is grounded in historical and social practices, making organisational practices important (Orlikowski & Baroudi, 1991). However, in contrast to the believe in a positivist stance, scientific knowledge is imperfect to a certain degree, which requires a critical view on our knowledge. Thus some researchers have more valid explanation of reality making knowledge relative (Zachariadis et al., 2013). The above stated results in an epistemological relativism that guides the understanding of what knowledge is. Consequently, to derive at valid knowledge it is essential to extend quantitative analysis as they oversimplify (Saunders et al., 2016).

3.1.3 Axiological Assumptions

As knowledge is based on historical and social practices, values play an important role in research (Saunders et al., 2016). Biases based on worldview or cultural experience exist and need to be minimised in research. Thereby it is necessary for the researcher to be as objective as possible (Saunders et al., 2016).

3.1.4 Research Philosophy

My objective understanding of ontology is grounded in the perception that we are able to measure the outcome of certain organisational practices. However, the underlying mechanisms that lead to the results can be difficult to observe and often only parts of these patterns can be accessed. This results in our knowledge often being imperfect making it necessary to assess and discuss the results we obtain critically. Nevertheless, the goal of scientific inquiry is to come as close to reality as possible even though this goal cannot be achieved. Due to the understanding that all measurement can fail, it is vital to apply several measurements and observations that whose individual downsides should counterbalance in order to achieve a good understanding of reality. Furthermore, it is important for the researcher to be as objective as possible to identify and balance potential bias. This can be achieved by critically assessing what we know. Therefore, my research philosophy is that of a *critical realist*.

3.1.5. Approach to Theory Development

In this thesis an abductive approach to theory development is followed. From known premises (i.e., established agile practices and maturity models) a testable conclusion (i.e., the measurement instrument) is generated. The data collection presented in this thesis is used to explore agile capabilities and patterns of mature agile organisations are operationalised in the survey tool, which is finally assessed in a pre-test and exposed to existing literature to critically discuss both the model and the literature. Thereby, the set of observations made during the research indicate the conclusion, but lack a logical guarantee (Saunders et al., 2016) due to for example the small sample size making the discussion of the results necessary.

3.2 Research Design and Logic

The following section describes how the overall design and logic of the study addresses the exploratory research objective of this thesis. It highlights how the application of a sequential mixed-method approach led to a comprehensive investigation of the issue. The usage of an sequential mixed-method research design for the development and testing of a new measurement instrument is recommended for example by Creswell, Fetter, and Ivankova (2004). The combination of qualitative and quantitative elements in this study allows an in-depth understanding of the problem as they can be highly informative (Saunders et al., 2016). However, as Zachariadis et al. (2013) explain, when qualitative work is used for preparation of a quantitative study, qualitative approaches should follow. Nevertheless, there are good reasons to apply mixed methods based on the purpose of compensation (Zachariadis et al., 2013), as well as to assess the reality (McEvoy & Richards, 2006).

The sequence of qualitative, followed by empirical quantitative methods allowed to a certain degree to offset the disadvantages of the individual parts, namely the missing generalizability of the findings from the qualitative stage and the in-depth insights lacking in quantitative research.

The overarching strategy of the study was informed by established scale development procedures as presented by DeVellis (2012) and concretised by Churchill (1979), Rossiter (2002), and MacKenzie, Podsakoff, and Podsakoff (2011). In a first step a conceptual definition of the focal construct (i.e., agile capability) was developed to ensure a clear understanding from the beginning of the study on, as this is necessary to determine what one actually attempts to measure and thereby to avoid confusion (Churchill, 1979; MacKenzie et al., 2011; Podsakoff, MacKenzie, & Podsakoff, 2016). Due to the abstract nature of agile capability, the traditional scale development method was extended by a standard method to develop maturity assessment for processes as introduced by Becker,

Knackstedt, and Pöppelbuß (2009). This seems reasonable due to certain similarities between maturity models and scales since, a "maturity model serves as the scale for the appraisal of the position on the evolution path" (Becker et al., 2009, p. 2013) of an organisation. The combination of scale and maturity model development was applied before by for example Wulf, Winkler, and Brenner (2015) to develop a scale to measure ITIL capabilities. Becker et al. (2009) recommend conducting a comparison of existing maturity models, which was used to inform the choice of scale. By using an established scale as a basis, it could be ensured that the developed scale is consistent and thereby allows numerical comparison and analysis of the pre-test results.

To develop the initial pool of items that present the focal construct, the significant frameworks that suggest practices to organisations on how to implement agility were identified. To identify similar and dissimilar items in the different frameworks, a method inspired by card sorting (Cataldo, Johnson, Kellstedt, & Milbrath, 1979) was applied, allowing a more focused initial pool of items. To outweigh the inexperience of the researcher with agile practices in organisations, interviews with experienced practitioners were conducted.

In the second part of this study, indication for the developed model's reliability and validity were obtained through an empirical pre-test as informed by DeVellis (2012). In this step, through the statistical analysis, the underlying dimensions of the construct were identified, and the scale was tested. In this study the coordinated application of first qualitative research methods followed by a quantitative pre-test allowed to achieve a relative comprehensive assessment of the model. Finally, the discussion provides a qualitative assessment of the findings from the pre-test to fulfil the requirement of completeness of mixed methods in critical realism (Zachariadis et al., 2013).

3.3 Research Methods

The previous sub-chapter explains how the different methods were combined and how they interact and consequently the reasoning for the mix. This section in contrast explains how the different methods were operationalised in the context of this research project as well as the choices made while conducting the study.

3.3.1 Construct Definition

Following the recommendations of Podsakoff et al. (2016) for improved conceptual definitions the focal construct of agile capabilities was defined. As a starting point, the dictionary definition of *agility* was used to generate an initial understanding of the term. Next, literature

definitions were collected, and the concept was compared with its 'opposite' of traditional software development methods. Additionally, the agile manifesto as a representation of practitioners' understanding of the term was reviewed. This resulted in a set of common attributes of the term collected in a matrix that allowed to derive a conceptual definition (chapter 2.3).

3.3.2 Identification of Frameworks

Based on an analysis of literature, several agile frameworks were identified. To search the literature about agile frameworks, a process inspired by a structured literature review (Webster & Watson, 2002) was applied. The search was conducted on Google Scholar and Copenhagen Business School Library Intranet. The search terms used were: "agile frameworks", "scaled agile", "scaled agile frameworks", "scaling agile", and "scaling agile frameworks" Further, snowballing was applied, meaning that literature referenced in an article about agility was also searched when in a relevant context. Further, the identified frameworks were searched based on the idea that literature often states several examples and thereby, an article about one of the frameworks might mention another one. The search was conducted based on a full-text search as for the identification it was not necessary that the frameworks were the main point of investigation, but rather to collect a pool of potential frameworks. Additionally, to the scientific literature search, the industry reports from VersionOne Inc. (2015, 2016, 2017, 2018) were analysed to identify additional frameworks as well as to generate an overview of the importance of the individual frameworks relative to each other. This resulted in the selection of six frameworks and one case study (Table 2).

3.3.3 Item Pool Generation and Initial Refinement

To generate the initial item pool, the practices proposed in the selected frameworks were identified by documenting every artefact, role, practice, etc. described in the frameworks. Here the individual differences between the frameworks influenced the process. As the SAFe framework presents the most comprehensive collection of practices (i.e., 116 items) it was the natural starting point. The other frameworks were checked to identify similar practices as well as to add missing items to the pool. This resulted in 184 items that could be identified initially (see appendix for a table with the items).

The practices were grouped in four domains: overarching, team, project, and program. This structure is similar to the one employed in SAFe and RAGE, leaving out the large solution configuration (SAFe) as it presents an extension of the program domain. The four groups were

necessary to make the amount of practices more manageable and was justified as the different frameworks often applied this similar structure as well as this represents a classical structure present in most (hierarchical and non-hierarchical) organisations.

For the purpose of identifying the similar and dissimilar items and thereby refining the item pool, the identified practices were sorted based on a method inspired by card sorting. This became necessary as several items addressed similar practices but with different title that still covered similar practices. This was done in the four domains individually. The participants of the card sorting study were three students at CBS, who were chosen based on the pragmatic reason of availability. The cards were prepared in handwriting with a description of the item on the back of the card. During the study it showed that limited domain knowledge of agile methods also limits the validity and reliability of the results. Therefore, the researcher had to critically review the results by comparing the description of the grouped items to ensure that the purpose of the items was similar and not just the title for instance. As card sorting in general allows the identification of similarity among a group of items and thereby the formation of groups (Nurmuliani, Zowghi, & Williams, 2004), it was considered an appropriate method to structure the documented items into similar groups. Card sorting with the additional critical review of the results consequently allowed to distil the similar practices (see appendix for results and documentation).

In general, it played an important role to identify practices that had a similar function despite differing in the details of implementation. For instance, the reflection at the end of an iteration is called PI Retrospective in SAFe and Sprint Review in Nexus, but essentially, they have the same purpose of reflecting on the process. These differences were addressed as by comparing the individual descriptions in the frameworks to each other and then develop a collective terminology that presents the condensed essence of the practices. Consequently, not every fine-groined difference between the frameworks is represented in the items, but the overarching ideas.

3.3.4 Scale Definition

To identify a suitable scale for measuring the agile capability of an organisation the maturity model development method introduced by Becker et al. (2009) was applied. The authors present seven requirements for a "well-founded design" (2009, p. 213) of maturity models. In the following, it is discussed how these requirements are addressed in this study.

First, Becker et al., (2009) require a comparison of existing maturity models. This resulted in identifying two types of models: commercial type models, and established software development

maturity models. The models were discussed as presented in chapter 2.5 as well as in Table 3. The second requirement namely that of an iterative procedure was fulfilled by assessing the models with experts and refining them based on the expert's input. The evaluation requirement is addressed in the discussion presented in chapter 5 of this thesis. The application of "well-founded and finely attuned" methods (p. 214) is presented in this Chapter and thereby aims to fulfil the fourth requirement. The problem relevance (requirement 5) as well as problem definition (requirement 6) are covered in the introduction in chapter 1. As the results of the development are presented in this thesis, the presentation is targeted at researchers mainly. To further target practitioners the format of a management report could be used as additional channel. Scientific documentation (requirement 8) and targeted presentation is again covered by documenting the development process in this thesis.

3.3.5 Expert Interviews

Expert interviews were conducted to address the issue that the researcher only has a limited perspective on the issue of agile due to its lack of practical experience with the focal construct. For this purpose, six experts were asked for their assessment of the item pool and the suitability of the scale.

The experts were contacted through LinkedIn. The experts were identified through their membership in the LinkedIn groups *Scaled Agile Framework® (SAFe®)*[5] and *Scaled Agile – SAFe Enthusiasts*. The groups consist of professionals that have an expressed interest in agile capabilities and were therefore considered a suitable starting point. Next, the group members were filtered for persons living in Denmark to allow in-person interviews. The remaining group were scanned based on their experience since having an expressed interest in a topic, does not mean that you can also be considered an expert with sufficient practical experience. Criteria to contact a person were years of experience and official certificates from agile frameworks. Important was that people were actually experienced with agile beyond a single team as this is important for the capability of an organisation. The remaining twenty people were contacted through direct messages in which the thesis project was introduced, and the contacted person were asked if they are available for an interview to share their experience with agile. This resulted in six experts available for interviews (Table 4). The current status of the model was sent to the experts before the interviews to improve the overall quality of the interviews' responses.

[5] This is the official group for the SAFe Framework managed by Scaled Agile Inc.

Table 4: Overview of Experts

Interviewee	Profession	Professional Experience with Agile	Certification
1	Consultant[6]	4 years of work experience with agile	SAFe 4.0 Agilist
2	Independent Agile Coach and Consultant[7]	> 10 years of work experience with agile	Scrum@Scale Practitioner; SAFe Program Consultant (SPC4.5); Certified LeSS Practitioner;
3	Independent Agile Trainer and Coach[8]	14 years of experience with agile (adoption and working);	SAFe 4 Program Consultant (SPC4)
4	Consultant[9]	6 years of experience with agile	SAFe 4 Agilist; SAFe 4.5 Program Consultant; SAFe 4 Product Owner/Product Manager
5	Agile Coach and Scrum Master[10]	7 years	SAFe 4.0 Practitioner; SAFe 4.0 Advanced Scrum Master; SAFe Agilist
6	Project Manager[11]	6 years of experience with agile	SAFe 4 Program Consultant; SAFe 4 Product Owner/Product Manager

In a first iteration, the initial item pool and scale were presented to an expert with four years of experience with SAFe. The feedback was collected over e-mail due to time constraints on the site of the expert. As this was the first iteration the disadvantages of collecting feedback over e-mail (namely limited follow-up possibility) was seen as being acceptable because the of the planned additional interviews. Furthermore, the expert was available for follow up questions. The remaining interviews were conducted face-to-face (four interviews) and over Skype (two interviews).

The interviews followed a brief interview guide to ensure that all topics were sufficiently covered while allowing room for the interviewee to present their experience (see appendix for interview guide). The interviews were structured as follows: start with introducing the research topic, followed by presenting the items and scale to the interviewees, and concluded with asking the interviewees to add any thoughts that were not covered sufficiently during the interview.

All the interviews were recorded and transcribed, as well as notes were taken during the interviews. For transcription the service of *Otter.Ai* was used for an initial draft of the transcripts and

[6] Email feedback

[7] Interview 2 & 5

[8] Interview 3

[9] Interview 4

[10] Interview 6

[11] Interview 7

this version was then reviewed by the researcher. The parts quoted in the survey were reviewed to improve the readability of the sentences. This revision was necessary as the quality of the transcript was limited despite the time saved through the service. The revision of the transcript further allowed the researcher to ensure that the important aspects of the interviews were correctly implemented in the development of the survey instrument (see appendix for interview transcripts). The main outcome of the interviews was the importance of organisational culture as explained further in chapter 4.1 and the importance of providing suitable descriptions for the different levels (see appendix)

An alternative method that could have been applied to identify the final pool of items could have been the Delphi method, where experts gather in a room and agree on a set of items. However, this was not suitable for this study as a lack of access to experts, especially at the same time, constrained this method. Nevertheless, the combination of presenting the item pool to a set of different experts addressed this constrained and at the same time allowed a usage of the knowledge of different experts.

3.3.6 Identifying Dimensions of Agile Culture

As the interviews revealed that the practices assessment needs to be supplemented by an assessment of the organisational culture, it became necessary to identify the components or dimensions of an agile culture (chapter 4.1.2). These were identified by analysing the principles and values in the frameworks and extended by the culture presented in the Spotify case study. The results of this analysis are presented in chapter 4.1. Additionally, the experts were asked about their impression of the cultural dimensions (see appendix for the full analysis).

3.4 Pre-Test

In the second part of this research a pre-test of the developed model was conducted. The main goal of this pre-test was to identify potential issues with the model in an empirical setting and to obtain an indication of the model's validity and reliability.

3.4.1 Data Sources

The survey was shared through personal network with eight people that have working experience in agile organisations and one organisation that uses agile methods. Additionally, the survey was shared in seven LinkedIn groups and six Facebook groups with a total number of 231,211 members.

Table 5: Overview of Social Media Groups Used to Distribute Survey

Network	Group Name	Members[12]	Link
LinkedIn	Scaled Agile Framework	35,810	https://www.linkedin.com/groups/4189072/
	Scaled Professional Scrum & The Nexus Framework (Scrum.org)	438	https://www.linkedin.com/groups/8454481/
	Agile Managers Forum	3,399	https://www.linkedin.com/groups/4080352/
	Product & Project FRAMEWORKS: PMBOK DA SAFe DevOps Scrum Lean Kanban XP LeSS SoS DSDM Nexus etc	1,662	https://www.linkedin.com/groups/8198168/
	Agile and Lean Software Development	156,656	https://www.linkedin.com/groups/37631/
	LeSS - Large-Scale Scrum	2,787	https://www.linkedin.com/groups/6968022/
	Scaled Agile -SAFe Enthusiasts	1,635	https://www.linkedin.com/groups/8315187/
Facebook	Enterprise Scrum	3,242	https://www.facebook.com/groups/EnterpriseScrum/
	Modern Agile	5,720	https://www.facebook.com/groups/modernagile.org/
	Agile Project Management	3,585	https://www.facebook.com/groups/791366830981941
	Agile Scrum Q&A Forum	4,185	https://www.facebook.com/groups/65503514201/
	Agile/Scrum Project Management	3,925	https://www.facebook.com/groups/AgileScrumProjects/
	Scrum	8,167	https://www.facebook.com/groups/scrumframework/
Total		231,211 [13]	

The social media groups were chosen because they allowed to contact a widespread and international group of professionals in the field of agile that would overlap with the desired target respondents. The groups consist of people with an expressed interest in agile methods with differing range of experience with scaled agile ranging from beginner to very experienced. The major disadvantage of this distribution method is that it cannot be controlled who really responses to the survey. However, as the survey is intended for organisations that work with agile with more than one team, anyone with sufficient insights about the organisation is a qualified respondent.

3.4.2 Data Collection

The survey was implemented and distributed with the survey tool *Qualtrics*, which was accessed through a license from CBS (see appendix for the implementation). Qualtrics was selected because it provided more suitable implementation options for the practices' assessment than other available tools. The survey consisted of three parts, namely the practices' assessment, statements to

[12] As of 20th of August 2019

[13] There certainly is an overlap of the members in the groups, but this number gives a rough estimate.

measure culture, and lastly demographics of the respondents and their organisation were collected. In the introduction, the research purpose of the study and an estimate of the time needed to complete the survey were given to the respondents. Further, the respondents were informed that all results are anonymous and only used for the course of this thesis. To increase motivation, respondents were informed that they would be able to download their responses at the end of the survey which could help them reflecting on their organisation's agile capability (see appendix for an example). Additionally, to increase trustworthiness the template from CBS and its logo were used as well as the student's email address clearly stated. This also served the purpose of providing a feedback channel for the respondents if they had any comments, recommendations, or issues.

The response collection was open from the 29th of July 2019 to the 27th of August 2019 for a total of twenty-eight days. Collected data were cleaned based on following criteria: as Qualtrics saves responses of people that only clicked the link, but have not answered any question, this was the first group of responses that were removed. They were identified by the missing mandatory confirmation after the introduction to the rating scale for the practices' assessment. Next, the group of responses that were missing substantial parts of the answers were removed. Cases that were missing only the demographics' part were kept. Furthermore, responses with a suspicious response pattern were removed. These responses were identified by having all the same level for the practices' assessment. One case was partially removed, because the statements were all answered the same. This case was only used in the analysis of the practices but excluded from analysis of the culture part and the comparison of the practices with the culture.

3.4.3 Measurement Model

The conceptualisation of agile capability in chapter 2 in combination with the interviews informed the operationalisation of agile capability that led to the development of a two-folded measurement model. The model consists of a formative measurement instrument consisting of practices that contribute to agility in an organisation, which is supplemented by measuring reflectively agile culture in organisations.

The pre-test was conducted based on the results from the model development as presented in chapter 4.1. In the first part of the survey the respondents were asked to rate the practices in their organisation, followed by statements that aimed to assessing the organisations agile culture through the application of a Likert-scale. Thereby, the rating took place in the form of self-rating (Rossiter, 2002). In the last part, demographics about the respondents and their organisation were collected. For

details about the practices' assessment model and the statements for the cultural assessment see chapter 4.1.1 and chapter 4.1.2.

The questions for the practices' assessment were summed to obtain a score for the organisation's practices following an additive logic (Lasrado, Vatrapu, & Mukkamala, 2017). Thereby, the highest theoretically possible score for practices is ninety. For the cultural assessment, the responses were obtained on an ordinal Likert-scale. To obtain a score the average is calculated. Consequently, the score for the culture assessment ranges from -2 to +2 with zero being the mid-point (Lasrado et al., 2017).

3.4.3 Data Analysis

For all statistical analysis of the data collected during the pre-test, *IBM SPSS Statistics Version 25* was used. Access to the software was provided by CBS.

The formative measurement model to assess agile capabilities from a practice perspective is based on the assumptions of a multiple regression (Diamantopoulos & Winklhofer, 2001). Therefore, in a first step it was checked if the data collected from the practices part in the survey fulfilled the eight assumptions of a linear regression (Laerd Statistics, 2015b). These assumptions are:

1. The dependent variable is measured at a continuous level.
2. The independent variables are measured at a continuous level (variables measured at an ordinal level must be treated as continuous variable).
3. There is independence of observations.
4. There is a linear relationship between
    a) the dependent variable and each of the independent variables
    b) the dependent variable and the independent variables collectively
5. Data shows homoscedasticity of residuals
6. No multicollinearity
7. No significant outliers, high leverage points, or highly influential points
8. Residual errors are approximately normal distributed

The first three assumptions were fulfilled based on the study design as the dependent variable (i.e., the score for agile practices) is measured on a continuous level, the independent variables (i.e., the practices) are ordinal values treated as continuous, and the observations are independent because

of the cross-sectional study design. The assumptions 4-8 were tested using SPSS Statistics. The results of these tests are presented in chapter 4.2 (for the SPSS Statistics output see appendix).

3.3.4 Validity assessment

Validity for formative measurement models is based on internal and external validity (Diamantopoulos & Winklhofer, 2001). According to Rossiter (Rossiter, 2002) the content validity of formative measurement models is always given as the items define the underlying construct. He even claims that "all that is needed is a set of distinct components as decided by expert judgment" (Rossiter, 2002, p. 315). Thereby, it is important to ensure in the development process of the model that the items cover the construct sufficiently. Consequently, it is important to be rigorous when developing the measurement tool based on scientific standards. This is given for the practices' assessment and the cultural assessment as the items were developed based on established frameworks as well as the item selection was confirmed by experienced practitioners. Furthermore, the definition of the concept of agile capabilities was the foundation for item development and thereby further supported content validity (Diamantopoulos & Winklhofer, 2001). The second aspect of internal validity is based on indicator collinearity, which is seen as an indicator for individual indicator validity (Diamantopoulos, Riefler, & Roth, 2008). If one of the items is a perfect linear correlation of the other items, this item most likely contains redundant information and could therefore be excluded (Bollen & Lennox, 1991). Multicollinearity is commonly measured through the variance inflation factor (VIF). The VIF for each item indicates the potential presence of collinearity. VIFs of up to 10 are regarded as acceptable in literature (Kleinbaum, Kupper, Muller, & Nizam, 1998). Therefore, items with a VIF higher than 10 indicate issues with collinearity for these items resulting in problems regarding validity of the individual items.

The assessment of external validity is limited to the feedback generated during the pre-test, as no control variable could be included in the survey, because no 'gold standard' measurement instrument for agile capability exists and a reflective measure as suggested by (Diamantopoulos & Winklhofer, 2001) and Hair, Hult, Ringle, and Sarsted (2014) could not be reasonably identified.

Based on the assumption that practices might lead to agile culture, the correlation of the practices' score and the cultural assessment was tested. A high correlation would indicate that this assumption would potentially be true.

3.4.5 Reliability assessment

For each practice and level additional more specified descriptions were included in the survey (see appendix). This increased reliability of the self-rating as all respondents had less room for interpretation. However, as the aim of the description was to allow the rater to interpret if their organisation has a similar practice that is called differently, this might decrease reliability of the findings. This can be seen as a trade-off between reach and reliability, because with more specific descriptions the reliability could be improved, but at the same time more raters would state that their organisation does not have this practice. This is a result of the many different options to implement agile practices. In general, the reliability of the results from the pre-test is diminished by the self-rating approach (Rossiter, 2002).

Due to time constraints given by the nature of a thesis (i.e., fixed hand-in date), a reliability assessment based on test-retest was not possible. However, this would have supported the reliability assessment of the formative model.

The reliability of the second part of the measurement model can be statistically assessed by Cronbach's alpha (MacKenzie et al., 2011), which measures internal consistency among the items. This is appropriate, because the reflective nature of the model implies that the individual items are consistent (DeVellis, 2012). However, when analysing the Cronbach's alpha, it is important to critically review the results by subjectively assess them in relation to theory especially with small sample size as it means that alpha is unstable. The result of a calculation of Cronbach's alpha of the items is presented in chapter 4.2.

4 Results

The purpose of this study is to investigate how the agile capabilities of organisations can be assessed. In order to develop a model that reliably measures agile capabilities, a series of methods were applied. The following section contains the results of the development of the measurement model (chapter 4.1) by presenting the results of the practices identification process, the scale to measure the individual practice's capability, as well as the indicators for the organisation's culture. The results of the empirical pre-test are presented in chapter 4.2. It contains the results of the empirical pre-test as well as the results of a quantitative reliability and validity assessment of the model.

4.1 Findings from Item Generation and Scale Development

In the following section the results of the development of the measurement model based on the analysis of the most relevant frameworks, the Spotify case-study, and the expert interviews are presented.

4.1.1 Measuring Agile Practices

Based on the analysis of six frameworks and one case study as well as the expert interviews, eighteen practices that lead to scaled agile capabilities were identified, resulting in a formative measurement model. These practices can be grouped into one of three domains: *team*, *program*, and *portfolio*. These domains are based on the structure that can be found in the SAFe and RAGE. This structure was confirmed by one of the interviewees as he stated that "organizations working with agile […] use these levels, as an internal model of how they are organized. So that's […] a good thing that you chip into that." (Interviewee 3, personal communication, July 4, 2019, 24:05). Additionally, Interviewee 5 confirmed that leaving out the 'large configuration' of SAFe is reasonable as it is mainly an extension of the program domain (min 1:06). The team domain contains the practices that are applied in the development team to ensure an agile development process. The program domain contains the practices that are applied on a program level to allow an agile development process with more than a single team by ensuring alignment among the teams as well as supporting practices used for additional coordination and support purposes for instance. The portfolio domain contains practices applied at a portfolio level to ensure that the portfolio's development effort goes in the same direction and meet strategic objectives.

The team domain's practices are *agile coach*, *product owner*, *daily meeting*, *iteration planning*, *iteration review*, *iteration retrospective*, and *innovation time*. The program domain's practices are *program owner*, *scrum-of-scrum meeting*, *demo day*, *program retrospective*, *develop on cadence*, *systems architect*, and *communities of practice*. The portfolio domain's practices are *portfolio strategy*, *coordinate portfolio value stream*, *lean budgeting*, and *discover opportunities*. These practices were all presented with an mouse-over that described the general purpose of the practice to allow the raters to interpret if their organisation has a practice that fulfils these description but that might be called differently in their organisation (see Table 6 for how the practices were described in the survey). This was necessary, because might give the practice a different name (e.g., 'agile coach' might be called 'scrum master').

Table 6: Description of Practices in Survey

Domain	Practice	Description in Survey
Team	Agile Coach	A person that supports the team to overcome obstacles and facilitates the development process. An agile coach does not delegate tasks. Example: scrum master.
	Product Owner	A person that provides the tasks to the team by communicating what needs to be done. However, a product owner does not determine the way a team works. It provides the connection to the business side and is responsible for the product backlog.
	Daily Meeting	Daily meeting is used to coordinate work and share information among team members. Team members briefly present what happened, what are obstacles, and what is planned for the day. Only information relevant for everybody is shared. Example: daily scrum
	Iteration Planning	During iteration planning (i.e., sprint planning) the goal for the next development iteration is defined. The product owner presents the requirements and the team forecasts how much customer value it can deliver in the next iteration.
	Iteration Review	During iteration review, results of the last iteration are presented to the product owner and stakeholder to get feedback. The product owner decides whether the requirements are fulfilled.
	Iteration Retrospective	During iteration retrospective, the team reflects on the last iteration. The team identifies areas of improvement and defines way to improve. The scrum master supports this reflection.
	Innovation Time	In innovation time the teams explore, develop, and experiment with ideas they think could be useful, but that are not prescribed from a higher authority. This could take place in dedicated time each iteration (or week) or for example every third iteration.

Program	Program Owner	The program owner works on similar tasks as a product owner, but on a higher level. They identify customer needs, prioritize features, and provide the program vision. The equivalent practice in LeSS is role is "Area Product Owner" and in SAFe.
	Scrum-of-Scrum Meeting	Every team sends one representative to this meeting. Similar to the daily meetings on team level, the participants present what happened, what are obstacles, and what is planned for the day. Only information relevant for everybody is shared.
	Demo Day	During demo day teams present their results developed in the last iteration. It provides the opportunity to gather feedback from a diverse group of stakeholders (e.g., business owners or customers). Further it helps teams to stay informed about what other teams are working on. In SAFe The equivalent practice in SAFe is "System Demo" and at Spotify During demo day teams present their results developed in the last iteration. It provides the opportunity to gather feedback from a diverse group of stakeholders (e.g., business owners or customers). Further it helps teams to stay informed about what other teams are working on. In SAFe The equivalent practice in SAFe is "System Demo" and at Spotify "Weekly Demo".
	Program Retrospective	During the program retrospective, the teams reflects on the program's processes. The participants identify areas of improvement and define ways to improve.
	Develop on Cadence	Develop on cadence ensures that development of solutions takes place in a predictable way by synchronising events such as scrum-of-scrum meeting, demo day, and program retrospective. This leads to a coordinated development cycle in a program.
	Systems Architect	The system architect ensures that new developed systems fit to the existing components, code, and architecture. It develops and communicates a shared technical and architectural vision.
	Communities of Practise	Communities of practise are organized groups of people who have a common interest in a specific technical or business domain. They collaborate regularly to share information, improve their skills, and actively work on advancing the general knowledge of the domain.
Portfolio	Portfolio Strategy	The portfolio strategy provides the overall guidance to the portfolio. Decisions made in the portfolio should be based on the portfolio strategy. It is derived from the organisation's strategy.
	Coordinate Portfolio Value Stream	A portfolio consists of different value streams. These can be organised for example based on internal or external user or the concrete product they are working on. This increases efficiency of the portfolio.
	Lean Budgeting	Lean budgeting allows more adaptable funding of projects than traditional approaches to budgeting. It gives flexibility by for example re-allocating budget.
	Discover Opportunities	Discover opportunities ensures that the portfolio keeps in touch with external development by for example observing the market/competition or emerging technologies. Thereby, it identifies potential opportunities to generate value.

To assess the capability of the different practices a generic scale from the COBIT 4.1 framework was adopted (see Table 3). COBIT 4.1 is based on a traditional CMMI scale. In contrast to other scales of the CMMI family and later versions of COBIT, it sees a practice as being performed only on higher levels, whereas other frameworks already define the first or second level as being an

effective practice (Pasquini & Galiè, 2013). Interviewee 5 stated: “I have the impression that the new one is optimised to give you the good results you want instead of give you the exact reality.” (2019, min 2:05).

Furthermore, the scale is extended by the option that a practice is ‘non-existent’ as this is important when measuring. Additionally, the meaning of the highest level was slightly altered based on Fontana et al.’s (2015) proposal that the highest level of agility should be one of continuously improving the practice not on external best practice, but on the individual organisation’s experience. This is in line with the Spotify case-study, which shows that agility means continuously develop practices internally based on experience made in the organisation with different ways of performing a practice. Additionally, Interviewee stated that “one of the most important things […] in an agile organisation is that we continuously improve and learn” (2019, min 11:45).

Table 7: Generic Description of Practices' Levels (Adopted from COBIT 4.1)

Level	**Description**
0: Non-existent	Lack of any recognisable practice in the area.
1: Initial/Ad Hoc	There are no standardised practices; instead, there are ad hoc approaches that tend to be applied on an individual or case-by-case basis. The overall approach is disorganised.
2: Repeatable but Intuitive	Practices have developed to the stage where similar practices are followed by different people undertaking the same task. There is no formal training or communication of standard procedures, and responsibility is left to the individual. There is a high degree of reliance on the knowledge of individuals and, therefore, errors are likely.
3: Defined	Practices have been standardised and documented. It is expected that these practices should be followed; however, it is unlikely that deviations will be detected. The practices themselves are not sophisticated but are the formalisation of existing practices.
4: Managed and Measurable	Compliance with practices is monitored and measured; action is taken where practices appear not to be working effectively.
5: Optimised	Practices are constantly refined, based on the results of continuous improvement and experience within the context they are applied making the organisation and the practices quick to adapt. Inspiration for how to improve can come from inside and outside the organisation (i.e., internal and external best practices).

The reasoning behind the scale is that if we think of how an organisation can make a transition from a traditional to more agile ways of organising it seems to be normal that at the beginning practices might only be performed on an ad-hoc basis, which leads to more repetitive patterns in the practices. To reach the next level, it helps organisations to actually write down or define these practices based on how they perform them, because this helps becoming aware of how the practices are performed and consequently supports reflection. In a next step, to actually see the impact of the practices on the organisations it is necessary to perform them following a defined pattern (i.e., *Level 4*). Only if a pattern is followed for a certain time, it is possible to assess the actual impact of this way of performing the practice. This results in improving the practice by adjusting it to the organisation's reality (i.e., *Level 5*). To be on a *Level 5*, it would additionally necessary to constantly improve practices for example by regular reflection on it in the retrospectives.

4.1.2 Practices and Culture

The expert interviews revealed that a focus on practices in assessing agile capabilities is not sufficient. For a complete picture of the agile capability of an organisation it is also necessary to assess the culture of an organisation. Interviewee 3 explains based on his experience:

> Just measuring the practices that wouldn't be enough, because we have seen a lot of organisations that do the practices: they do the daily stand-up, they have sprint planning, they also have product owner roles, but they're not really working agile.
> (2019, min 11:45).

However, in the context of transforming larger organisations, Interviewee 2 states that "you need both to understand how to think differently, and you also need to have different practices" (2019, 29:24) for a successful implementation of agile. Furthermore, in the context of measuring agile capability Interviewee 3 stated: "if you only assess this the practices, then you miss some important things here" (Interviewee 3, 2019, 36:43), which indicates the importance of measuring both practices and culture. Interviewee 2 even stated that "I don't think there should be any practices necessarily" (2019, 00:00) to measure agility, as for him they do not indicate agility. He even states, "I have seen, very successful teams without a product owner or a scrum master" (min 32:45). On the contrary, "it is actually more common" (min 15:54) to him that organisations adopt practices that the organisation thinks are agile but lack the agile mindset". This seems to be in-line with Interviewee 6 who said:

> You are on the right track, if you not only measure 'how well do you perform the practices as specified in the methodology?' I do think it's definitely worthwhile to try to get data points on 'Do you follow the methodology?' But I think where it can become really, really interesting is to kind of cross reference: 'How do you follow the methodologies?' versus 'how well are you doing on the underlying principles?'
> (Interviewee 6, personal communication, July 18, 2019, min 7:18).

Interviewee 4 explains that organisations that want to make a successful agile transformation should focus more on the values and principles of agility but sees the practices as "the way to start" (personal communication, July 9, 2019, min 6:25). Interviewee 2 in its second interview explains that the SAFe approach to change the mindset is "not to focus too much on it, but […] start out by changing the structures and the practices" (personal communication, July 12, 2019, min 3:03). All of these statements indicate that to appropriately measure agility, a combination of practices and culture needs to be assessed, since they seem to influence each other. Furthermore, the statements indicate that to become more agile practices are a reasonable way to start the transformation.

4.1.3 Measuring Agile Culture

Consequently, in a second part of the measurement model, the culture of the organisation regarding agility is assessed. Analysing the values and principles underlying the frameworks lead to identification of seven distinct dimensions of agility that in agile organisations are supposed to be high (Table 8). The values are *adaptability*, *autonomy*, *continuous improvement*, *empiricism*, *systems thinking*, *transparency*, and *value focus*.

These dimensions were covered by two to three statements (see Table 9). Each of these dimensions were measured on a 5-point Likert-scale ranging from *strongly disagree* to *strongly agree* (i.e., *strongly disagree*, *disagree*, *neither agree nor disagree*, *agree*, and *strongly agree*). Thereby, culture is measured on a reflective scale. Translated into numerical values the culture was measured on a scale from -2 to +2 and the overall score was the average of the responses.

Table 8: Cultural Dimensions in Frameworks, Agile Manifesto, and Interviews[14]

	SAFe	LeSS	DA	Nexus	S@S	RAGE	Spotify	Agile Manifesto	Interviews
Adaptability	x		x		x	x	x	x	x
Autonomy	x		x				x	x	x
Continuous Improvement	x	x	x		x		x	x	x
Empiricism	x	x	x		x		x		
Systems Thinking	x	x	x				x		x
Transparency	x	x	x		x		x		x
Value Focus	x	x	x		x	x	x	x	x

Table 9: Measurement Statements for Agile Culture

Dimension of Agile Culture	**Measurement Statement**
Adaptability	a) Processes in my organisation can be adjusted to changing requirements quickly. b) My organisation can implement user feedback quickly, even if that changes the initial plan.
Autonomy	a) Decision making is decentralised: my team can make decisions without consulting higher authorities. b) My team is able to cover the whole development cycle (from ideation to deployment) for a product without support from outside the team. c) Projects can be started without management approval.
Continuous Improvement	a) Employee development is supported, even in areas that are not directly related to the job. b) An error has occurred: my organisation makes sure the lesson is learned (instead of just solving the problem). c) Topics that require improvement, are actually improved and not just documented.
Empiricism	a) Decisions are made based on practical evidence (not theory). b) We experiment a lot. c) My organisation prefers ‘Trial & Error’ over extensive planning.
Systems Thinking	a) If it is in my organisation's interest, our team would "lend" our best programmer to another team, even if that means delays for my team. b) My KPIs encourage me to consider impact for the whole organisation when I make a decision.
Transparency	a) Failures (e.g., individual mistakes) are openly shared. b) Financial data of my organisation is unconditionally available to all employees.
Value Focus	a) The most important thing is to provide measurable value to the user. b) Results are more important than process or documentation.

[14] For statements indicating the dimension in a framework see appendix.

4.2 Results of Pre-Test

The following section presents the results of the pre-test. A total of seventy-four responses were recorded. Out of these fifty-four cases were eliminated from the data set: twenty-eight cases did not include any responses as Qualtrics also records all incomplete cases. In twenty-five cases the responses for the practices were incomplete. One case showed a suspicious response pattern for the practices and missed the culture part. Additionally, one case showed a suspicious response pattern for the culture part (i.e., all answers were *strongly disagree*), which was consequently excluded from the part of the analysis where the cultural part was involved, but kept for the practices analysis since no suspicious pattern was observed for this part. All in all, the pre-test generated twenty valid responses for the practices part of the questionnaire ($N = 20$) and nineteen valid responses for the culture part ($N = 19$).

4.2.1 Qualitative Feedback

Members of the social media groups where the survey was shared commented on the survey[15]. One person indicated that measuring agility is impossible or at least counter-productive to the agility of an organisation:

> I am afraid the general consensus in the Agile community is still that trying to measure agility is like trying to measure how wet water is. Or maybe better compared to measuring happiness (hint: research shows that measuring happiness actually makes people unhappy). It is impossible to measure agility; it is impossible to report a status-quo. Trying to do so will result in a less agile organisation. (K. van der Pasch, comment on Facebook, August 21, 2019)

This view was also shared by a member of a group on LinkedIn who summarised his comments by stating:

> Hopefully you are beginning to realize by now that the idea is either a lot harder than you thought or completely futile. I'd go for the latter, if it was a viable idea, we'd already have an 'Agile Maturity' assessment." (P. Oldfield, comment on LinkedIn, August 8, 2019).

[15] Screenshots of all comments can be found in appendix.

4.2.1 Results of Practices' Assessment

The mean score for all practices was 40.60 (*SD* = 15.288). The highest mean of 3.05 (*SD* = 0.759) was observed for the 'Daily Meeting' practice, followed by a mean of 2.85 (*SD* = 1.226) for the 'Iteration Planning' practice, and a mean of 2.55 (*SD* = 1.468) for the 'Systems Architect' practice. The lowest mean of 1.70 ((*SD* = 1.525) was observed for the 'Develop on Cadence' practice, followed by a mean of 1.80 (*SD* = 1.436) for the 'Program Retrospective' practice, and a mean of 1.90 was observed for 'Innovation Time' (*SD* = 1.373) and 'Scrum-of-Scrum Meeting' (*SD* = 1.586).

Table 10: Results of Practices Assessment

Domain	Practice	Mean	SD	VIF
Team	Agile Coach	2.15	1.348	8.253
	Product Owner	2.75	1.020	23.577
	Daily Meeting	3.05	0.759	6.057
	Iteration Planning	2.85	1.226	50.348
	Iteration Review	2.50	1.357	39.844
	Iteration Retrospective	2.60	1.429	45.086
	Innovation Time	1.90	1.373	7.350
Program	Program Owner	2.15	1.309	8.401
	Scrum-of-Scrum Meeting	1.90	1.586	63.318
	Demo Day	2.10	1.619	32.167
	Program Retrospective	1.80	1.436	100.639
	Develop on Cadence	1.70	1.525	53.561
	Systems Architect	2.55	1.468	20.862
	Community of Practice	2.20	1.322	42.759
Portfolio	Portfolio Strategy	2.30	1.490	76.805
	Coordinate Portfolio Value Stream	2.10	1.447	28.203
	Lean Budgeting	1.95	1.234	32.578
	Discover Opportunities	2.05	1.146	17.930
Overall		40.60	15.288	-

Testing the Assumptions of Multiple Linear Regression

All relationships are approximately linear as assessed by visual inspection of the plots of studentized residuals versus unstandardized predicted values, and the dependent variable (i.e., practices' score) versus the individual independent variables (see appendix).

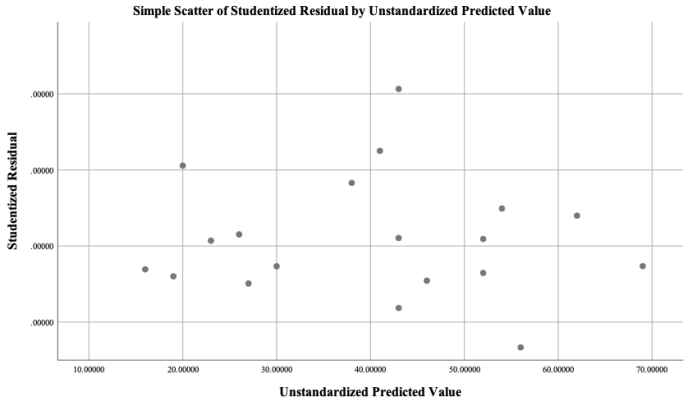

Figure 1. Simple Scatter of Studentized Residual by Unstandardized Predicted Value

There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values (Figure 1).

An assessment of multicollinearity of the model (among the items) was conducted to investigate if two or more of the practices are highly correlated with each other. Only the practices 'Agile Coach', 'Daily Meeting', 'Innovation Time', and 'Program Owner' were below the suggested threshold of VIF < 10 (see Table 10). Hence, as suggested by (Hair et al., 2014), the practices were grouped into the domain 'team', 'program', and 'portfolio' as informed by the frameworks' structure. The grouping of the practices into the individual domains resulted in significant lower VIFs among the practices in the domains as well as between the domains (see Tables 11-14). As the acceptable level is VIF < 10, all the domains as well as the practices contributing to the domains are relevant.

The grouping of the practices resulted in a higher-order model that consists of the three domains as the lower-order component that together form the higher-order component 'agile capability'. Since there is no indication in literature that one of the domains contributes more to the agile capability of an organisation, all three domains were regarded as contributing equally to the overall score. This was achieved by weighting the scores of the domains.

Table 11: VIF and Correlations for Practices in Team Domain

Practice	VIF	Correlations						
		Agile Coach	Product Owner	Daily Meeting	Iteration Planning	Iteration Review	Iteration Retrospective	Innovation Time
Agile Coach	1.461	1.000						
Product Owner	1.879	0.526	1.000					
Daily Meeting	1.445	-0.008	0.153	1.000				
Iteration Planning	2.834	0.110	0.095	0.291	1.000			
Iteration Review	5.216	0.072	-0.057	0.332	0.712	1.000		
Iteration Retrospective	7.813	0.005	-0.108	0.359	0.775	0.895	1.000	
Innovation Time	1.922	0.293	0.282	-0.045	0.366	0.480	0.488	1.000

Table 12: VIF and Correlations for Practices in Program Domain

Practice	VIF	Correlations						
		Program Owner	Scrum-of-Scrum Meeting	Demo Day	Program Retrospective	Develop on Cadence	System Architect	Communities of Practice
Program Owner	2.708	1.000						
Scrum-of-Scrum Meeting	3.350	0.565	1.000					
Demo Day	4.224	0.663	0.639	1.000				
Program Retrospective	4.131	0.521	0.753	0.801	1.000			
Develop on Cadence	6.023	0.287	0.683	0.588	0.644	1.000		
System Architect	2.904	0.420	0.522	0.507	0.454	0.759	1.000	
Communities of Practice	3.063	0.103	0.437	0.482	0.466	0.788	0.618	1.000

A second indicator for issues with correlation is the correlation between individual items with a correlation > 0.7. In the team domain, the correlation between ‘Iteration Planning’ and ‘Iteration Review’, between ‘Iteration Planning’ and ‘Iteration Retrospective’, and ‘Iteration Review’ and

'Iteration Retrospective' were above this threshold. This could indicate that these practices are related, which fits to their role in the frameworks and thereby would indicate a certain degree of nomological validity for the model. However, this could also indicate that the respondents were not able to appropriately differentiate these practices, which would make a refinement of their description necessary to ensure reliability of the assessment of these practices.

In the program domain, the correlation between 'Program Retrospective' and 'Scrum-of-Scrum Meeting', between 'Program Retrospective' and 'Demo Day', between 'System Architect' and 'Develop on Cadence', and between 'Communities of Practice' and 'Develop on Cadence' were above the threshold of 0.7. As for the team domain, the correlations between 'Demo Day', 'Program Retrospective', and 'Scrum-of-Scrum Meeting' might be explained by their actual relation to each other. However, for the 'Develop on Cadence' correlation with 'System Architect' and 'Community of Practice' no such obvious explanation can be given. 'Develop on Cadence' also has the highest multicollinearity (VIF = 6.023) in the program domain.

Table 13: VIF and Correlations for Practices in Portfolio Domain

Practice	**VIF**	**Correlations**			
		Portfolio Strategy	Coordinate Portfolio Value Stream	Lean Budgeting	Discover Opportunities
Portfolio Strategy	7.873	1.000			
Coordinate Portfolio Value Stream	4.503	0.864	1.000		
Lean Budgeting	2.744	0.752	0.533	1.000	
Discover Opportunities	1.785	0.638	0.505	0.597	1.000

Table 14: VIF and Correlations for Team, Program, and Portfolio Domain

Domain	**Mean**	**SD**	**VIF**	**Correlations**		
				Team	Program	Portfolio
Team	15.26	4.700	1.376	1.000		
Program	12.34	6.977	2.853	0.508	1.000	
Portfolio	12.60	6.889	2.324	0.298	0.749	1.000
Overall	40.20	15.582	-	-	-	-

In the portfolio domain, the correlation between 'Portfolio Strategy' and 'Coordinate Portfolio Value Stream', and between 'Portfolio Strategy' and 'Lean Budgeting' is above the threshold. 'Portfolio Strategy' also has the highest multicollinearity (VIF = 7.873) in the portfolio domain.

For the higher-order construct, only the correlation between the program and portfolio domain were above the threshold of 0.7.

4.2.2 Analysis of Culture Assessment

The mean score for all agile culture statements was 0.161 (*SD* = 9.503), indicating that the average response was *neither agree nor disagree*. The scale had a high internal consistency, as determined by a Cronbach's alpha of 0.813, based on a recommended Cronbach's alpha of > 0.7 (DeVellis, 2012).

Table 15: Results of Culture Assessment

Dimension (Statement)		**Mean**	**SD**
Adaptability	a)	-0.053	1.129
	b)	0.368	0.895
Autonomy	a)	0.474	0.841
	b)	0.211	1.228
	c)	-0.421	1.071
Continuous Improvement	a)	0.211	1.134
	b)	0.053	1.129
	c)	0.000	1.00
Empiricism	a)	0.316	1.376
	b)	0.263	1.195
	c)	0.158	1.068
Systems Thinking	a)	0.211	1.084
	b)	-0.211	1.032
Transparency	a)	0.263	1.098
	b)	-0.263	1.408
Value Focus	a)	0.632	1.116
	b)	0.526	1.020
Total		2.737[16]	9.503
Cronbach's Alpha	0.813		

[16] Mean for the scale as a whole (sum of the average of each item).

The suitability of the data set to conduct an exploratory factor analysis with all statements was assessed. Inspection of the correlation matrix showed that all variables had at least one correlation coefficient greater than 0.3. The overall Kaiser-Meyer-Olkin (KMO) measure was 0.349, which is unacceptable according to the classification of Kaiser (1974). The analysis of the KMO measures for the individual variables (Table 16) showed that only two variables are above the minimum threshold value (KMO > .5) (Laerd Statistics, 2015c). Together, this indicates that the data set is not appropriate for an exploratory factor analysis.

Table 16: KMO measures for Cultural Statements

Variable	KMO Measure
1a)	0.307
1b)	0.269
2a)	0.451
2b)	0.241
2c)	0.799
3a)	0.473
3b)	0.356
3c)	0.366
4a)	0.452
4b)	0.205
4c)	0.406
5a)	0.446
5b)	0.161
6a)	0.376
6b)	0.249
7a)	0.281
7b)	0.571

As the seventeen statements measured seven dimensions of agile culture (i.e., adaptability, autonomy, continuous improvement, empiricism, systems thinking, transparency, and value focus), the results from the statements were aggregated as otherwise the dimensions with three statements would have had a higher impact than the dimensions measured through two statements. This aggregation took place by averaging the results (i.e., dividing the sum of the two (or three) statements in a dimension by two (or three)). This allowed to obtain a more balanced score for the seven

dimensions (Table 17). The highest mean for the aggregated score of 0.579 (*SD* = 0.786) was observed for the 'Value Focus' dimension. The lowest mean of 0.0 was observed for 'Systems Thinking' (*SD* = 0.646) and 'Transparency' (*SD* = 1.080). The scale with aggregated dimensions had lower, but still acceptable internal consistency, as determined by a Cronbach's alpha of 0.767.

Table 17: Cronbach's alpha for Culture Dimensions

Dimension	**Mean**	**SD**	**R^2**	**Corrected Item-Total Correlation**	**Cronbach's Alpha if Item Deleted**
Adaptability	0.158	0.851	0.393	0.359	0.764
Autonomy	0.088	0.736	0.429	0.581	0.722
Continuous Improvement	0.088	0.908	0.592	0.681	0.693
Empiricism	0.246	0.902	0.419	0.387	0.760
Systems Thinking	0.0	0.646	0.341	0.268	0.774
Transparency	0.0	1.080	0.584	0.524	0.733
Value Focus	0.579	0.786	0.476	0.652	0.706
Overall	1.579	3.859			
Cronbach's Alpha	0.767				

As all the dimensions are supposed to measure the same underlying construct (i.e., agile culture), it is expected that they show a relative high correlation with the sum of all the other dimensions, as determined by the 'Corrected Item-Total Correlation'. The threshold-value is 0.3 for indicating that an item is measuring not the same construct (Laerd Statistics, 2015a). For the 'Systems Thinking' dimension the coefficient correlation is only 0.268, which indicates that this dimension might measure a different construct. Consequently, if 'Systems Thinking' was excluded from the model, the Cronbach's alpha would slightly increase to 0.774. Excluding any other dimension would decrease Cronbach's alpha slightly. Furthermore, if the variation in the dimensions is based on differences in the construct 'agile culture' than the items should be able to explain the variation in the other items. Therefore, it is preferred that the items have large values for the squared multiple correlation (R^2), as it indicates the items can be explained by each other (Laerd Statistics, 2015a). Again, the lowest score of 0.341 can be observed for 'Systems Thinking'. Together, this indicates that the dimension 'Systems Thinking' needs further attention.

4.2.3 Analysis of Relationship of Practices and Culture

For the purpose of assessing the convergent validity of the two models, there relationship was explored. In a first step, the data points from both the practice and culture assessment were visually inspected for linear relationship and outliers.

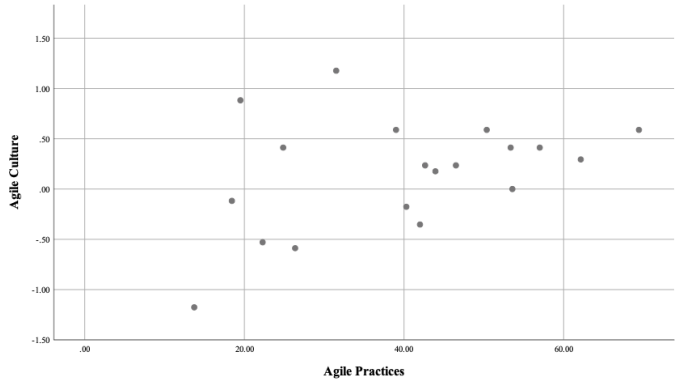

Figure 2. Simple Scatter of Culture by Practices

The visual inspection showed that there could be a linear relationship with outliers characterised by relative high values for agile culture. Due to the small sample size, the outliers were kept for further analysis. Next, the results from both the practice and culture part were tested for normality to assess if the data is appropriate for further analysis. Agile practices scores were normally distributed with a skewness of 0.013 (*SE* = 0.524) and kurtosis of -0.898 (*SE* = 1.014). Agile culture scores were also normally distributed with a skewness of -0.581 (*SE* = 0.524) and kurtosis of 0.621 (*SE* = 1.014). Based on the Shapiro-Wilk's test, which is explicitly suitable for small sample size (Laerd Statistics, 2018), agile practices and culture scores were normally distributed ($p > .05$), too.

Table 18: Assessment of Normality

		Statistic	Std. Error	z-score
Agile Practices	Skewness	0.013	0.524	0.025
	Kurtosis	-0.898	1.014	-0.886
Agile Culture	Skewness	-0.581	0.524	-1.109
	Kurtosis	0.621	1.014	0.612

Further, both scores for both agile practices and agile culture were normally distributed, as assessed by visual inspection of Normal Q-Q Plots.

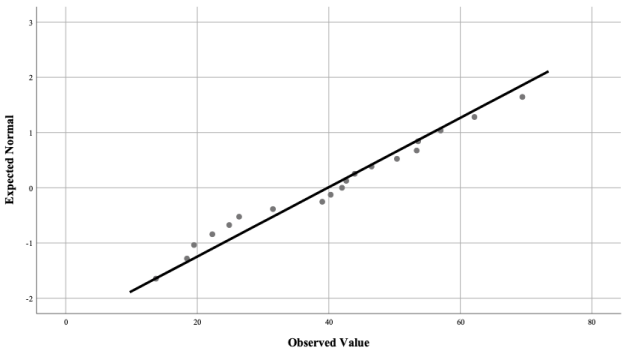

Figure 3. Normal Q-Q Plot of Agile Practices

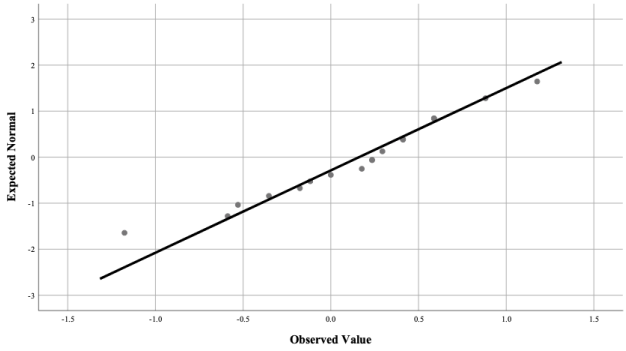

Figure 4. Normal Q-Q Plot of Agile Culture

A Pearson's product-moment correlation was run to assess the relationship between agile practices and agile cultures among nineteen organisations. Preliminary analysis showed the relationship to be linear with both variables normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). There was a no statistically significant moderate positive correlation between the score for agile practices and agile culture, $r(17) = .381$, with agile practices explaining 15% of the variability in agile culture (Table 19).

Summarizing, the analysis indicates that there is a mild correlation between the scores for agile culture and agile practices for the small sample (N=19) from the pre-test. If we assume that both models measure agile capability, this indicates a certain degree of convergent validity for the models.

Table 19: Pearson Correlation for Practices and Culture

		Agile Practices	Agile Culture
Agile Practices	Pearson Correlation	1	0.381
	Sig. (2-tailed)		0.108
Agile Culture	Pearson Correlation	0.381	1
	Sig. (2-tailed)	0.108	

In order to further analyse the relationship between agile practices and culture, the correlation of the individual practice domains on culture were assessed. Visual inspection showed that there could be a linear relationship for agile culture by program (Figure 6) and portfolio domain (Figure 7), but not for team domain (Figure 5).

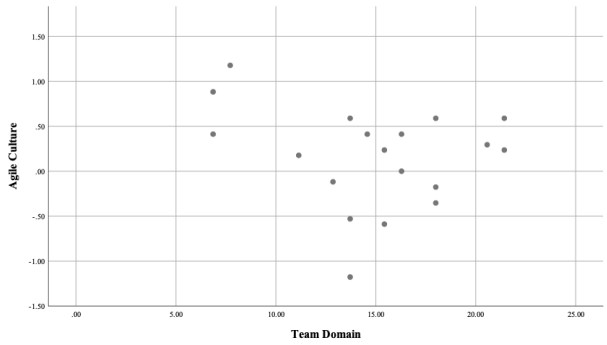

Figure 5. Simple Scatter of Culture by Team

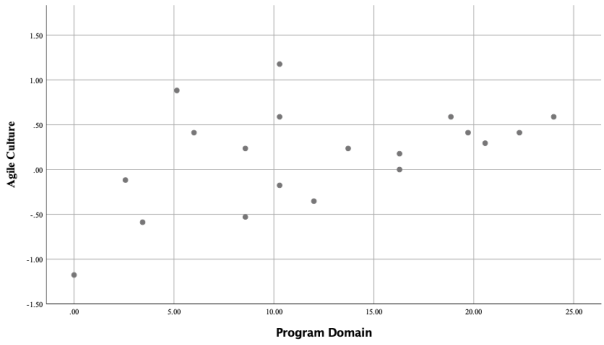

Figure 6. Simple Scatter of Culture by Program

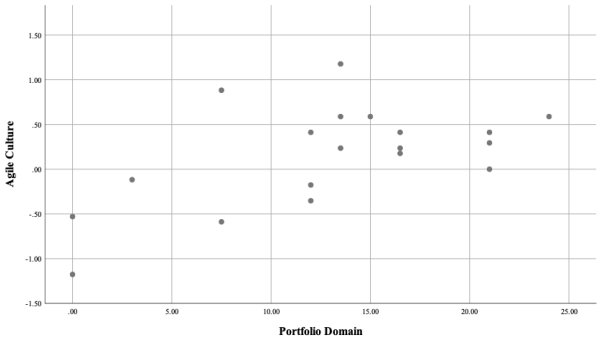

Figure 7. Simple Scatter of Culture by Portfolio

Next, the results from program and portfolio domains were tested for normality to assess if the data is appropriate for further analysis (Table 20). Program domain scores were normally distributed with a skewness of 0.078 (*SE* = 0.524) and kurtosis of -0.992 (*SE* = 1.014). Agile culture scores were also normally distributed with a skewness of -0.491 (*SE* = 0.524) and kurtosis of -0.286 (*SE* = 1.014). Based on the Shapiro-Wilk's test, which is explicitly suitable for small sample size (Laerd Statistics, 2018), agile practices and culture scores were normally distributed ($p > .05$), too.

Table 20: Assessment of Normality for Program and Portfolio Domain

		Statistic	Std. Error	z-score
Program Domain	Skewness	0.078	0.524	0.149
	Kurtosis	-0.992	1.014	-0.979
Portfolio Domain	Skewness	-0.491	0.524	-0.936
	Kurtosis	-0.286	1.014	0.282

Further, scores for both program and portfolio domain were normally distributed, as assessed by visual inspection of Normal Q-Q Plots (Figure 8 and Figure 9).

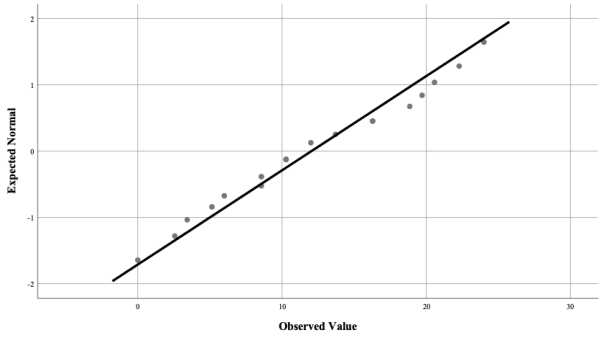

Figure 8. Normal Q-Q Plot of Portfolio Domain

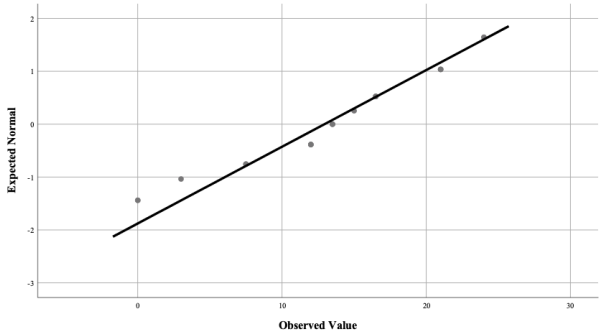

Figure 9. Normal Q-Q Plot of Program Domain

A Pearson's product-moment correlation was run to assess the relationship between agile culture and program domain as well as between agile culture and program domain among nineteen organisations. Preliminary analysis showed the relationships to be linear with all variables normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). There was a no statistically significant moderate positive correlation between the score for program domain and agile culture, $r(17) = .453$, with agile practices explaining 21% of the variability in agile culture (Table 21).

Table 21: Pearson Correlation for Agile Culture and Program Domain

		Agile Culture	Program Domain
Agile Culture	Pearson Correlation	1	0.453
	Sig. (2-tailed)		0.051
Program Domain	Pearson Correlation	0.453	1
	Sig. (2-tailed)	0.051	

There was a statistically significant strong positive correlation between the score for portfolio domain and agile culture, $r(17) = .565$, with agile practices explaining 32% of the variability in agile culture (Table 22).

Table 22: Pearson Correlation for Agile Culture and Program Domain

		Agile Culture	Program Domain
Agile Culture	Pearson Correlation	1	0.565
	Sig. (2-tailed)		0.012
Program Domain	Pearson Correlation	0.565	1
	Sig. (2-tailed)	0.012	

5 Discussion

The following chapter contains a discussion of the findings of this study and thereby answers the research question of *how can organizations reliably, comprehensively, and parsimoniously measure their agile capability?* In sum, the results show that measuring agile practices should take place on the three different levels of team, program, and portfolio. A scale adopted from the COBIT 4.1 framework could be used to indicate the maturity of the practices. This measurement of agile practices should be complemented by assessing the culture in the organisation to obtain a comprehensive view of the agile capability of an organisation. In the first sub-chapter the findings are discussed in the light of existing theory. Next, the implications for practice are presented.

5.1 Link to Theory

5.1.1 Practices and Culture

The results of this study show that agile capabilities are defined on the two dimensions of organisational practices and culture. In the light of capabilities as "repeatable patterns of action" (Wade & Hulland, 2004, p. 109) that form competitive advantage according to the RBV on organisations, agile capabilities can be conceptualised. The grounding in the RBV leads to the focus on the frameworks and their practices as they provide a foundation for identifying established repeatable patterns in the form of prescriptive practices. Extending the focus from practices to include culture in the assessment is necessary as the interviews revealed that the effectiveness of practices is not guaranteed by fulfilling them, but also influenced by underlying organisational culture. It can be said that visible practices are only part of the capabilities that needs to be extended by the suitable organisational culture. The importance of a suitable agile culture is further underlined as it is part of the frameworks as represented by 'agile mindset' or 'agile values' underlying the frameworks.

To illustrate this point, think of a meeting situation where a team aims to reflect about potential areas of improvement. Even if the meeting is conducted in a way that is suggested by external frameworks and works well in many other organisations, it could be that the goal of the meeting is missed because people refuse to be transparent about their failures. This might be because the employees, even though they are asked about failure, fear personal consequences such as embracement for instance and consequently do not share all aspects of the failure. Without the appropriate identification of the problem, it is impossible to find an appropriate solution. This

however fits the findings from a study by Misra, Kumar, and Kumar (2010) that concluded that adopting agile may require changes in organisational culture.

This interplay of culture and practices can also be understood in the light of Feldman and Pentland (2003) that describe capabilities as having an ostensive and performative aspect, where the practice (and its definition or related instruction) provides the ostensive aspect and the culture the performative aspect. The main argument here is that even if the practices are formally and correctly applied, it can happen that due to the organisational culture they are not effective. This is underlined by one of the interviewees stated that an organisation might be very agile without any or very few of the practices.

Additionally, it seems to be that the two aspects influence each other. However, it remains unclear in which way this influence takes place. One interviewee revealed that he has seen both organisations with an agile culture being effective in their way of working without any of the practices, but also organisations that apply all the practices presented in one of the frameworks without being agile. Based on this, it is indicated that agile capability can be formed by the practices, but without a suitable culture this might not be effective, which would indicate a moderating relationship (Hair Junior et al., 2014). From an agile transformation perspective, implementing agile practices could be a way to change the culture, as indicated by the expert interviews. On the other hand, it could also be that an agile culture leads to suitable practices that could also be different from the practices presented in the frameworks.

In terms of the measurement model for agile practices this would result in a level five assessment without being in the lower levels before, which to a certain degree would violate the assumptions of the scale. To make this more complicated, the results from the pre-test indicated a slightly positive correlation between practices and culture, which would mean that they both increase and decrease in a similar way and thereby not fitting to any of the above stated assumptions. Analysing the correlation between the three domains and organisational culture showed that portfolio domain has the highest correlation of the three with the results from the cultural assessment. This could indicate that culture is more influenced by the portfolio domain than the team or program domain. One reason could be that two of the practices (i.e., ‘Lean Budgeting’ and ‘Discover Opportunities’) are likely to fulfil the requirements of the taxonomy of ISD by Conboy (2009) most of the time. Another interpretation could be that agile culture is determined more by the higher level in an organisation, which would indicate that for an agile transformation of an organisation, a top-down approach might be more suitable than a bottom-up approach in terms of cultural change.

However, there is no clear evidence whether top-down or bottom-up is more favourable (Conboy & Carroll, 2019).

In summary, it can be concluded that for adopting agile it is important to be aware of both practices and cultural aspects as they influence each other and the outcome of a transformation.

5.1.2 Practices for Agile Transformation

The assumption of this research is that measuring agile capability of an organisation provides a certain value. This was challenged by the responses to the survey that stated measuring agility in fact makes an organisation less agile. Referring to the argument before, one could say that in order to enable agility beyond a single team measuring agility actually supports the implementation and thereby it indirectly contributes to agile capabilities. As interviewee 3 stated: "one way of using assessments is to look at it today and then make an assessment. And then later on, take up again, to see if we have improved something" (2019, min 6:15). This highlights that it is important how an assessment tool would be used, which fits to the findings of Conboy (2009) who found that methods might not be agile in every case, but that it depends on the actual implementation. Another aspect here is that the difference between measuring agility on the one hand is seen as decreasing agility, on the other hand the frameworks stress that empiricism and experimentation are important aspects in agile development. Since experimentation requires a certain degree of documentation and measuring outcome to make an appropriate decision and the need to continuously improve internal practices are both inherently agile, we can see these as true assumptions. If one sees the implementation of agile practices as an ongoing experiment, the conclusion is that a certain degree of documentation and measuring is inherent to the implementation of agile practices. To make this point clearer, think about the situation that an organisation might want to continuously improve (an inherent characteristic of an agile organisation), but does not measure and document its progress (the experiment). This would mean it decides on the practices that works best for them (or the way they implement the practice) based on gut feeling. Based on this, it can be concluded that measuring as such is not a problem, but the actual implementation of the measurement. This is also in line with the agile manifesto as it does not forbid documentation, but instead states that it should not overwhelm.

Assuming that measuring agile capability in fact decreases agility of an organisation would mean that the concept of the frameworks implies that they cannot result in an agile organisation. This is, because implementing the practices in the organisations requires a certain degree of control over

the implementation process, which following the above stated logic, violates agile principles. This string of thought leads to the next discussion point.

In a formative measurement model, the practices define the underlying construct (Rossiter, 2002) and therefore, it is necessary that all practices individually and aggregated contribute to agile capability. However, comparing the practices to the ISD taxonomy from Conboy (2009) indicates that not all the practices under all circumstances contribute agility to an organisation.

As some practices do not provide real agility in the sense of ISD, it is possible to conclude two things: First, as the practices are well established in the frameworks there seems to be a necessity for it. However, 'necessary' does not make it agile. In the sense of the 'alignment enables autonomy' from the Spotify case-study, it is indicated that some ‚non-agile' practices might be necessary for an organisation to enable agility at scale, which in turn makes them an important part of the overall measurement of agility. This could also be fitting to the concept of organisational ambidexterity that states that organisations need to be aligned and efficient while simultaneously being adaptive to change to be successful (Raisch & Birkinshaw, 2008). Thereby, for larger organisations to benefit from agile teams it might be necessary to have practices in place that in itself cannot be seen as agile, but instead enable agility and thereby contribute to agility indirectly. On a more abstract level one could say that being agile alone is not sufficient. This would be a justification for having practices in an organisation that are not agile and at the same time in line with the ambidextrous view. At the same time, one could conclude that the frameworks in fact do not provide instructions for agile software development, but instead provide instructions how the idea of contextual ambidexterity, which is characterised by an organisational context where individuals decide how to deal with different demands (Gibson & Birkinshaw, 2004) can be implemented in the software development process. This point is also supported by the findings from Sailer (2019) who argues that Scrum facilitates ambidexterity.

This in fact provides value to the organisation, as we learn from the ambidexterity literature (He & Wong, 2004). However, in the context of this thesis, this reasoning provides certain issues as it would falsify the underlying assumption that agile capabilities can be measured formatively from measuring agile practices. Consequently, measuring the practices provides an indication for the maturity of the implementation stage of the practices, but they might lead to ambidexterity instead of agility.

Based on the above stated, it can be concluded that the developed model measures the compliance of an organisation with the frameworks, which not necessarily means agility. If the

assumption that the frameworks actually do not address agility in organisations the conclusion is that the model is not measuring agility. However, if the assumption that the practices in the frameworks lead to more agility is correct, the model can be seen as a measure for agility.

5.1.3 Assessing the Practice Maturity

Looking at the way the levels of agile practices are assessed, it seemed that CMMI based scales with their focus on defining the practices might be inappropriate. However, there are two reasons for the choice of this scale: First, due to their popularity the general understanding of their logic is widespread, which increases reliability of the measurement. Second, CMMI based scales provide an established way of describing the maturity of an organisational IT capability, meaning their scale describes how organisations can achieve higher maturity (Paulk et al., 1993). In the case of agility for which some argue that one of the requirements is 'no documentation' – which is not true as explained in the agile manifesto (Kent Beck et al., 2001) – it makes sense to have the steps from ad hoc, over documenting and managing it, towards the highest level where the way the practice is performed is constantly improved based on an organisation's experience. Following this logic, one could argue that only on *Level 5* agility of an organisation is ensured. Another aspect is that organisations adopting one of the frameworks might benefit from first following the prescriptive practices in the frameworks and then develop based on their experience with the practices their own interpretation of the practices (*Level 5*). This makes the developed instrument mostly applicable for organisations that are in the transition process as it actually measures the progress in the transition towards agile practices.

Furthermore, the idea of a maturity model is to provide guidance through presenting best practices. In the field of agility however, "there is only 'good practice that may be best in a given context'" (Oldfield, Facebook comment, 2019). This perspective was also shared by Interviewee 1 who stated: "I do not think that e.g., 'follows best practice' equals effective" (2019). In terms of the proposed scale this is reflected in *Level 5* as it requires that organisations find their own best practice based on their experience; the levels below only support organisations in achieving this.

The rating scale developed for the practices only consists of one attribute. Thereby, the rating becomes coarser compared to a scale with multiple attributes. However, there is a trade-off between comprehensive and detailed in the sense that a finer groined scale would require more effort from the perspective of the rater. As the response rate for the pre-test was already rather low considering the

generally high interest in the topic and the widespread distribution of the survey, a more detailed assessment would potentially require too much time from the respondent.

The findings from the pre-test revealed that in order to measure agile practices the structure of the three domains team, program, and portfolio is appropriate. This structure can be found in SAFe and RAGE, but also reflects the classical organisation (PMI, 2013). Other frameworks such as LeSS do not cover the portfolio domain explicitly but focus on the team and program level. However, they do not exclude or deny the existence of the portfolio level. Thereby, this three-level structure forming a higher-level construct seems to be an appropriate structure for assessing agile practices.

5.1.4 Correlation Among Practices

The high correlation between some practices in the team domain could be interpreted as indicating a certain degree of nomological validity, because the correlated practices represent the practices related to iterations. However, one would not expect that 'iteration review' and 'iteration retrospective' are highly correlated, because they might take place in a similar setting, but fulfil a different function (i.e., one focuses on the product, the other on the process). Consequently, the description of the practices and the corresponding level descriptions should be reviewed.

The high correlation between some practices in the program domain are more difficult to explain. It is conspicuous that 'Program Retrospective' is correlated with 'Scrum-of-Scrum Meeting' and 'Program Retrospective', which are somewhat the equivalent practices to the correlated practices in the team domain. This could indicate two things: either the practices' description needs review, or the practices are in fact related when it comes to practical implementation in the twenty organisations. However, for the 'Develop on Cadence' correlation with 'System Architect' and 'Community of Practice' no theoretical explanation can be given. Hence, reliability could be limited. 'Develop on Cadence' also has the highest multicollinearity in the program domain. From a theoretical perspective one could argue that most if not all of the practices in the program domain aim at generating 'cadence' in the development effort of several teams. Therefore, it makes sense that this practice shows the highest overall multicollinearity with the other practices. However, it is questionable if this item should remain in a survey when it is assumed that it works as an overall indicator of the program domain instead of one dimension or aspect of it.

The high correlation between 'Portfolio Strategy' and 'Coordinate Portfolio Value Stream' might indicate that 'Portfolio Strategy' is not sufficiently different from 'Coordinate Portfolio Value Stream' from the practical implementation point of view, even though there is no such relationship

between the two practices from a theoretical point of view. This would require reviewing the description of the practices. Overall, the high correlation between the practices in contrast to 'Discover Opportunity' might be based on the fact that 'Discover Opportunity' can be seen as more optional when implementing agile on portfolio level, while the other practices emerge from existing traditional practices.

The high correlation between program and portfolio domain practices might be theoretically justified, as these two domains represent the scaling practices. Organisations that are more likely to be advanced with agile on a team level, while the scaling of agile beyond a single team often represents the new aspect, which is covered by the two domains collectively.

5.1.5 Agile Dimensions

The analysis of the dimensions of agile culture showed that 'systems thinking' might be a different concept. In light of the ISD taxonomy provided by Conboy (2009) this seems to be appropriate as 'systems thinking' does not directly contribute to one of the first criteria of agility. However, 'systems thinking' provides an important function to an effective working of scaled agile implementation as it increases alignment and thereby enables autonomy, which in turn is an important aspect for agility. Without autonomy parts of the benefits of agility cannot be achieved, because necessary changes could not be implemented quickly for instance. Furthermore, 'systems thinking' might help organisations to react to change more quickly as it provides the foundation for decisions that might harm one team in the short term but benefit the team and the organisation as a whole in the longer term. Therefore, it is suggested to investigate further the relationship of 'systems thinking' and agile or drop this dimension from the measurement model in subsequent studies on agility. As 'systems thinking' is part of the SAFe and LeSS principles, is indirectly addressed in DA as well as in the Spotify case-study, this finding could be another indicator that the frameworks provide more than agility, which influences the validity of the measurement tool.

5.1.6 Agile and RBV

The focal construct of this thesis agility of organisations was investigated through the lens of the RBV on organisations. This allowed treating agility as an capability (Wade & Hulland, 2004) that can provide competitive advantage to an organisation (Barney, 1991). This lens provided guidance during the research project as it motivated the initial choice to focus on the practices as a source of agile capability. This choice was based on the definition of capabilities as "repeatable pattern of

action" which seems to be reflected in practices. However, as the research conducted reveals, practices are not the core of agility and it is not even ensured that practices lead to agility. Thereby, I would conclude that agility and agile capability should be researched through a different perspective.

5.2 Implication for Practice

This thesis shows that for an organisation to become agile it is necessary to be aware of both practices and organisational culture implying that an agile transition requires more than just adopting certain practices. Instead, organisations should also ensure that adaptability, autonomy, continuous improvement, empiricism, transparency, and value focus are the foundation for their actions.

The measurement tool provides value to practice as it allows organisations to assess themselves in a parsimonious way that does not require much instruction or outside help. However, it is still important that organisations use it in an effective way for example through connecting it to a process. This could be implemented by using it in a regular time-interval and analyse the results for example in retrospectives. Thereby, it could serve as a tool that helps organisations reflect on their agile capabilities by indicating less mature practices or potential for improvement in their culture.

Furthermore, the tool could be used to improve when problematic areas are identified. For example, when an organisation decides that it wants to become better in identifying innovations, they could focus on improving the practices 'innovation time' and 'discover opportunities'. Additionally, the practices identified provide a comprehensive summary of the different frameworks' practices and thereby making them more accessible, especially for organisations with limited resources.

5.3 Limitation of Study

The most significant limitation of the study seems to be the missing reference variable as it is used in other formative measurement pre-tests (e.g., Bruhn, Georgi, & Hadwich, 2008) and suggested by (Diamantopoulos et al., 2008). Including a control variable would have allowed to assess statistically the convergent validity of the survey (Hair Junior et al., 2014). Additionally, a control variable would have helped to further understand the relationship between agile practices and culture, as their contribution to the variable could have been assessed. The issue here is that when measuring concepts such as 'job satisfaction', it is straight forward to ask the respondents about their subjective opinion of overall job satisfaction, because in the end the concept of 'job satisfaction' is based on the individual's perception. For agile capabilities it is more difficult to come up with a reasonable

reference variable, as it is less about personal perception. Especially when the rater is unexperienced and has only limited insights into different organisations it would be difficult to argue that the subjective perception of the respondent is valid and reliable; and not biased by other factors such as vanity for instance.

Overall, the structure of the practices' assessment is based on SAFe. Also, the interviewees all had a focus on SAFe through their qualification. Thereby, the selection of the items could be biased towards this specific framework. This was to a certain degree balanced by analysing five other frameworks and the Spotify case-study, which added additional aspects to the practice selection (e.g., Discover Opportunities in the portfolio domain). Furthermore, as SAFe is the most popular among the frameworks (VersionOne Inc., 2015, 2016, 2017, 2018), the focus on SAFe seemed to be an acceptable limitation, because it also increases relevancy for practice of this study. Generally speaking, there seems to be a lack of established agile practices in literature, which lead to the decision to use these frameworks as foundation in a first place.

Another limitation has its foundation in the unclear relationship between agile practices and culture. Thereby one important aspect of agile capability could not be explained. Having a better understanding of this relationship would enable further analysis of the pre-test's results.

The next limitation of the study is related to the small number of valid responses, which means that a different sample might result in different results. Despite this being anticipated beforehand, this has implications as the generalisability of the findings is limited due to the potentially inappropriate coverage of the population resulting in a relatively high standard error. Thereby, the results from the statistical analysis have relatively low power and can only be seen as indications that require additional assessment by for example relating them to theory. Further, certain statistical analysis such as exploratory factor analysis of the items reflecting agile culture were not possible due to the small sample size.

The findings from the organisational culture assessment are only measured based on the perception of single employees. Hence, this assessment might be biased by the rater's individual experience in the organisation with is added to the general reliability issue with self-raters (Rossiter, 2002).

5.4 Future Research

Based on the discussion above a number of areas for future research evolve. First of all, due to the issues detected it would make sense to implement the above discussed findings from the pre-

test and improve the measurement model on this basis. This should be followed by another pre-test with a larger sample size. An increased number of responses for the adjusted pre-test could be achieved by a longer response collecting time. More responses would allow additional statistical analysis that could not be conducted with the small number of responses.

To further increase the validity of the measurement model it would be necessary to assess a number of organisations through a qualitative study and compare the results to the assessment through the model.

Another area of future research possibility is based on the relationship of practices and culture. Despite the importance of both areas that was shown in this study, further research needs to be conducted to fully understand their relationship. Another source of future research area is derived from the expert interviews. As Interviewee 4 explained that agile has to fit to the overall context of the organisation and in his experience, agility is not always the best solution, future research should investigate the circumstances under which agile is beneficial for an organisation. Furthermore, an agile measurement model's relevancy could be improved by connecting it to the potential outcome of an agile transformation. As at the moment many see agile as a silver bullet that solves many problems in organisations, it could be beneficial to first assess what the desired outcome of an agile transformation is to see if agile could help addressing the issue.

6 Conclusion

This study sought to shed light on how agile capabilities in the IT departments of organisations can be measured. Agile capabilities provide a way for organisations to deal with the challenges they are exposed to by a changing environment and thereby have the potential to be a resource for organisations leading to competitive advantage. Agile software development methods evolved as an answer to the issues related with traditional plan-driven software development methods. To guide organisations in the transition from these traditional approaches to an agile way of working, several frameworks have been developed. These frameworks propose several practices that aim at supporting this agile transition. For these practices to be effective, the organisational culture needs to be suitable.

The conceptualisation of agile capabilities in this thesis guided the development of a measurement instrument. This instrument has its foundation in the two aspects of practices and culture that determine the agile capability of an organisation. To identify the relevant practices several scaling agile frameworks were compared. The assessment of the practices was based on an adapted scale from the COBIT 4.1 framework that is part of the traditional IT maturity assessment family and related to the CMMI. The main adjustment in the scale was that the highest level implies continuous improvement based on the reality of the organisation instead of just following best practices. The agile practices from the frameworks were grouped into three domains, which was supported by the expert interviews, the structure found in the frameworks, as well as by the empirical pre-test.

The goal of the empirical pre-test was to uncover potential issues with the operationalisation of the measurement instrument as well as to obtain data for the statistical analysis of the instrument's validity and reliability. The analysis of the results from the empirical pre-test of the measurement instrument showed that overall the measurement instrument needs additional improvement. The analysis of the results from the cultural assessment indicate that the concept of systems thinking is different from the other subdimensions of agile culture. This is interesting, because the frameworks as well as the experts considered systems thinking an important part of agile culture. Another interesting finding was that the portfolio domain has the strongest correlation with agile culture.

However, the small sample size of the pre-test means that this insight would need further investigation by researchers. The discussion showed that the frameworks guide organisations in implementing contextual ambidexterity instead of agility in the strict sense of established definitions of agility. The main contribution to the knowledge base as well as for practice is that agile practices and culture jointly contribute to agile capabilities of an organisation and that agile capabilities should be measured on the three organisational levels of team, program, and portfolio. Next steps in this

research area should include refining the measurement instrument based on the findings from this study and thereby addressing the weaknesses of the measurement instrument at this stage.

References

Agarwal, A., Shankar, R., & Tiwari, M. K. (2006). Modeling the Metrics of Lean, Agile and Leagile Supply Chain: An ANP-Based Approach. *European Journal of Operational Research*, *173*, 211–225. https://doi.org/10.1016/j.ejor.2004.12.005

Agile. (2010). In *Oxford Dictionary of English* (3rd ed.). Oxford University Press.

Alqudah, M., & Razali, R. (2016). A Review of Scaling Agile Methods in Large Software Development. *International Journal on Advanced Science, Engineering and Information Technology*, *6*(6), 828. https://doi.org/10.18517/ijaseit.6.6.1374

Ambler, S. W. (2011). Examining the Agile Manifesto. Retrieved March 11, 2019, from http://www.ambysoft.com/essays/agileManifesto.html

Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (1998). *Critical Realism: Essential Readings*. London: Routledge.

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, *17*(1), 99–120.

Bass, J. M. (2016). Artefacts and agile method tailoring in large-scale offshore software development programmes. *Information and Software Technology*, *75*, 1–16. https://doi.org/10.1016/j.infsof.2016.03.001

Beck, K. (1999). Embracing Change with Extreme Programming. *IEEE Computer*, (10), 70–77.

Beck, Kent, Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., … Thomas, D. (2001). Manifesto for Agile Software Development. Retrieved February 3, 2019, from http://www.agilemanifesto.org

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009). Developing Maturity Models for IT Management. *Business & Information Systems Engineering*, *1*(3), 213–222. https://doi.org/10.1007/s12599-009-0044-5

Benner, M. J., & Tushman, M. L. (2003). Exploitation , Exploration , and Process Management : The Productivity Dilemma Revisited. *The Academy of Management Review*, *28*(2), 238–256. https://doi.org/10.2307/30040711

Boehm, B., & Turner, R. (2004). Balancing Agility and Discipline: Evaluating and Integrating Agile and Plan-Driven Methods. In *Proceedings of the 26th International Conference on Software Engineering (ICSE'04)* (pp. 718–719). IEEE. https://doi.org/10.1109/icse.2004.1317503

Boehm, B., & Turner, R. (2005). Management Challenges to Implementing Agile Processes in

Traditional Development Organizations. *IEEE Software*, *22*(5), 30–39. https://doi.org/10.1109/MS.2005.129

Bollen, K., & Lennox, R. (1991). Conventional Wisdom on Measurement: A Structural Equation Perspective. *Psychological Bulletin*, *110*(2), 305–314. https://doi.org/10.1037/0033-2909.110.2.305

Brown, S. L., & Eisenhardt, K. M. (1995). Product Development: Past Research, Present Findings , and Future Directions. *The Academy of Management Review*, *20*(2), 343–378.

Bruhn, M., Georgi, D., & Hadwich, K. (2008). Customer equity management as formative second-order construct. *Journal of Business Research*, *61*(12), 1292–1301. https://doi.org/10.1016/j.jbusres.2008.01.016

Buchalcevova, A. (2018). Application of Methodology Evaluation System on Current IS Development Methodologies. *International Journal of Information Technologies and Systems Approach*, *11*(2), 71–87. https://doi.org/10.4018/ijitsa.2018070105

Canna, C. N., Richardson, I., Noll, J., Razzak, M. A., & Beecham, S. (2018). Scaling Agile across the Global Organization: An Early Stage Industrial SAFe Self-Assessment. In *ACM/IEEE 13th International Conference on Global Software Engineering* (pp. 116–125). https://doi.org/10.1145/3196369.3196373

Cao, L., Mohan, K., Xu, P., & Ramesh, B. (2009). A Framework for Adapting Agile Development Methodologies. *European Journal of Information Systems*, *18*(4), 332–343. https://doi.org/10.1057/ejis.2009.26

Cataldo, E. F., Johnson, R. M., Kellstedt, L. A., & Milbrath, L. W. (1979). Card Sorting as a Technique for Survey Interviewing. *The Public Opinion Quarterly*, *34*(2), 202–215.

Chan, F. K. Y., & Thong, J. Y. L. (2009). Acceptance of Agile Methodologies: A Critical Review and Conceptual Framework. *Decision Support Systems*, *46*, 803–814. https://doi.org/10.1016/j.dss.2008.11.009

Chow, T., & Cao, D.-B. (2008). A Survey Study of Critical Success Factors in Agile Software Projects. *Journal of Systems and Software*, *81*, 961–971. https://doi.org/10.1016/j.jss.2007.08.020

Christopher, M. (2000). The Agile Supply Chain - Competing in Volatile Markets. *Industrial Marketing Management*, *29*, 37–44.

Churchill, G. A. J. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research*, *16*(1), 6473.

CMMI Product Team. (2010). *CMMI for Development, Version 1.3*.

Cockburn, A., & Highsmith, J. (2001). Agile Software Development: The People Factor. *Computer*, (11), 131–133. https://doi.org/10.1109/2.963450

Conboy, K. (2009). Agility from First Principles: Reconstructing the Concept of Agility in Information Systems Development. *Information Systems Research*, *20*(3), 329–354. https://doi.org/10.1287/isre.1090.0236

Conboy, K., & Carroll, N. (2019). Implementing Large-Scale Agile Frameworks: Challenges and Recommendations. *IEEE Software*, *36*(2), 44–50. https://doi.org/10.1109/MS.2018.2884865

Conboy, K., & Fitzgerald, B. (2004). Toward a Conceptual Framework of Agile Methods: A Study of Agility in Different Disciplines. In *Proceeding of ACM Workshop on Interdisciplinary Software Engineering Research (WISER)* (pp. 37–44). Newport Beach, CA, USA: ACM.

Creswell, J. W., Fetter, M. D., & Ivankova, N. V. (2004). Designing A Mixed Methods Study In Primary Care. *Annals of Family Medicine*, *2*(1), 7–12. https://doi.org/10.1370/afm.104.INTRODUCTION

DeVellis, R. F. (2012). *Scale Development - Theory and Application* (3rd ed.). SAGE Publications.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*(12), 1203–1218. https://doi.org/10.1016/j.jbusres.2008.01.009

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index Construction with Formative Indicators: an Alterantive to Scale Development. *Journal of Marketing Research*, *38*, 269–277.

Dikert, K., Paasivaara, M., & Lassenius, C. (2016). Challenges and Success Factors for Large-Scale Agile Transformations: A Systematic Literature Review. *Journal of Systems and Software*, *119*, 87–108. https://doi.org/10.1016/j.jss.2016.06.013

Dingsøyr, Torgeir, & Moe, N. B. (2014). Towards Principles of Large-Scale Agile Development: A Summary of the Workshop at XP2014 and a Revised Research Agenda. In T. Dingsøyr, N. B. Moe, R. Tonelli, S. Counsell, C. Gencel, & K. Petersen (Eds.), *Agile Methods. Large-Scale Development, Refactoring, Testing, and Estimation* (Vol. 199, pp. 1–8). Berlin: Springer. https://doi.org/https://doi.org/10.1007/978-3-319-14358-3_1

Dingsøyr, Torgeir, Moe, N. B., Fægri, T. E., & Seim, E. A. (2018). Exploring Software Development at the very Large-Scale: a Revelatory Case Study and Research Agenda for Agile Method Adaptation. *Empirical Software Engineering*, *23*(1), 490–520. https://doi.org/10.1007/s10664-017-9524-2

Dingsøyr, Torgeir, Nerur, S., Balijepally, V., & Moe, N. B. (2012). A Decade of Agile Methodologies: Towards Explaining Agile Software Development. *Journal of Systems and Software*, *85*, 1213–1221. https://doi.org/10.1016/j.jss.2012.02.033

Disciplined Agile Inc. (2019). Disciplined Agile. Retrieved September 13, 2019, from http://disciplinedagiledelivery.com

Duboc, L., Rosenblum, D. S., & Wicks, T. (2007). A Framework for Characterization and Analysis of Software System Scalability. In *Proceedings of the the 6th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on The foundations of software engineering* (pp. 375–384). ACM. https://doi.org/10.1145/1287624.1287679

Duncan, R. B. (1976). The Ambidextrous Organization: Designing Dual Structures for Innovation. *The Management of Organization*, *1*, 167–188.

Dybå, T., & Dingsøyr, T. (2009). What Do We Know about Agile Software Development. *IEEE Software*, *26*(5), 6–9. https://doi.org/10.1109/MS.2009.145

Easterby-Smith, M., Thorpe, R., & Jackson, P. (2015). *Management & Business Research*. *The British Journal of Psychiatry* (5th ed.). London: SAGE Publications. https://doi.org/10.1192/bjp.112.483.211-a

Essmann, H., & Du Preez, N. (2009). An Innovation Capability Maturity Model -Development and Initial Application. *World Academy of Science, Engineering and Technology*, *3*, 435–446.

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing Routines Organizational as a Source of Flexiblity and Change. *Administrative Science Quarterly*, *48*, 94–118. https://doi.org/10.1038/nn1246

Fontana, R. M., Fontana, I. M., Da Rosa Garbuio, P. A., Reinehr, S., & Malucelli, A. (2014). Processes Versus People: How Should Agile Software Development Maturity be Defined? *Journal of Systems and Software*, *97*, 140–155. https://doi.org/10.1016/j.jss.2014.07.030

Fontana, R. M., Meyer, V., Reinehr, S., & Malucelli, A. (2015). Progressive Outcomes: A Framework for Maturing in Agile Software Development. *Journal of Systems and Software*, *102*, 88–108. https://doi.org/10.1016/j.jss.2014.12.032

Fraser, P., Moultrie, J., & Gregory, M. (2002). The Use of Maturity Models / Grids as a Tool in Assessing Product Development Capability: a Review. In *IEEE International Engineering Management Conference*. Cambridge UK.

Gallagher, K. P., & Worrell, J. L. (2008). Organizing IT to Promote Agility. *Information Technology and Management*, *9*, 71–88. https://doi.org/10.1007/s10799-007-0027-5

Gibson, C. B., & Birkinshaw, J. (2004). The Antecedents, Consequences, and Mediating Role of Organizational Ambidexterity. *Academy of Management Journal*, *47*(2), 209–226.

Greening, D. R. (2010). Enterprise Scrum: Scaling Scrum to the Executive Level. In *Proceedings of the Annual Hawaii International Conference on System Sciences* (pp. 1–10). https://doi.org/10.1109/HICSS.2010.186

Gren, L., Torkar, R., & Feldt, R. (2015). The prospects of a quantitative measurement of agility: A validation study on an agile maturity model. *Journal of Systems and Software*, *107*, 38–49. https://doi.org/10.1016/j.jss.2015.05.008

Gupta, A. K., Smith, K. G., & Shalley, C. E. (2006). The Interplay between Exploration and Exploitation. *The Academy of Management Journal*, *49*(4), 693–706. https://doi.org/10.2307/20159793

Hair Junior, J. F., Hult, G. T. M., Ringle, C. M., & Sarsted, M. (2014). *Primer On Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Thousand Oaks: SAGE Publications.

He, Z.-L., & Wong, P.-K. (2004). Exploration vs. Exploitation: An Empirical Test of the Ambidexterity Hypothesis. *Organization Science*, *15*(4), 481–494. https://doi.org/10.1287/orsc.1040.0078

Henderson, J. C., & Venkatraman, N. (1993). Strategic Alignment: Leveraging information technology for Transforming Organizations. *IBM Systems Journal*, *32*(1), 472–484.

Hill, M. D. (1990). What is Scalability? *ACM SIGARCH Computer Architecture News*, *18*(4), 18–21.

Hitt, M. A., Keats, B. W., & DeMarie, S. M. (1998). Navigating in the New Competitive Landscape : Building Strategic Flexibility and Competitive Advantage in the 21st Century. *Academy of Management Executive*, *12*(4), 22–42.

Hobbs, B., & Petit, Y. (2017). Agile Methods on Large Projects in Large Organizations. *Project Management Journal*, *48*(3), 3–19.

Horlach, B., Böhmann, T., Schirmer, I., & Drews, P. (2018). IT Governance in Scaling Agile Frameworks. In *Multikonferenz Wirtschaftsinformatik* (pp. 1789–1800). Lüneburg.

Hossain, E., Babar, M. A., & Paik, H. (2009). Using Scrum in Global Software Development: A Systematic Literature Review. In *2009 Fourth IEEE International Conference on Global Software Engineering* (pp. 175–184). https://doi.org/10.1109/ICGSE.2009.25

ISACA. (2012). COBIT 5: A Business Framework for the Governance and Management of Enterprise IT. *Isaca*.

ISO/IEC. (2006). ISO/IEC 15504.

IT Governance Institute. (2009). COBIT 4.1 - Control Objectives, Management Guidelines and Maturity Models. https://doi.org/10.1002/bse.3280020501

Jakobsen, C. R., & Johnson, K. A. (2008). Mature agile with a twist of CMMI. *Proceedings - Agile 2008 Conference*, (April), 212–217. https://doi.org/10.1109/Agile.2008.10

Jakobsen, C. R., & Sutherland, J. (2009). Scrum and CMMI - Going from good to great: Are you ready-ready to be done-done? *Proceedings - 2009 Agile Conference, AGILE 2009*, 333–337. https://doi.org/10.1109/AGILE.2009.31

Javdani Gandomani, T., & Ziaei Nafchi, M. (2015). An empirically-developed framework for Agile transition and adoption: A Grounded Theory approach. *Journal of Systems and Software*, *107*, 204–219. https://doi.org/10.1016/j.jss.2015.06.006

Kaiser, H. F. (1974). An Index of Factorial Simplicty. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/BF02294054

Kalenda, M., Hyna, P., & Rossi, B. (2018). Scaling Agile in Large Organizations: Practices, Challenges, and Success Factors. *Journal of Software: Evolution and Process*, *30*, 1–24.

Keplinger, W. (2007). Agility in Information Systems Development: Characterisation, Motivation and Conceptualisation. *The Information Systems Student Journal*, *2*, 25–28.

Kerin, R. A., Varadarajan, P. R., & Peterson, R. A. (1992). First-Mover Advantage: A Synthesis, Conceptual Framework, and Research Propositions. *Journal of Marketing*, *56*, 33–52. https://doi.org/10.2307/1251985

Kishore, R., Swinarski, M. E., Jackson, E., & Rao, H. R. (2012). A Quality-Distinction Model of IT Capabilities: Conceptualization and Two-Stage Empirical Validation Using CMMi Processes. *IEEE Transactions on Engineering Management*, *59*(3), 457–469. https://doi.org/10.1109/TEM.2011.2165287

Kleinbaum, D. G., Kupper, L. L., Muller, K. E., & Nizam, A. (1998). *Applied Regression Analysis and other Multivariable Methods* (3rd ed.). Belomnt, CA: Thomson Brooks/Cole Publishing.

Kniberg, H. (2014a). Spotify engineering culture (part 1). Retrieved March 17, 2019, from https://labs.spotify.com/2014/03/27/spotify-engineering-culture-part-1/

Kniberg, H. (2014b). Spotify engineering culture (part 2). Retrieved March 17, 2019, from https://labs.spotify.com/2014/09/20/spotify-engineering-culture-part-2/

Kniberg, H., & Ivarsson, A. (2012). Scaling Agile @ Spotify - with Tribes, Squads, Chapters & Guilds.

Kochikar, V. P., & Ravindra, M. P. (2007). Developing the Capability to Be Agile. *Organization*

Development Journal, *25*(4), 127–134.

Kohlegger, M., Maier, R., & Thalmann, S. (2009). Understanding Maturity Models Results of a Structured Content Analysis. *Proceedings of I-KNOW '09 and I-SEMANTICS '09*, (September), 51–61.

Kolt, K. (2014). The Agile Manifesto Updated. Retrieved September 20, 2004, from https://agilesista.com/2014/09/26/the-agile-manifesto-updated/

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Laanti, M. (2014). Characteristics and Principles of Scaled Agile, 9–20. https://doi.org/10.1007/978-3-319-14358-3_2

Laanti, M., Similä, J., & Abrahamsson, P. (2013). Definitions of Agile Software Development and Agility. In F. McCaffery, R. V. O'Connor, & R. Messnarz (Eds.), *Systems, Software and Services Process Improvemen. European Conference on Software Process Improvement 2013. Communications in Computer and Information Science* (Vol. 364, pp. 247–258). Berln, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-39179-8_22

Laerd Statistics. (2015a). Cronbach's alpha using SPSS Statistics. Retrieved September 3, 2019, from https://statistics.laerd.com/

Laerd Statistics. (2015b). Multiple Regression Using SPSS Statistics. Retrieved September 1, 2019, from https://statistics.laerd.com/

Laerd Statistics. (2015c). Principle Component Analysis. Retrieved September 2, 2019, from https://statistics.laerd.com/

Laerd Statistics. (2018). Pearson's Product-Moment Correlation Using SPSS Statistics. Retrieved September 3, 2019, from https://statistics.laerd.com/

Lasrado, L. A., Vatrapu, R., & Mukkamala, R. R. (2017). Whose Maturity is it anyway? The Influence of Different Quantitative Methods on the Design and Assessment of Mautrity Models. In *Proceedings of the 25th European Conference on Information Systems (ECIS)* (pp. 2918–2927). Guimaraes, Portugal.

Lee, G., & Xia, W. (2010). Toward Agile: An Antegrated Analysis of Quantitative and Qualitative Field Data on Software Development Agility. *MIS Quarterly*, *34*(1), 87–114.

Leppänen, M. (2013). A Comparative Analysis of Agile Maturity Models. In R. Pooley, J. Coady, C. Schneider, H. Linger, B. C., & M. Lang (Eds.), *Information Systems Development: Reflections, Challenges and New Directions* (pp. 329–343). New York: Springer. https://doi.org/10.1007/978-1-4614-4951-5

Lieberman, M. B., & Montgomery, D. B. (1988). First-Mover Advantages. *Strategic Management Journal*, *9*, 41–58.

Little, J. (2007). The Nokia Test. Retrieved February 23, 2019, from http://agileconsortium.blogspot.com/2007/12/nokia-test.html

Lukasiewicz, K., & Miler, J. (2012). Improving Agility and Discipline of Software Development with the Scrum and CMMI. *IET Software*, *6*(5), 416–422.

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

Maier, A. M., Moultrie, J., & Clarkson, P. J. (2012). Assessing Organizational Capabilities: Reviewing and Guiding the Development of Maturity Grids. *IEEE Transactions on Engineering Management*, *59*(1), 138–159. https://doi.org/10.1109/TEM.2010.2077289

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87.

Mature. (2009). In *Oxford English Dictionary (OED)*.

Maturity. (2009). In *Oxford English Dictionary (OED)*.

McConnell, S. (2000). Cargo Cult Software Engineering. *IEEE Software*, (2), 11–13. https://doi.org/10.1109/MS.2000.10012

McEvoy, P., & Richards, D. (2006). A Critical Realist Rationale for Using a Combination of Quantitative and Qualitative Methods. *Journal of Research in Nursing*, *11*(1), 66–78. https://doi.org/10.1177/1744987106060192

Misra, S. C., Kumar, V., & Kumar, U. (2010). Identifying some critical changes required in adopting agile practices in traditional software development projects. *International Journal of Quality and Reliability Management*, *27*(4), 451–474. https://doi.org/10.1108/02656711011035147

Nerur, S., Mahapatra, R., & Mangalaraj, G. (2005). Challenges of Migrating to Agile Methodologies - Organizations Must Carefully Assess Their Readiness Before Treading the Path of Agility. *Communications of the ACM*, *48*(5), 72–78. https://doi.org/10.1145/1060710.1060712

Nurmuliani, N., Zowghi, D., & Williams, S. P. (2004). Using Card Sorting Technique to Classify Requirements Change. In *12th IEEE International Requirements Engineering Conference (RE'04)*.

O'Reilly III, C. A., & Tushman, M. L. (2008). Ambidexterity as a Dynamic Capability: Resolving the Innovator's Dilemma. *Research in Organizational Behavior*, *28*, 185–206.

https://doi.org/10.1016/j.riob.2008.06.002

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information Systems Research*, *2*(1), 1–28.

Paasivaara, M. (2017). Adopting SAFe to Scale Agile in a Globally Distributed Organization. In *IEEE 12 International Conference on Global Software Engineering (ICGSE)* (pp. 36–40).

Paasivaara, M., Behm, B., Lassenius, C., & Hallikainen, M. (2018). Large-Scale Agile Transformation at Ericsson: a Case Study. *Empirical Software Engineering*, *23*(5), 2550–2596. https://doi.org/10.1007/s10664-017-9555-8

Paasivaara, M., & Lassenius, C. (2011). Scaling Scrum in a Large Distributed Project. In *2011 International Symposium on Empirical Software Engineering and Measurement* (pp. 363–367). https://doi.org/10.1109/esem.2011.49

Papadopoulos, G. (2015). Moving from Traditional to Agile Software Development Methodologies Also on Large, Distributed Projects. *Procedia - Social and Behavioral Sciences*, *175*, 455–463. https://doi.org/10.1016/j.sbspro.2015.01.1223

Pasquini, A., & Galiè, E. (2013). COBIT 5 and the Process Capability Model. Improvements Provided for IT Governance Process. *Proceedings of FIKUSZ '13 Symposium for Young Researchers*, 67–76.

Paulk, M. C. (2001). Extreme Programming from a CMM Perspective. *IEEE Software*, *18*(6), 1–8.

Paulk, M. C., Weber, C. V., Garcia, S. M., Chrissis, M. B., & Bush, M. (1993). *Key Practices of the Capability Maturity ModelSM, Version 1.1. Technical Report CMU/SEI-93-TR-025 ESC-TR-93-178*. Pittsburgh, Pennsylvania.

Pernstål, J., Feldt, R., & Gorschek, T. (2013). The Lean Gap: A Review of Lean Approaches to Large-Scale Software Systems Development. *Journal of Systems and Software*, *86*(11), 2797–2821. https://doi.org/10.1016/j.jss.2013.06.035

PMI. (2013). *The Standard for Portfolio Management* (3rd ed.).

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2016). Recommendations for Creating Better Concept Definitions in the Organizational, Behavioral, and Social Sciences. *Organizational Research Methods*, *19*(2), 159–203. https://doi.org/10.1177/1094428115624965

Porter, M. E. (1980). *Competitive Strategy*. New York: The Free Press.

Putta, A. (2018). Scaling agile software development to large and globally distributed large-scale organizations, 141–144. https://doi.org/10.1145/3196369.3196386

Raisch, S., & Birkinshaw, J. (2008). Organizational Ambidexterity: Antecedents, Outcomes, and

Moderators. *Journal of Management*, *34*(3), 375–409.

Reifer, D. J., Maurer, F., & Erdogmus, H. (2003). Scaling Agile Methods. *IEEE Software*, *20*(4), 12–14. https://doi.org/10.1109/ms.2003.1207448

Rigby, Darrell K.; Sutherland, Jeff; Noble, A. (2018). Agile at Scale. *Harvard Business Review*, *11*, 88–96.

Rossiter, J. R. (2002). The C-OAR-SE Procedure for Scale Development in Marketing. *International Journal of Research in Marketing*, *19*(4), 305–335.

Sailer, P. (2019). Project management methods as a way to ambidexterity. *International Journal of Managing Projects in Business*. https://doi.org/10.1108/IJMPB-05-2018-0094

Sambamurthy, V., Bharadwaj, A., & Grover, V. (2003). Shaping Agility through Digital Options: Reconceptualizing the Role of Information Technology in Contemporary Firms. *MIS Quarter*, *27*(2), 237–263.

Saunders, M., Lewis, P., & Thornhill, A. (2016). *Research Methods For Business Students* (7th ed.). Harlow, Essex: Pearson. https://doi.org/10.1002/ejoc.201200111

Scaled Agile Inc. (2019). Scaled Agile Framework. Retrieved September 13, 2019, from https://www.scaledagileframework.com

Schmalensee, B. R. (1985). Do Markets Differ Much? *The American Economic Review*, *75*(3), 341–351.

Schoemaker, P. J. H., & Amit, R. (1993). Strategic Assets and Organizational Rent. *Strategic Management Journal*, *14*(1), 33–46.

Schwaber, K., & Beedle, M. (2001). *Agile Software Development with Scrum* (1st ed.). Upper Saddle River, NJ, USA: Prentice Hall PTR.

Schwaber, K., & Scrum.org. (2018). Nexus Guide - The Definitive Guide to scaling Scrum with Nexus: The Rules of the Game, 0–11.

Sheffield, J., & Lemétayer, J. (2013). Factors associated with the software development agility of successful projects. *International Journal of Project Management*, *31*(3), 459–472. https://doi.org/10.1016/j.ijproman.2012.09.011

Sidky, A., Arthur, J., & Bohner, S. (2007). A disciplined approach to adopting agile practices: The agile adoption framework. *Innovations in Systems and Software Engineering*, *3*(3), 203–216. https://doi.org/10.1007/s11334-007-0026-z

Strode, D. E., Huff, S. L., & Tretiakov, A. (2009). The impact of organizational culture on agile method use. *Proceedings of the 42nd Annual Hawaii International Conference on System*

Sciences, HICSS, (May 2014). https://doi.org/10.1109/HICSS.2009.436

Sutherland, J. (2001). Agile Can Scale: Inventing and Reinventing Scrum in Five Companies. *Cutter IT Journal*, *14*(12), 5–11.

Sutherland, J., & Scrum Inc. (2019). The Scrum@Scale Guide.

Sutherland, J., Viktorov, A., Blount, J., & Puntikov, N. (2007). Distributed Scrum: Agile Project Management with Outsources Development Teams. In *Proceedings of the 40th Hawaii International Conference on System Sciences*.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic Capabilities and Strategic Management. *Strategic Management Journal*, *18*(7), 509–533. https://doi.org/10.1142/9789812834478_0002

The LeSS Company B.V. (2019). Large-Scale Scrum. Retrieved September 13, 2019, from https://less.works

Thompson, K., & CPrime. (2013). Recipes for Agile Governance in the Enterprise.

Tousignant, D. (2018). How Agile are you? Free Agile Maturity Assessment. Retrieved July 13, 2019, from https://capeprojectmanagement.com/agile-self-assessment/

Turetken, O., Stojanov, I., & Trienekens, J. J. M. (2017). Assessing the Adoption Level of Scaled Agile Development: a Maturity Model for Scaled Agile Framework. *Journal of Software: Evolution and Process*, *29*, 1–18. https://doi.org/10.1002/smr

Turner, R., & Jain, A. (2002). Agile Meets CMMI: Culture Clash or Common Cause? In D. Wells & L. Williams (Eds.), *Extreme Programming and Agile Methods — XP/Agile Universe 2002* (pp. 153–165). Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-45672-4

Tushman, M. L., & O'Reilly III, C. A. (1996). Ambidextrous Organizations: Managing Evolutionary and Revolutionary Change. *California Management Review*, *38*(4), 8–30.

Vaidya, A. (2014). Does DAD Know Best, Is it Better to do LeSS or Just be SAFe? Adapting Scaling Agile Practices into the Enterprise. *Thirty-Second Annual Pacific Northwest Software Quality Conference*, 21–38.

Vallon, R., da Silva Estácio, B. J., Prikladnicki, R., & Grechenig, T. (2018). Systematic Literature Review on Agile Practices in Global Software Development. *Information and Software Technology*, *96*(April 2017), 161–180. https://doi.org/10.1016/j.infsof.2017.12.004

van Oosterhout, M., Waarts, E., & van Hillegersberg, J. (2006). Change Factors Requiring Agility and Implications for IT. *European Journal of Information Systems*, *15*(2), 132–145. https://doi.org/10.1057/palgrave.ejis.3000601

VersionOne Inc. (2015). *9th Annual State of Agile Survey*.

VersionOne Inc. (2016). *10th Annual State of Agile Survey*.

VersionOne Inc. (2017). *11th Annual State of Agile Survey*.

VersionOne Inc. (2018). *12th Annual State of Agile Survey*.

Vijayasarathy, L. R., & Turk, D. (2008). Agile Software Development: A Survey of Early Adopters. *Journal of Information Technology Management*, *XIX*(2), 1–8.

Vinekar, V., Slinkman, C. W., & Nerur, S. (2006). Can agile and traditional systems development approaches coexist? An ambidextrous view. *Information Systems Management*, *23*(3), 31–42. https://doi.org/10.1201/1078.10580530/46108.23.3.20060601/93705.4

Wade, M., & Hulland, J. (2004). Review: The Resource-Based View and Information Systems Research: Review, extension, and Suggestions for Future Research. *MIS Quarterly*, *28*(1), 107–142.

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), 13–23.

Wernerfelt, B. (1984). A Resource-based View of the Firm. *Strategic Management Journal*, *5*(2), 171–180. https://doi.org/10.1002/smj.4250050207

Wind, J., & Mahajan, V. (1997). Issues and Opportunities in New Product Development: An Introduction to the Special Issue. *Journal of Marketing Research*, *34*, 1–12. https://doi.org/10.2307/3152060

Winter, S. G. (2003). Understanding Dynamic Capabilities. *Strategic Management Journal*, *24*, 991–995. https://doi.org/10.1002/smj.318

Wulf, J., Winkler, T. J., & Brenner, W. (2015). Measuring IT Service Management Capability: Scale Development and Empirical Validation. *12th International Conference on Wirtschaftsinformatik*, 630–644.

Zachariadis, M., Scott, S., & Barrett, M. (2013). Methodological Implications of Critical Realism for Mixed-Methods Research. *Management Information Systems Quarterly*, *37*(10), 1–25.

List of Appendices

Due to size, appendices are to be found in a separate document.