Evaluation of Self-Assessment Tools

A Case Study

Kristin Egeland
Victoria Værnø Tutturen

TASK DESCRIPTION

An evaluation of a self-assessment tool with respect to use in a Norwegian hospital organization will be performed. Criteria are developed to perform an evaluation of a self-assessment tool, and the evaluation will be conducted with respect to these criteria. This is a qualitative study.

ABSTRACT

While self-assessment tools are widely used, little research has been conducted regarding the evaluation of such tools. In addition, there is a widespread misconception regarding self-assessment tools, where they are considered and used as if they were measuring instruments, while they in fact provide nothing close to valid measures. In this thesis we provide a conceptualization of self-assessment tools as facilitators of a process where its users build an image of the organization. This image then forms the basis for changing the organization. The Psychiatric Health Services Division at IHT is experiencing major changes in their context and are therefore open to introducing tools that can aid them in the changes this requires them to make. Specifically, they have been looking at the LESAT, a self-assessment tool developed at MIT. Such tools, taken from another context, should be evaluated before adopting them to an organization. We propose three criteria that serve as a framework for evaluating self-assessment tools. To help the division in their evaluation, we have conducted a case study, evaluating the LESAT for use in their organization. The result of the evaluation was that the LESAT is inappropriate for use in the division, and its design is found to focus attention towards generating numerical scores rather than encourage reflection and help users improve on their image of their organization.

SAMMENDRAG

Selv om selvevalueringsverktøy er mye brukt, finnes det lite forskning på evaluering av slike verktøy. I tillegg er det en utbredt misforståelse tilknyttet selvevalueringsverktøy, hvor de blir sett på og brukt som måle-redskaper, mens de i realiteten ikke gir i nærheten av gode måleresultater. I denne studien gis en konseptualisering av selvevalueringsverktøy som fasilitatorer av en prosess hvor dets brukere bygger et bilde av organisasjonen. Dette bildet kan så benyttes som et utgangspunkt for å endre organisasjonen. Divisjon Psykisk Helsevern på Sykehuset Innlandet opplever store endringer i sine omgivelser og har interesse for verktøy som kan hjelpe dem i endringene dette medfører at de må gjennomføre. De har spesifikt sett på LESAT, et selvevalueringsverktøy utviklet på MIT. Slike verktøy som er hentet fra en annen kontekst burde evalueres før de tas inn i en organisasjon. Vi legger fram tre kriterier som kan benyttes som et rammeverk i slike evalueringer av selvevaluerings-verktøy. For å hjelpe dem i evalueringen har vi gjennomført en case-studie hvor vi evaluerer LESAT for bruk i divisjonen. Resultatet av evalueringen var at LESAT er uskikket for bruk i divisjonen, og at verktøyets design drar fokus mot generering av tallsvar fremfor å motivere til refleksjon og hjelpe dets brukere utvikle bildet de har av organisasjonen.

TABLE OF CONTENTS

List of Tables

List of Figures

1 INTRODUCTION

The primary goal of this thesis is to evaluate a management tool, specifically a self-assessment tool, with respect to use in a Norwegian hospital trust division. The hospital trust division is the Psychiatric Health Services Division of Innlandet Hospital Trust (IHT). The self-assessment tool is the LAI Enterprise self-assessment tool (LESAT), developed by the Lean Advancement Initiative (LAI) research consortium at the American university Massachusetts Institute of Technology (MIT). The LESAT is a self-assessment tool that supports organizations in implementing a set of best practices identified by the LAI. Throughout the thesis, self-assessment tools will be referred to as SATs.

This research is motivated by the request of our external project mentor Clas Waagø-Hansen. He is the assistant division director of the Psychiatric Health Services Division at IHT. IHT is a local health enterprise in the Norwegian public specialist health care services that provides services in the Norwegian counties of Oppland and Hedmark. Through recent reforms, the Norwegian public health care sector has been subject to many changes. For instance, the municipalities are now taking on a larger part of the responsibility for psychiatric health care services. One of the consequences is that psychiatry in the specialist health care services receives less funding, requiring that they continually rethink how to best fulfill their new responsibilities. Clas Waagø-Hansen requested that we evaluate the LESAT for use in the division to support current and future change processes in the division.

In this thesis an SAT will be evaluated, or assessed. The words assess and evaluate will be used interchangeably throughout the thesis as these words are very close synonyms. The dictionary meaning of "assess" is "evaluate or estimate the nature, ability, or quality of", while the meaning of "evaluate" is "form an idea of the amount, number, or value of; assess" (Oxford Dictionaries, 2014).

1.1 MOTIVATION FOR EVALUATING SATS

SATs are widely used and discussed in academia and in the commercial world. However, evaluation of these tools has received very little attention. Most of the literature about these tools is published to present a tool made by the same author, and there is little available research conceptualizing what SATs are and how they can be evaluated. Beyond the literature about SATs however, there is an abundance of literature issuing warnings about potential issues with adopting management theories and tools.

Importing tools from other contexts without evaluating or adapting them to the new context can be problematic. For example, Norwegian researchers issue sharp warnings against blindly adopting tools based on theories from other contexts such as lean, as valuable Norwegian practices may be replaced or disrupted in the process (Levin, Nilssen, Ravn, & Øyum, 2012). Also, there is the possibility that the tool may fall within the category of what is referred to as a management fad. A management fad can be described as a concept or technique that becomes very popular, but later is found to fall short of the expected benefits and are quickly discontinued or only used to a very modest extent (Ponzi & Koenig, 2002). Introducing such concepts and techniques simply because they have a good reputation may therefore not bring with it the desired results. However, avoiding research, insights and tools from other countries and industries completely would be unwise. Norwegian organizations need to draw on global resources while also retaining and enhancing good Norwegian practices. To achieve this a process of evaluation and perhaps revision needs to take place, also for tools. Literal translation, such as from English to Norwegian, is only a very small part of this process. It may not even be necessary as Norwegians are typically proficient English speakers. There could however be underlying assumptions to the tool that do not apply to Norwegian hospitals. Issues like the language and ill fitted assumptions about the realities of the organization can make these tools less helpful than expected.

Additionally, the quality of a tool should not be judged by the presentation by its author. There are for example many SATs that are presented as measuring devices that the organization can use to measure for example how innovative it is or how

ready it is for change. However, ability to measure is a characteristic that is rarely held by such tools (see chapter 2). This may result in organizations to use these tools for purposes where an SAT is not a good choice, or use the results from the self-assessment in a dysfunctional way.

This motivates conducting systematic and critical evaluations of SATs with respect to the context where they will be used. As little research is done on the evaluation of SATs there is a need for a conceptual understanding of what these SATs are, and how they can be evaluated. In this thesis we therefore provide a conceptualization of SATs and a framework for evaluating them. We will further demonstrate such an evaluation in practice by evaluating an SAT for the context of a public Norwegian health care organization. Through this case study we hope to shed light on the issues related to evaluating and adopting such SATs, and to show how SATs are used in practice.

1.2 RESEARCH QUESTIONS AND THESIS STRUCTURE

In order to evaluate the LESAT for use in the Psychiatric Health Services Division at IHT, a set of criteria for evaluating SATs is needed. The first research question to be considered is therefore as follows:

1. What criteria should be used to evaluate an SAT?

This is a theoretical exercise that will be performed by first providing a conceptualization of what SATs are, and then developing criteria for such tools from evaluation literature.

These criteria can then be utilized to answer the second research question:

2. How well does the LESAT fulfill the criteria for SATs with respect to use in the Psychiatric Health Services Division at IHT?

Kaplan and Jarzabkowski (2006) found that the use of management tools in practice is typically much more of a messy, and even political, process than the idealized processes expected by the authors of the tools. SATs are often associated with a

presentation of the author(s) explaining how it will improve your organization, and enable its users to achieve some elaborately described goal. We cannot take these idealized processes as the basis for our evaluation, and must therefore test the tool in practice and in the context it is to be evaluated for. The second research question will therefore be answered by testing the LESAT through a pilot project in the Psychiatric Health Services Division at IHT.

The structure of the thesis is as follows. A theoretical conceptualization of SATs is presented in chapter 2. To support the evaluation of the LESAT, a set of criteria has been developed for SATs in chapter 3. Chapter 2 and 3 are theoretical chapters, forming a basis for the rest of the thesis and answering the first research question. These chapters contain long summaries, as these summaries contain sufficient information to enable the reader to read the rest of the thesis after reading the summaries of these chapters. The research strategy is described in chapter 4, followed by presentations of our case SAT and organization. The characteristics of the Psychiatric Health Services Division at IHT are identified in chapter 5. This is followed by an analysis of the LESAT in chapter 6, where the underlying theory of the LESAT is explained and the tool itself is described. After this case presentation, we present our findings in chapter 7. Chapter 8 is the evaluation of the LESAT with respect to the division and the criteria. Chapter 9 is a discussion of our findings. Finally, a conclusion of the research is presented in chapter 10.

2 ABOUT SATs

The SATs that we consider within our scope are tools to aid in self-assessment processes in connection to change on the organizational level. During a self-assessment the evaluator(s) determines a representation of the status quo of the organization. This representation forms the basis for action and change in behavior. These actions and behavioral changes can have an impact on the organization. When we refer to SATs in this thesis, these are the tools that we refer to. This scope consideration excludes all SATs that are not concerned with the organizational level, such as SATs for assessing your current medical situation or your personality traits. Further, as these SATs are meant to drive change in the organization, it also excludes SATs with other purposes than this. This includes questionnaires issued by external entities in connection to research projects.

Little is said about the SATs we are considering in general. What is written about SATs is typically by authors presenting tools they themselves have designed. We therefore start by considering the fundamental issues related to the role of such a tool in organizational development. Then we provide an overview of what an SAT is, followed by what components such a tool normally consist of.

2.1 FUNDAMENTAL CONSIDERATIONS ABOUT SATs

An SAT is a tool used to facilitate a self-assessment related to organization change efforts. The role envisioned for an SAT in the organizational change process will depend on one's social ontological position, or how we understand social entities - in this case organizations. In this discussion it is useful to make a distinction between people's representation of the organization, or image of it, and the organization itself. This representation may affect the organization through change in behavior or initiated action based on the image. This in line with the ontological position of critical realism (Bhaskar & Lawson, 1998). Critical realists subscribe to the view that there is a real world, including the social world, which exists no matter whether or how well we know and understand it (Bhaskar & Lawson, 1998, p. 17). In other words, there is an external reality to which we can direct our attention, which is also separate from our descriptions of it. Critical realists accept terms that account for

regularities in the natural or social order that may not be directly observable into theoretical accounts. For example, a concept such as "authority" can be used in theoretical accounts, even though they may only be observable through their effects. The natural and social worlds differ in that unlike the natural world, the social world depends upon human action for its existence and is socially constructed (Fairclough, 2010, p. 204). For example, humans can understand natural structures such as gravity, but we have no power over it. An organization on the other hand is a social structure, that humans can reproduce and transform (Bhaskar, 1998, p. 36). A distinction is made between construal and construction of the social world, where construal is our representation of it, or our descriptions of it (Fairclough, 2010, p. 204), and construction is the actual changing or reinforcement of the social reality that exists independently of any one actor. Because the mechanisms of the social world can be identified and described, one is given the opportunity to introduce changes to transform the status quo (Bryman, 2012, p. 29). In other words, our construals of the social world can have socially constructive effects that change the nature of the social world. The role of an SAT will then be to facilitate the construal of the organization, the representation of it, which serves as a foundation for organizational change.

2.2 BREAKDOWN OF SATS

To better explain these tools, we will in this section break down the concept of "SAT". The implications of the terms "self", "assessment" and "tool", will be considered.

2.2.1 SAT: "*SELF*"

In a self-assessment the people who are the subject of the evaluation (the evaluand), also perform the assessment. The evaluator also being the evaluand makes self-assessments somewhat of a niche in the world of evaluation. It has important implications for what can be expected from the use of these tools.

As the evaluators are, or are a part of the evaluand, the evaluators will have potentially high stakes in the evaluation results. Participants will be susceptible to self-report bias (Donaldson & Grand-Vallone, 2002), and there is no guarantee that they will have experience or knowledge of research methodology or evaluation

principles. In a study showing how management tools were actually used in organizations, Kaplan and Jarzabkowski (2006) showed that use of tools is a highly political, symbolic and socially interactive process. The resulting self-assessment report or other conclusions taken on the basis of a self-assessment will therefore be affected by methodological sources of uncertainty that researchers or others strive to overcome by using external evaluators. Using the results as generalizable and valid data for research or benchmarking activities external to the organization would therefore not hold to an academic standard. These results should therefore be restricted to use for internal development, and not as non-biased information about the organization. Also, use of these results for incentive purposes would increase the stakes the evaluators have in the results, and we would expect an increase in the bias and politicization of the tool. We would therefore argue that collecting self-assessment outcomes for use externally to the context they are generated in would be unwise.

The goal is in other words not to get externally trustworthy results, but to develop your organization. Having the organization itself perform the evaluation allows people internal to the organization to learn from the process. The learning from this process comes both from increased experience in assessment, and the knowledge acquired about the organization in the process. Also, assuming the results are used internally to the organization, allowing the evaluand to be the evaluator reduces the gap between those who evaluate, and those who use the results. Taking this argument to the extreme would be to say that if everyone in the organization took part in the evaluation, the gap between the evaluators and those who use the results would be closed. With a *self*-assessment, some of those who are going to realize the agreed upon changes implied by the new construal of the organization have also been part of construing that image. This does not necessarily mean that they all agree with the resulting representation, but they have been part of the process of creating it and therefore have an understanding of it. The constructive effects of the construal may therefore become stronger in a self-assessment than for an assessment performed by an external actor.

7

2.2.2 SAT: "*ASSESSMENT*"

When performing an assessment, the evaluator makes an effort to build an image of the organization, and also makes a judgment based on that image and the their view of what the organization should look like. These images, or construals (see chapter 2.1), are built by people internal to the organization through participation in the assessment process. The evaluators individually or collectively expand, adjust and coordinate their images of the organization. Change happens as these images become a foundation for envisioning and motivating action and change in behavior, or in other words when these construals have constructive effect.

Evaluators construe individual or shared images of the organization by considering information about the topics to be evaluated. Information must be gathered, its relevance evaluated, and processed, not necessarily in that order. This can happen by participants simply reflecting on the issues at hand, drawing on their own knowledge, interests and experience. In contrast, it can also happen by defining the concept in question and what information is relevant to this concept, and then trying to measure it. For example, to make a judgment about job satisfaction, participants can draw upon their own knowledge and experience, or they can attempt to measure the concept of job satisfaction. However, we note that to assess and to measure are not equivalent activities. To assess is defined in the dictionary to be "evaluate or estimate the nature, ability, or quality of", while measure is defined "ascertain the size, amount, or degree of (something) by using an instrument or device marked in standard units" (Oxford Dictionaries, 2015) (Oxford Dictionaries, 2015). Still, the SATs within our scope may easily be mistaken for measurement instruments. Acknowledging that there is a theoretical possibility of having tools that serve as a form of measurement instrument within our scope, we choose to exclude such tools now. Such a tool would need very different attributes than those we have observed in the literature thus far. Also, for organizational change purposes we argue that an SAT that relies on reflection of participants, and where measurement results would be input to the assessment process rather than the output, can be a good choice. To illustrate this we provide a brief explanation of measurement in social science and its complexities below.

Although it may seem intuitive that an SAT could serve as a measurement instrument, and measurement may appear to be a superior approach to assessing the organization, measuring concepts in social science is seldom a straightforward endeavor. In fact, the positive effects of using measurement may often not be worth the extra effort needed. In some rare cases where a concept is possible to quantify in a reasonably direct way, it may be measured directly. For example, if you wish to measure household income or a number of children, this is relatively straightforward to measure. However when concepts are not easily quantified, the researcher must choose measurable indicators to stand for the concept. An indicator "is something that is devised or already exists and that is employed as though it were a measure of the concept" (Bryman, 2012 p. 164). These indicators are normally devised by the researcher based on commonsense understanding of the concept or from anecdotal or qualitative evidence (Bryman, 2012, p. 164). Sometimes multiple indicators are deemed necessary to measure one concept (Bryman, 2012, p. 166). Concepts we are interested in for organizational change are often at a high abstraction level and complex, for example organizational competitiveness and quality of products and services. Finding suitable indicators for such concepts is a complex and difficult task, and indicators used in one organization may be meaningless in another. Measuring such concepts therefore requires a lengthy and potentially futile process to find suitable indicators, a process each organization may need to repeat for themselves.

Furthermore, using measurement necessitates complying with quality criteria that most self-assessments today do not satisfy. One of the main arguments for using measurements is that the results can be expected to say something true about the concept that is measured. In social research, the quality of the resulting data is often evaluated by considering the reliability and validity of measurements. Here, reliability is concerned with the consistency of a measure of a concept and validity refers to the matter of whether an indicator (or set of indicators) really measures the concept that is meant to measure (Bryman, 2012). To be able to provide reliable and valid results, considerable effort must go into securing the quality of the measurements. Most self-assessments today typically inquire about large concepts, such as "productivity" and "innovation", and some have lower-level "indicators" to help answer the questions. However, these indicators are far from the customized, detailed, complex and

quantifiable indicators that would satisfy the quality criteria above. This does however not mean that SATs cannot provide quantified results. Actually, many SATs do just this, adding to the potential to confuse these assessments with measurement instruments. By asking the evaluators to give a score to these concepts, many such tools leave the quantification process to the evaluator's opinion. Here, an important distinction must be made between measuring a concept and quantifying a concept. The difference is that quantification does not have to come from a measurement process. Thus, the fact that a concept has been quantified, or "given a score" does not necessarily mean that the resulting numbers satisfy any quality criteria.

By concluding that the SATs that we consider in our scope are not measurement instruments, we are left with the following question: if these tools are not measurement instruments, what are they? As seen in chapter 2.1, SATs can facilitate the construal process of a common representation of the organization. In the same chapter we explain how construing an image of an organization is an important process, because this image may have constructive effect as it forms the basis for action. Facilitating such a construal is therefore an important task. Moreover, it is a task that can be achieved through the reflection of participants on topics concerning the organization, and do not imply a need for measurement. Therefore, we assert that the SATs we are considering are tools to help organizations in the process of construing the image of the organization and how it should be. Furthermore, we argue that by allowing for reflection, and not just strict measurement, will enable the organizations to consider a wider range of topics. Not all concepts are measurable, at least not without a very time consuming and tedious measurement process including many indicators. However, this is not to say that they cannot be assessed or that such an assessment would not be useful. When organizations use the SATs we consider, the image of the organization is built through reflection, where participants draw on their knowledge and experience. By leveraging the participants' experience, knowledge and analytic skills, an assessment process can provide answers about the organization that may be hard to acquire through measurement data. For example, participants in an assessment process may be able to say something about vague topics such as "coherence of strategies" that would be hard, time-consuming and potentially not possible to find valid indicators for.

2.2.3 SAT: *"Tool"*

The kind of SAT examined in this thesis has the form of a document with a set of talking points, questions or the like. They are typically a document with a list of questions or statements, often in the shape of a test, that participants use to for example rate their organization. The designs we have identified for SATs that are within the scope of this thesis are described in detail below in chapter 2.3.

An important aspect of an SAT is its material property, as opposed to being purely conceptual. As a document with written words, phrases and formulations, lines and dots and perhaps even numbers, it is a consistent object that can be handed from person to person without losing its shape or content, in contrast to for example a spoken thought. It can be pointed to, waved in front of peoples' faces, talked about and changed - and it will stay changed in the same way until someone else comes along to change it again. An SAT's material property leads to a conserving effect, where information is encoded and conserved across people, distance and time.

A consequence of an SAT's material property is that the author can, just as they can through writing a textbook, communicate their ideas, both insights and agendas, through the tool. These ideas, the insights and agenda, will be referred to as the underlying theory of the tool. Through using the tool, the participants may acquire new knowledge from the introduced theory, for example knowledge about other successful organizations or new perspectives that change the way the evaluators view and analyze their organization. Bhaskar (1989, p. 2) argues that to change the social world around us (in this case the organization), we need to understand it. He argues that to understand it we need to identify the structures at work, which is what the social sciences aim to do. Introducing theory about the structures at work in an organization can therefore be a fruitful way to help the evaluators understand their organization. This means helping the evaluators construe an image of the organization that serves as a foundation for initiating actions and behavior that do in fact result in a changed organization. This effect clearly depends on the theory, and that the theory and communication of it helps the evaluators develop a better understanding of their organization.

Using a tool to facilitate a self-assessment can also have a focusing and guiding effect. When the process is centered on a tool, the agenda is already set by the contents and design of the tool. By agreeing to use the tool, the participants let the tool set the "rules of the game". Although participants may arguably decide to oppose the tool, or parts of it, the tool will be the source of a focusing and guiding force. The process itself may even be set and facilitated by the tool, by for example giving the users the task of scoring their organization or finding improvement projects. These focusing and guiding effects may reduce the time, effort and focus otherwise needed to structure the process, and to agree on topics and tasks for the participants.

Combining the conservational, focusing and guiding effects an SAT can have, introducing a tool can be an effective way to influence people to behave in a certain way. Introducing tools is for example a common way to standardize processes in an organization. Perhaps all doctors need to tick off on a checklist to ensure standardized and complete discharge procedures. Creating a tool can be an influential way to spread knowledge, as the material property ensures the information encoded into the tool stays intact (though this is not a guarantee for conservation of meaning or initial intention). A tool's focusing and guiding properties may also reduce the complexity associated with making use of the tool's creator's knowledge. Consider the engineering student's challenge of helping her grandmother make a family heritage website. The grandmother would not be able to make a website if this required understanding all the paths one could take to this end. However, the student could make a list of all the steps that need to be taken, perhaps including some pictures of the boxes to click, that the grandmother could follow without needing to understand all the choices. The grandmother could then use the student's list to make more websites, enjoying the tool's conserving property as well.

Information is coded into the tool and conserved through writing. This means that the information is mediated through language, or discursively mediated. Directions for use and the concepts to be assessed and their descriptions, all that constitutes the theory the author has coded into the tool, is therefore discursively mediated. We have argued that the theory coded into the tool is very important, as it may influence the image construed by the evaluators and therefore the foundation for action and change of behavior. This calls for a focus on the discourse embedded in the tool and its

effects. The field of critical discourse analysis (Fairclough, 2010) (CDA) provides such a focus, seeing the socially constructive effects of discourse as a central concern. Through discourse, different versions of the social world is constructed and reconstructed (Bryman, 2012). Through dialogue, we change or reinforce certain ways of talking about concepts (develop or reinforce a discourse), while exposure to new ways of talking about a concept may change our views and behavior in relation to it. Following the ontological position of critical realism, CDA draws a distinction between our understanding of social entities and the entities themselves. Fairclough (2010, p. 204) outlines the difference between development of our understanding (construal) and of the social entity itself (construction): the world is discursively construed in many ways, but which construals come to have socially constructive effects depends upon important aspects such as power relations. Discourse shapes what can be said and who can say it, and determines social practices by realizing rules, identities, contexts, values and procedures (Grant, Iedema, & Oswick, 2009). The practice of discourse "rules in" certain ways of talking about concepts that are judged to be acceptable, legitimate and intelligible. It also limits and restricts the way we talk about and conduct ourselves in relation to the concepts, and how we construct knowledge about them (Grant et al., 2009, p. 216). CDA provides us with analytical concepts that can help us explain the socially construal and constructive effects of the discourse embedded in the tool, and will be employed in our thesis.

The conserving property of a tool may also allow for it to become a common point of reference and a mediating artifact. Consider that tools can be pointed to, and can be changed or filled in, passed from person to person, then changed or filled in again. It is a common object, residing outside of each person's mind but also understood in some way by all users. Unlike a thought conveyed through conversation or action, the tool is a constant artifact available to several people at once. This aspect only becomes relevant when the tool is available and used by several people that make out a heterogeneous group, as will for example be the case when people come from different divisions or belong to different professions. This is very often the case, and the way a tool is used as a mediating artifact between people or groups is both interesting and important. This leads us to consider theory about boundary objects. A boundary object is an object that resides between social worlds and enables

cooperation (Star, 2010). The concept of boundary objects was created in reaction to other theories about cooperation, which the authors found to be too focused on consensus (Star, 2010; Star & Grisemer, 1989). Star (2010) relates that through her fieldwork among scientists and on cooperation in heterogeneous groups she found that even though consensus was rarely reached, and when it was it was often fragile, cooperation would still continue. So, while discourse analysis sheds light on the effects of discourse found in the tool and in the organization, conceptualizing the tool as a boundary object informs how the tool is used as an artifact of cooperation in heterogeneous groups.

Boundary objects inhabit several social worlds and satisfy the information requirements of each (Bowker & Star, 1999). They are "weakly structured in common use and more strongly structured in individual site-use" (Bowker & Star, 1999, p. 297). When local groups work on the object, they maintain the vaguer identity of the object, while at the same time making it more specific and more tailored to local use within the social world (Star, 2010). A group with participants from different social worlds can for example have a common map or diagram, but have more detailed versions or understandings within their local worlds. The social worlds may not fully agree on what the common representation is saying, but having it enables them to communicate and work together. Using this approach, groups can cooperate without consensus, and tack back-and-forth (Star, 2010) between the vague and the local form of the object (Star, 2010, p. 605). An example of such a common object is the ICD, or the International Classification of Diseases. The ICD enables information to be moved across the globe between conflicting medical beliefs and practices systems (Bowker & Star, 1999, p. 290). More relevant to our context, we can employ the example of a shared project schedule. In the Project, many different groups have to cooperate to meet all deadlines and achieve the set milestones. In this case, they can have shared project schedule with all their common deadlines. Different groups using the shared project schedule can have more detailed and specialized plans locally. Local groups can communicate and change the common schedule to coordinate with other local groups and to meet local requirements. In this way, the local groups tack back-and-forth, adapting the common project schedule and their local plan.

Furthermore, this can be done without understanding what the other groups are doing, or agreeing with how they setup their schedules.

2.3 DESIGNS OF SATs AND THEIR COMPONENTS

There are a vast number of available SATs, and these take on different forms. They are also made to assess a number of different concepts, for example incentive systems or concepts found in production system theories such as TQM or lean. Maturity models is one form of SAT that has been proposed for a range of different activities including, but not limited to, quality management, supplier relationships, software development and innovation (Fraser, Moultrie, & Gregory, 2002). Maturity models, or maturity grids, include a number of maturity levels. Many also have descriptions for each level, which is how they "codify what might be regarded as good practice" (Fraser et al., 2002, p. 244). For example, Bhasin (2011) presents an SAT for "measuring how lean" an organization is. The tool asks you to give the rating between one and ten, and provides descriptions of the highest and lowest rating for each concept to be assessed. Keep in mind that the term "measurement" is here used in a very informal way (see chapter 2.2.2), probably referring to that the evaluators are asked to quantify each concept. Some maturity models also take similar approaches to the tool presented by Bhasin (2011), though with slightly varying designs. For example, some resemble Likert questionnaires where only the highest level, or ideal, is described for each concept (Fraser et al., 2002).

During our exploration of the vast and varied offerings of SATs developed in the industry and academia, it has become clear that the basic structure of these tools seems to be common, and several other components are shared. The design of the specific SAT that we evaluate in our thesis is only one of the possible designs for SATs that fall within the scope of our general discussion. In the following, we strip down SATs to what we have seen to be common of all SATs and then add components gradually to better be able to explain the functionality of each component. This will also give an overview of the different kinds of SATs that fall within our scope.

15

2.3.1 DESIGN 1: THE BARE MINIMUM (BASIC COMMONALITY)

In addition to being a possible complete SAT in itself, Design 1 is the point of departure for all of the other designs. This is the commonality we have identified in the range of SATs encountered in our search. The bare minimum is simply a list of themes, or concepts, that the participants are to consider and then assess for the organization. This assessment can either be done individually or in a group.

Theme	Notes
Incentive systems	Example: "We only have bonuses, but I don't think they are good enough motivation. Perhaps we should create a couple new incentives.."
...	
...	
Continuous improvement	Example: "I don't know what this is!"

FIGURE 1: DESIGN 1: BARE MINIMUM

The bare minimum design can in itself reap many of the benefits we have argued that SATs may provide. In other words, you do not have to add any extra components to be able to reap these benefits. The bare minimum can focus and guide the assessment processes. Furthermore, even though this is the stripped-down version, the tool can still transfer knowledge from its authors about how to see the organization. By design, what the author considers relevant to focus on in an organization is coded into the tool by introducing some concepts and excluding others. Theory, here used in a wide sense, is therefore coded into (even this relatively simple) tool.

2.3.2 DESIGN 2: AN SAT WHERE THE CONCEPTS ARE ACCOMPANIED BY AN IDEAL

This form is similar to the bare minimum, but for every theme, or concept, there is a concrete description of the goal. For example, if the theme is "quality work" the ideal could be "continuous improvement is a part of the culture". Though these are typically best practices, we acknowledge that these ideals could either be theoretical ideals, best practices, or ideals set out by the organization itself. It could also be a

combination of these, with for example both the theoretical ideal and something specific for the organization.

Theme	Ideal	Notes
Incentive systems	Coordinated incentive systems support the company strategy	"We only have bonuses, our strategy requires more enthusiastic people. Perhaps we should create an incentive for enthusiasm."
...	...	
...	...	
Continuous improvement	Continuous improvement is part of the culture	"We have a long way to go before everyone thinks work with standardization is cool..."

FIGURE 2: DESIGN 2: BARE MINIMUM WITH IDEAL

Including the component that presents an ideal to each theme/topic has some implications. By staking out the directions of the decision process, you both direct and focus the process. The complexity of the reflection (or discussion if it is done in a group) is reduced from including a consideration of all possibilities in connection to the theme, to a comparative activity where people compare their organization to the description of the ideal. The inclusion of an ideal introduces a highly normative element, as the word "ideal" indicates. This may serve as a strong guiding force in the process, reducing the options for what directions to peruse and thereby what actions are considered. Furthermore, including ideals increases the amount of information that can be coded into the tool. This means that including an ideal increases the amount of knowledge that can be transferred from the authors of the tool to the participants in the assessment process. For example, if the ideal is made on the basis of a theory (for example lean), the participants of the assessment can be provided with more guidelines as to how to follow the theory than with the bare minimum.

2.3.3 DESIGN 3: AN SAT WITH LEVELS ASSOCIATED TO THE CONCEPTS/THEMES

For this design, the SAT consists of a number of themes with associated levels. Such levels go under many names, where examples are capability levels or maturity levels. The most natural way to envision this is a set of concepts with ideals, and with levels where the lowest level is far from the ideal, and the highest level is the ideal. For the sake of argument however, this need not be the case and levels is seen as a separate component that may be added to the SAT without an ideal. We call the bare minimum design with levels Design 3. How many levels are used may vary. The levels can be presented in several forms. They can either just be assigned numbers, or they can have generic description. Generic descriptions are descriptions that are not specific to each theme, so the different levels have the same names for all the themes. For example, the highest level can be "world class" and the lowest level can be "below industry standard".

Theme	Level1 Below Industry Standard	Nivå 2 Industry Standard	Nivå 3 World Class
Incentive Systems	☐	✔	☐
...	☐	☐	✔
...	✔	☐	☐
Continuous Improvement	☐	☐	✔

FIGURE 3: DESIGN 3: SAT WITH LEVELS

Adding levels enables the tool to influence the assessment process beyond guiding the topic of conversation. Adding levels to the SAT serves as a process guide. The participants can either be asked to just assign their current state, or they can also be asked to assign a future or desired level. Having the participants determine the level of the organization for each topic may have a focusing effect because it demands that people make a choice. Also, the setting a level visualizes the participants' judgments, which may have a motivating effect. Furthermore, if the assessment is done in a

group, having levels may concretize discussion as people discuss the organization's level or defend independent individual choices.

2.3.4 DESIGN 4: (DESIGN 2+3) SAT IDEAL AND LEVELS

This is a more complex design where both levels and an ideal are included. When combining the two components, the ideal is the description of the best level. This design can take on many different forms, either with a description of each level for each theme, or with descriptions for only some of the levels while the descriptions for other levels are replaced by a number. For example, you can have five levels where only the lowest and highest levels are described for each theme. It is also possible to only describe the ideal for each theme, and then have the other levels be generic descriptions of how close you are to the ideal.

Theme	Level 1	Level 2	Level 3 / Ideal
Incentive Systems	Incentive systems are ad hoc and in conflict	Incentive systems are coordinated	Coordinated incentive systems support the strategy
Continuous Improvement	Improvement initiatives are ad hoc	Improvement initiatives are systemized	Continuous improvement is a part of the culture

FIGURE 4: DESIGN 4: LEVELS AND IDEAL

The combination of ideals and levels may have a further focusing and guiding effect on the self-assessment process. The more levels have descriptions, the more information can be coded into the tool. Following, by adding descriptions to more levels the authors may further guide the participants in the process. This may make the tool easier and more comfortable to use, as the complexity in the reflection process is further reduced. It also makes it possible to transfer more knowledge to the participants, for example enabling them to follow the theory in greater detail without being familiar with it. Inclusion of levels and ideals may aid in the process of making a change initiative plan based on the theory embedded in the tool, as descriptions of levels can provide guidance. However, attention should be drawn to that the more

detail is added to the tool, the less freedom is provided to the participants. Including extensive details and descriptions may reduce the reflection process to one concerned with matching what is described in the tool to what the participants can find in their organization.

2.4 SUMMARY

In this chapter we have described our fundamental ontological assumptions, considered central aspects of SATs, and outlined the different forms of SATs that fall within the scope.

The role envisioned for an SAT will depend on one's social ontological position. As we make a distinction between people's representations of the organization and the organization itself, we look to the ontological position of critical realism. Critical realists prescribe to the view that there is an external reality to which we can direct our attention, which is separate from our descriptions of it. In line with this, a distinction is made between construal and construction of the social world, where construal is our representation or descriptions of it, and construction entails actually changing or reinforcing the social reality independent from individual actors. Following, the role of the SAT is to facilitate the construal of the organization, the representation of it, which will then serve as a foundation for organizational change.

One central aspect of SATs, is the fact that those evaluating are also a part of what is being evaluated. This means that the evaluators will be susceptible to self-report bias, and that the assessment process may become a political, symbolic and socially interactive process. Care should therefore be taken when utilizing the results of such assessments, and collecting self-assessment outcomes for use external to the context that they are generated is argued to be unwise. Further, we argue that the goal of a self-assessment to develop the organization and not to get externally trustworthy results. Doing the assessment internally allows people to learn from the process, and it means that those who are going to realize the agreed-upon changes have been a part of it construing the image these decisions are based on.

We then consider how a self-assessment is a process of assessment. The assessment process is described as one where an image of the organization is construed, and a judgment is made on the basis of this image. SATs are often mistaken for measuring instruments. However, it is important to keep in mind that measuring concepts that are not directly quantifiable, which topics typically covered in SATs are not, brings with it a set of quality criteria that self-assessments do not fulfill. Moreover, the role of the SAT is to facilitate the construal of the organization and this task does not imply a need for measurement. In self-assessments participants leverage their experience, knowledge and analytic skills. Allowing for reflection and not just strict measurement will enable consideration of a wider range of topics. So, if the output of such a self-assessment is a set of numbers, these numbers should not be seen as measurement results.

The last central aspect of SATs to be considered is its material property. In other words, we look at the effects of it being a tool. The material property of an SAT leads to a conserving effect, where information is encoded and conserved across people, distance and time. As a consequence, the author of the tool can communicate ideas or theory through the tool. Through using the tool, the participants may therefore acquire new knowledge from the introduced theory, which may help the participants understand the underlying structures of their organization. Furthermore, using the tool may have a focusing and guiding effect, as all participants have agreed to use the tool and therefore allows it to set the "rules of the game". The combination of introducing theory, and focusing and guiding the process enables the tool to influence the participants to behave in a certain way. Uncovering the information that is coded into the tool is therefore an essential part of understanding the effect of the tool. As this information is discursively mediated, the socially construal and constructive effects of the discourse embedded in therefore important to look at when evaluating a tool. Furthermore, for different social groups to cooperate without necessarily reaching consensus, the tool should ideally serve as a "boundary object", enabling cooperation through serving as a common artifact inhabiting several social worlds.

Furthermore, we look at how SATs can have various designs. However, they all seem to share some common components. To describe the different components, and give

an overview of the possible designs that SATs can have, we present designs where each outline one potential form of an SAT. Design 1 is referred to as the bare minimum, and is the basic commonality of all forms of SATs. It is simply a list of themes, or concepts, that the participants are to consider and assess for the organization. Design 2 is an SAT where the concepts are accompanied by an ideal. This design is similar to the bare minimum, but for every theme, or concept, a concrete description of the goal is included. Design 3 also resembles the bare minimum, but for each concept, or theme, there are associated levels that the participants are meant to use to score the organization in the assessment. The last design, design 4, is a combination of design 2 and design 3, which entails that the design has both levels and an ideal associated to each topic, or concept.

3 CRITERIA FOR SATS

In this thesis we seek to evaluate an SAT, which as previously stated is a tool that facilitates a self-assessment. To lay the groundwork for systematically and critically analyzing an SAT, we will establish a set of criteria. Rebolloso, Fernández-Ramirez, Canton, and Pozo (2002), Rebolloso, Fernández-Ramírez, and Cantón Andrés (2008), and Rebolloso, Fernández-Ramírez, and Cantón Andrés (2009) have done some research on evaluation of SATs, with focus on the use of one such SAT in a university. However, beyond this, little research has been done to develop criteria for SATs specifically. We therefore base our criteria on the criteria presented in Rebolloso et al. (2002). However, while their criteria are found to be a good foundation, they are not elaborate or detailed enough to serve as a systematic and comprehensive guide for our evaluation. We are therefore brought to attempt to expand on these criteria. To do so, a thorough understanding of what the tool is to facilitate, self-assessments, is key. We therefore need in-depth knowledge self-assessments and what make them good. Criteria for self-assessments are therefore explored in the first section of this chapter. Little is written about criteria for *self-*assessments specifically, but the field of evaluation, or assessment, is considerable. To understand what criteria are relevant for self-assessments, we combine our conceptual understanding of SATs from chapter 2 with theory from the field of evaluation. After criteria for self-assessments are developed, we consider criteria for SATs. The criteria for SATs from Rebolloso et al. (2002) are expanded with the components of the criteria for self-assessments that can be attributed to a facilitating SAT.

This chapter is structured as follows. We first look at criteria for the process that is to be facilitated by the SAT, namely self-assessments. Here, we consider three different authors' criteria for evaluations, and discuss the relevance of the criteria for self-assessments. The evaluation criteria that are found to be suitable to self-assessments are referred to as the self-assessment criteria (not to be confused with criteria for SAT*s*). We then go on to consider criteria for SAT*s*. We start off with presenting the SAT criteria that resulted from the research of Rebolloso et al. (2002). Further, we explain the role of the tool in a self-assessment process. This forms the basis for a discussion about which components of the self-assessment criteria can be attributed to

the tool. Finally, we go through each of the SAT criteria of Rebolloso et al. (2002), and elaborate on the content so that all relevant self-assessment criteria are considered.

3.1 LITERATURE REVIEW: THE FIELD OF EVALUATION

3.1.1 STUFFLEBEAM'S CRITERIA

When considering criteria for assessments (in our case self-assessments), we are evaluating evaluations. The evaluation of evaluations is coined meta-evaluation by Scriven (1969). A still influential article defining the start of meta-evaluation as a field was Stufflebeam's (2011) classic article, originally published in 1974. He introduced a comprehensive framework for conducting meta-evaluations in a systematic way. His framework covers the whole process of meta-evaluation, from assessing how the evaluation is designed and carried out, to how the results are dispersed and used by the audience. The framework includes a complete set of evaluation criteria, grouped into three main categories. First of all, evaluations must produce findings that are useful to stakeholders, and the first category is therefore usefulness (Stufflebeam, 2011, p. 101). The usefulness category assures that the evaluation findings are informative to the audience and makes a desirable impact on their work. Second, seeing as the findings must also be worth more than the cost of obtaining them, the second category is cost/effectiveness. Both the categories of usefulness and effort are very relevant for self-assessments. Performing a self-assessment is supposed to help drive change that will have a positive impact on the organization. As described in chapter 2, self-assessment is the process of construing a common representation of the organization, which again can have socially constructive effects. As for the effort category, if the self-assessment is to have a positive impact on the organization, the resulting effect of the assessment must be worth the effort that is put into it. If the assessment takes up more effort than what the positive impact is worth, the entire purpose of the self-assessment is unfulfilled.

Stufflebeam's last category is technical adequacy. The technical adequacy category is taken from accepted criteria for research because both research and evaluation reports must contain sound information (Stufflebeam, 2011, p. 101). The criteria in this

category are internal validity, external validity, reliability and objectivity. These technical adequacy criteria are the same criteria that Bryman (2012) presents as technical criteria for quantitative research in his book on social research methodology. There is an ongoing debate about the use of these criteria for social research. For example, Lincoln and Guba (as cited by Bryman, 2012, p. 49) propose that alternative criteria for assessing qualitative research are required. They propose the concept of trustworthiness as a criterion of how good a qualitative study is, which consists of credibility, transferability, dependability and confirmability (Bryman, 2012, p. 49). This debate is attempted reconciled in the influential book "Theory-driven Evaluations" by Chen (1990). Chen (1990, p. 27) tries to reconcile the debate about what criteria should apply for evaluations by arguing that there is no one best method that can universally apply to every evaluation situation. For example, Chen (1990) points out that the classic criteria of internal validity typically employed for quantitative methods, and that of credibility used for qualitative methods, both essentially have the same ultimate goal of producing trustworthy results. Chen (1990) tries to overcome the need for having two labels for the same basic goals by naming his evaluation criteria by their purpose rather than by what research methods are used to fulfill them (Chen 1990, p. 60). Thus, to avoid stepping into an extensive discussion about social research methods, we look to Chen's criteria for evaluations.

3.1.2 CHEN'S CRITERIA

Chen (1990, p. 34) proposes what he calls theory-driven program[1] evaluation as an alternative perspective to evaluations. As mentioned above, Chen (1990, p. 34) tries to reduce the focus on research methodology in his meta-evaluation criteria, in favor of the goals to be reached. The meta-evaluation criteria he synthesized are responsiveness, objectivity, trustworthiness, and generalizability (Chen 1990, p. 58). He does not claim that the criteria are exhaustive, but rather a synthesis of the criteria

---

[1] Chen refers to theory-driven program evaluation in his book. However, the line between the field of evaluation and the field of program evaluation is somewhat blurred, and the concepts of evaluation and program evaluation are often used interchangeably in the literature. Program evaluation literature is however typically concerned with evaluations on the organizational or national level, not for example of a single person, which suits our purpose. We will refer to program evaluation as evaluation in this chapter, to not create an unnecessary perceived difference between the two.

of most concern to many evaluators. We note that these criteria have been named so that they can be confused with the more common uses of these words. The list below describes the contents of Chen's (1990) criteria. To avoid confusion, when we use Chen's (1990) criteria they will be referred to as his criteria, for example "the objectivity criterion".

The responsiveness criterion is strongly related to Stufflebeam's (2011) category of usefulness and is considered to cover this criterion. Responsiveness refers to the need for evaluations to be relevant and useful to the needs and concerns of all stakeholders, whose views should be equally represented. However, "all stakeholders" is a much stronger consideration than Stufflebeam's (2011) "useful to some stakeholders". This reflects an increased awareness about ethical matters such as fairness in the field of evaluation. An evaluation needs to be timely to be responsive, meaning that stakeholders must get the information while there is still time to use it. The results of the evaluation should be relevant for stakeholders' decision-making process, and it should be broad, taking in the whole range of program consequences. On the same grounds as the argument above for why Stufflebeam's (2011) usability is relevant for self-assessments, responsiveness is a reasonable criterion for a self-assessment.

The objectivity criterion refers to the expectation that the evaluators' personal preferences or loyalties should not influence evaluation results. Objectivity also refers to that results yielded by certain research techniques or procedures should be replicable by others by using the same techniques and procedures. This is highly connected to the traditional criteria of validity and reliability. As argued in chapter 2.2.1, the results of the self-assessment will be highly dependent on what experiences the participants bring with them, and what reflections they are capable of. Therefore, the results will not be replicable by others. A fundamental attribute of self-assessments therefore compromises the assessment's objectivity. Therefore, if one chooses to perform a self-assessment, this criterion should not be used to evaluate it.

The trustworthiness criterion is defined as an assurance that an evaluation will provide convincing evidence that can be trusted by stakeholders in their utilization of evaluation results. In his attempt to separate between the discussion of methods and

values, Chen (1990) emphasizes that under a set of given conditions the trustworthiness criterion can be fulfilled by a variety of methods and strategies. The trustworthiness criterion seems reasonable for self-assessments. As seen in chapter 2.2.2, the purpose of the self-assessment is to construe an image that forms a basis for action. Bhaskar (1998, p. 2) argues that to change the social world we need understand it, which entails identifying the structures at work (see chapter 2.2.3 for further explanation). It is therefore important to make sure that the construed image is based on an understanding of the structures of the organization. Demanding that the assessment include some form of strategy that ensures that the process yields convincing evidence is therefore considered to be a sensible criterion.

The generalizability criterion refers to the extent to which evaluation results can be applied to related future circumstances in which stakeholders are interested (Chen, 1990). The purpose of the self-assessment is to construe an image that can form the basis for decision-making (see chapter 2.2.2). In other words, the result of a self-assessment is a created image of the current state of the organization, and hopefully action that changes the organization for the better. This will effectively mean that the assessment results will become outdated. Therefore it is counter intuitive to demand of all self-assessments that their results have to remain relevant. In other words, focusing on making the results generalizable may be an unnecessary restriction of the self-assessment process. This generalizability criterion is therefore not considered relevant to self-assessments.

3.1.3 GUBA AND LINCOLN'S CRITERIA

As a fundamentally different approach to evaluations, Guba and Lincoln (1985); (Guba & Lincoln, 1989, 2001) have proposed an approach to evaluation coined fourth generation evaluation. In fourth generation evaluation, all meta-evaluation criteria are developed from the basic assumptions of the constructivist paradigm (Guba & Lincoln, 2001, p. 1). This is an ontological position that asserts that social phenomena and their meanings are continually being accomplished by social actors (Bryman, 2012, p. 33). The criteria are therefore centered around this social construction of reality. We have argued that self-assessments facilitated by the tools in our scope is a process of construing an image of the organization. By leveraging the participants'

experience, knowledge and analytic skills, a self-assessment process can provide answers about the organization that may be hard to acquire with more rigid methods (see chapter 2.2.2). As critical realists with focus on the process of construing an image of the organization, we are interested in the construction that receives emphasis by Guba and Lincoln's (1985, 1989, 2001). However, we subscribe to an ontological position that separates between the process of construal and that of social construction (see chapter 2.1), meaning that we do not regard construing an image to be the same process as actually constructing the social world. Rather, we regard the constructivist concept of construction (Guba and Lincoln, 1985, 1989, 2001), as a process of construal (Fairclough, 2010). From a critical realist standpoint then, the criteria of fourth generation evaluation can provide insight into the construal process that we consider to be very central to self-assessments. Criteria of fourth generation evaluation are presented below.

The criteria found within the framework of fourth generation evaluation are coined authenticity criteria. These are fairness, ontological authenticity, educative authenticity, catalytic authenticity and tactical authenticity (Guba & Lincoln, 1989, 2001). Fairness is achieved by ensuring that all competing views are accessed, exposed and taken into account during the assessment (Guba & Lincoln, 2001, p. 7). Ontological authenticity is the extent to which the constructions of those being evaluated, and those evaluating, have become more sophisticated and informed (Guba & Lincoln, 2001, p. 7). Educative authenticity requires participants in the process to have become more understanding of others' constructions (Guba & Lincoln, 2001, p. 7). Catalytic authenticity is the extent to which action is stimulated and facilitated by the evaluation, and tactical authenticity is the degree to which participants are empowered to take the action that the evaluation implies (Guba & Lincoln, 2001, p. 7). In short, the assessment process and results should provide organizational learning, promote fairness and understanding of different views, and support peoples' ability to act on the results. All of these criteria can be achieved by a self-assessment, as they do not require an external evaluator. In fact, Guba and Lincoln (1985, 1989, 2001) emphasize that including stakeholders in the process is important. The authenticity criteria are considered relevant and central to investigating the process of construing an image that enables actors to take action.

3.1.4 SUMMARY OF SELF-ASSESSMENT CRITERIA

In this chapter, we have identified several evaluation criteria that are relevant for self-assessments, making out our self-assessment criteria. Firstly, we have found Stufflebeam's (2011) usefulness and effort to be important for self-assessments. The usefulness criterion ensures that findings are informative to the audience and makes a desirable impact on the work and is considered covered by Chen's (1990) responsiveness criteria. Including the effort criterion means considering whether the findings are worth the cost of obtaining them.

Secondly, we include Chen's (1990) responsiveness and trustworthiness to our set of self-assessment criteria. The responsiveness criterion refers to the need for an evaluation to be relevant and useful to the needs and concerns of all stakeholders, whose views should be equally represented. This is strongly related to Stufflebeam's (2011) usefulness criterion, and it is considered covered by the responsiveness criteria. The trustworthiness criterion ensures that an evaluation provides convincing evidence that can be trusted by stakeholders when they use the evaluation results.

Thirdly, Guba and Lincoln's (1985, 1989, 2001) authenticity criteria are found to be relevant for self-assessments. The authenticity criteria are fairness, ontological authenticity, educative authenticity, catalytic authenticity and tactical authenticity. Fairness is achieved by ensuring the consideration of all competing views. Ontological authenticity is the extent to which the constructions of those evaluating or being evaluated have become more sophisticated and informed. Educative authenticity considers whether participants have become more understanding of others' constructions. Catalytic authenticity is the extent to which action is stimulated and facilitated by the evaluation. Tactical authenticity is concerned with the degree to which participants are empowered to take the action implied by the evaluation.

3.2 CRITERIA FOR SAT*S*

With a presentation of self-assessments, SATs (SATs) and evaluation theory, and a synthesis of criteria for self-assessments, we are ready to consider criteria for SATs. As previously stated, the purpose of an SAT is to facilitate a self-assessment of an organization. Overall, a good SAT should help the organization achieve a good self-assessment. This means that the criteria for the tool will depend on what is considered a good self-assessment, and the ways in which a tool can contribute in this process. There is little research providing criteria for SATs. One group of scholars however, Rebolloso, Fernández-Ramirez et al. (2002) and Rebolloso et al. (2008, 2009), have contributed considerably to the literature on SATs and provided criteria for these.

3.2.1 EXISTING CRITERIA FOR SATs

Rebolloso et al. (2002) developed criteria specifically for evaluating SATs. These criteria have been shown to cover the main issues experienced by evaluators using an SAT and have a low degree of overlap (Rebolloso et al., 2002). The focus of their research was a case SAT used to assess a quality system introduced in a Spanish university. The quality system is a version of the theory Total Quality Management that has been introduced to European universities. The SAT evaluated by Rebolloso et al. (2002) and Rebolloso et al. (2008, 2009) helps the organization benchmark itself against a set of best practices, a variant of Design 4 (see chapter 2.3.4). They have developed criteria for SATs through literature analysis and empirical research where they study the use of their case tool in practice. In their paper, Rebolloso et al. (2002) explored five criteria: "comprehensibility", "clarity", "usefulness", "relevance" and "effort". These were identified and developed with consideration of evaluation theory and empirical research where they identified the main challenges people experienced when using the tool. Weight was given to the criterion's testability in practice and that they contribute to the SAT gaining supporters and a reputation for its usefulness and advantages (Rebolloso et al., 2002). They conclude that for further research they advise combining these criteria so that there are only three main criteria. Both the criteria comprehensibility and clarity, and the criteria usefulness and relevance, were so interdependent that they should be seen as one (Rebolloso et al., 2002). Building on their research, we use the term "comprehensibility" to refer to the combination of comprehensibility and clarity. This is because in order for users to comprehend the

meaning of the tool's content it must be formulated clearly, making clarity a sub-criterion for comprehensibility. To satisfy the unified criterion of comprehensibility, the concepts introduced in the tool must be understandable to those performing the assessment (comprehensibility), and the concepts need to be clearly expressed and explanations provided need to be clear, simple and exhaustive (clarity) (Rebolloso et al. 2002). The term "usefulness" refers to the combined criteria of usefulness and relevance, with relevance being a sub-criterion of usefulness. The original usefulness implies that the concepts introduced in the tool should be useful for improving the organization, and relevance concerns the need for the concepts to be key to understanding the functioning of the organization (Rebolloso et al., 2002). The effort criterion refers to the amount of effort required by the evaluation committee to use the tool (Rebolloso et al., 2002).

We consider these three overarching criteria to be a solid foundation, however they are not specific enough to guide our evaluation of a tool. We seek more specific criteria to guide our understanding of what makes a tool facilitate a good self-assessment. In the following chapter we therefore build on our understanding of SATs and self-assessments to further develop the criteria identified by Rebolloso et al. (2002).

3.2.2 Tool as One Component of a Self-Assessment Processes

We have made a distinction between the SAT and the self-assessment, where the role of the tool is to facilitate the latter. It is important to keep in mind that the tool cannot be held solely responsible for ensuring a good self-assessment, as the tool is only one element in the process. To place SATs in a context of use, Rebolloso et al. (2008) describes self-assessments as consisting of two key elements: input and throughput. Input includes the SAT and information provided to the evaluators, for example reports that should be read before or during the assessment. The throughput is the second key element and consists of the process and the committee. Here, the process captures how the self-assessment is performed, including aspects such as the dynamics of assessment meetings and communication between committee members and other stakeholders. The committee is the group of people chosen to perform the self-assessment. Both the information provided, the process and the committee are

31

very important factors in the self-assessment (Rebolloso et al., 2008). When further developing the criteria of Rebolloso et al. (2002), it is important to keep in mind what can be attributed to the tool, and what is partially or wholly dependent on other factors.

3.3 EXPANDING CRITERIA FOR SATS

We now have a synthesized set of criteria for self-assessments. We also have a set of criteria developed specifically for SATs that are not elaborate or detailed enough to serve as a systematic and comprehensive guide for the in-depth evaluation we wish to perform. In order to justify using the criteria for SATs developed by Rebolloso et al. (2002) as a framework, or criteria that can be expanded on, we show that the criteria of comprehensibility, usefulness and effort can be made to include the parts of the self-assessment criteria that can be attributed to the tool. Once this is established, the next chapters will consider in more detail what parts of the self-assessment criteria that can be attributed to the tool and how the tool can satisfy these.

3.3.1 MAPPING CRITERIA FOR SATS TO CRITERIA FOR SELF-ASSESSMENTS

The comprehensibility criterion for the tool is a criterion that is needed first and foremost because we are dealing with a tool, and not because it is a self-assessment. Comprehensibility is however a prerequisite for the tool to have an active role in the self-assessment. The tool must be comprehensible if it is to support the achievement of any of the self-assessment criteria and is therefore associated with all of them.

The effort (Rebolloso et al., 2002) criterion for tools is highly linked to the effort (Stufflebeam, 2011) criterion for self-assessments. As an artifact that focuses and guides the process, the effort associated to the entire self-assessment is highly dependent on the effort associated with using the tool.

The usefulness criterion for the tool has a very wide scope, demanding that the concepts introduced in the tool should be relevant and useful for improving the organization (Rebolloso et al., 2002). We here understand "concept" as meaning the information coded into the tool by the author as well as how this is presented. Apart from the self-assessment criterion of effort (Stufflebeam, 2011), the criteria identified

for self-assessments are responsiveness (Chen, 1990), trustworthiness (Chen, 1990) and the authenticity criteria (Guba & Lincoln, 1989). The main source of influence for the tool to aid in the achievement of these criteria for the self-assessment is the concepts introduced in it and presentation of these. As the usefulness criterion for SATs (Rebolloso et al., 2002) includes the consideration of the usefulness and relevance of the concepts introduced in the tool, this criterion encompasses many of the relevant questions to ask when considering if the tool facilitates a good assessment-process. Except for the effort criterion (Stufflebeam, 2011), we find that for all the criteria that were found relevant for self-assessments, the components of these that can be attributed to the tool, can be considered under the usefulness criterion for SATs.

All the components of the self-assessment criteria that can be attributed to the tool map into the criteria for SATs developed by (Rebolloso et al., 2002), so these criteria are considered to be a suitable framework. In the next chapters, how aspects of the tool can support the achievement of the self-assessment criteria will be investigated. This investigation results in an expansion of the proposed criteria for SATs, and a summary of the expanded criteria is provided at the end of chapter 3.

3.3.2 COMPREHENSIBILITY

Comprehensibility is a prerequisite for the tool to facilitate the assessment. The information coded into the tool by the author is discursively mediated, and so comprehensibility becomes essential for the tool to have any effect on the assessment process. In other words, if the tool is completely incomprehensible to all users, one could just as well not have a tool, and the achievement of the self-assessment criteria would be completely dependent on the other elements of throughput and input.

For the theory coded into the tool to have an influence on the construal process of the assessment, the concepts introduced by the tool must be comprehensible. They must be presented in a way that makes the theory clear for the users so that the presentation of the content does not lead to a large variety in the interpretations of the theory. Further, to guide and focus the process in a way that does not require too much effort, the structure of the tool and how to use it should be intuitive.

Furthermore, if the tool is to serve as a boundary object, it is key that all the users of the tool comprehend its content. Recall from chapter 2.2.3 that boundary objects are common artifact that inhabit different social worlds and enable cooperation without necessitating consensus. To have this function, the tool needs to meet the information requirements of all its users (see chapter 2.2.3). The compensability must therefore be evaluated with respect to the users of the tool. What a manager with business education background finds to be clear and easy to understand will be very different from what a doctor or politician may find comprehensible. If each group of people is to comprehend the tool, at a minimum the language used must be accessible to all.

3.3.3 EFFORT

This criterion considers the effort needed to use the tool (Rebolloso et al., 2002), and supports the achievement of the effort (Stufflebeam, 2011) criterion for self-assessments. Although some aspects of the effort criterion for self-assessments may be attributed to the process or the participants, this is a criterion that can be largely attributed to the tool. To satisfy the effort criterion for SATs, the tool must be designed so that the recourses that goes into using the tool does not outweigh the positive results of using the tool. Furthermore, the quality of the assessment depends on the committee members' active participation and attitudes (Rebolloso et al., 2009). This is a reason to reduce the effort required by committee members when using the tool. The dependence on committee members' willingness to participate constructively is especially relevant for self-assessments, where those performing the assessment are typically employees and other stakeholders with busy schedules and who do not necessarily have positive attitudes to the assessment going into it.

3.3.4 USEFULNESS

Most of the components of the criteria identified for self-assessments that can be attributed to the tool fall under the usefulness criteria for SATs. We here go through the components of criteria for self-assessments that are affected by the tool, and discuss what this means for the usefulness criteria.

The responsiveness (Chen, 1990) criterion for self-assessments, and the closely related usefulness (Stufflebeam, 2011) can be highly affected by the tool. These criteria consider the desirability and force of the assessment's impact. This includes issues such as timeliness and breadth of the assessment's consequences, issues that are not wholly dependent on the tool. We consider timeliness to be independent of the tool, and attributed to the process. Also, the breadth of consequences will also depend on the committee and process. However, breadth of consequence, force and desirability of the assessment's impact can be affected by the tool. We have argued that for the assessment to result in change, underlying structures of the organization need to be identified. Theory coded into the tool can support the process of construing an image of the organization by providing new knowledge and perspectives. How well this image serves as a foundation for change therefore depends on the theory that is communicated through the tool. In other words, to meet the requirements of the usefulness criterion, the theory that is coded into the tool must be helpful for the construal of the image of the organization. It is therefore very important to consider the theory coded into the tool when evaluating the usefulness of an SAT. As the theory is discursively mediated, we advise the use of discourse analysis to evaluate the theory coded into the tool, and the potential effect this may have on the construed image and on the organization.

Responsibility for ensuring the trustworthiness criterion (Chen, 1990), can also only partially be attributed to the tool. The quality of the information provided into the process will affect the evidence the evaluators base their reflections on, impacting the trustworthiness of the evaluation results. As the assessment is based on reflection and not measurement, the knowledge and experience of the participants will also serve as evidence and therefore impact trustworthiness. However, the tool does have the potential to impact the trustworthiness of the assessment through the theory coded into it. We have argued that through the assessment, evaluators try to identify and understand underlying structures of the organization and construe their image based on this (see chapter 2.2.3). The theory coded into the tool can be more or less helpful in the process of uncovering the realities of the organization. If the theory introduced in the tool helps uncover the realities of the organization, this increases the trustworthiness of the self-assessment. Again, this calls for analyzing the underlying

theory with respect to how suited it is to explain the underlying structures of the organization in question.

We have argued that the authenticity criteria (Guba & Lincoln, 1989) should be met by a self-assessment to ensure a good construal process. The construal process takes center stage in these criteria. Parallel to the representations of the organization being construed through reflection and knowledge sharing, these images affect the participants' representations of the organization and therefore their ability to act on the new understanding (Guba & Lincoln, 2001).

Ontological and educative authenticity (Guba & Lincoln, 1989) concerns the change in individuals' representation of the organization that results from the self-assessment. Individuals' construal should become more sophisticated and informed, and they should have a better understanding of the representations others in the organization have. The achievement of these self-assessment criteria will be highly dependent on the process. The assessment leading to participants becoming more sophisticated and informed, learning from each other and understanding others' views, adjusting views and testing assumptions requires the process to be set up as an effective learning arena where sharing views and knowledge is supported. However, these things can to some extent be attributed to the tool as well. In the same way as argued above for the trustworthiness and responsiveness criteria, the theory coded into the tool can contribute with insights and knowledge and therefore support participants in their process of becoming more informed and sophisticated. Furthermore, the presentation of the concepts in the tool may have effect on how helpful the tool is. To support these criteria, the tool should focus and guide the assessment in a way that encourages reflection. This reflection should be directed towards the information coded into the tool so that the participants can learn from it, and towards others' views of the organization so they can learn from each other, and towards information available to them about the organization. The discourse embedded in the tool, how the concepts are presented and how the tool is designed as to focus and guide the process can affect the self-assessment process. We therefore argue for an investigation of how the discourse and presentation of the concepts in the tool affects the process, and if it

supports a reflection process where the participants become more sophisticated, informed and understanding of others' views.

Catalytic and tactical authenticity (Guba & Lincoln, 1989) concerns the consequences of the self-assessment in terms of it resulting in action being taken. This is also mainly the responsibility of the committee members and the process, even though it can be helped or hampered by the tool. The tool can and should impact the result of the self-assessment, and these results should lead to action. What can be attributed to the tool on this front is the impact of the theory coded into the tool, which should be suited to the realities of the organization. Further, we have argued that tools, in our case SATs, can serve as boundary objects. If the SAT does effectively serve as a boundary object, this will enable collective action. Boundary objects facilitate the actors in the process of constructing a common representation of the different actors' knowledge and interests, enabling collective action (Kaplan & Jarzabkowski, 2006, p. 13). SATs can be used as an artifact that people can point to and fill inn with their individual and collective conclusions. This is a process of tacking back-and-forth (Star, 2010) where people go back-and-forth from their local understanding and reasoning, and the common representation that is being built. This can be a process of negotiation, consensus or another. The tool then becomes an artifact used to establish a common representation, which may be more or less based on consensus, which can serve as a common platform for deciding on what actions to take. If the tool serves as an effective boundary object, it will facilitate action, increasing its usefulness. This calls for the investigation of the tool's potential to serve as a boundary object. To serve as a boundary object, it needs to satisfy the information requirements of all those social groups that are included in the process (Bowker & Star, 1999). Evaluating this therefore requires understanding the informational requirements of those who are going to use the tool.

The fairness criterion is achieved by ensuring that all competing views are accessed, exposed and taken into account during the assessment (Guba & Lincoln, 2001, p. 7). In his responsiveness criterion, Chen (1990) also requires assessments to fairly represent the views of all the stakeholders, and address all of the stakeholders' most important concerns. However, equally representing the concerns and views of all

stakeholders is not practically possible, and we therefore limit our evaluation to consider the representation of the most important concerns of the most important stakeholders. This self-assessment criterion cannot be fully attributed to the tool. A large part of ensuring the fairness of the self-assessment lies in ensuring a process that allows for a fair representation of views, a fair representation of stakeholders in the evaluation committee and the information they are provided with. However, through the concepts introduced and the presentation of these, the tool can have significant impact on what views are represented and how. It must be acknowledged that tools can be used as political instruments in favor of some stakeholder groups (Kaplan & Jarzabkowski, 2006), and it is not desirable for the tool to systematically lead to some stakeholder groups being favored.

The discourse embedded in the tool provides a path to considering the extent to which SATs meet or conflict with the interests and concerns of important stakeholders. We have argued that SATs can serve as boundary objects. By conceptualizing boundary objects as texts, Oswick and Robertson (2009) explore how boundary objects can bring a particular discourse into the context of those using it. Through the construction of the tool, a particular discourse becomes embedded within it. Oswick & Robertson (2009) found that even though objects such as tools may be used to facilitate constructive cooperation, they can also be imposed on particular groups to their disfavor. By "ruling in" or "ruling out" certain ways of talking about concepts, conducting oneself in relation to them or constructing knowledge about them, discourse acts as a powerful ordering force (Grant et al., 2009, p. 216). Interests can by extension become embedded in tools. As interests become embedded in the tool, pre-existing power relationships are discursively replicated, sustained or challenged (Oswick & Robertson, 2009, p. 188). This means that discourse embedded in the tool may affect power relationships to the favor or disfavor of important stakeholders. For an evaluation of SATs, we advise consideration of how through discourse, the tool affects power relationships, views of the organization and what is considered important and legitimate.

3.4 CRITERIA FOR SATS

In conclusion, the criteria we will use for evaluating SATs are comprehensibility, usefulness and effort. Tradeoffs will have to be made between these criteria. For example, minimizing effort could be done by not introducing the tool at all, but this would also minimize the usefulness of the tool. Also, increasing comprehensibility by reducing the complexity of the concepts introduced in the tool could prevent complex but important questions from being asked, also compromising usefulness.

The table on the following page summarizes the evaluation criteria that will be used for the evaluation of the SAT in this research project.

Criteria	Description
Comprehensibility	The structure of the tool and how to use it is intuitive. Explanations provided are clear, simple and exhaustive. The language used in the tool matches the user's competence level. The concepts presented in the tool are clearly expressed and not ambiguous.
Effort	Employing the tool in the organization requires less resources than the benefits gained. The level of effort needed by the self-evaluation committee to discuss and respond to the tool is not so high that it negatively affects the participants' active participation and attitudes.
Usefulness	The concepts chosen for assessment are applicable and key to understanding the functioning of the organization. The theory coded into the tool should be suited to explain the underlying structures of the organization.

The presentation of the concepts in the tool should support a reflection process where the participants become more sophisticated by learning from the theory presented in the tool, and informed and understanding of others' views.

The tool should enable participants to take action. Serving as a boundary object, the tool may enable collective action without necessitating consensus. To serve as a boundary object, the concepts in the tool need to be abstract enough to be applicable across sites, but not so flexible that the concept comes to mean completely different things to different users. The content should satisfy the information requirements of all the users.

The self-assessment should satisfy the most important stakeholders' most important concerns. For the tool, this means that the recommendations of the tool and influence of the discourse embedded in the tool should be in line with important stakeholder values and interests. |

TABLE 1: CRITERIA FOR SATS

4 RESEARCH STRATEGY

Through this thesis we have had a goal of delivering useful insights to the hospital and, if reasonable, adapt the tool to their organization. We have also wished to contribute to the academic literature. An appealing research strategy was therefore Action Research (AR), where the researchers participate in creating working solutions for their research subjects as well as writing up their reflections for the academic community. Because of this dual role of the researcher, choosing this research strategy has implications for what precautions need to be taken.

In this chapter we will give an overview of AR and the implications of choosing this approach. We then give a crude overview of our overall research strategy to set the stage for a detailed examination of our methods and key choices in the processes before a critique of research strategy.

4.1 ABOUT ACTION RESEARCH AND ITS IMPLICATIONS

The essence of action research is that researchers are involved in the processes they study and take an active part in finding solutions in cooperation with the research subjects. AR researchers employ similar research methods as conventional researchers. However, while both AR and conventional researchers aim to add scientific information to their fields of research, AR involves an inherent focus on action. AR researchers aim to provide solutions to problems for the research subjects through their work. A central concept in AR is that of workability. Workability is concerned with "whether the actions taken in the AR process actually result in a solution to the problem" (Greenwood & Levin, 2007, p. 68). Especially from the perspective of the local stakeholders, workability is a central aim of any AR project (Greenwood & Levin, 2007, p. 100). Still, it is worth noting that successful workability only shows that something worked, it does not automatically create a credible understanding of why it worked (Greenwood & Levin, 2007, p. 100). As noted by Greenwood and Levin (2007, p. 100), "workability is the key data point, but not the end point of an AR process. It does, however, show that they provided a practical solution to a particular problem." They further note that moving from workability to achieve credible knowledge requires further analysis, searching

literature and considering known cases for other approaches the create similar outcomes (Greenwood & Levin, 2007, p. 100). For the researchers this means that throughout the process action has to be followed by reflection, and these reflections should follow the logic of social science research.

The role of the researcher is a key issue in an AR project. The researcher has two main tasks that can be a considerable challenge to balance. They need to solve pertinent problems with the local problem owners (action) and at the same time rigorously scrutinize the experiences from the field engagement (research) in order to communicate research-based findings (Levin, 2012). The role of an AR researcher in the process of solving problems is what Greenwood and Levin (2007, pp. 124-125) call the friendly outsider. The researcher brings with him, or her, a set of analytical frameworks and the general advantage of being an outsider (Greenwood & Levin, 2007, p. 120). Friendly is here meant in a special sense, involving the ability to reflect back to the local group things about them in a way that is supportive rather then negatively critical or domineering (Greenwood & Levin, 2007, p. 125). It here becomes the job of the researcher to open up lines of discussion, and make the tacit knowledge that guides local conduct evident (Greenwood & Levin, 2007, p. 125). The researcher is also responsible for running the AR process, a major challenge being the design of adequate arenas for communication about the problems to be solved (Greenwood & Levin, 2007). Designing and partaking in these arenas is a challenging task (Levin, 2012).

The second main task of the AR researcher is to rigorously scrutinize the experiences and write up the findings from the research. They have to comply with relevant academic practices, norms and values in order to be able to contribute to the general body of knowledge (Levin, 2012, p. 138). This means that deep empathic and political involvement must be coupled with critical and reflective research, which requires the researcher to treat their own experiences at 'arm's length' (Levin, 2012, p. 134). To achieve a trustworthy contribution to the academic literature, Levin (2012) identifies four essential factors in the practice of action research: standardized methods, research partnering, awareness of own biases and alternative explanations. These are techniques that help the AR researcher create a distance to the actions, emotional

involvement and political strategizing they have been engaged in (Levin, 2012). Standardized methods are those techniques that are often described as social science methods. Just as with other social research, AR researchers must follow accepted social science procedures for data construction and analysis to ensure academic rigor. Research partnering, awareness of own biases, and alternative explanations are further techniques the researcher should employ to strengthen the distance to the research subjects needed for academic integrity of research results.

4.2 OVERALL RESEARCH STRATEGY

The overall research strategy consist of the following steps: develop a set of criteria from existing literature, complete several rounds of interviews and discussions to help us construct an image of the case organization, conduct a pilot where participants from the organization use the tool, and finally evaluate the tool based on the criteria and the findings.

All the techniques we have drawn on are associated with qualitative social research. In qualitative research, the emphasis is usually on words rather than quantifications (numbers) and the focus is on understanding the social world by examining how the participants interpret it (Bryman, 2012). We have argued that the role of SATs for organizational change is to facilitate its users' individual and collective construal process of their organization (see chapter 2). To be able to understand and observe this process, we need data gathering and analysis techniques that help the researchers understand individuals' interpretations, or construals, of the organization. Individuals' construals are complex and not easily accessible, but by observing how people talk about things and how they react, we can get insight into their view of the world around them. Qualitative research techniques such as unstructured and semi-structured interviews, and coding are therefore well suited to our purpose.

Further, we have combined a relatively strong inductive approach with a relatively strong deductive one. A deductive approach to theory is one where theory, or predefined concepts, guide the research, and a inductive approach is one where theory and concepts is the product of research (Bryman, 2012, p. 8). Our purpose on behalf of the hospital has been to evaluate a specific SAT for use in our case organization.

To do this in a systematic way we developed a set of criteria, and these have been based on purely theoretical grounds. These criteria are developed so that their content is set, providing a framework to guide the evaluation of the tool. Concepts that become fixed once they is developed, are called definitive concepts (Bryman, 2012, p. 388). The criteria are such definitive concepts that guide our research as we look for data that can be related to them, making this a deductive approach. At the same time, our emphasis on the key role of the construals of the organization by its members led us to an inductive approach. Emphasizing the need to understand the organization's individuals' in order to understand the organization implies acknowledging that our own understanding of the organization's needs is not enough. This leads to a contradiction: how can we decide on criteria beforehand and at the same time acknowledge that we do not sit with the answers of what the organization needs? Blumer (as cited in Bryman, 2012, p. 388) argues that the employment of definitive concepts entails the application of a straightjacket on the social world. The way we have attempted to handle this issue is to have a mainly inductive approach to the data gathering and analysis. We have attempted to construe our image of the organizations' reactions to the tool without the criteria being the main guiding concepts. Rather, the individuals' main concerns and reactions have guided our further exploration. This is reflected in our data presentation, where the main findings from the field are presented before these are mapped to the criteria. The final evaluation chapter opens for inclusion of those findings that are not easily mapped to the criteria, in an attempt to include all the important findings from the data in the evaluation. Note that we do not claim that these two processes have been kept isolated from each other, or that this was our intention. The theory, both the criteria and other theory used in this thesis, has inspired our data gathering and analysis. However, we have actively and consciously let the individuals' construals and main concerns guide our search for a deeper understanding of the situation where the tool is introduced to the organization.

In line with the principles of AR research, the participant's and our own increased understanding led to a series of actions being taken during the course of our research. By allowing such actions, AR enabled us to learn more within our set time frame than what would have been possible without the opportunity of making such interventions. Imagine simply introducing the tool as it is into the organization without allowing for

action to be taken. Such a strategy would have yielded clear, but very simple results. The evaluation would have been a clear and complete failure of the tool on all fronts, because the participants were not experienced enough in the English language to use it in any purposeful way. In line with AR principles, we took several steps to increase the tool's workability in order to dig deeper into the levels of the issues and potential related to the tool. We call these steps intervention points and they will be described in detail.

In the following chapters we describe the methods used for data construction and analysis, our roles as friendly outsiders and process facilitators, the intervention points and their implications, and in the end a critique of the research strategy.

4.3 DATA COLLECTION (SOCIAL RESEARCH METHODS)

For data collection, interviews were held with employees in the division and a pilot study was conducted where the SAT in question was used by a test-group. Two rounds of interviews were conducted before the pilot project, and one last round of interviews was held right after the pilot. All interviews except the first two and the entire pilot project was audio recorded. These records were then transcribed.

4.3.1 TAKING THE ROLE OF AR RESEARCHERS

As AR researchers, we needed to focus on solving pertinent problems together with the case organization, while also rigorously scrutinizing the experiences for writing up research-based findings. This requires being very alive to the development of our own understanding, our interventions in the process, and how we affected the research subjects. The intervention points, which typically included making a change to the tool (for example translating to Norwegian), came after we encountered what we experienced to be a roadblock of some kind. When we intervened, it was very important for us to gain a solid understanding of what the research subjects and ourselves were thinking before we increased the workability. To handle these issues, our most employed strategy was to thoroughly discuss and write notes on our findings the very same evening after the interviews. In these sessions we read through our notes from the day, worked on developing theories together based on what we had

experienced, tried to come up with conflicting or alternative theories based on the same data, and listened to parts of the day's audio-files.

Another main concern was how to behave during the interviews and pilot. A main goal was to avoid imprinting our view of the world, the tool and the organization on the research subjects, only to hear our own ideas reflected back to us in the interviews and pilot later. This was especially important because of our interest in the construal processes facilitated by the tool. If the research subjects changed their views of the organization because they were convinced by us, or they fed our understanding of the organization back to us to please us, we would only be observing our own influence. A simple solution to these issues would be to have a strict interview scheme and written instructions for the participants in the pilot, and not join in the conversation. However, the world of health care was very new to us, and there were very many things that they talked about as completely natural that were very strange to us. We therefore needed the freedom to ask follow-up questions. Also, in line with an AR strategy, we wanted to allow the research subjects' construal of the organization and what issues they considered pressing to guide our data collection. This demands a flexible data collection strategy, one which we applied all the way down to the interview level. A strategy for balancing our level of engagement was to focus intently on how we phrased our questions. The meant that we were careful not to ask guiding questions or comment more than necessary. For instance, we were careful not to build sentences starting with "so, what you are saying is…" or "so, in your organization, [something] would perhaps be…". Instead, we would say "and when you say [something], what do you mean by that?" or "could you explain what you mean by that?". In addition to being very aware of how we phrased our questions, we were also cautious not to talk too much. In order to learn as much as we could about how the research subjects view their organization, and how the tool facilitated conversation and discussion, we actively employed silence. By allowing for some awkward silences we were able to see how the conversations developed, and whether interesting conversation came about if people felt "forced" to talk.

Being engaged in conversation without being enthusiastic conversation partners who fill in sentences when words are left unsaid, and who refuse to answer questions like

"what do you think? You probably know a lot about this", could quickly have felt anti-social and uncomfortable. To balance the need for freedom of enquiry, social agreeableness and avoiding imprinting our views on our research subjects, we chose to play with open cards. This implied having dialogues such as this when one of us wanted to say something that was rooted in a theory we had at the moment:

> Victoria: So do you think that (interrupted by Kristin)
>
> -
>
> Kristin: I think I know what you want to ask, you sure you want to?
>
> -
>
> Victoria: Haha no, thanks. Another glitch there guys, our goal is to understand these things, as you know. We are not here to hear our own thoughts reflected back to us. We have one view of the situation, and it is really easy to adopt others' image and build on it. It is less natural to dig down and figure out how you really see things. We want give you room to go through that process and then give us your image, instead of us giving you ours. So then we have to avoid saying things like 'is it not the case that...?' or 'don't you think...?'. And we have a long way to go before we learn this stuff so we have to keep reminding each other.
>
> (Facilitating individual pilot round)

Such comments were very well received, and most often the conversation went right back to normal as if nothing had happened. Over the course of the year we became increasingly comfortable with using each other. We asked each other if we should mention or ask about certain things, and stopped each other from saying too much.

Another issue that was tough to balance was how much "help" to give the research subjects when we were discussing the tool or theory related to the tool. When we met roadblocks, we needed to understand how much of a problem they were. The roadblocks were important findings and we had to be careful not to supply explanations for everything the research subjects wanted to know. Rather, we had to explore these issues when they arose. For example, if they said that a word was hard to understand, we would ask, "can you elaborate on that?", "what exactly is it that is

hard about it?" or "what do you mean by 'hard'?". With this lack of a helpful attitude that we needed to take, creating a positive and supportive environment became critical. This was done by consequently and frequently reminding them that their view was the most valuable one. Also, when there was something they did not understand or master, we made sure they understood that the fault lay with the tool and not with their competence.

4.3.2 SAMPLING: CHOICE OF PARTICIPANTS

In social research, the process of selecting participant is referred to as sampling. Here, the universe of all possible units that can be selected is referred to as the population, while the segment of the population that you select for the investigation is referred to as the sample (Bryman, 2012, p. 187). In quantitative research it is normal to use a sampling technique referred to as probability sampling. Here, the sample is selected using random selection, meaning that each unit in the population has the same chance to be selected. However, as we conduct qualitative research we employ a sampling technique referred to as purposive sampling. Purposive sampling is a technique where the aim is not to sample research participants on a random basis (Bryman, 2012, p. 418). Instead, the participants are selected based on their relevance to the research question. This requires a clear idea of what criteria are used to determine the relevance of potential participants (Bryman, 2012, p. 418).

As our research question is concerned with evaluating the SAT for use in the case organization, our criteria for sampling participants are based on who would, or could, participate in the use of the SAT if it were to be introduced in the organization. On this basis we made a set of criteria to determine the relevance of potential participants. Firstly, the group of participants combined had to be familiar with the whole organization and all of its departments. Secondly, the participants had to be available to meet at the same location for at least one day. Thirdly, we wanted both genders to be represented in the group. Lastly, we asked for participants from different backgrounds. Specifically, we requested that at least one would have a background from economy and management and that the group were to include somebody with a medical background and practical experience from the case organization.

In addition to participants with the characteristics outlined above, we also wanted a union representative to be included in the pilot. Union representatives are central to change processes in Norway (chapter 5.5) and including a representative provides insight into this central stakeholder group. Additionally, such a representative will have experience in securing the interests of the employees and participation in organizational change process, which will add to the value of the inclusion of such a participant.

These are the people that we got the opportunity to talk to:

Title	Description
Department Head 1	Psychiatric nurse, leader of a psychiatric hospital department
Department Head 2	Psychiatrist, leader of DPC
Unit Leader 1	Psychiatric nurse, leader of a Unit in a DPC
Staff Employee 1	Educated as social worker, both clinical and management background from the case organization.
Staff Employee 2	Educated as ergonomist, long experience with both clinical and management work in the case organization
Staff Employee 3	Psychiatric nurse, long experience with clinical and management work in the case organization
Staff Employee 4	Long experience with management in the case organization
Union Representative 1	Nurse, representative from the Nurse's Union

TABLE 2: LIST OF PARTICIPANTS

4.3.2.1 CHOICE OF PARTICIPANTS FOR THE DIFFERENT STAGES OF THE PROJECT

The main purpose of the first interview round was to become familiar with the organization. It was therefore important to talk to somebody who knew the organization well. We got the opportunity to talk to Department Head 1, Department Head 2 and Unit Leader 1. These department heads were leaders from departments at different geographical positions, which was helpful to give a more representative image of the organization.

In the second interview round we tried to map out the lifecycle processes of the case organization and identify the most important stakeholders. We wanted participants who could participate in the pilot, and with an overview of the organization. We therefore talked to Staff Employee 2, Staff Employee 3 and Staff Employee 4, who were all willing to participate in the pilot and had experience from the case organization as well as responsibilities that required them to have an overview over the organization.

When choosing the participants for the pilot, we wished for a group between four and seven. This was because we wanted to have a potentially consensus making process and a heterogeneous group, but few enough people to make room for everyone to participate actively and naturally. We were given access to five participants. Not all the participants from the second interview round were available on the day of the group pilot round. Staff Employee 3 was not able to participate, and Staff Employee 1 came in as a replacement. In addition, the Department Head 1 was able to join us for the process. Lastly, as mentioned above, one of the participants was included to stand for the union representatives. Union Representative 1 was therefore one of the participants in the pilot. In this group, one participant was male and the other four were female. To keep the gender of the participants hidden, we will refer to all of them with the gender-neutral term "he".

The purpose of the third interview round was to ask questions about the pilot, and about their experience with the SAT. We therefore have the same participants here as in the pilot.

4.3.3 DATA COLLECTION OVERVIEW

Table 3 provides a chronological overview over the interviews and the different elements of the pilot. One initial round of interviews was conducted in the fall of 2014, and the pilot preparation interviews, the pilot itself and the follow-up interviews were all conducted during the spring of 2015.

ID	Date	Participant(s)	Purpose	Time (h)
1	Nov. 13th 2014	Department Head 2	Interview and guided tour: Get to know organization	4
2	Nov. 13th 2014	Unit Leader 1	Interview: Get to know organization	1
3	Nov. 14th 2014	Department Head 1	Interview and guided tour: Get to know organization	2
4	Feb. 11th 2015	Staff Employee 3	Interview: Prepare for pilot	2
5	Feb. 11th 2015	Staff Employee 4	Interview: Prepare for pilot	1
6	Feb. 11th 2015	Staff Employee 2	Interview: Prepare for pilot	1.5
7	Feb. 12th 2015	Staff Employee 3 Staff Employee 4	Interview: Prepare for pilot	1
8	Mar. 12th 2015	Staff Employee E	Pilot: Facilitated individual use of SAT	3.5
9	Mar. 18th 2015	Staff Employee 2 Union Representative 1	Pilot: Facilitated individual use of SAT	4
10	Mar. 19th 2015	Department Head 1, Staff Employee 1, Staff Employee 2, Staff Employee 4, Union Representative 1	Pilot: Facilitated self-assessment using SAT in a group	7
11	Mar. 23th 2015	Union Representative 1	Interview: Follow-up after pilot	2
12	Mar. 24th 2015	Staff Employee 1	Interview: Follow-up after pilot	2
13	Mar. 24th 2015	Staff Employee 2	Interview: Follow-up after pilot	1
14	Mar. 25th 2015	Department Head 1	Interview: Follow-up after pilot	1
15	Apr. 10th 2015	Staff Employee 4	Interview: Follow-up after pilot	1

TABLE 3: OVERVIEW OF DATA COLLECTION ACTIVITIES

4.3.4 INTERVIEWS

Qualitative researchers emphasize that it is impossible to observe and report on the social context around you without also making interpretations of it. Furthermore, the researcher often "embodies the view of social reality as a constantly shifting emergent property of individual's creation" (Bryman, 2012, p. 36). This has certain implications on how interviews are to be performed. In a qualitative interview setting, going off topic is often encouraged. It is also permitted for the interviewer to depart significantly from the plan of the interview. Both the schedule and the guide can be disregarded. If the interviewer finds the prepared questions to be inadequate as the interview goes on, new ones can be made on the go. These are things that you cannot do in a quantitative interview, because it would compromise the standardization of the interview process, which will directly compromise the reliability and validity of the measurements.

The data collection for this thesis has involved three rounds of interviews, where all these were conducted as qualitative interviews. The first round was conducted in the fall of 2014, as part of a preliminary study. These interviews were done with the intention of gaining a better understanding or case organization. The second round of interviews was done to prepare for the pilot, where we tested the tool for use in the case organization. Lastly, the third round of interviews was conducted as follow-up interviews after the pilot.

In the first interview round (ID 1, 2, 3) we planned semi-structured interviews. We had an extensive range of topics we wanted to ask about, most of them inspired by a literature review, and therefore chose to bring an interview guide. However, we anticipated that we had not yet covered all the most important topics, and wished to have the opportunity to follow-up interesting information that was uncovered during the interviews. This is well in line with semi-structured interviews. In this interview form one uses a list of questions or a set of topics to be covered, but questions can also be asked following information that is given (Bryman, 2012). However, one of the interviews in this round (ID 2) ended up being spontaneous, as we were given the chance to interview one extra person when we arrived at the case organization. This spontaneous interview was a performed as an unstructured interview. This is a form

of interview that is very similar in character to a conversation (Bryman, 2012). The choice to do this interview differently was partly motivated by the fact that this interviewee held a different position in the organization than the two other subjects. This meant that the guide that was developed for the two others could not be used well in this interview. However, we were very interested in his insights, and could think of a lot of questions to ask him. In compliance with the characteristics of an unstructured interview, we asked very open questions, and followed up with questions about several topics that we realized he had insight into during the course of the interview.

All of the interviews in the first round were conducted in an informal setting. The goal of the visit was not primarily rigorous data collection. Rather, we sought to get a feeling of the organization and test out some of our assumptions. We also wanted to build relationships to the subjects in the expectation of working with them again later in our research. To achieve an informal tone, we chose not to record the interviews, and wrote extensive notes instead. The quotes presented from these interviews have been approved by the subjects. They have been sent the quotes, along with the context we intended to use them in. This was mainly an ethical consideration, and at the same time a practical way to check if we understood the subjects right.

The second and third rounds of interviews were both done in connection with the pilot, and were both recorded and transcribed. The second round (ID 4-7) was conducted with the intention of gathering the information that we needed to facilitate a good pilot. To achieve this we held semi-structured interviews, as we did in the first round. We did this to try to create an arena where we could learn to better understand the organization. The interviews in the third round (ID 11-15) were also held as semi-structured interview. In this round we wanted them to relate to us how they had experienced the pilot, and their thoughts about the SAT. We therefore asked open questions where they were prompted to share what they thought about the SAT and the process. By not asking narrow questions and pointing them in certain directions, we intended to open for them to structure their answers after how they thought. Then, after asking several a very open questions like "so, after this process, what do you think?" we asked some specific questions about aspects that were observed to be

problematic in the pilot. In addition, we made some personal questions for the participants who had said something in the pilot rounds that we did not fully understand, or we wanted them to expand on. The group pilot round was on a Thursday, and the follow-up interviews were held on the following Monday to Wednesday. For these interviews we came to the offices of the participants. The break over the weekend after the group pilot round was inserted intentionally, to give them time to think about what they thought, and not feel like it took up too much time all at once.

4.3.5 THE PILOT

Having a pilot allowed us to test the SAT in the case organization. The pilot consisted of two stages. The facilitator's guide for the SAT relates that the self-assessment should be conducted individually before it is done as a group. Therefore, the first stage was an individual assessment. Here, we facilitated as the participants individually went through the tool. This was followed by the second stage: a group pilot round. Here all the participants were gathered in one room to do the assessment together. Between these stages, we made sure that there were several days available that we could use to learn from the prior stage, so that the next stage would provide us with us with as much insight as possible.

As the tool is comprehensive, and we had limited time, we anticipated that we would not have time to go through all the practices. The tool contains different sections (I, II and III) with different subsections (A, B, C, …) that cover very different aspects of the organization. Consequently, it would not be a representative process if we were to start at the top and then work our way down until we ran out of time. Therefore, we started out having them do the first practice of each subsection. This meant that we for section I would do I.A.1, then I.B.1, then I.C.1, etc. Section III was gone through in a similar fashion. However, section II has a different structure and the same principle was therefore applied a bit differently here. Section II has a set of practices that are to be repeated for each of the lifecycle processes identified in the organization. This means that each practice should be repeated for every lifecycle process. So, instead of starting with the first practice of every subsection, we here started with going through all the practices for one lifecycle process. Ideally, they

54

would be able to go through the whole tool in this way, and then go back up and do the second practice of every subsection. However, the practices took more time than expected to cover, and only a few were covered.

4.3.5.1 FIRST STAGE: THE INDIVIDUAL PILOT ROUND

The first stage was the individual assessment where three out of the five participants were able to partake. One participant did the individual assessment alone, with us facilitating over the phone (ID 8). The other two who participated in this stage did the individual assessment in the same room with us present as facilitators (ID 9). The first participant got the tool in its original form. This meant that all the text was in English. This proved to be a significant challenge. He was not able to understand the content, and we were asked to translate the content orally, so that he could properly understand what was written in the SAT. This took up a lot of time, and was a quite frustrating process for the participant. So, for the next round of individual assessments, where the two participants were present, we presented them with a Norwegian version of the SAT. Since they were to make individual assessments, they were told not to discuss what the different concepts in the SAT meant, or could be interpreted as, in the division. We wanted to observe their individual reflections and different interpretations, since this was considered to be the purpose of the individual assessment. However, as the process went on we saw that they were able to talk to each other and better understand the content together. By allowing them to communicate, we got insight into what they were thinking. In addition, they were also encouraging each other to dig further into their views.

4.3.5.2 SECOND STAGE: THE GROUP PILOT ROUND

The second stage was the group pilot round. Here, all five participants were able to take part. The day was started with a half hour introduction to the SAT and the agenda for the day. Then, the participants were asked to familiarize themselves with the words described in the dictionary in the back of the SAT. This dictionary contains words that have a special meaning in the tool, and that need to be understood to be able to understand the content. As we had translated the content of the tool to Norwegian, we had also translated the dictionary. Next, the participants were to go through the practices. Two of the participants had not done the individual assessment, and were therefore not familiar with the practices. We therefore gave the group a few

55

minutes to read the practices so that all of them could contribute to the discussion. Following this, they were asked whether they had any questions about any parts of the practice before discussing the capability levels. We were interested to see if they could come up with any change initiative ideas in this process. However, as this was a pilot, and the participants did not have the mandate to initiate improvement initiatives based on their assessment, coming up with such change initiative ideas were not seen as a natural part of the process. This was seen as a weakness in out research design, and therefore, after deciding on the capability levels for the practice, we asked them to come up with such ideas. During this discussion, we wrote down what improvement plans they had proposed to give this more focus. The process of going through each practice as described was much slower in the group pilot round than in the individual around. Thus, it soon became apparent that the approach with going through the first practice of each subsection for every section would not work here. When one third of the group pilot round had gone by, we had still only gotten through two practices, both in section I. Therefore, to ensure that all sections of the tool were covered, we chose to use about the same amount of time on every section, and did not complete the first section before moving to the second. The practices that were covered were I.A.1, I.B.1, II.A.1, II.B.1, II.C.1, II.D.1, II.E.1, III.A.1 and III.B.1.

4.3.6 INTERVENTION POINTS

During our investigation of the tool with respect to using it in the organization, we encountered several issues that would have led us to simply discard the tool if we had not had the opportunity to intervene further. These issues were related to part of the tool being customized for production companies, the tool being originally in English, and that some of the vocabulary was unknown to the participants even in Norwegian. Concluding with the tool being inappropriate for the organization on the basis of it being in a foreign language would not have been an interesting research result. A perhaps useful analogy for this is picturing the issue of evaluating the tool with respect to the organization as an onion. By recognizing problems (layers of the onion), documenting them and then intervening to improve them (pealing off layers), we were able to gain further insight into the research question. In the AR process, taking actions in to find a solution to the problem is the process of increasing workability (Greenwood & Levin, 2007, p. 68). We did not end up with a solution

with high workability in the end because we were evaluating the tool and therefore needed to carefully documented and scrutinize each layer. In this chapter, the interventions are described in chronological order.

4.3.6.1 INTERVENTION POINT 1: CHANGED "LIFECYCLE PROCESSES" IN THE TOOL

One of the sections in the SAT is dedicated to analysis of the organization's so-called "lifecycle processes". According to the tool's guide, these processes need to be identified for your specific organization before using the tool. However, one set of such "lifecycle processes" was already written down in the tool because this section had been customized to a specific industry where the "lifecycle processes" are mostly the same. Changing these was not controversial use of the tool, but it required our intervention. We brought these "lifecycle processes" into the first three interviews (ID 1, 2, 3) and worked with the participants to find a workable solution. The participants established that the "lifecycle processes" would have to be changed in order for testing the tool to result in anything other than confusion and frustration. During the next four interviews (ID 4, 5, 6, 7) we worked on finding alternative "lifecycle processes" that would fit the organization. Together with the participants we found that using the responsibilities for hospitals as defined in the Specialist Health Care Law as a basis would provide a reasonable and relevant basis for further evaluation of the tool. Further, going through the tool's definitions of "lifecycle processes", the participants identified one more relevant "lifecycle process" (management) that was not covered by the law. We then changed the tool (specifically changed the pdf-file that is the tool) to display these new processes instead of the old ones.

4.3.6.2 INTERVENTION POINT 2: IDENTIFIED "EXTENDED ENTERPRISE" AND "TIME HORIZON"

Another prerequisite for the assessment process according to the tool's guide is to have established the "extended enterprise" of the organization and the "time horizon" of the change initiatives that should result from using the tool. The "extended enterprise" is defined as "all organizations along the multiple value streams that contribute to providing value to the enterprise stakeholders. The extended enterprise may include customers, suppliers, government, and other entities that might have indirect influence over enterprise activities" (Valerdi, Nightingale, and Blackburn, 2008).

57

Understanding, identifying and determining the "extended enterprise" and "time horizon" was done together with participants as parts of the fourth, fifth, sixth and seventh interviews. A "time horizon" of three years was determined. We brought a computer to these interviews and together with the participants drew up a diagram of the "extended enterprise" after the framework presented in the tool's guide. The diagram served as a way to concretize the task and make it easy for the participants to take charge and directly change the solution. First we asked them to come up with the most important stakeholders on their own to experience their initial thoughts. Then they were presented with the diagram based on a literature review of the Norwegian health care sector and earlier interviews, and filled in the ones they mentioned that were not put down already. Then we discussed the diagram and changed it back-and-forth until the participants were satisfied.

4.3.6.3 INTERVENTION POINT 3: TRANSLATING TO NORWEGIAN

During the first facilitated individual pilot round (ID 8) it became clear that there was a need to translate the tool to Norwegian. The participant was highly uncomfortable discussing based on an English text. Although he understood almost all of the individual words, the language barrier was so high that he felt the meaning became lost. He insisted on having the text read back to him in Norwegian before voicing his opinions with regards to the score valuations. He did not trust his own translation from English to Norwegian, and we quickly understood that pressuring him to voice his initial reflections on the English text would make him feel vulnerable and uncomfortable.

Having local participants feel uncomfortable and confused is not reconcilable with good AR research. A central part of AR is the generation of learning arenas. In these arenas, "communication between insiders and outsiders aim to produce learning and open up the process of reflection for the involved parties" (Greenwood & Levin, 2007, p. 95). To enable such a good communicative reflection process requires all participants to actively participate in reflection. It is therefore key that all participants feel confortable and capable to do so. We learned through this first facilitated individual pilot round that leaving the tool in English would make the tool a complete failure in the organization. Further, gathering five important, busy and insightful

people for an entire day only to have them struggle together with a foreign language would be a waste of their valuable time.

We had five days to translate the tool to Norwegian before facilitating the next individual pilot rounds. In an attempt to peel off only one layer of the onion at a time, we translated as literally as possible. Some words did not have one clear translation to Norwegian, and we had to use some common sense and our understanding of the tool's underlying theory to decide on a Norwegian word. However, most of it was directly translatable, and when the English version was for example ambiguous and vague, our Norwegian version was just as ambiguous and vague. When we had to make a choice between a natural Norwegian way of saying something, and leaving the interpretation up to the reader with a somewhat more awkward Norwegian, we would go for the latter. Figure 5 and figure 6 are examples of a translated part of the tool.

Level 2			Level 3			Level 4			Level 5		
Leaders understand and promote the interaction and relationship across boundaries.			Leaders are working across boundaries, and their work is evaluated based on enterprise performance.			Leaders focus on enterprise-level value creation, and demonstrate "enterprise thinking" through their practices and behavior.			Leaders leverage the synergies across the extended enterprise for the benefit of all stakeholders.		
	C	D		C	D		C	D		C	D

FIGURE 6: EXTRACT FROM LESAT

Nivå 2			Nivå 3			Nivå 4			Nivå 5		
Ledere forstår og promoterer interaksjon og forhold på tvers av grenser.			Ledere jobber på tvers av grenser, og arbeidet blir evaluert basert på enterprisens prestasjon.			Ledere forkuserer på verdiskapning på enterprisenivå, og viser "enterprisetenkning" gjennom sin praksis og oppførsel.			Ledere utnytter synergier på tvers av den utvidede enterprisen til fordel for alle interessenter.		
	C	D		C	D		C	D		C	D

FIGURE 6: EXTRACT FROM A TRANSLATED VERSION OF LESAT

To avoid negative reactions to the language because of some unnatural translations, we often gave the participants the English version if they reacted to a statement we felt was extra strange because of our translation.

4.3.6.4 INTERVENTION POINT 4: PROVIDING EXPLANATIONS

The tool's guide recommends facilitators to provide a half-day lecture about the tool's underlying theory, but we chose to only present explanations of some central concepts and to be available for questions throughout the process. This was both for practical and research-based reasons. The guide recommends one day before the collective round for introducing the participants to the theory associated to the tool, and to individually perform the assessments. We could only gather all five participants for one day, and to avoid forcing them through individual theoretical lectures we supplied information as needed during the individual and collective rounds. Also, we did not wish to push the tool's underlying theory into their construals of the organization before they were introduced to the tool. We needed them to be inclined to react if something in the tool went against their interests. We expected them to be less critically reflective on what they observed in the tool if we had held the very positive, convincing presentation of the theory that the tool's guide recommended beforehand. However, to use the tool the participants needed to be familiar with certain concepts. For instance, they needed access to the meaning of concepts such as "enterprise thinking" and "extended enterprise". With our approach the participants were only given a short introduction to the major concepts in the underlying theory of the SAT. To make up for this, we were available so that if things were found to be confusing due to a lack of understanding of the underlying theory, we were there to explain.

We were very careful when providing explanations. In much the same way as when we translated to Norwegian, we here aimed to peel away one more layer of the onion at a time to learn about the potential of the SAT without the identified obstacle. Therefore, when the participants were not able to discuss a practice in a way they found satisfactory and asked us for help, we intervened very gradually. First, if they did not understand some word, we pointed them to the dictionary at the end of the SAT. Here, central words are explained and we had translated them to Norwegian. If they did not understand this explanation, or their questions were not answered by the dictionary, we explained further. We would usually try to reformulate into more natural Norwegian, and if this did not work for the participants we would give a short explanation of the relevant theory. If these answers were not sufficient, we provided an example from the industry, often one used in the literature related to the tool. For

instance, when the participants were confused about the meaning of upstream and downstream stakeholders, we explained what these would be for an airplane manufacturing enterprise. We provided our thoughts on what a concept could mean in their context in less than a handful of situations. This was done when a discussion went quiet or unmotivated for so long that people seemed to have lost their way back into the reflection process. The idea was that if the practice led into a valuable discussion when an explanation was provided, further research could go into how to change the language to fully exploit this potential. An example of this is when the discussion stagnated about the standardization of processes. We here intervened and pointed out that the procedures in the division could be considered as such "standardized processes". Once the practice was interpreted as being about the procedures, the discussion became alive and focused again.

4.4 DATA ANALYSIS

To analyze our data, we use the technique referred to as coding, a standard approach to analysis of qualitative data. The process of coding, according to Corbin and Strauss (2008, p. 160), can be initiated with open coding. Open coding is defined as "breaking data apart and delineating concepts to stand for blocks of raw data. At the same time, one is qualifying those concepts in terms of their properties and dimensions" (Corbin & Strauss, 2008, p. 195). This is a process where the researcher, through a brainstorming approach to analysis, put interpretive conceptual labels on the data (Corbin & Strauss, 2008, p. 160). Corbin and Strauss (2008, p. 160) make a distinction between higher-level and lower-level concepts. Here, higher-level concepts are called categories, or themes, and tell us what a group of lower-level concepts indicate. However, they note that all concepts, regardless of level, arise out of the data. Corbin and Strauss (2008, p. 160) explain how when conceptualizing data, the researcher scrutinizes it in an attempt to understand the essence of what is expressed in the raw data. The researcher can then attempt to describe that understanding with the conceptual name. Furthermore, Corbin and Strauss (2008, p. 195) introduce the technique of comparative analysis. They explain comparative analysis as the process of comparing incidents against each other to consider their similarities and differences. If the incident is found to be conceptually similar to previously coded incidents, they can be given the same conceptual label and therefore

be put under the same code. As more incidents are added to the same code, more can be learned about this code (Corbin & Strauss, 2008, p. 195). Theory can be developed by showing how the themes and concepts systematically interrelate (Corbin & Strauss, 2008, p. 55).

As advised by Corbin and Strauss (2008), we appointed the different blocks of the text with lower-level concepts and attributed them to categories. This was done by initially writing in the margins of the transcribed interviews, and then categorizing the concepts in an Excel document, where each category was given a sheet. Comparative analysis was used as we added incidents to existing codes. Furthermore, the concepts were analyzed with an inductive approach, in an attempt to build theory. This process was supplemented with a more deductive approach to analysis, as our inductive presentation of findings is followed by an evaluation according to the criteria found in Chapter 3.

4.5 CRITIQUE OF RESEARCH STRATEGY AND PRECAUTIONS

Traditional research criteria to consider in critique of research are generalizability, validity and reliability (Tjora, 2012). Generalizability is the question of whether the research results can be generalized beyond the context where the research was performed (Bryman, 2012, p. 47). In action research, instead of referring to the generalizability of findings, Greenwood and Levin (2007, p. 66) write about the transferability of knowledge into new contexts. As mentioned in chapter 3.1.1, trustworthiness (Guba & Lincoln, 1989) parallels the traditional research criteria of validity and reliability. Levin (2012, p.146) argues that the trustworthiness criterion covers the grounds of action research and is better suited than the traditional validity and reliability criteria to judge results from such a research strategy.

4.5.1 TRANSFERABILITY

The transferability of knowledge to a new context relies on understanding contextual factors of both the place of inquiry and the new context where the knowledge is to be used (Greenwood & Levin, 2007, p. 66). A critical assessment of whether the two contexts have sufficient process structures in common to make it worthwhile to link

them should be performed. To assess the transferability of our findings, the context where the knowledge will be applied must therefore be considered. It follows that we cannot say anything certain about the applicability of our findings without knowing the context it is to be transferred to. However, we expect there to be several contexts with significantly similar process structures where parts of our results may be useful. Because of our access to participants from different areas of the organization, as well as being representative of a group that could be charged with the task of performing a self-assessment, we expect the pilot results to be very useful for the entire Psychiatric Health Division at IHT. We further assume that several aspects of our findings can be useful in other Norwegian health organizations where they are considering including a new SAT. However, there are contextual factors that have affected our findings, and these must be critically compared to a potential new context.

When judging the transferability of our findings, the pilot process should also be considered. Due to resource restrictions we were not able to test the tool in the exact way that its authors advice for its use. In the tool's guide, a three-day seminar is outlined and proposed. For our pilot, we were not able to gather all the participants for three days. We were however able to hold a one-day seminar where they were all present, and facilitate individual rounds for three of the participants before the seminar. To account for this reduction in time and information (some of which was intentional for research reasons, see chapter 4.3.6.4) we were available for questions throughout the process. Additionally, we set aside time after each discussion of a practice where we urged the participants to come up with change initiatives reflecting their score evaluations. Although we altered the process to try to minimize effects of this change in the process, the process in the pilot may not be fully representative of the process that would take place if the organization introduced the tool. However, we note that in such a busy organization our time-reduced process could perhaps be representative of how the tool would be used in the organization. The main message stands clear, before using the research results from this thesis into a new context, the context of this thesis as well as the new one should be critically compared.

4.5.2 TRUSTWORTHINESS

To achieve trustworthy research results as AR researchers, we need to comply with relevant academic practices, norms and values in order to be able to contribute to the general body of knowledge (Levin, 2012, p. 138). Further, deep involvement with the research subjects and process must be coupled with critical and reflective research. To achieve a trustworthy contribution to the academic literature, Levin (2012) identifies four essential factors in the practice of action research: standardized methods, research partnering, awareness of own biases and alternative explanations. Explanations of these techniques and our employment of them are described below.

4.5.2.1 STANDARDIZED METHODS

To achieve academic integrity Levin (2012, p. 145) argues for the use of standardized methods. The standardized methods we have employed are described earlier in this chapter and can all be categorized under the umbrella that is qualitative social research methods. In our research we have employed semi-structured interviews and conducted a pilot study where we facilitated and mainly observed. All interviews and pilot rounds, except the very first three interviews, were audio recorded and transcribed. These transcriptions were then coded following the recommendations of Corbin and Strauss (2008).

4.5.2.2 RESEARCH PARTNERING

One approach to increase the academic integrity of action research is the use of research partnering (Levin, 2012, p. 144). Levin (2012, p. 144) explains how the use of research partnering allows for one researcher to take and active role, while the other can contribute less and simply observe. In addition, he notes that two people sense more than one, and having two individual perceptions and interpretations is useful in the process of trying to make sense of experiences. Furthermore, in action research the researcher is required to both contribute to practical problem solving and contribute to research. In this endeavor, Levin (2012, p. 144) argues that being accompanied by a colleague can be very useful, as it gives you the opportunity to interpret and discuss solutions in the field, before making decisions. In our research, we used such research partnering. We conducted every single interview together, even though it was tempting at times to separate the work-load. It was quite challenging to facilitate the process, and at the same time not share our views and biases and we

would on occasion stop to deliberate on what actions should be done. More than once we made use of the opportunity to discuss our next move before conducting it. For example, one could ask the other "should I say that this could be interpreted as…" In addition, we reaped the benefits of being two researchers in the field when it came to analyzing our finding, as we often sat with different thoughts about the experiences of events after they had occurred. This led us into many interesting discussions that enabled us to build on each other's observations and interpretations to gain new insights.

4.5.2.3 ALTERNATIVE EXPLANATIONS

In addition, Levin (2012, p. 145) argues for the use of alternative explanations. The process of finding alternative explanations is described as an approach where the researchers try to come up with more than one model for explanation, by forcing themselves to think alternatively. This was a challenging activity and one which took several rounds of interviews to feel comfortable with. After a day of conducting interviews or facilitating a pilot round, we set aside an hour for generating alternative explanations to what had happened and why, that day. We went through our notes from the day and tried to build theories and then find others that could match the data but at the same time be contradictory to the previous theory. This was a challenging activity, and at times we did not have the necessary creative capacity to make meaningful alternative explanations. However, over the course of the research we found the process to result in several important discussions. For example, we assumed that the most likely reason for one of the participants to be more engaged in the individual pilot round (together with another participant) than during the group pilot round was that the group pilot round had more people involved in the discussion. However, after realizing that this may be a false theory after finding several alternative explanations, we asked the participant in the next interview if the participant experienced a difference between the two situations. It turned out that the participant was very alive to his reduced involvement, and it was because the first situation had introduced the tool and everything was new. In the second, the tool was old news and it was frustrating and slightly boring to have to repeat conceptual discussions about how to interpret it. With practice we became aware of the difference between being constructive and creating explanations together, and then going into another mental mode of critiquing that explanation. We found it very

65

useful to help build each other's explanations and not be too critical until the explanation was almost completely constructed. This took time, and we only felt competent in this process towards the pilot and the last rounds of interviews.

4.5.2.4 UNDERSTANDING BIASES

Further, to counter the 'distortions' that come from individual's biases, and increase the academic integrity of their research, researchers must have a systematic approach to deal with this (Levin, 2012, p. 144). Levin (2012, p. 144) recommends a strategy to achieve such a systematic approach where the researchers write down what they consider as the important issues and potential findings, prior to embarking on the research. He argues that such a list of biases will enable the researchers to be aware of them when actions are taken or issues are analyzed. This was an important procedure for us. We did not start using this technique before after the third interview, but wrote bias documents before the fourth interview and onwards. Before the pilot we wrote down what we predicted would happen for the different practices before we facilitated any of the individual or group assessments. The alternative explanations approach was also useful for uncovering biases, as the example above illustrates. The bias documents included notes on how we thought the participants would react, and what they would answer. By writing such a document, we had a way of controlling our bias. When what we expected happened, we considered more thoroughly during the data analysis if we were the ones who had influenced them or if it came to be on their initiative. The bias documents also became valuable additions to documentation of what we believed between the different intervention points. To use our metaphor again - once we pealed away a part of the onion, it was hard to go back and remember what construal of the organization we used to have. Our experience is that these bias documents along with the generation of explanations and alternative explanations helped to facilitate our reflection process in a systematic and critical manner, and have been very useful.

5 CHARACTERISTICS OF THE PSYCHIATRIC HEALTH SERVICES DIVISION AT IHT

The goal of this thesis is to evaluate the LESAT, an SAT, for use in the Psychiatric Health Services Division at IHT (referred to as the division). We have found that in order for an SAT to be well suited for use in a specific organization, it needs to fulfill the set of criteria outlined in chapter 3. These criteria are "comprehensibility", "usefulness" and "effort". This means that the questions asked and recommendations given should be comprehensible and useful, and that using the tool should not require too much effort. An evaluation based on these criteria must be done for the specific organization, because what is found to be useful in one organization for example, will not necessarily be the same as in another. To be able to consider whether the tool fulfills the criteria in the division, the characteristics of the division need to be understood. This chapter therefore presents characteristics of the division.

The description of the characteristics of the division will cover seven main topics. The chapter begins with an overview of the formal structure. The second section of this chapter describes the characteristics attributed to the organization being part of the Norwegian public sector and regulated by law, and some consequences of this. The next two sections consider structural and cultural aspects of the organization. These sections are followed by the fifth section, which is concerned with the political economy that the division exists within. To fulfill the criteria of usefulness the most important concerns of the most important stakeholders should be addressed. In this thesis we define stakeholders with the help of Freeman and Gilbert as cited by Mitchell, Agle, and Wood (1997) as "any group or individual who can affect or is affected by a business". The stakeholders in question are the stakeholders of the evaluation. However, the SAT considers possible changes to the entire organization, meaning that the consequences of using the tool may affect all of the organization's stakeholders. The organization's stakeholders are therefore also the stakeholders to the self-assessment. Therefore, to be able to evaluate the SAT with respect to the usefulness criteria, the stakeholders of the division will be identified in the chapter 5.6. Lastly, chapter 5.7 presents change processes and a discussion of change management in the division before the summary of the chapter in chapter 5.8.

The division is an integrated part of its mother organization IHT. There is not much theory that is made for our organization specifically. For each of the sections in this chapter, we will therefore provide general descriptions of Norwegian hospitals, the Norwegian public sector and specialist health care services, while descriptions of psychiatric health care and our specific organization will be provided where this has been found theoretically or empirically. In the cases where it is Norwegian hospitals or hospitals in general we have information about, we will refer to these. When we have specific information about the division, we will refer to the division.

5.1 FORMAL STRUCTURE

5.1.1 FORMAL STRUCTURE OF THE NORWEGIAN HEALTH CARE SERVICES

The Norwegian public sector has three political levels: the state, the counties and the municipalities. The administration of the public health care services is clearly divided into specialist health care services and primary care services. Here, the specialist health care services include care provided at hospitals, outpatient clinics and hospital pharmacies, while primary health services are services such as general practitioners, nursing care and community based services. The specialist health care services is governed and regulated by the state, while the primary care is governed by the municipalities and regulated by the state. Even though our focus will be on the specialist health care services, it is important to have a general understanding of the entire system. Below is an overview of the health care services.
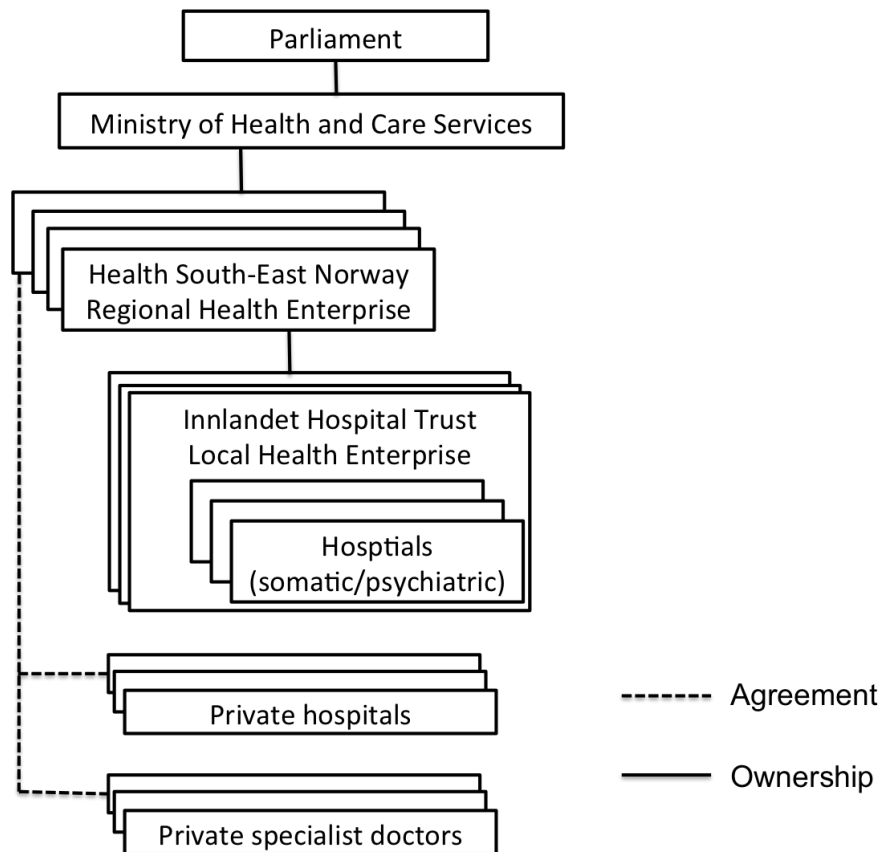
FIGURE 7 STRUCTURE OF THE SPECIALIST HEALTH CARE SERVICES. UPDATED FROM JOHANSEN (2006)

The specialist health care services is also structured in three main levels. The Ministry of Health and Care Services is at the state level. The next level is the regional level with four health regions. Each region has a Regional Health Enterprise (RHE) (Norwegian: regionalt helseforetak). The RHEs are: Health Northern Norway, Health Central Norway, Health Western Norway and Health South-East Norway. The RHEs are owned by the Ministry of Health and Care Services and are governed and regulated by the state. However, they have independent boards and are separate legal subjects. The RHEs are responsible for specialist health care services in their region and are owners of the next administrative level below, the Local Health Enterprises (LHE) (Norwegian: lokalt helseforetak). The local LHEs run hospitals, outpatient clinics and hospital pharmacies. The LHEs are also separate legal subjects. As we can see from figure, IHT is a LHE. Our case organization is a division of IHT, responsible for providing specialist psychiatric health care services.

The Health Care Services Minister appoints the boards of the RHEs, and the directors of the RHEs and boards of the LHEs are appointed by the regional boards (Byrkjeflot,

2004). The central government manages the RHEs through appointing members for the RHE boards, through laws and regulations, policy objectives, a formal performance management system, budgets, and steering documents (Lægreid, Opedal, & Stigen, 2005). The RHEs are given mission statements (Norwegian: oppdragsdokumenter) from the Health Care Services Minister, with guidelines for what should be prioritized. This mission statement is then passed down to every LHE owned by the RHE. Also, the RHEs have to work out a report to the Department annually, containing information about the year that has passed as well as a plan for the upcoming years. All the LHEs that are owned by the RHE should be included in this report (Departementenes sikkerhets- og serviceorganisasjon, 2014).

5.1.2 FORMAL SYSTEM OF NORWEGIAN PUBLIC PSYCHIATRIC HEALTH CARE SERVICES

Psychiatric health care services in Norway are provided by the municipalities, private organizations (non-profit), private specialists and the state. The municipalities are responsible for offering services for people with psychiatric health problems that are mild and expected to last for a short period of time. This includes providing services such as prevention, treatment, care and rehabilitation to these patients. Medical conditions that are treated at this level can be mild depressions, mild anxiety and day-to-day follow-up of chronic conditions such as schizophrenia. If the patient experiences a need that goes beyond the offering of the municipalities, the patient should be referred to the specialist health care services. The specialist health care services only provide a temporary service, and their view on their role in patient treatment is well illustrated by comment made by Staff Employee 4, where he explains how the municipalities have the responsibility for the patients.

> Staff Employee 4: It is they [the municipalities] who own the patients. The municipalities are the primary service (Norwegian: "førstelinjetjeneste"). When somebody gets sick it is the municipalities that have the responsibility. The municipality that the person lives in has the responsibility for the patient, and when the municipality cannot solve the task for the patient, then it is the specialist health care services, the secondary service (Norwegian: "andrelinjetjeneste"), which is us. […] The primary service is the

municipality, who really owns the patient, as you live there, and then the
secondary service is the specialist health care services.

When a general practitioner refers a patient to the specialist psychiatric health care
services, the patient is referred to a district psychiatric center (DPC). The DPC is
therefore referred to as the way in and out of the specialist psychiatric health care
services. Each DPC has a responsibility for a certain area of the country. This means
that there will be a certain number of people that the DPC needs to have sufficient
services for. Under each DPC there is a number of units providing different services.
For instance, there can be units for acute or crisis services, out-patient clinics, and
units for ambulant services (SNL AS, n.d). Ambulant services means that personnel
will go out to the patient and provide treatment outside the health care property. There
will also be a separate unit providing psychiatric health care services for children and
young adults. Treatment of these patients are separated from the treatment of adult
patients because special laws apply to how health personnel can treat them. For
instance, there are different laws about how you can restrain an adult versus how you
can restrain a child (Helsepersonelloven, 1999; Psykisk helsevernloven, 1999).

However, the patient does not always have to go through the general practitioner and
a DPC to get treatment from the specialist health care services. If the patient has an
acute need for psychiatric care, they can be brought directly to a hospital with an
"immediate help" (Norwegian: øyeblikkelig hjelp/ ø-hjelp) department. "Immediate
help" is mentioned specifically in the law, which states that patients in need of
"immediate help" shall receive it without having to wait (Spesialisthelsetjenesteloven,
1999). This will for example be the case when a patient has tried to commit suicide or
is considered suicidal, is psychotic or has suffered an acute change in behavior. It is
regulated by law that RHEs should appoint a number of health institutions or
departments in the health region that are to be responsible for patients in need of
immediate psychiatric help and to provide them with the necessary treatment.
(Spesialisthelsetjenesteloven, 1999) This "immediate help" is separated from the term
"acute help", which is not defined in the law, but means that the patient is in need of
help shortly. The patient will then be prioritized highly but the unit is not required by

law to see the patient immediately. In practice this means that the patient could be put on a waiting list.

5.1.3 FORMAL STRUCTURE OF INNLANDET HOSPITAL TRUST

Our case organization is, as previously mentioned, the Psychiatric Health Services Division of the LHE Innlandet Hospital Trust (IHT). IHT is a LHE owned by Health South-East Norway RHE. It is responsible for providing specialist health care in 49 municipalities. IHT operates six somatic hospitals, two psychiatric hospitals, DPCs, mental health clinics for children and adolescents (Norwegian: barn-og ungdomspsykiatri, or BUP), departments for habilitation, rehabilitation centers, departments for addictive behavior (alcohol, drugs and gambling) and 29 ambulance stations. In total it is an organization with about 7900 employees. (Sykehuset Innlandet, 2014)

IHT is structured into a set of divisions. Some of these divisions are hospitals, and some divisions are grouped by services such as the Psychiatric Health Services Division. Figure 2 is an illustration of the organizational structure.
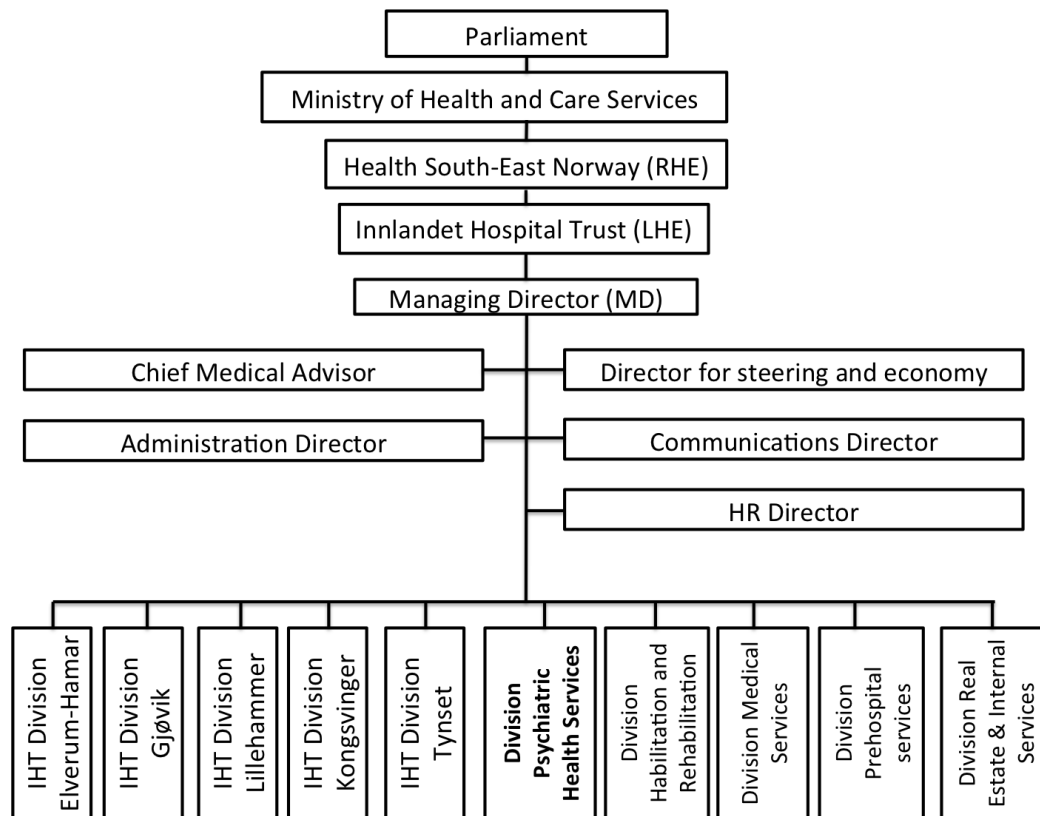
FIGURE 8 STRUCTURE OF INNLANDET HOSPITAL TRUST (IHT)

5.1.4 FORMAL STRUCTURE OF THE PSYCHIATRIC HEALTH SERVICES DIVISION

The Psychiatric Health Services Division is our case organization. This is the largest division under IHT and spans across several institutions. The division includes Reinsvoll Hospital and Sanderud Hospital, two purely psychiatric hospitals. It also includes five DPCs, two mental health out-patient clinics for children and adolescents (out-patient BUP) and one inpatient clinic for children and adolescents (inpatient BUP) (Sykehuset Innlandet, 2014). Unlike the two purely psychiatric hospitals, the other entities are geographically located in or together with other divisions of IHT. Gjøvik DPC is for example located together with another division, the IHT Division Gjøvik, a hospital in the city of Gjøvik. The organization's structure is illustrated below. The bottom row of entities is the departments, and each department has units. All units, departments and divisions in IHT have a leader.
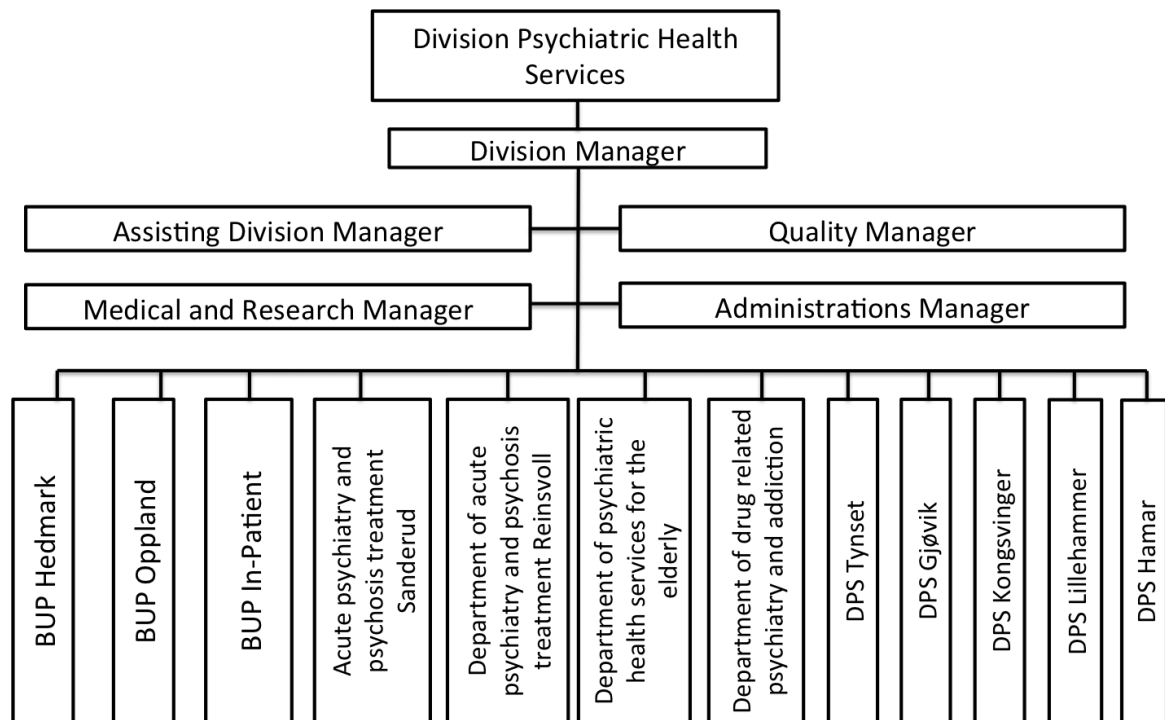
FIGURE 9 STRUCTURE OF THE PSYCHIATRIC HEALTH SERVICES DIVISION AT IHT

5.2 HOSPITALS ARE PUBLIC ORGANIZATIONS

With few exceptions, hospitals in Norway are owned by the state and therefore a part of the public sector. IHT is a public organization in the Norwegian welfare state, and is governed as part of the public bureaucracy. This has strong implications for the way the organization operates. The Norwegian health care services is based on the principle of equal access to services. This principle implies that all inhabitants should have the same opportunities to access health services regardless of social or economic status and geographic location (Johansen, 2006). IHT has made this principle their official vision, illustrating the importance of the principle, "Innlandet Hospital Trust will provide good and equal health services to all who need it, when they need it, independent of age, residence area, ethnic affiliation, gender and economic situation" (Sykehuset Innlandet, 2014).

5.2.1 FINANCING IN THE NORWEGIAN PUBLIC HEALTH CARE SECTOR

The way public organizations are financed separate public organizations from private organizations. The organization is not profit driven, but dependent on public financing (Edgren & Roald, 2001). Unlike private for-profit enterprises, IHT's primary goal is

74

not to earn money but to ensure the population's access to health care services. Also, in contrast to private production companies for example, hospitals are not free to choose how much to "produce" and are heavily restricted in their choices of what to offer. This is very relevant for us because in much of the literature on change management and organizational theory, such as Kotter's (1996) book on organizational change, concepts such as "market" and "competition" are central. With the financing system found in the public sector, many such terms become somewhat unclear or irrelevant, and may even take on different meanings.

Several factors separate the division from businesses in the private market. The first is the absence of market feedback through sales. In the private sector, businesses receive market feedback through sales. When a business offers a product or service that is profitable, meaning that they cost less to offer than what customers are willing to pay, the business will earn more money. A public hospital on the other hand is given a certain budget, which it can spend on providing services. So, while the decisions of for-profit enterprises for example are based are on market profitability, the guiding force of public health care organizations is principles and rules. Concepts such as "competition" take on a different meaning when organizations do not compete with each other in the "market". So, in public organizations, the feedback from the market becomes absent or unclear (Jacobsen, 2012). Another factor complicating the financing in the Norwegian public health care services is that the entities within the specialist health care services cannot go into voluntary liquidation (Lægreid et al., 2005). However, while organizations cannot go bankrupt, they face a limited budget. With the budget as a constraint, the hospital aims to meet the needs of the population to an as large extent as possible. There will always be a demand for services that can improve the health of the inhabitants, especially since they do not have to pay for it. During the interview, Department Head 1 noted the following.

> Department Head 2: If we increase the offering by providing more services, the demand goes up.

Market feedback can be an effective mechanism to promote prioritization of profitable offerings, and over time several quasi-market mechanisms have been

introduced to the Norwegian health care services. This has been done to mimic a form of market feedback as a steering device. For instance, the somatic health care gets funding partially based on "DRG-coding". In this system, internal prices are set for the services an organization can provide. The organizations document what services they have provided, and are later compensated based on the internal prices. In psychiatric health care, the DRG-coding system has not been introduced. However, a quasi-market mechanism has been introduced in the area of outpatient treatments. Outpatient treatments are associated to earnings for the division, and these earnings make up a significant part of the division's balance sheet. Even with some quasi-market mechanisms however, concepts such as "competition" in a "market" with customers that pay for products and services are not directly transferrable to the public sector. Funding still comes from the state and the end users have little concern regarding the cost of the services and products offered. Union Representative 1 illustrates this point:

> Union Representative 1: There is activity based financing. It is the state that gives, and then one decides how much things cost, and then you go and send bills back-and-forth. But the owner is the state. [...] There has not been focus on earning money for us before now. Now we also have to see how we can earn more money. It may not be wrong but. It is kind of strange, given that it is the state that gives and the state that. It is really just swopping money. It is! And the enterprise can never go bankrupt. It can be placed under administration, and the board replaced and all this, but.

5.2.2 POLITICAL LEADERSHIP

The public health care service has a political leadership and a company-like structure. Through the Ministry of Health and Care Services, the state and therefore the elected parliament is the owner of the RHEs, who in turn own the local LHEs. The company-like structure of the specialist health care service was set up following the the 2002 hospital reform in order to improve efficiency and flexibility (Byrkjeflot, 2004). The principle idea is that by organizing the health care sector in enterprises and by using management principles, the political leadership's day-to-day management can be reduced. This is then meant to increase the focus of the political leadership on

76

principal issues relating to priorities and hospital structure. Additionally, managers of the enterprises are given more responsibility and freedom (Lægreid et al., 2005). The idea is that the Ministry should plan and draw up the framework that the enterprise leadership is to practice within.

In practice however, this system has been found to lack some predictability and allow a higher level of detail in the political steering than intended. (Lægreid et al., 2005) found that the parliament on several occasions engaged in issues that were formally supposed to be determined by the executive boards of the RHEs. This is relevant from a change management perspective because enterprise executives' room for maneuverability is not stable. Change initiatives can be set in motion by executives, but later become a target of political intervention. Media pressure and lobbying can influence the ministry to intervene in situations that should have been discussed or clarified at target-setting time (Lægreid et al., 2005). This issue is illustrated by the resignation of the RHE Health Central Norway's chairman, Marthe Styve Holte in October 2014. According to Health Central Norway, he resigned because of disagreement with the owner about important steering principles (Haukås, 2014). He said the ministry micromanaged the board and did not grant enough autonomy to let the board do their job right (NTB, 2014). Frustration with this lack of predictability in the division is reflected in a statement from Department Head 1 about what is needed in order for Innlandet Hospital to handle resistance associated with big decisions.

> Department Head 1: And the central politicians need to keep their hands off (Norwegian: fingrene av fatet)! Not travel from shore to shore promising things that are completely against the processes that IHT are responsible for, it is completely insane.

5.2.3 HEALTH ENTERPRISES ARE HIGHLY REGULATED BY LAWS

Health enterprises are bureaucratic organizations that are governed by a range of laws and regulations. Laws and regulations limit what an organization will be allowed or able to change. For example, laws and regulations limit how an organization can be structured, meaning that laws and regulations may have to be altered in order for the organization to be able to change (Jacobsen, 2012, p. 111). Even though laws can be

changed over time, the changes proposed by the SAT should work within the existing laws and regulations.

There is a number of laws governing Norwegian hospitals. The Health Enterprise Law (Norwegian: "helseforetaksloven") clarifies that the purpose of the health enterprises is to provide good and equal specialist health care services to all that are in need, when they need it (Helseforetaksloven, 2001). The Specialist Health Services Law (Norwegian: "spesialisthelsetjenesteloven") outlines what should be done to promote health in the population and contribute to providing well-suited care of high quality and high availability (Spesialisthelsetjenesteloven, 1999). The Patient and User Rights Law (Norwegian: "pasient- of brukerettighetsloven") clarifies that the purpose of the health enterprises is to provide good and equal specialist health care services to all that are in need, when they need it (Pasient- og brukerrettighetsloven, 1999). The Health Personnel Law (Norwegian: "helsepersonelloven") exists to contribute to safety for patients and quality in the health and care services, in addition to ensuring trust in health personnel (Helsepersonelloven, 1999).

These laws have different purposes and in conjunction, they make a comprehensive set of laws that the Health Enterprises must abide by. They regulate many aspects of the work done in hospitals and other health care facilities. For example, The Specialist Health Services Law controls what services the facilities should provide and sometimes how they should provide them. As mentioned in section 5.1.2, the law requires them do provide what is called "immediate help". Each RHEs are to point out the necessary amount of health institutions or departments in these institutions with the duty of delivering such "immediate help" to patients in need of psychiatric health care. The law also states that there has to be one responsible leader at each level. This is what is referred to as unitary management. Also, The Health Personnel Law states that an organization that provides health and care services must be organized in such a way that the health care personnel is able to perform their statutory duties. The law goes on to outline several duties that health care personnel must abide by. In addition to the laws mentioned above, there is also a special law for psychiatric health services, The Psychiatric Health Care Law (Norwegian: "psykisk helsevernloven") is

specifically concerned with admitting and treating patients against their will, termed "force" (Psykisk helsevernloven, 1999).

5.3 STRUCTURAL CHARACTERISTICS

Structure affects an organization by creating focus, coordination and stability (Jacobsen & Thorsvik, 2007). The way an organization is structured can be the target of a change effort. It may also affects the way an organization responds to change efforts and determine what changes are possible. Therefore, understanding the structure of an organization becomes important when considering the opportunities for, and the implications of, change efforts. To describe the structure of specialist health services organizations, we present Henry Mintzberg's (1983) idealized professional bureaucracy model and consider the possibilities and limitations implied by seeing the organization as a professional bureaucracy. Additionally, to gain further insight into the workings of hospitals, we consider the infrastructure of medical work. Here the structure of the organization is modeled in terms of the standards and systems that produce order and regularity in the organization of the work.

5.3.1 HOSPITALS AS PROFESSIONAL BUREAUCRACIES

Hospitals are often described in terms of Mintzberg's (1983) idealized professional bureaucracy model. Mintzberg's (1983) professional bureaucracy is characterized by a highly trained workforce where employees in the operative core are given a high level of autonomy in their daily work. The high level of autonomy is necessary because none other than the professionals have the knowledge and skills to decide how their work should be done (Jacobsen & Thorsvik, 2007). Delegation of decision power to the professionals secures high professional quality of the work performed, and increases the speed of problem solving and task performance. The Norwegian public specialist health care services system shares many of these characteristics. Organizations in this sector have a highly professionalized operative core consisting of several professional groups. The professions are given a high level of autonomy, determined by their authorization. Of all the professions found in a hospital, doctors have the largest set of decisions they are authorized to make. They have six years of university education and five more years of on-the-job training to become a specialist. In psychiatric health care, the main group of specialist doctors are psychiatrists.

Doctors are the ones who mainly diagnose patients and connect diagnosis to a treatment. This gives them a strong influence on what the organization provides (Sveri, 2004). Nurses are also a large professional group in this system with three years of higher education, and for specialized nurses two more years of on-the-job specialist training. In psychiatric health care, psychologists have also become an important authorized professional group, also with six years of education and optional further specialization. These professional groups, along with several other authorized personnel, make up the main workforce in the division.

The need for professionals to have high levels of autonomy leaves little room for an organization to impose coordination and standards on the professionals. "Complex work processes cannot be formalized by rules and regulations, and vague outputs cannot be standardized by planning and control systems." Mintzberg's (1983, p. 211). This is important to note when considering change efforts aimed at changing the way the work is done. Attempts at such control mechanisms could very easily become misguided, and end up programming the wrong behaviors and measuring the wrong outputs. Instead, much of the standardization and coordination is achieved in professional bureaucracies by hiring people who are educated in the professions they need (Mintzberg, 1983). Through the education and training, professionals acquire standardized skills and knowledge. For instance, physicians learn to diagnose and order treatment for patients through what Mintzberg (1983) refers to as a pigeonholing-process. This means that instead of making unique decisions for every patient, the physician places the patient in one of a set of categories, with associated treatment plans that have been learned through their education and training. The professionals also learn what to expect from their colleagues. The training institutions and other organizations where the professionals interact outside of the organization, such as educational institutions and union activity, therefore play a major role in determining the organization's professional standards and ways of coordinating. Because of this, organizations that employ a given profession exhibit similar ways of doing things, ways imposed on them from the outside of the organization (Mintzberg, 1983). This effect is clear in Norwegian hospitals. The work is performed in strikingly similar ways across hospitals, even across country borders. A Danish doctor (with Norwegian authorization, but this is only a formality) can for example accept a

weekend shift in a hospital in the North of Norway and will recognize the way the work is performed and organized (Sveri, 2004).

However, even though the content in the education and training of the professionals is mostly influenced by external actors, the individual organizations will have influence over how the professionals' work is performed. During our visit to the division we found an interesting example of how education and training can be used to change the way the work is done within an organization. At Gjøvik DPC, the Unit for General Psychiatric Health Services (Norwegian: Enhet for Allmennpsykiatri) is introducing a compulsory education in cognitive environment therapy (Norwegian: Kognitiv Miljøterapi), an alternative form of treatment that does not involve medication. The whole department will be educated through a set of courses and homework, ending in an exam. In the words of Unit Leader 1:

> Unit Leader 1: Educating the entire staff is what creates real change. You have to include everyone, the treatment group as well, so that you don't have a few people who only consider medical treatment alternatives.

Along with explaining how the work is standardized and coordinated, Mintzberg (1983) outlines several challenges for the organization structured as a professional bureaucracy. The strong affiliation the employees have with their profession can make the organization they work in seem almost incidental to the individual. It can be seen as just a convenient place to practice their skills. Employees can become loyal to their profession and not to the place where they happen to practice it (Mintzberg, 1983). Employees' loyalty to the standards of their profession also make organizational steering a challenge. The professions will typically value their professional standards higher than economic considerations. For example, a doctor in a hospital is probably not comfortable asking "do we have enough money to save this patient's life?" (Jacobsen & Thorsvik, 2007). Health care organizations may therefore have trouble controlling costs, as it is the decisions made by the professionals in their work that drive costs (Sveri, 2004; Jacobsen & Thorsvik, 2007).

The differences in the standards and skills among the professional groups in professional bureaucracies also pose a challenge. Rivalry and conflicts easily arise between professional groups, leading to problems with cooperation such as sub-optimization (Jacobsen & Thorsvik, 2007). In the specialist health care services system, this rivalry has typically been most prominent in discussions about what the professional groups should be authorized to do, who should be represented in the leadership, and which leaders have the final say. However, authorizations are determined on a national level, and individual organizations therefore do not have the power to decide in these matters. Rivalry related to formal and informal leadership positions on the other hand will be under the influence of the specific organization. The professional rivalry issue was the target of the so-called unitary management reform in 2002 (Sveri, 2004). The reform replaced the system of having one head nurse and one head doctor in a leadership team, with having one leader on each hierarchical level. This has had an impact on the so-called professional struggle (Norwegian: proffesjonsstrid), and the state of this struggle in the division today is illustrated by the answer given by Staff Employee 4 when asked about the professional struggle.

> Staff Employee 4: Yes I remember that used to be a big issue, but it is not very relevant anymore. After the unitary management reform it, much of the professional struggle went away, at least in our division. (...) Most of the struggle was about 'who really decides?' and 'who is the boss?' and the doctors claimed that no one other than them can be leaders. The system now is that on the top you have the division director and assisting director, and one of those have to be a psychiatrist. At the department level the leader has to have a specialist psychologist advisor and a psychiatric advisor, unless the leader has one of those professions themself. (...) In our division almost all of the leaders have backgrounds as nurses actually. The doctors either want to be leaders in a high position or treat patients. We don't have a single doctor in level four leadership (the lowest leadership position), we have two in level three (department leaders), and then we have Clas, the assisting director for the division.

Coordination beyond the professionals learning what to expect from each other through education and training, is dependent on professional's willingness to work cooperatively. Professional bureaucracies not being integrated entities complicate further coordination between professionals in the organization. The professions are collections of individuals who come together to draw on common resources and support services but otherwise want to be left alone (Mintzberg, 1983). As long as the pigeonholing process described above works, this may not pose a problem. However, when cases fall through the cracks between the standard programs, the lack of other coordination mechanisms becomes a problem. In this context, Mintzberg (1983) specifically mentions the case of hospital patients who fall between the medical specializations of psychiatry and somatic medicine. This may result in repeated transfers in search of the right department, a time consuming process when time is critical. The dependency on professional's willingness to cooperate and the reliance on external factors such as education make professional bureaucracies hard to change. Change happens slowly, and the administrators or professionals who want to change the way work is performed in the organization may have to move carefully in small, hardly discernible steps (Mintzberg, 1983).

5.3.2 INFRASTRUCTURE OF THE MEDICAL WORK

Mintzberg's (1983) model of professional bureaucracies has made considerable contributions to our understanding of hospitals. However, Vinge (2005) notes that Mintzberg's (1983) structural models are ideals formed to understand the dilemma between specialization and coordination. To further elaborate on this information, he works to examine how the work is organized in practice. Vinge and Knudsen (2002) understand the structure of hospitals by examining the medical work's infrastructure. An organization's infrastructure is the set of different standards and classification systems that structure the work. It is the set of standards and systems that produces order and regularity in the organization of the work (Vinge & Knudsen, 2002).

Vinge and Knudsen (2002) explains the infrastructure of hospitals by examining the way the medical work of physicians is organized and conducted in practice. Even though their analysis focused on Danish hospitals, they argue that these elements are very similar in all the Nordic countries. Consequently, their insights can be used in to

gain insights into the Norwegian specialist health care services system as well. Vinge and Knudsen (2002) identify three key elements of the infrastructure of the medical work in hospitals. The first element is the specialist doctor education (Danish: "speciallægeuddannelsen"), the additional education and on-the-job training that is required for a doctor to become a specialist doctor. The education produces a hierarchical classification system and includes an on-the-job training system where junior doctors circulate between hospitals and divisions to gain experience. At the start of their career, doctors will therefore stay only a short time in each position. This results in a nomadic work life with short stays in each organization. Many hospitals therefore have a large part of the core workforce is temporarily employed. The second element of the infrastructure is the way the work is organized by the help of certain tools. Health workers use tools such as the work schedule and task list to determine who should do what and where. The work is divided into positions and these positions are filled by people at given times with the help of planning tools such as the work schedule. Positions are fixed, with a fixed set of tasks belonging to it, but the people in the positions are not fixed, meaning that the position is filled by different people at different times. Physicians will fill different positions according to what is required of them at specific times. When a physician's workday is over, their position will be filled by someone else who will pick up the work where the first one left off. The last element identified in the infrastructure is the timing of the work, or pace of the day, defined by a set of fixed daily meetings. There are three daily coordination meetings, and physicians time their activities according to these meetings.

Considering these elements and how they relate to each other is important when considering potential changes to the system (Vinge & Knudsen, 2002; Sveri, 2004). Vinge and Knudsen (2002) argue that a power balance is built into the infrastructure by defining who gets to fill what positions, and this power balance creates robustness in the system. The system works in such a way that a doctor eventually gets to fill more attractive positions, and changes in the infrastructure may therefore threaten those who have worked their way up. Robustness is strengthened because the infrastructure is a stabile system of roles, functions, educational paths and meetings, making it difficult to make an isolated change to one of these elements (Vinge & Knudsen, 2002). The nomadic workforce together with how tasks are to be

fulfilled by positions and not a specific employee means that the organization of the work is less dependent on the people performing it, and more dependent on the infrastructure (Vinge, 2005). For example, even though the doctors have a high degree of autonomy in their work with the patients, they are not necessarily in a position to decide how their work-day is organized.

These three elements, along with their interdependencies, create a robust and standardized medical infrastructure that must be considered when trying to carry out a change in a hospital setting. Vinge and Knudsen (2002) and Sveri (2004) argue that when considering changes in the specialist health care services, the infrastructure of the work must be examined closely. The infrastructure determines the way the work is actually done. A prerequisite for changing the infrastructure is the ability to discuss it in relationship to change initiatives (Vinge & Knudsen, 2002). Change initiatives should therefore be discussed and formulated in a language that relates specifically to the infrastructure and is accessible to the employees who are to implement the change.

5.4 NATIONAL AND ORGANIZATIONAL CULTURE

An organization's culture can influence change efforts in a number of ways, and is therefore important to consider. There are many ways to define an organization's culture. Jacobsen and Thorsvik (2007, p. 120) find that most definitions agree that culture is a set of experiences, thoughts and beliefs that are common for a group of people in a given social setting. Shared beliefs and values provide guidelines for how individuals should behave in the organization, and these unwritten rules can have a strong influence on change efforts. Changing an organization's culture is often a difficult if not impossible task, and typically takes a long time. Culture therefore has an important role in deciding what changes are possible to implement in an organization (Edgren & Roald, 2001). Culture can both hinder and support change efforts. At the same time culture can also be affected by change efforts, or be a potential target for change in itself. For example, a culture where employees leave work at exactly four o'clock every day could hinder a change effort where the implementation depends on employees working late hours for a period of time. This culture could also be a target of change, where the organization wants the employees

to become more flexible with their schedule. This culture could also be affected by a change in the structure, for example if the employees would now be required to communicate with a team in a different time zone and therefore have to be available in the evening. In the following we present an overview of major cultural aspects that may influence change efforts in the division.

5.4.1 ORGANIZATIONAL CULTURE AND CONTEXT

The culture in the organization will be affected by its context. For instance, organizations are an integrated part of the national culture that influences both the environment in which it does business and the people it employs. Dominating values, norms and practice in organizations often reflect and adapt to the national culture (Jacobsen & Thorsvik, 2007, p. 133). Christensen, Lægreid, and Stigen (2007, p. 116) describe Norwegian culture as having relatively high level of trust and strong collectivist and egalitarian values. The general level of trust in public institutions, such as hospitals and the political leadership, is higher than in most other countries.

The established industrial practices are influenced by the national values, such as the relatively high level of trust mentioned above. Norwegian organizations have relatively low levels of internal conflict (Levin et al., 2012, p. 27). They are also identified as having a low level of respect for authorities and a high degree of informal structures, where the formal position of managers does not necessarily provide them with authority (Edgren & Roald, 2001). In the Norwegian specialist health care services, the combination of high levels of trust and informal structures can be seen in the relationships between political and managerial executives, and between executives and their subunits. Christensen et al. (2007) found that even though these relationships are defined in formal hierarchical terms, they are in reality largely based on trust. The appreciation of trust and informality in the division is illustrated by the importance attributed to informal relationships in the organization by Union Representative 1.

> Union Representative 1: My contact with leaders is often... we talk about other things than work. You can call that irrelevant talk, but it is actually a very good way of building relationships. When you have that looser form about

other things than work, you get to know people. I am on the employee-side of things, and to ensure good cooperation, it is very important to get to know people without there always being problems. Often when we build good relationships when there are no problems, then we are also able to work well when there are difficult problems to solve. You have to work well 'on green' to work well 'on red'! [Green and red are color codes in a tool they use to label goals and projects as running smoothly or in critical need of attention.] If we achieve that, then it is much easier for me to contact people informally. Employees come to me with issues. So when I have good relationships to for example HR people and leaders, I can give them a call and ask 'Listen here, are things like this and this?' And then we can discuss it informally. So we have trust in each other, and then we solve that issue in a good way because we worked well 'on green'.

Christensen et al. (2007) found that this high level of mutual trust contributes to goal acceptance by subunits and individuals in the organization, contributes to allowing high levels of autonomy, and reduces the need for external and formal steering devices.

Culture is also strongly connected to informal power. Culture determines who is considered legitimate when it comes to evaluating if things are being done the right way. This informal power combined with the relatively high level of informal structure in the specialist health care services can contribute to some groups acquiring strong positions that are not fully reflected in the formal organizational hierarchy. Consequently, when trying to introduce a change effort in these organizations, the people with informal power must be considered. Edgren and Roald (2001) describe a merger between two Norwegian hospitals that failed because of employee resistance, and identify several reasons for the failure. Informal structures where here a part of the identified problem. They found that the formal position of the change manager did not automatically give them an authority position. Edgren and Roald (2001) found the impact of informal leaders to be decisive in the cultural resistance at the hospital, and predict that cultural resistance can only be reduced if the majority of formal and informal leaders accept the change.

Furthermore, there are other cultures in the division's environment that influence the organizational culture as well. Employees are members of groups external to the organization that influence their values, norms and practices. In the Norwegian health care sector, such external influences include long, comprehensive and standardized educations, along with strong unions and associations. Hall (2005) found that health care professions typically develop strong cultures through shared educational experiences and the socialization process during training. Such influences from the environment can contribute to subcultures within the organization (Jacobsen & Thorsvik, 2001, p. 186). Specialist health care organizations are large complex organizations with many people in different groups, which will lead to different interests and perspectives. The different educational experiences, training and socialization within the professions shape the way people think and understand the world (Hall, 2005). The different professions may therefore have trouble understanding each other's point of views. Coming to agreements within the organization about change efforts, may therefore become a challenging task. This can pose a challenge when trying to implement changes, as the different ways of seeing the world may lead to more potential reasons to resist a change effort (Jacobsen, 2012, p. 99).

5.5 POLITICAL ECONOMY

Klev and Levin (2012) recognize the importance of political economies for change processes. When using the expression "political economy" they refer to "a rather general description of the political arrangement of social and political institutions, legislations and traditions that condition the organization of working life" (Klev & Levin, 2012, p. 56). They note that the political economy will differ in various countries, which means that a change activity that is feasible in one country might not be possible in another. They specifically point to the legitimacy, power and role of trade unions, and to expectations about the level of participation and democracy. In addition, they point to the role of state institutions in shaping conditions for change, and of national values and traditions such as described in chapter 5.4.

In Norway, worker's unions have a central role and there is a close relationship between enterprises, the authorities, and the unions. The relationship between enterprises, authorities and unions is characterized by a high level of trust and mutual cooperation (Levin et al., 2012). This is nicely illustrated for the division with the quote in chapter 5.4.1 from Union Representative 1 about trust. These relationships have been developed over more than a century, where both employer associations and workers' unions have played an important role. Today, the employer associations and workers' unions typically have centralized general agreements specifying terms of cooperation and workers' conditions. Workers' right to elect union representatives at the workplace is for example included in the agreements between workers' unions and employers (or their associations). The authorities have played an important role over the course of history by passing laws that secure workers' rights, such as the right to be organized and determining a minimum level of representation in the enterprise (Levin et al., 2012).

Levin et al. (2012) refers to the relationship between enterprises, authorities, and unions as the "Norwegian model". At the enterprise-level, Levin et al. (2012) describes the Norwegian model as the "Norwegian cooperation model". Today, over half of Norway's employees are organized in workers' unions, and are therefore offered work conditions negotiated by their unions. Furthermore, enterprises often do not differentiate between those that are organized and those who are not, and therefore offer the same conditions also to those not organized. An important point to make is that the relationship between enterprise leaders and unions is characterized by respect, making it possible to have both conflicts and cooperation at the same time (Levin et al., 2012). For example, cooperation between union representatives and enterprise leaders to find solutions for structural changes can happen simultaneously with an intense conflict regarding salaries between the same parties. The central role of union representatives in the division is illustrated by the comments of Department Head 1 and Department Head 2.

> Department Head 2: The leaders have the formal authority to make changes, but not unconditionally. Even when just changing one job description I have to inform the union representatives and confer with them about it. The agreement

89

between the parties gives them the right to contribute their opinion about potential changes.

> Department Head 1: The union representatives are a part of all change processes. We just closed down a unit, and representatives were there at every meeting, during every decision and evaluation. The employees trust the representatives to represent their interests.

Cooperation between enterprise leaders and union representatives is common, and this cooperation bears a high level of legitimacy as illustrated by Staff Employee 4.

> Staff Employee 4: And then we have the division-level union representatives that are elected on the highest level. They take part in the leader gatherings and in the leadership training. Not all of the gatherings but many of them. So they should absolutely know what is going on. The division director has a meeting with them one Monday every month, contact forum it's called. Then they have issues they want to discuss, and then the division director informs them about what he is doing, change processes, budget, for all those things they are a part of it. They don't have any right to decide, they are to ensure the employees' interests and are the employees' mouthpiece into processes and to the director. It is co-determination in that sense, in that they have a channel in. So it is very important to have good working relationships to the union representatives.

A culture has emerged where employees expect a high level of power sharing between leaders and employees on all levels (Levin et al., 2012). This means to both having the power to influence how they perform their work, and having a say in decisions that affect them. Employees expect to have a say in major decisions, and union representatives are commonly included in decision-making processes. The relationship goes both ways, and leaders typically report that including union representatives in decision processes related to change efforts lead to useful input from employees (Levin et al., 2012).

This is further illustrated by the answer to our question about who would be consulted when deciding on introducing a tool such as the LESAT as for example tool for the department leaders to use on a yearly basis. We were told the employees would expect to have been included that decision process through their union representatives:

> Staff Employee 4: It obviously affects the employees when the leaders get a new tool to lead by. So the representatives should absolutely be a part of, at least be involved, in the process. How much they would participate in using the tool I don't know, but they should at the very least be oriented about what is going on. And they should have the right to be consulted about the choice of tool and so on. They would probably be put off it they were not involved in that. If they later learned that a management tool was employed that they were not informed about before its use was determined finally, well, they would let us hear it.

In the division, as throughout the specialist health care services, representatives for patient groups are also included in the divisions' affairs. User representatives are represented in the IHT board, as well as included through monthly meetings with executives and represented in all larger projects and change initiatives:

> Staff Employee 4: In all large processes we have user representatives. We have a user committee that has monthly meetings with the top management, so users are highly involved in what is going on at Innlandet Hospital Trust. [...] We don't start have big projects without user representatives.

5.6 STAKEHOLDERS

To fulfill the criteria of usefulness, the most important concerns of the most important stakeholders should be addressed. Furthermore, the stakeholders of the organization are referred to frequently in the SAT. It is therefore important to know who they are, both for us as researchers and for the participants. The stakeholders to the self-assessment are also the stakeholders of the organization. An identification and analysis of the division's stakeholders is presented in this chapter.

To identify the stakeholders and investigate their importance, a list of the main groups and organizations that can affect or be affected by the division is presented. This list was supplemented by inputs from interviews at Gjøvik DPS and Sanderud Hospital (see chapter 4), where the informants gave feedback on the stakeholders we had identified and added some more. The stakeholders are sorted into groups, groups that either describes what kind of organization they are or their relationship with the division. The importance of the identified stakeholder groups is then considered using the Stakeholder Salience Model. The list of stakeholders is presented in the following section, followed by a presentation of the Stakeholder Salience Model and an analysis of the identified stakeholder groups.

5.6.1 STAKEHOLDERS TO THE PSYCHIATRIC HEALTH SERVICES DIVISION AT IHT

- Owners and external leadership
    - The state (Ministry of Health and Care)
    - Health South-East Norway RHE
    - IHT leadership
- Internal leadership
    - Administrative leadership of the Psychiatric Health Services Division at IHT
    - Steering committee
- Patients and relatives
- Employees
    - Non-medical employees
    - Physicians
    - Psychologists
    - Nurses
    - Physical therapists
    - Medical receptionists
    - Social workers
    - Occupational therapists
    - Other caregiver professions
        - (Norwegian: vernepleiere, hjelpepleier, barnepleiere)
- Society
    - Population (as taxpayers, voters and potential patients)
    - Municipal local authorities
    - County local authorities
    - Media

- Other entities in IHT
  - Hospitals
- Entities in the municipalities
  - Primary health services
    - General practice, pregnancy and antenatal care, health clinics for mother and child, school clinics, mental health care, preventive medicine, nursing homes, rehabilitation, physiotherapy, communicable disease control, environmental health, health promotion
  - Norwegian Labor and Welfare Administration
  - Child Protective Services
- External contributors to patient's health care
  - Other LHEs
  - Private clinics and hospitals
- Unions/associations.
  - Workers' unions
    - Doctors union, nurses union, psychologist association, etc.
  - Patient associations
    - Association for Mental Health (Norwegian: Mental helse), Association for Relatives of Psychiatric Health Patients (Norwegian: LPP), Association for Patients with Substance Abuse (Norwegian: RIO), etc.
- Academia
  - Educational institutions (Norwegian: forsknings-og undervisningsinstanser)
  - Research centers
- Suppliers
  - Pharmacies (We put them in this group because they supply the division with medicine)
  - Technology (IT-company, machines)
  - Medical (pharmaceuticals) consultants

5.6.2 STAKEHOLDER SALIENCE MODEL

To find an approach to the stakeholder analysis, we have considered the research by Mitchell et al. (1997). Based on a comprehensive literature review of the field of stakeholder theory, they propose a model for grouping stakeholders according to stakeholder salience. Stakeholder salience is here defined as "the degree to which managers give priority to competing stakeholder claims" (Mitchell et al., 1997, p. 854). Through their research they develop a theory that goes beyond identifying

stakeholders to consider how stakeholder salience can explain who and what managers actually pay attention to. The result is a framework where stakeholder salience is determined by considering the stakeholder's power, legitimacy and urgency. Power is here determined by the extent to which you can gain access to coercive, utilitarian or normative means to impose your will on the relationship. In this context, coercive power is power based on physical resources of force, violence, or restraint. Utilitarian power, is based on financial or material resources, and normative power is based on symbolic resources. Legitimacy is a separate attribute from power, and it is emphasized that you can have legitimacy without power. However, if you do have both, you get authority. Legitimacy is here defined as "a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions" (Mitchell et al., 1997, p. 866). Urgency is the third attribute in this model. It is defined as "the degree to which stakeholder claims call for immediate attention" (Mitchell et al., 1997, p. 867). A stakeholder is considered to hold the attribute of urgency when the two attributes of time sensitivity and criticality are held. Time sensitivity relates to what extent a delay in attending a claim or relationship will be seen as unacceptable to the stakeholder. Criticality is the importance of the claim or the relationship to the stakeholder. (Mitchell et al., 1997)

Mitchell et al. (1997) group stakeholders based on how many of the three attributes of power, legitimacy and urgency, are held by the stakeholder. If only one of the attributes is held, the stakeholder is considered to be a "latent" stakeholder. Stakeholders that fulfill the criteria of two of the attributes are referred to as "expectant" stakeholders. This is because these are stakeholders that "expect something". The stakeholders that have the combination of all three attributes are called "definitive" stakeholders. (Mitchell et al., 1997)

5.6.3 STAKEHOLDER SALIENCE ANALYSIS

The salience of the stakeholder groups identified for the division are explained below in terms of the stakeholder power (P), legitimacy (L), and urgency (U).[2]

Stakeholder group	P	L	U	Comment
Owners and external leadership	x	x	x	The owners and the leadership will have utilitarian power, since they control the finances. The state also has a great deal of legitimacy in power of the political leadership being democratically elected. As owners and leaders, this group will have some urgency. This will mostly be the case for the leadership of IHT. As the direct leadership to the division, they are unlikely to accept delays to their requests, and will consider their demands to be critical. This means that they have both have criticality and time sensitivity, and therefore urgency. The state and the RHE will also have criticality, since they will consider their requests to be critical, but time sensitivity will be less than that of the LHE. This is because they need to go through more levels and it is to be expected that the requests will come with some delay.
Internal leadership	x	x	x	The internal leadership will also have utilitarian power, since they control the distribution of finances in the organization. Seeing as they are the leaders of the organization, they can also be expected to hold a certain degree of legitimacy. Leadership will also have urgency, because they will not tolerate delays and will see their claims as critical to the organization, thus fulfilling the criteria of having time sensitivity and criticality.
Patients and relatives	x	x	x	The patients are given legal rights from the Patient and User Rights Law, which gives them a certain amount of power.

---

[2] When X-es are surrounded in brackets, this signifies that they have the corresponding attribute, but to a less extent than the others who are considered to have it. Also, when performing this analysis, there were several stakeholders that have power that come from rights given to them by law, e.g. patient rights. This cannot be characterized as being utilitarian, normative or coercive, however, we still chose to include it.

				The patient will also have some legitimacy, since it is considered a legitimate claim in the Norwegian society to have a say in the treatment of your own body. Finally, the patient will have a very high level of urgency because the request for care will be both critical and time sensitive. A delay may cause a worsened condition (time sensitivity), and a failure by the organization to act may have critical consequences (criticality).
Employees	x	x	x	Employees have coercive power, in the form of restraint, since they can refrain from doing their work if they disagree with the way things are done. The different professions will also have a normative power, concerned with defining what is the acceptable way of doing the work. The employees also have legitimacy, because employees' opinions have a high level of legitimacy in Norway. Employees also have urgency. Time sensitivity and criticality comes from the fact that employees are often dependent on the organization to be able to fulfill their tasks satisfactorily. Also, they are often dependent on the organization's response to solve issues that relate to their work-life.
Society	x	x		Society is a group with several different kinds of stakeholders. The population has power through voting. The local authorities at the municipal and county level bear legitimacy because they represent and are responsible for the well being of the people in their area. The media also has some legitimacy, since their mandate is to comment on the ongoing in the country. The media also has a normative means of inflicting its will, since they can influence the views of many people by making a case very visible. This can then put pressure on the organization. The media therefore has power.
Other entities in IHT	x	x	x	The hospitals at IHT have power, since they can refer patients to the division. They also have legitimacy by being part of the same organization and working towards the same

				ends. Finally they have urgency, because they are in need of timely, precise and correct communications. If they need information about a patient for example, this will be both time-sensitive and critical.
Entities in the municipalities	x	x	x	The entities in the municipalities can be divided into two sub-groups, a medical and a non-medical. The private health services is a medical provider. The Norwegian Labor and Welfare Administration and the Child Protective Services are not medical providers, but still cooperate with the division to help the patient. The entities in the municipalities all have urgency, because they may need to exchange information quickly about patients. The information needed may be critical to the treatment of the patient, and will therefore be seen as both time sensitive and critical to the entities in the municipalities. The medical providers also have power, because they can refer patients to the division. So do the non-medical, as law demands cooperation and communication with these entities. They all have legitimacy, the medical providers because they are working to care for the same patients, the non-medical because they also have the same goal of helping the patients.
External contributors to patient's health care	x	x	x	As with the municipalities, all these external contributors have urgency, because they may need to exchange information quickly about patients. Also here the information needed may be critical to the treatment of the patient, and will therefore be seen as both time sensitive and critical to the external contributors. They also have power, because they can refer patients to the division. Further, they have legitimacy, because they are working to care for the same patients.
Workers' unions and patient associations	x	x	x	The unions and associations have the same attributes to some extent. Unions have power because they represent the professions, who have a powerful position in the organization. The amount of power to the union will depend

97

				on the strength of the profession. In Norway, unions are also considered highly legitimate actors. Patients' associations' power also depends on the size of the association. For instance the Cancer Association in Norway is very powerful. These associations are also legitimate in much the same way as unions. Both patient associations and workers' unions groups expect to be included in all major processes in the organization. This is considered critical to this stakeholder group, and they will react if they are not included and informed. Further, a delay in attending to this demand, by informing these groups after a decision has been made, will be considered unacceptable. It follows that both the attributes of time sensitivity and criticality are held, and they therefore hold the attribute of urgency.
Academia	x	x		Academia has utilitarian and normative power through their control over what is taught to the future professionals. They also have legitimacy, because their way of accumulating knowledge is respected and considered legitimate.
Suppliers	x		(x)	Suppliers have utilitarian power, since they control material resources. If they wish, they can withhold resources. Suppliers also have some degree of urgency, as they need information from the division to plan their activities.

TABLE 4: STAKEHOLDER SALIENCE ANALYSIS

From this analysis, it becomes clear that the division has a large set of important stakeholders. All groups have at least two of the attributes and many of the groups were found to have power, legitimacy and urgency. It is not surprising that the division has so many important stakeholders. This is an organization with a responsibility to the community. Many actors will depend on the division doing its job, since what they provide is so vital to patients.

5.7 Change Processes in the Division

The LESAT is concerned with best practices for large-scale changes in organizations. A short overview of recent and future change processes in the division and important considerations with regards to change management is therefore presented.

The combination of being a professional bureaucracy where the professionals, especially doctors, bear the highest legitimacy, having a co-determination model and high levels of informal power structures, and a large range of stakeholders makes change in the division hard to manage. In the division, leaders must walk a fine line between undisputable, top-down decisions, and involving stakeholders in the process. Through their unions or associations, employees and patients expect to be heard and informed in all major decisions affecting them (see chapter 5.5). The division is geographically dispersed over 49 municipalities and has strong stakeholder interests in many directions from patient groups, municipalities, professional groups, politicians and local populations. Change initiatives will seldom be good for everyone at once, and as many people are affected by the division's actions, projects can meet resistance from many directions. The balance between involving stakeholders in the process, and having to make tough decisions that not everyone agrees with and make them with legitimacy, is of utmost centrality in change processes in the division.

The division was recently put through a large, resource intensive project called Project Psychiatric Health Services 2011-2013 (spoken of as "the Project"), and involvement has been mentioned as one of the most important success factors. The Project (Project Psychiatric Health Services 2011-2013) was one of the largest recent changes in the division. There was a general agreement between our interviewees that this project was completed, and changes implemented, in a surprisingly successful way.

> Department Head 1: For the Project many things are on green, it has been very good. And it has meant a lot. We have a completely different basis now to work on all the other challenges that we have.
> (Practice I.A.1 group pilot round)

When asked about the most important success factors behind the Project, the answers were related to choice of the project leadership team and involvement. On the one hand, involvement of employees and user groups was considered very important for the acceptance of the changes in the organization and from stakeholders such as patient associations and municipalities. On the other hand, having a project leader who was able to make and stand up for tough decisions was given much of the credit for the project being successful. Also, including an experienced doctor on the leadership team was seen as critical. In chapter 5.3.1 we described the division as a professional bureaucracy, where the loyalty of the professionals is often more tuned in to their professional values than to the organization they belong to. The representation of professional values in the project leadership team is seen as central for the legitimacy of the decisions made.

> Staff Employee 3: She [project leader] stood in the middle of all these conflicts with a straight back.
> Staff Employee 4: That was alpha omega. [...]
> Staff Employee 3: And of course, she was not a medical professional, so the Project's success rested very much on the shoulders of Harald too. He is an experienced old psychiatrist who has seen it all. With great professional legitimacy. If he hadn't been there, the Project would not have been a success.
> Staff Employee 4: Yes.

The Project resulted in several plans for major change initiatives ("Master Plans"). The Master Plans are still being followed up and implemented, along with the other yearly orders to change from the division's owners. This is useful to have in mind when reading the participants' comments about transformation and change initiatives in the data presentation.

With respect to future changes, being able to change, including large-scale changes, is very important in the division. Campaigns, new political environments (indirect owners), budget cuts and a constant pressure from the public and political leaders to improve necessitate having the ability to change. As Department Head 1 remarked during the group pilot round:

Department Head 1: With respect to the Project, we are well on our way, lots of things on green. But new things keep coming along as well, the Patient Safety Campaign that is being rolled out throughout, implementation of lots of new guidelines, technical/professional guidelines (Norwegian: faglige retningslinjer), and we don't have all that in place. But we don't know what challenges we will get in two years, that we have to work on then. We don't live in an industry where we see the future. We have to enable ourselves to have a system, so that we can make the changes that come in a good way.

5.8 SUMMARY

In this chapter we have explored structural and cultural characteristics of Norwegian hospitals, and considered the political economy that the division is placed within. In addition, we have explored what stakeholders are of importance to the division, and important considerations with regards to change management.

The first structural aspect to be considered was the formal structure of both the Norwegian health care sector and division. This description placed the division in the context of the Norwegian health care sector. As a part of the specialist health care services in Norway, the division is a part of the public sector. This means that it has political leadership and receives public funding. Furthermore, as professional bureaucracies with a robust infrastructure, hospitals are organizations that may prove difficult to change.

The culture in Norwegian hospitals has also been described, by considering effect of the Norwegian national culture, subcultures. The Norwegian culture contributes a high level of trust and a relatively high level of informal structures. Subcultures are found to both make the organization less robust, but at the same time present a challenge. Subcultures are largely connected to the different professions and the organization has little influence over these cultures. In addition, by considering the political economy we have explained the importance of worker's unions in the Norwegian context, where the cooperation between enterprise leadership and union representatives is legitimate and their inclusion in change processes and major

decisions is expected. In the specialist health services, user representatives are also a part of this co-determination system. The division has very many stakeholders, and that they all were found to be important to consider. Campaigns, new political environments (indirect owners), budget cuts and a constant pressure from the public and political leaders to improve necessitate having the ability to change. Balancing involvement with top-down decisions and ensuring that the professionals consider the decisions legitimate are important concerns for change management in the division.

6 THE LAI ENTERPRISE SAT (LESAT)

The Lean Advancement Initiative (LAI) at the Massachusetts Institute of Technology (MIT) developed the SAT chosen for this thesis. LAI is a research consortium with collaborative partnerships with industry, government and academia. LAI has developed a theory of organizational change they call the Enterprise Transformation Paradigm. Within the paradigm, an enterprise is defined as a "complex, integrated, and interdependent system of people, processes and technology with a distinct mission that creates value as determined by its key stakeholders based on that mission" (Nightingale & Srinivasan, 2011). Because of this wide definition, we will use the words organization and enterprise interchangeably throughout the rest of the thesis. As a part of LAI's work with the Enterprise Transformation Paradigm, the LAI Enterprise SAT (LESAT) was developed. The LESAT was created to help organizations understand their current state and identify opportunities for the future (Nightingale & Srinivasan, 2011). The LESAT is the SAT we will be considering in this thesis. The latest version of LESAT, LESAT Version 2.0, is designed with focus on manufacturing and product-oriented enterprises, but it is supposed to be widely applicable also for government and healthcare organizations (Nightingale et al., 2012, p. 9).

To fully understand the LESAT and its role in the Enterprise Transformation Paradigm, the paradigm needs to be explained. We will therefore start by introducing Dunphy's (1996) conceptualization of change theories. This will be followed by a description of the history of LAI and the context in which the change theory was developed. Using Dunphy's (1996) conceptualization of change theories, we will then shed light on the key elements of LAI's Enterprise Transformation Paradigm. Following this, the LESAT will be described in detail, starting with its purpose and use, followed its main structure, content and then our customizations of it.

6.1 UNDERLYING THEORY OF AND CONTEXT OF THE LESAT

Dunphy (1996) identifies five main elements that make up a change theory. The first element is *a basic metaphor* of the nature of the organization. A change theory also has *a definition of the role of the change agent*. The third element is *an analytical framework* that is used for understanding the organizational change process. The analytical framework specifies key variables critical to the change process, and the processes that link the variables. *An ideal model of an effectively functioning organization* suggests direction for change and the values to be used in evaluating the success of the change intervention (survival, growth, workforce satisfaction, etc.). The last element is an *intervention theory* that specifies when, where and how to intervene so as to move the organization closer to the ideal (Dunphy, 1996, p. 543).

6.1.1 DEVELOPMENT OF THE ENTERPRISE TRANSFORMATION PARADIGM

The development of change theories is strongly affected by the historical and social context it takes place in (Dunphy, 1996). LAI is a research team at MIT that has been conducting research in organizational change since 1993 (Nightingale & Srinivasan, 2011). The research was initially centered on introducing the concept of "lean production", or just "lean" to the aerospace industry. This concept originated from studies of Toyota, a car manufacturing company that concurrently achieved low costs, high quality and production flexibility. MIT has historically had a central role in developing theory about lean through a series of efforts since the late 1980s. The first effort was concerned with understanding the factors leading to Toyota's success, leading to the book by Jim Womack, Daniel Jones, and Daniel Roos, "The Machine that Changed the World". Womack and Jones later tried to generalize the understanding of Toyota's production system, leading to a new book in 1996 where they presented 5 principles of lean thinking. These principles were drawn from a study of the automotive industry, and focused on waste elimination and identifying value from the perspective of the customer (Nightingale & Srinivasan, 2011). The five principles are listed below.

- "Specify value from the standpoint of the end customer by product family.
- Identify all the steps in the value stream for each product family, eliminating whenever possible steps that do not create value.

- Make the value-creating steps occur in tight sequence so that the product will flow smoothly toward the customer.
- As flow is introduced, let customers pull value from the next upstream activity.
- As value is specified, value streams are identified, wasted steps are removed, and flow and pull are introduced, begin the process again and continue it until a state of perfection is reached in which perfect value is created with no waste." (Nightingale & Srinivasan, 2011)

Working towards generalizing lean theory to the aerospace industry, LAI established 5 lean enterprise principles in 2004 that built on the original 5 lean principles. This new set of principles were made to broaden the focus from waste elimination to include value creation, where terms such as "stakeholder value" and "enterprise perspective" were introduced, while lean-specific terms such as "pull", "flow" and "waste" were removed. LAI continued to work on broadening the applicability of their research (now changing its name from Lean Aerospace Initiative to Lean Advancement Initiative). What emerged was a change theory they called the Enterprise Transformation Paradigm. For this paradigm, LAI has developed the Enterprise Transformation Framework, a set of what Nightingale and Srinivasan (2011) refer to as "overarching" enterprise principles and transformation methodologies that are to work independently of the industry and organization (Nightingale & Srinivasan, 2011). From the founding in 1993, LAI has evolved from a focus on lean processes and tools in the aerospace industry to what they call holistic enterprise transformation and architecting in a broad range of industries including service industries and health care (Nightingale & Srinivasan, 2011).

6.1.2 ENTERPRISE TRANSFORMATION FRAMEWORK

The Enterprise Transformation Framework developed for the Enterprise Transformation Paradigm is illustrated below (figure 10). The Enterprise Transformation Framework consists of five interlinked parts. These parts are the 7 Enterprise Principles, the Enterprise Transformation Roadmap, and the tools LESAT, Enterprise Architecture (EA) framework and Enterprise Strategic Analysis for Transformation (ESAT). The EA framework helps you to examine the enterprise

through multiple perspectives or views. These are the strategy view, policy/external environment view, process view, organization view, knowledge view, information view, product view and services view. The ESAT provides a process that should be followed in the planning cycle to analyze and improve the overall enterprise performance. This process incorporates the use of the LESAT and EA framework (Nightingale et al., 2012).
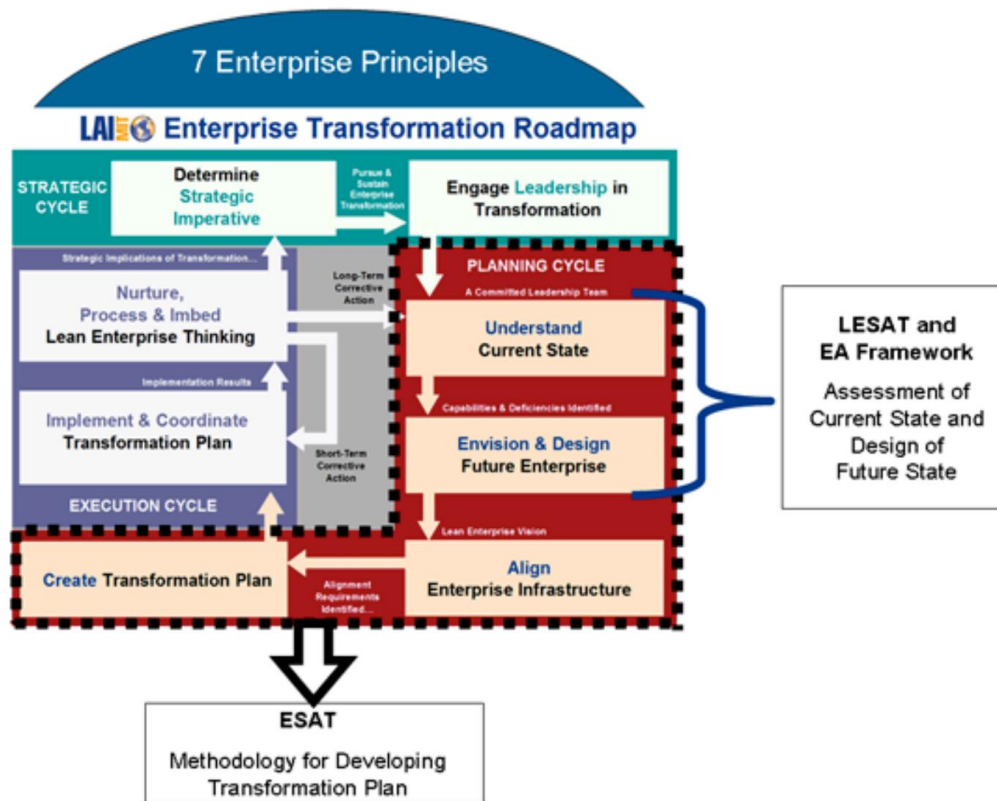
FIGURE 10: ENTERPRISE TRANSFORMATION FRAMEWORK (NIGHTINGALE, 2009)

The 7 Principles of Enterprise Transformation are the cornerstones of the Enterprise Transformation Paradigm, and are therefore central in the Framework. These principles, described by (Nightingale & Srinivasan, 2011), are summarized in the list below.

- Adopt a holistic approach to enterprise transformation.
- Secure leadership commitment to drive and institutionalize enterprise behaviors.
- Identify relevant stakeholders and determine their value propositions.
- Focus on enterprise effectiveness before efficiency.
- Address internal external enterprise interdependencies.

- Ensure stability and flow within and across the enterprise.
- Emphasize organizational learning

The second central piece in the Enterprise Transformation Framework is the Enterprise Transformation Roadmap. This is a step-by-step guide that organizations should follow according to the paradigm. It describes the steps to achieve a successful transformation, and that an organization should keep following to continually improve. The Roadmap is divided into three cycles: A strategic cycle, a planning cycle and an execution cycle, and each of these cycles consist of several steps. The Roadmap is presented in figure 11.
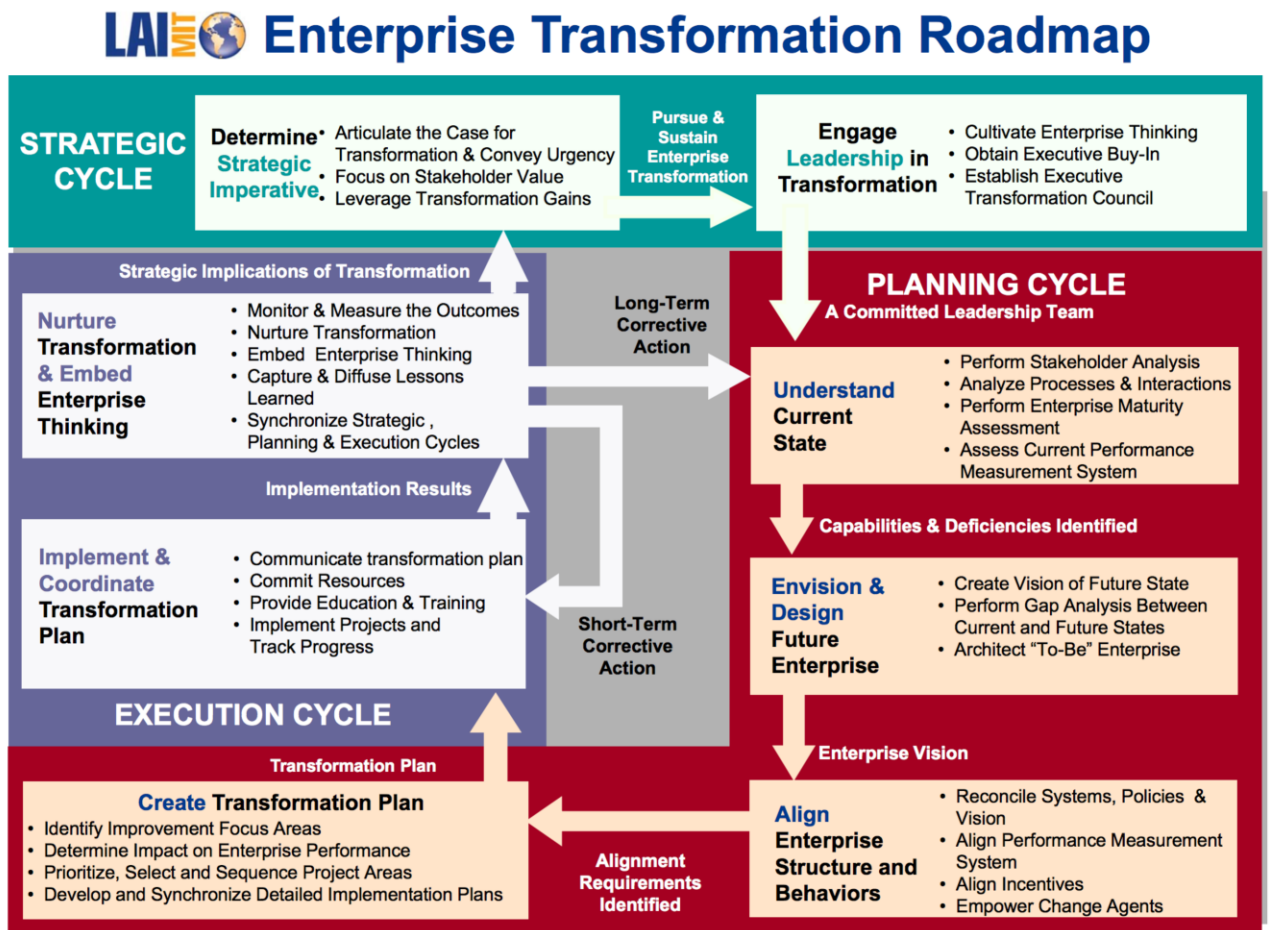
FIGURE 11: ENTERPRISE TRANSFORMATION ROADMAP (NIGHTINGALE & SRINIVASAN, 2011)

6.1.3 ANALYSIS OF THE ENTERPRISE TRANSFORMATION PARADIGM

The Enterprise Transformation Paradigm will further be analyzed based on Dunphy's (1996) description of the key elements that make up a change theory.

6.1.3.1 ORGANIZATIONAL METAPHOR

Enterprise Transformation Paradigm's organizational metaphor can be found in the description of their subject for change - the enterprise. As mentioned above, an enterprise is described as "a complex, integrated, and interdependent system of people, processes and technology with a distinct mission that creates value as determined by its key stakeholders based on that mission" (Nightingale & Srinivasan, 2011). This gives the image of organizations existing with a distinct purpose, defined by a mission determined by the needs of the stakeholders. Also central to the Enterprise Transformation Paradigm is the focus on delivering value by considering the "extended enterprise", defined as "all organizations along the multiple value streams that contribute to providing value to the enterprise stakeholders. This may include customers, suppliers and others that might have indirect influence over enterprise activities and its ability to deliver value" (LESAT Version 2.0, 2012, p. 41). The nature of the organization can therefore be described as follows: A purposeful system of people, processes and technology that achieves value for its multiple stakeholders in a changing environment, and is an active part in a system of organizations that contribute to providing value to the stakeholders.

6.1.3.2 CHANGE AGENTS

When considering the role of the change agent we look to the explanation used by LAI. The definition of a change agent found in the Enterprise Transformation Paradigm is based on the definition of Womack, Jones, and Roos (1990). Here the change agent is "an individual who acts as a driving force in the change effort by planning, managing, and championing the implementation process. This individual must have knowledge about the enterprise and a clear understanding of the future vision in order to motivate and educate others in the organization" (LESAT Version 2.0, 2012, p. 39). The role of the change agent is first and foremost held by the senior leadership. "Enterprise transformation begins with the commitment of the senior leadership team, which must invest the resources needed to change the way the enterprise works on the large scale. At the same time, leadership must require the

personal dedication of all stakeholders to make local changes on an ongoing basis" (Nightingale & Srinivasan, 2011). The role of the change agent is divided, and the role one takes depends on the position they have in the organization. The senior management is responsible for designing the main direction of the change, while other stakeholders such as employees become a form of change agent as they join the effort and ensure that the changes are implemented and continually improved.

6.1.3.3 ANALYTICAL FRAMEWORK

The analytical framework is a model for understanding the organizational change process, ideally including key variables critical to the change process and the processes that link these variables. In the Enterprise Transformation Paradigm, organizational change is understood as a combination of episodic and continuous change. Nightingale and Srinivasan (2011) describes episodic change as typically top-down change efforts aimed at changing the entire organization, and continuous change as bottom-up efforts more focused on local work practices. The view of change underlying the Enterprise Transformation Paradigm is illustrated by Nightingale and Srinivasan's (2011) predictions for change efforts, "With an engaged leadership team, the transformation can emanate outward in concentric circles to reach every part of the enterprise. Then it becomes both top-down and bottom-up, with continual iteration." The key variables critical to change and the processes linking these are reflected in the way Nightingale and Srinivasan (2011) models the enterprise. The enterprise is modeled as a set of "enterprise processes", called the "enterprise process architecture". The process architecture is divided into three main categories called enterprise-level processes. These are the three categories of processes and practices that together make up the enterprise:

1. Enterprise transformation leadership
2. Lifecycle processes
3. Enabling infrastructure processes

Leadership processes are those activities necessary to achieve change. These activities are reflected in the Roadmap, such as developing a new strategic direction, forming alliances and resource deployment. Lifecycle processes are the core business processes, the activities that add value over the lifetime of its products and services. The enabling infrastructure processes provide resources, information and services to

the lifecycle and leadership processes. Analyzing the enterprise's processes architecture is seen as key in successful transformation efforts.

6.1.3.4 IDEAL MODEL

The Enterprise Transformation Paradigm's ideal model is reflected in the Enterprise Transformation Framework. Nightingale and Srinivasan (2011) say that enterprises should be designed in such a way that it can deliver the value required from it by the relevant stakeholders. This organizational design processes is done by considering the different views in the EA framework. Although the ideal construction of an organization will depend on the value it should deliver to stakeholders, there are several "best practices" associated with each of the enterprise-level processes. These "best practices" can for example be found in the LESAT. In the Enterprise Transformation Paradigm, an effectively functioning organization also continually transforms itself in a never-ending spiral of improvement (Nightingale & Srinivasan, 2011).

6.1.3.5 INTERVENTION THEORY

Lastly, the intervention theory of the Enterprise Transformation Paradigm is communicated through the Roadmap. The when, where and how to intervene is described as a series of steps, where the use of tools such as LESAT, ESAT and EA framework is recommended. Also, the 7 Enterprise Principles illustrate the essence in the intervention theory, explaining the main aspects that must be in place to be able to transform the enterprise. It is for example emphasized that all employees should adopt a holistic view of the enterprise transformation process, and that focus should be on effectiveness before efficiency and organizational learning should be emphasized. The Enterprise Transformation Framework in its entirety also illustrates the intervention theory in a more general form. The Framework includes methods and tools for helping organizations "drive genuine enterprise transformation" (Nightingale & Srinivasan, 2011). Consequently, the Enterprise Transformation Framework, with the roadmap as a step-by-step guide, provides a concrete implementation guideline for the Enterprise Transformation Paradigm.

6.2 THE LESAT

6.2.1 USE AND PURPOSE OF LESAT

The LESAT is an SAT in the form of a free downloadable pdf-file. The LESAT has the form of a checklist with 43 "practices", and for each practice the organization gives itself a score on a scale from one to five, five being the best. These levels are called "capability levels" for the associated practice. For each practice then, the capability level 5 is the description of a best practice for the enterprise. In this way, the list of practices with the associated capability levels represents the ideal model of the change theory. With its list of topics (practices) with associated capability levels and an ideal described for each topic (level five), the form of the LESAT falls under general form called Design 4, outlined in 2.3.4. The LESAT is a tool for the organization to compare itself to this ideal. The organization gives two scores for each practice - one for the current state and one for the desired. Leaving the scoring of desired states to the organization enables the organization to decide which practices are deemed the most important for the organization. Then, by comparing the desired capability level with the current capability level for each practice, the organization can use the LESAT to determine what practices to prioritize in their change efforts (Mize & Nightingale, 2002, p. 9).

Scores are given as follows. For each capability level of a given practice, there is an explanation that helps the organization understand what score to give themselves. Figure 12 below is an example of one of the 43 practices with its associated capability levels. When the users find that their organization's current state matches the description in one of the capability levels, they circle the "C" (for current state) in the matching box. The same applies for the capability level the organization aims to reach; the user then circles the "D" (for desired state).

111

Each practice has a title describing it, such as "Analyze Enterprise Processes and Interactions" (LESAT Version 2.0, 2012, p. 11), with a comment under the title describing what the enterprise will accomplish if they employ that practice, like "Understand process interdependencies" (LESAT Version 2.0, 2012, p. 11). For each practice, a description is given of what an organization looks like at every capability level from 1 to 5. Each practice also comes with a list of "indicators", comments to help users understand what to look for when comparing their organization to the capability level descriptions, for example "The practice and language of process analysis (such as value stream mapping) are used to understand important enterprise processes" (LESAT Version 2.0, 2012, p. 11).

EP #	ENTERPRISE PRACTICES	Capability Levels				
		Level 1	Level 2	Level 3	Level 4	Level 5
I.C.1	**Analyze Enterprise Processes and Interactions** *Understand process interdependencies*	There is no understanding or a limited understanding of the need for process mapping and analysis. The documented process flow differs from the actual flow.	Core enterprise processes are mapped and have been analyzed.	Mapping and analysis of current processes allows the identification of critical interactions. Significant opportunities for eliminating waste and creating value are identified and aligned with the strategic objectives.	Depth and breadth of knowledge of enterprise processes exposes interdependencies across the enterprise.	Continuously evolving enterprise processes and their interdependencies are evaluated across the extended enterprise.
		C | D	C | D	C | D	C | D	C | D
	Indicators (Examples)	• The practice and language of process analysis (such as value stream mapping) are used to understand important enterprise processes. • Current value streams of major customers/product lines have been mapped, and hand-off points and interfaces clearly defined. • Enterprise leadership actively manages processes that have interactions across functions.				

FIGURE 12: PRACTICE I.C.I (LESAT VERSION 2.0, 2012)

When performing the LESAT, the organization is meant to have a designated facilitator to make sure that evaluation is done in the correct way. For this purpose MIT has made a facilitator's guide. Here, it is suggested to do the assessment in a series of consecutive sessions over two or three days. If three days are preferred, the following schedule could be used. The first day is started by introducing the participants to the LESAT and training them in its use. They will here be taught the Enterprise Principles and the assessment process. Following this introduction, the participants will individually perform the assessment using the LESAT score sheets. The facilitator will then collect these results at the end of the day. On the second day, the facilitator will present the results of the initial analysis to the respondents and facilitate a discussion, where the respondents are to reach a consensus on final scores. On the third day, with the facilitator's guidance, the participants should interpret the assessment results and identify and prioritize actions to be made. (Nightingale & Srinivasan, 2011)

During this process, the participant will need to come to a consensus about what scores apply for the current state, and the desired state. The organization also needs to agree on how to use the results to make decisions. As an example, the organization can choose to prioritize improving the practices where there is a large gap between current and desired state. Using the LESAT should focus the discussion of how to prioritize resources and support a common understanding of the organization's state and desired future.

6.2.2 OVERALL STRUCTURE

The overall structure of the LESAT follows the way enterprises are modeled - the enterprise process architecture. The practices and associated capability levels are grouped into the three enterprise-level processes: leadership, lifecycle and enabling infrastructure processes.

This categorization into sections and their subsections is illustrated in figure 13 below.
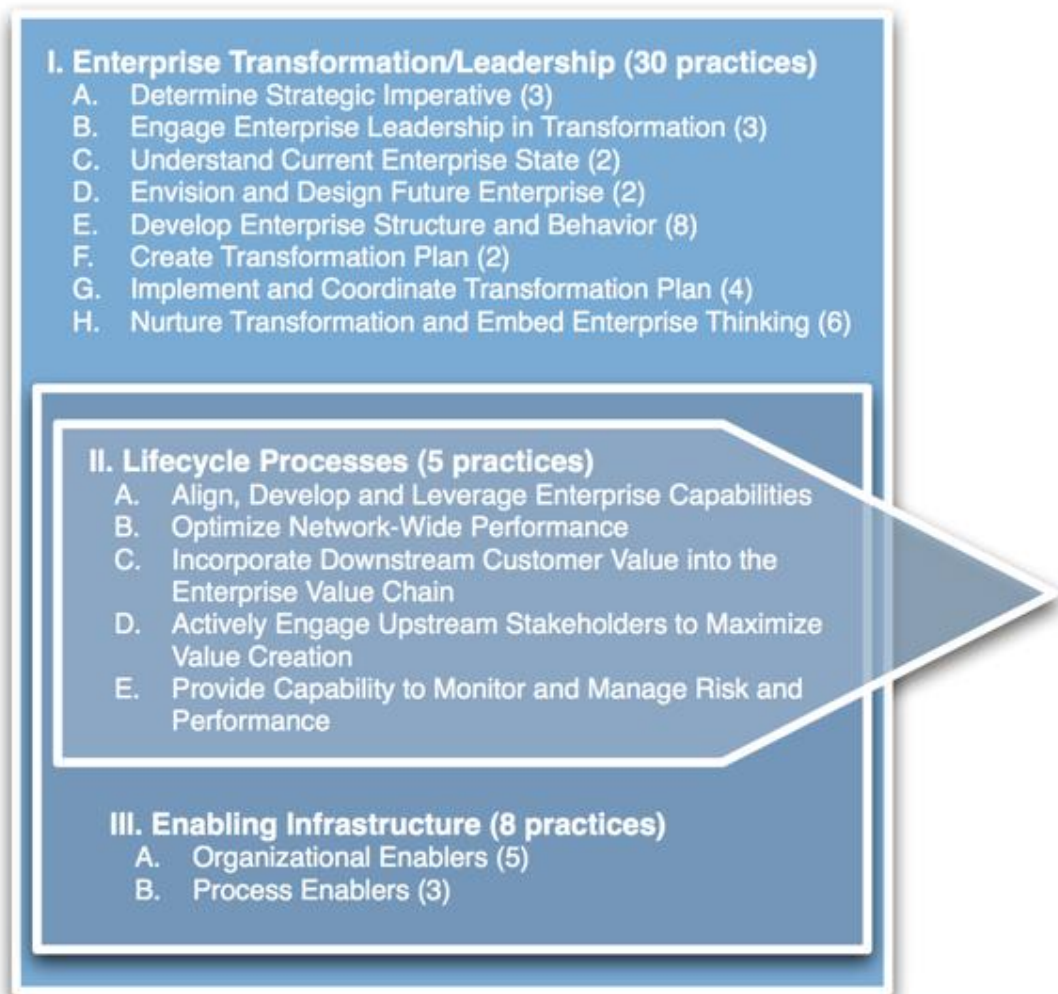
FIGURE 13: LESAT ORGANIZATION AND NUMBER OF PRACTICES (LEAN ADVANCEMENT INITIATIVE AT MIT)

Section I has practices that relate to the enterprise's leadership processes. Specifically, as the LESAT is a tool for driving change, these practices relate to the enterprise's transformation leadership processes. The practices found in this section are highly connected to the Roadmap. The LESAT links each group of practices in section I to a section in the Roadmap, as can be seen from the figures 13 and 11. For example section I.H, "nurture transformation and embed enterprise thinking" (LESAT Version 2.0, 2012, p. 19) is also the title of the last step in the Roadmap. This section can therefore be seen as an assessment of the organization's implementation of the intervention theory of the enterprise transformation paradigm. (Nightingale & Srinivasan, 2011)

114

Section II includes the practices relevant for the enterprise's lifecycle processes. The enterprise lifecycle processes are the core business processes, and what these processes are will depend on how the enterprise delivers value to its stakeholders. In section II, the five practices in this section (II.A-E in figure 13) should be assessed for every single such lifecycle process. This means that section II is repeated for every lifecycle process. For example, if one of the enterprise lifecycle processes is "product development", then product development should be assessed with regards to all five practices in section II.

The LESAT's focus on manufacturing-oriented enterprises is especially noticeable in section II. In the LESAT, a predefined set of lifecycle processes is included in the tool. This predefined set of lifecycle processes is very oriented towards manufacturing enterprises. The set of lifecycle processes in the LESAT Version 2.0 are as follows:

1. Program management
2. Requirements definition
3. Product development
4. Supply chain management
5. Production
6. Distribution and sales

Because all five practices in section II are assessed for each of these six lifecycle processes, section II has a different form than section I and III. As the five practices are repeated for each lifecycle process, extra lines are placed under these practices to allow a new score for each lifecycle process. On the next page is an extract from the LESAT, showing the practice II.A with some of the lifecycle processes underneath. (Nightingale & Srinivasan, 2011)

EP #	ENTERPRISE PRACTICES	Capability Levels				
		Level 1	*Level 2*	*Level 3*	*Level 4*	*Level 5*
II.A	**Align, Develop, and Leverage Enterprise Capabilities** *New opportunities build upon enterprise-enabled capabilities and lead to development of new ones*	Capabilities are understood only within individual enterprise elements. Improvements are *ad hoc* and focused on individual competencies. There is little apparent match between capabilities and enterprise strategy.	Potential opportunities arising from core capabilities have been recognized and acted upon within individual enterprise elements. Capabilities of individual enterprise elements are partially visible to the whole enterprise.	Capabilities of individual enterprise elements are understood and used across the enterprise. Enterprise strategy leverages existing capabilities.	Capabilities are integrated and enhanced across the enterprise with the focus on achieving an optimal combination of core competencies that are aligned with enterprise strategy to create competitive advantage.	Strategic plans and enterprise capabilities are dynamically aligned to ensure efficient creation of value for enterprise stakeholders over the entire product lifecycle.
II.A.1	**Program Management**	C D	C D	C D	C D	C D
	Indicators (Examples)	• The portfolio of programs is a balanced reflection of the full range of core enterprise capabilities. • The program selection and management process benefits from knowledge of the competitive environment to identify and exploit opportunities arising from the enhanced capabilities of the enterprise.				
	Evidence					
	Opportunities					
II.A.2	**Requirements Definition**	C D	C D	C D	C D	C D
	Indicators (Examples)	• Product and lifecycle requirements are defined in a clear and concise manner, based on needs of different stakeholders, the competitive environment, and capabilities existing across the enterprise.				
	Evidence					
	Opportunities					
II.A.3	**Product Development**	C D	C D	C D	C D	C D
	Indicators (Examples)	• The product development process realizes the enterprise strategy by delivering product designs that are timely and relevant. • The development process and product designs leverage distinctive enterprise capabilities.				
	Evidence					
	Opportunities					

FIGURE 14: PRACTICE II.A (LESAT VERSION 2.0, 2012)

Section III is the last section, with practices that relate to support processes. The support activities are enabling processes that provide resources to the organization. The rest of the organization serves as the internal customer for these services. LESAT emphasizes the importance of these activities. These activities can easily be overlooked, because they enable, rather than directly lead to, success. This section assesses the extent to which the enabling infrastructure processes support the leadership and lifecycle processes. The practices are divided into two subsections, organizational enablers and process enablers. Organizational enablers are meant to support units of an enterprise to become efficient at executing their functions. This includes best practices such as having an enterprise performance measurement system that supports enterprise transformation, and having information systems and tools that facilitate the flow of information and knowledge in the enterprise. Process enablers are practices that facilitate the enterprise transformation implementation. This includes practices such as standardizing processes, and using common tools and systems so that the enterprise is assured compatibility and reduces costs.

6.2.3 EDITED LIFECYCLE PROCESSES

As described in chapter 4.3.6.1, we identified the lifecycle processes for the Psychiatric Health Services Division at IHT, using the responsibilities for hospitals as defined in the Specialists Health Care Law as a basis. The first responsibility outlined in this law is patient treatment, the second is education of health professionals, the third is research and the fourth is the education of patients and relatives (Spesialisthelsetjenesteloven, 1999). In collaboration with the employees in the division, we identified the following lifecycle processes:

1. Patient Treatment
2. Education of Patients and Relatives
3. Education and Training of Health Personnel
4. Research
5. Management and Leadership

To exchange the manufacturing oriented lifecycle processes in the pdf document with our organization specific processes we edited the document to include the new lifecycle processes.

7 DATA PRESENTATION

This chapter is the result of a predominantly inductive process where we have observed as participants use the LESAT. We focused on their main concerns and the observations we made that we found to be of importance to the process and the participants' use of the tool. In chapter 8 the findings from this chapter, along with other relevant data, is used to conduct an evaluation of the tool with respect to the deduced criteria from chapter 3. This chapter has four main sections, and each section has a summary. The first four sections are structured as categorizations of our main observations during the pilot rounds. The last section, chapter 7.5 is categorizations of participant's reflections (meta-comments) about the process and the tool from the participants during and after the pilot rounds.

In the following presentation, quotes are cited directly from the transcribed audio files. To respect the participants in the pilot, they are cited anonymously. When their title, position or experience is relevant they are cited with their original semi-anonymous titles that correspond with the list of participants from chapter 4. The pilot participants are presented in table 6.

Title	Description
Department Head 1	Psychiatric nurse, leader of a psychiatric hospital department
Staff Employee 1	Educated as social worker, both clinical and management background from the case organization.
Staff Employee 2	Educated as ergonomist, long experience with both clinical and management work in the case organization
Staff Employee 4	Long experience with management in the case organization
Union Representative 1	Nurse, representative from the Nurse's Union

TABLE 5: PARTICIPANTS IN THE PILOT

For those quotes where title, position or experience is not relevant, or of very little relevance to the point made, we use the following randomized aliases for our five participants (see table 7). The same title is used for the same participant throughout the thesis. As mentioned in chapter 4, four participants were female and one was male, and we will refer to all of them with the gender-neutral term "he".

Alias
Participant 1
Participant 2
Participant 3
Participant 4
Participant 5

TABLE 6: RANDOMIZED PARTICIPANT ALIAS LIST

Further, in some cases participants wished to remain completely anonymous. We have created double aliases for these situations:

Double Alias	Situation
Participant E	One participant performed the individual pilot round when the tool was still in the original English language. Where we use this person's quotes that are related to the tool first being in English, he will be referred to as Participant E.
Participant X	This alias will be used when it would otherwise become clear who was who. If this alias is used several times in a dialogue, numbers will be given to the X's to separate the different people: X1, X2, etc.

TABLE 7: DOUBLE ALIAS LIST

7.1 CHALLENGES FOR THE PARTICIPANTS INTERPRETING THE LESAT

Most of the pilot duration was spent on discussions about how to interpret the practices in the tool. This means finding answers to what the content means, and what it means for the division. Using the LESAT requires the participants to make sense of the elements in the tool. One needs to make sense of the practices, capability level descriptions, statements, sentences, phrases and words in the tool. We will refer to such elements of the tool as "concepts introduced in the tool". The difficulties the participants had interpreting the tool was a major issue. They spent much time and effort on this process during the pilot rounds, and fronted this as a major concern during the pilot rounds and the follow-up interviews.

7.1.1 FOREIGN TONGUE (ENGLISH) MADE THE PARTICIPANTS UNABLE TO USE THE TOOL

The most basic interpretation was from English to Norwegian. Participant E used the original English version of the LESAT (all first practices in every subsection - I.A.1, I.B.1, etc.) during his facilitated individual pilot round. The participant's competence in the English language was representative for what can be expected of employees in the division. During the individual pilot round Participant E had severe trouble understanding the sentences in the tool, even when the individual words were known. The result of using the English tool was that we had to translate the practices word for word from English to Norwegian before the participant would determine his answers.

When a word is completely foreign, one is not able to make sense of it or use it in any sensible way. Several words in the LESAT were completely foreign to Participant E, for example "boundaries". When pushed to provide his interpretation of the practices, we discovered that even though he often grasped the main ideas, many words were found difficult to translate. The LESAT has a relatively advanced use of vocabulary, and words such as "cultivate" were completely unknown. Below are dialogues illustrating the participant's difficulties with the language.

Dialogue 1

> Participant E: [Reading practice I.A.1, capability level 2. In the LESAT it says, "Enterprise transformation is relegated to lower levels of the enterprise and application is fragmented"]
>
> OK, so this has something to do with lower levels? This is very.
>
> (Individual Pilot Round with English LESAT)

Dialogue 2

> Participant E: [Reading practice I.B.1, capability level 1. In the LESAT it says, " Lack of enterprise perspective leads to rigid boundaries that foster local optimization."]
>
> Yes the first level here. I can't see the connection [Norwegian: sammenhengen] at all.
>
> (Individual Pilot Round with English LESAT)

After Participant E's individual pilot round, we translated the parts of the LESAT we expected to have time to go through in the other pilot rounds (see chapter 4.3.6.3). The change in how Participant 4 experienced the tool was dramatic. He commented during several of the discussions throughout the group pilot round that the difference was so large that he wanted to completely disregard his earlier answers and thoughts.

Dialogue 3

> Participant E: My earlier answer is definitely wrong as I answered the English questionnaire, I am not on three! That is certain.
>
> (Discussion about practice I.B.1)

Dialogue 4

> Participant E: Translating it to Norwegian was definitely useful, because then I understand a little more of what I am reading! Even though it is. It is still a little.
>
> (Discussion about practice II.B.1)

Dialogue 5

Participant E: Um, I think I was on a completely different planet, so I don't think we should take my earlier assessment into account either. I am a lot better in Norwegian. (Laughs) I don't understand this either though! (Discussion about practice II.C.1)

7.1.2 LANGUAGE AND CONCEPTS IN THE LESAT WERE HARD TO EMPLOY.

Translating the tool to Norwegian did not solve the problem of people struggling to make sense of it. More than half of the amount of time in the group pilot round was spent on discussion concerning what the concepts introduced in the tool meant, and meant to in the division.

Dialogue 6

Participant 5: I think it is hard. Sometimes I have to. That logic. Understanding, what do they really mean? Until you make your own understanding of what you think is right. That's how it is! You kind of have to translate, what does this mean for us? (Individual pilot round, practice III.A.1)

We were able to observe the process of interpreting concepts, communicating the interpretations and keeping the interpreted concepts in mind while reading and discussing the practices, in various ways. Often the participants would voice their interpretations, explaining what they thought something meant. They would also very often read aloud to each other and fill in their own translated version of concepts instead of reading literally off the original:

Dialogue 7

Participant 5: [reading from I.A.1 title-box] If the "changes are integrated in the strategic planning process."
[LESAT: "Integrate *enterprise transformation* into strategic planning process"]
(Practice I.A.1)

Sometimes we had to infer their interpretations, typically when the participants seemed to all be thinking about the same thing without saying it aloud. When we could not tap into their interpretations we would ask about it, often after a discussion was finished so as not to disturb or influence them more than necessary.

7.1.2.1 THE LESAT LANGUAGE IS FOREIGN TO THE DIVISION

When expressing that making sense of the tool was hard, one of the terms used by the participants was "foreign". Even though the tool had been translated to the participants' native tongue, the language was still often experienced as foreign.

Dialogue 8

> Participant E: It is hard when it has been translated to Norwegian as well, because there are very many hard words here that we don't use.
>
> -
>
> Participant 1: Yes, that are not really translatable.
>
> -
>
> Participant E: And that is much easier to see now that it is translated, no wonder I didn't understand it before! I still don't understand it.
>
> -
>
> Participant 1: It is a technical language [Norwegian: Det er et fagspråk].
> (Discussion about practice III.A.1)

The division has a large, fully developed vocabulary that they often referred to as "our language" or "health language". We will refer to this set of vocabulary as the division language. These are the words and expressions that they are used to maneuvering in, and that are filled with meaning for them. Words are for example "patients", "level four leaders", "municipalities", "procedures" and more, words that will not invoke a reaction such as "what do you mean by that?" from co-workers. The division language includes very concrete words such as names that everyone know about, such as Innlandet Hospital Trust, and more abstract words that refer to a category of something such as "municipalities" or "partner instances". The vocabulary also includes highly abstract concepts such as "on green", which comes from Health South-East-wide risk analysis system and means that something is on the right track and does not need more attention. Other abstract concepts are "making money" which

means providing and documenting more outpatient appointments, and "user competence" which refers to the knowledge a patient has about their own situation and experiences.

Much of the vocabulary used in the LESAT is terms that the participants did not already have a meaning associated to. Describing the LESAT as foreign therefore makes sense. This required the participants to spend a large amount of time on filling the vocabulary used in the LESAT with meaning, as Participant 3 put it:

Dialogue 9

> Participant 3: It was hard. I had to think all the time, translate in my head, what is that and what is that? [...] You guys used "the enterprise" all the time, and when you say that, I have to the entire time think "oh yes, the division". "Yes this was the Hospital. This was that and this was that." But for you guys, you are used to maneuvering in it. I am not used to maneuvering a language like that at all. I think it was a very heavy language.
> (Follow-up interview)

There were many words in the LESAT that were unknown to the participants and therefore tough to interpret and use in their context, as seen in the dialogue above. Many of these words are a part of an academic vocabulary, especially academic fields such as economics and organizational studies. Many of the terms that the participants struggled with, such as "stakeholders", "flow", "intellectual capital" and "enterprise" took us by surprise. As Participant 3 said in the follow-up interview, to us these are very natural words filled with meaning that we have been trained to use with ease. However, this was not the case for the participants. In the same way as the division has its own language, so does the LESAT. When commenting on the LESAT's foreignness, the participants labeled the LESAT language as "market language", "technical language" and "industry language". A distinction can be made between vocabulary that was new and unknown, and vocabulary that was known but only carried meaning in contexts such as private market or industry.

An example of a term that was experienced as new and unknown was "enterprise". The enterprise must be defined before using the LESAT, and is the main subject of all the practices. In this pilot the enterprise was defined as the division, meaning that all practices were to be considered with respect to the division. However, the participants often had discussions where the "enterprise" ended up meaning IHT or Health South-East such as in the dialogue below. However, they always discovered after some time that they had gone off track and reminded each other that the enterprise was the division, but the word was clearly difficult to use.

Dialogue 10

> Participant 1: I am thinking that it is hard to get to level four with the way the organization is structured today.
>
> -
>
> Participant 5: Yes but with enterprise-level, then I don't mean the entire hospital, I mean the division.
>
> -
>
> Participant 1: Oh! You are thinking the division?
>
> -
>
> Participant 5: Yes, "demonstrate division thinking through their practice" [reading aloud from practice I.B.1 and filling in for "enterprise" with his own words]
>
> -
>
> Participant 1: Yes, that is a little less hairy goal then.
> (Practice I.B.1)

Other terms were hard because they were different from the division language, but the participants were familiar with the words when they were used in other contexts. There were many words that they only understood in an industrial or private business context. These were terms and phrases such as "competitive advantage" and "customer's business model". "Upstream" and "downstream stakeholders" were concepts that were both unknown to the participants, and easier for them to interpret in an industrial context. To illustrate how foreign and strange these concepts were to

the participants, a longer dialogue is presented from when the participants used the translated (to Norwegian) glossary to answer practice II.B.1 II.C.1 and II.D.1.

Dialogue 11

Participant 5: [Reading from II.B.1 title-box] "Optimize extended enterprise performance." Hm.

-

Participant 3: What is this about "incorporate downstream customer value into the enterprise value chain"?

-

Participant 5: Where does that say?

-

Participant 3: II.C, the title.

-

Participant 3: What is downstream? [Reads from the LESAT glossary] "Upstream stakeholder" (laughs), now I am going to start using a new language at the next meeting!

-

Participant 5: (Laughing) There is too much downstream here!

-

Participant 3: Haven't you heard about upstream stakeholder!? That's a totally normal thing, I'll tell them.

-

Participant 5: (Laughing) You are behaving like a downstream stakeholder! (...) [Reading from glossary] "Someone with a role later in the lifecycle", that could be the municipality. If we think in terms of services. But they are on both sides. In terms of patient treatment, the municipalities are before and after in a treatment.

-

Participant 2: It says "the product flows from upstream stakeholder to downstream end user", I can understand this in the industry. If you produce a phone then. But I can't really convert it into what we are doing, that is patient treatment. But I am sure it is just something I don't understand.

-

Participant 3: I think this one was really hard. And then it could all end up being kind of Lotto what we end up answering. (...)

-

Participant 3: If we are going to use this tool in the division, you guys should redo the entire language. It is so obvious that it is used in completely different places, that it is a system from outside that is pressed down, used, here. You have to use the language that is used here. This becomes horribly market oriented. I mean we discuss, using "production" in the division, I mean as a patient, are you a part of a production? Yes but.
(Practice II.B.1, II.C.1)

7.1.2.2 COMPLEX SENTENCES INCREASE DIFFICULTY

In addition to being a vocabulary that was foreign, many nuances are packed into each sentence in the LESAT. This increased the difficulty of interpreting the contents.

Dialogue 12

Participant 5: This tool really has to be converted so that it becomes understandable for every day use.

-

Participant 1: This is an industrial language. (...)

-

Participant 5: But I think it's really hard to understand! When it says, "Lack of enterprise perspective leads to rigid boundaries that foster local optimization." I have to really pay attention! It's a really heavy sentence structure.
(Practice I.B.1)

The language is also somewhat vivid and figurative. In the formulation above, words like "rigid boundaries" and "foster" can provide vivid images in the mind, but not give an indication of what to look for. "Rigid" for example may bring to mind non-flexible materials, perhaps a strong wall, without giving further indication of what makes a boundary rigid and thereby what to look for in the organization.

7.1.2.3 FOREIGN ABSTRACT CONCEPTS BECAME AMBIGUOUS

The LESAT is made with an aim to be applicable to many different organizations, and many the concepts are therefore on an abstract level. These words need to be interpreted to the organization that uses it, meaning that organization-specific definitions of the abstract concepts need to be found. The users have to figure out what the concepts can correspond to in their organization, and if several choices arise they have to decide on which ones to use. Before the pilot, we helped the participants define division-specific definitions for two such concepts, "lifecycle processes" and "stakeholders" (see chapter 4.3.6.1). The participants had to determine division-specific definitions for the other abstract concepts introduced in the tool during the pilot rounds. It became clear that the participants' ability to use abstract concepts that were in the division language was much better than the abstract concepts that were foreign to them. When asked about who "partner instances" could be, we were given a list of actors such as municipalities, other hospitals, child protective services and so on. For most of the abstract concepts in the LESAT however, no such abstract-to-specific translation was readily accessible to the participants. Even when using the LESAT glossary, which is supposed to give directions for what to look for and how to decide what goes into a concept, the range and scope of what the concepts could mean in the division was seldom clear to the participants.

Dialogue 13

> Participant 4: It is very hard to put this into a concrete context. There are so many examples for what this could be. It could be on different levels, if you are on the department level then it's about one thing, if you are on the hospital level then it is something completely different.
> (After a long discussion about practice III.A.1)

Unlike abstract concepts from the division language, the foreign abstract concepts in the LESAT did not become one undisputed category of specific instances in the division. In other words, very many of the abstract concepts in the LESAT became ambiguous when attempts were made to use them in the division. The large range of what abstract concepts in the LESAT could represent in the division was experienced

as frustrating and the participants often struggled to determine division-specific definitions to use as a basis for discussion.

An example of a concept for which the participants generated several potential division-specific definitions was "upstream stakeholders". Over the course of the discussion the participants proposed the municipalities, the professional groups in the division, the entities within the division that has handled a patient before another entity, the referrer (Norwegian: henviser), and the patients family and environment. All of these proposals are reasonable division-specific definitions of the division's upstream stakeholders, depending on your view.

7.2 Participants' Interpretation Process

The process of making sense of the concepts in the LESAT was characterized by a process where participants only focused on and employed parts of the content, considered small subsets of the potential division-specific definitions of concepts, and focused their attention on the things that sparked recognition.

7.2.1 Participants had Selective Focus

One of the most striking features of the participant's sense making process was how attention was focused. The participants were asked to go through the first practice in the sub-sections. This means that the available information about the practices included: a section header if it was the first sub-section in the section, a sub-section header with a list of "diagnostic questions" to help users get a feeling of what the sub-section was about, a practice title-box, and five capability levels with descriptions that could each include several sentences (see figure 15, the start of subsection I.A and the first practice I.A.1). However, focus was rarely cast outside of the capability level descriptions.

7.2.1.1 Participants Focused On Capability Level Descriptions
Only once did the participants comment on content from outside the capability level or title-box descriptions when discussing how to make sense of a practice. Text in the title-box was only mentioned four times. Throughout the individual pilot rounds and the group pilot round the participants directed their focus almost exclusively towards

the capability levels. Formulations in the capability level descriptions were often read aloud and referred to during the discussions.

SECTION I: ENTERPRISE TRANSFORMATION/LEADERSHIP

Definition: Develop, deploy, and manage enterprise transformation plans throughout the organization, leading to: (1) long-term sustainability, (2) acquiring competitive advantage, and (3) satisfaction of stakeholders along with a continuous improvement in all three outcomes.

I.A. Determine Strategic Imperative – The decision to pursue an enterprise transformation is strategic in nature and affects all organizational practices and processes in the enterprise. The enterprise is continually striving to eliminate waste and enhance relationships with all stakeholders.

Diagnostic Questions	
	• Are enterprise leaders familiar with the dramatic increases in competitiveness that many companies have realized as a result of transforming?
	• Are enterprise leaders fully aware of the potential opportunities (i.e., growth, profitability, and market penetration) that can be realized within their own organization as a result of transforming?
	• Has a suitable strategy been identified to use resources freed up by improvements?
	• Does "stakeholder value" strongly influence the strategic direction?
	• Has full leverage of the extended enterprise stakeholders been incorporated into the strategic plan?
	• Has a common vision been communicated throughout the enterprise and within the extended enterprise?
	• Has a compelling case been developed for transformation?

		Capability Levels			
EP #	**ENTERPRISE PRACTICES**	*Level 1*	*Level 2*	*Level 3*	*Level 4*	*Level 5*
I.A.1	**Integrate Enterprise Transformation into Strategic Planning Process** *Transformation is a key enabler for achieving strategic objectives*	Enterprise transformation efforts are ad hoc.	Enterprise transformation is relegated to lower levels of the enterprise and application is fragmented.	Enterprise transformation plans are formulated, but not integrated into the strategic plan.	Coordination and synergistic relationship exists between transformation and strategic planning.	Strategic plans leverage the results of transformation improvements to achieve enterprise objectives.
		C D	C D	C D	C D	C D
Indicators (Examples)		• Enterprise transformation implementation is included explicitly in the enterprise strategic plan. • Strategic planning makes allowance for anticipated gains from transformation improvements.				

FIGURE 15: PRACTICE I.A.1 WITH SURROUNDING INFORMATION

131

The participants explained the exclusive focus on capability level descriptions to be because the capability level descriptions are easier to understand than the surrounding content. Participant 3 neatly summed up the participants' comments about the capability level descriptions as opposed to the surrounding content:

Dialogue 14

> Participant 3: I felt I was reading things five times over, and still I did not understand how to think. I understood it when I could guide myself through the different levels though, then I understood where it was going in a way. But in the title and the other stuff, I thought it was hard to understand what is being asked about here.
>
> (During the follow-up interview when asked about how easy it was understand the tool)

This implies that the meaning of the surrounding content was very difficult to grasp, and that they focused their attention towards what was easier to understand. The capability level descriptions are also at a lower abstraction level than the surrounding content. The capability level descriptions include specific things you can look for in your organization. It does not require an overall understanding of the entire practice to consider each capability level individually and try to find a match for the description in your organization. Considering specific statements in the capability level descriptions therefore requires holding fewer concepts in mind than understanding the entire practice, which could explain the focus on capability levels.

7.2.1.2 PARTICIPANTS SELECTIVELY FOCUSED ON SUBSETS OF POTENTIAL DIVISION-SPECIFIC DEFINITIONS OF CONCEPTS

Selective focus was also evident in discussions about division-specific definitions of concepts. During the discussion about upstream stakeholders for example (see chapter 7.1.2.3), many division-specific definitions of upstream stakeholders were identified, such as municipalities, professional groups, the patients' family and more. However, only one of these division-specific definitions, the municipalities, was used by the participants to determine capability levels. Of many ways to see a concept, or many division-specific definitions, the participants usually found or chose a subset to focus on. This reduced the complexity of what was to be evaluated, and also the potential

for what could be explored. Given the heavy task of maneuvering in the new, foreign and abstract language of the LESAT, this seems like a very natural reaction.

7.2.1.3 PARTICIPANTS' INTERPRETATIONS BECAME DECOUPLED FROM CONTENT OF THE LESAT

The participants employed a selective focus all the way down into the capability level descriptions, where within one capability level description only parts of it would be given attention. Throughout the pilot rounds the participants often directed attention towards individual sentences, or even parts of sentences, while leaving out surrounding content.

Dialogue 15

> Participant 5: I am thinking that we are at level two. That core capability is recognized, and acted upon individual elements of the division.
> [LESAT level 2: "Potential opportunities arising from core capabilities have been recognized and acted upon within individual enterprise elements. Capabilities of individual enterprise elements are partially visible to the whole enterprise."]
> (Practice II.A.1)

In the dialogue above Participant 5 reduces the nuances in the capability level description by reading, "potential opportunities arising from core capabilities" as "core capabilities" (also note that he fills in "division" for "enterprise").

Further, the participants rarely focused on the holistic and abstract meaning of the practices. Each practice has an overall meaning with a very high abstraction level. Breaking them down into capability level descriptions reduces the abstraction level, but the overall meanings, communicated through the combination of all capability level descriptions, titles and surrounding content, are very abstract. Discussing these overall practice meanings and translating them to become division-specific was a task the participants found very hard, often did not accomplish, and gave surprisingly little focus. The following is an excerpt from a dialogue where this was actually attempted and where the participant seemed to capture the overall meaning of the practice well.

Dialogue 16

        Participant 1: But, is this here more, like, theoretical? Where one thinks that
the organization should continuously change their strategic plans? And when
you initiate improvement processes, then you have to make sure that they
don't live their own life and that results from them, in general, are included in
strategic plans. At a very general level, is that kind of what this is about? (...)
And that it shouldn't be that someone sits on the top and makes a strategy,
while others are pursuing improvement work that is never reflected in a
strategy.
        (Practice I.A.1)

The above dialogue was one of very few attempts at explaining to each other what
they thought the overall practice was about. Even in this dialogue, the other
participants did not give much attention to this explanation, but rather became silent
before the discussion continued with focus on the capability level descriptions without
much influence from Participant 1's comment.

For most of the practices, the understanding of the overall meaning was fragmented.
The interpretations were often pieced together from parts of the capability level
descriptions, and these parts were often taken somewhat out of their context. The
foreign and complex language, along with all the abstract concepts that needed to be
defined and then remembered and employed, made the overall meanings of the
practices very hard for the participants to interpret on both an abstract and division-
specific level.

Despite the participants' serious efforts to make sense of the tool, this was often not
achieved. Nuances, context and overall coherence in the descriptions in the LESAT
were lost as participants selectively focused on capability level descriptions, subsets
of the capability level descriptions or even sentences. We will further refer to such
resulting interpretations as being decoupled from the contents of the tool. Selective
focus also reduced the potential scope of the division-specific implications the
practice could have by the participants focusing on a small subset of potential
division-specific definitions of concepts. The participants were very aware of being

on shaky interpretive ground and that this had an effect on the way they answered the practices throughout the pilot rounds.

Dialogue 17

> Participant 3: It does something to the answers too, as you have to use so much, "what is really meant here?" You know, the point of having titles and headings in such a tool is to. But to get that result, then the language has to be so clear and evident [Norwegian: klart og tydelig] that you don't sit there wondering about, "what is really asked about here?" Because I think that does something to the answer.
> (Follow-up interview)

Selective focus was one of the central mechanisms we observed that very often resulted in a decoupling of the interpretations from the content of the tool. We observed several interpretations, discussions and conclusions during the pilot rounds that were very far from what we considered relevant to the practice they discussed.

7.2.2 PARTICIPANTS WERE GUIDED BY WHAT THEY RECOGNIZED

Our first impression of the discussions was that the interpretation process was a highly unstructured one. It was not a discussion that started with reading the entire practice, discussing what it meant in overall terms, then what the concepts introduced in that practice meant in the division and then discussing the capability levels. Instead, the participants picked up what we at first experienced as almost random pieces of a practice and build their interpretations from there. We realized however that the concepts, sentences and other pieces that the participants picked up were in fact not random at all, but were guided by what they recognized. This implies words, sentences, phrases and ideas that they could relate to and that gave them associations. Some parts of the practice descriptions sparked recognition and were picked up and expended on, while other parts were ignored or commented on with a humoristic or thoughtful tone.

The discussions often quickly became about what parts of the division could be division-specific definitions of concepts in the practice. When they had decided what

this element should be (usually only one division-specific definition, see chapter 7.2.1.2), they would stop the interpretation process and start matching the element to the pieces of the capability level descriptions that seemed to fit. In the following example the participants discuss practice is III.B.1, which is about process standardization, and how process standardization and continuous revision of the standards are employed in the enterprise.

Dialogue 18

> [Reading in silence]
>
> Participant 5: I think we are on level three.
>
> -
>
> Participant 4: Selected processes. We have that. That doesn't mean all of them.
> [LESAT Level 3: "*Selected processes* are standardized across the enterprise."]
>
> -
>
> [...]
>
> -
>
> Participant 1: What such processes do we have?
>
> -
>
> Participant 5: We have, the Ledelsens Gjennomgåelse. Maybe? [A twice-a-year evaluation and reporting procedure that is performed in every unit and department and sent to IHT]
>
> -
>
> Participant 4: Yes, that is what we report with. We review it continually, I mean it is twice a year. We measure ourselves on what we have done earlier and. And it is done across the division too for that matter. Because it is done in all divisions in all in the IHT. So I guess Level 4.
> [LESAT Level 4, what is given focus is in *italics*: "Process standardization and reuse is consistently *employed across the enterprise*. Process standards are *continually reviewed* to ensure highest performance."]

The division has standardized processes, which they call "procedures" in the division's language, for almost everything that is done. We therefore expected the participants to talk about their procedures during this practice. However, they ended

up talking about one single procedure, the leadership reporting system, and then mapping the capability levels to this one standardized procedure.

We found this selective focus on what first sparked recognition surprising at first. We realized after a while however that we also read the practices in this way, but that most of the content in the LESAT did spark recognition with us. This difference in our understanding of the practices and the participants' is not surprising, given that we have studied the theory behind the LESAT and the division for the last year. The participants have a very deep understanding of the division, but the concepts introduced in the LESAT were very new, and much fewer pieces of the LESAT's content sparked recognition for them. After the dialogue above was concluded, we intervened and proposed an alternative interpretation, asking what they thought about interpreting "standardized processes" as "procedures". This sparked a very effective discussion where sentences and capability levels were read and employed in their entirety now that sentences like "process standards are continually reviewed to ensure highest performance" fit to the concept they were discussing.

7.2.2.1 SELECTIVE FOCUS GUIDED INTERPRETATIONS TOWARDS FAMILIAR TOPICS
This guiding force from recognition was evident when it came the participants' understanding of the practices in their entirety. For example, practice II.A.1 was discussed and used quite easily by the participants, and it seemed that an understanding was established for the meaning of the entire practice that was relatively in line with the theory behind the tool. However, the scope of what the participants considered relevant to look at in the division was limited. The practice is about core competences and the enterprise's handling of these. However, the participants quickly focused - and stayed focused - on the formal competence/education of employees and user competence (division language for what patients know about their own situation and can contribute with in planning their treatment). Core competence is a rather abstract and business-academic term, but competence is a word that carries strong associations to important issues in the division. The formal competence of employees and user competence are hot topics in the division at the moment. It is therefore not surprising that it was these division-specific concepts that came into focus when the participants were presented with this practice containing the word "competence".

137

For most of the practices, this process of picking up recognizable pieces and making these the basis for discussion led to the participants mostly discussing established issues and topics in the division that the they were all already aware of and had discussed before. This is illustrated by the answer Department Head 1 gave when asked if they discussed anything that was new to him in the pilot:

Dialogue 19

> Department Head 1: No, I don't think so. Because this here, discussing where we are and where we are going, we have to do that all the time when making goal documents and the like. But it could be that the questions, or topics, or whatever they were called, oh yes practices, were questions that we are not used to asking ourselves. But in the discussion, well we discussed those things we usually discuss.

7.2.2.2 PARTICIPANTS RECOGNIZED AND FOCUSED ON DIFFERENT THINGS

What sparked recognition differed between the participants. The participants had different experience, views, and positions in the division, and what they saw in the practice descriptions was very different. What was given attention to in the LESAT, what division-specific definitions were suggested for the concepts and what examples from the division was brought forth, and what they argued that the descriptions in the LESAT implied for the division was all different between the participants. The participants were also highly aware of this, especially the fact that they all sat in different positions in the division and therefore knew about different things.

Dialogue 20

> Staff Employee 4: Well, we are on different levels here. Department Head 1 is in a department, so he can answer for the department and more out into the unit leader level. And we are sitting in the staff and see things in a different way. And you are a little all over, Union Representative 1!
> (Before practice I.A.1)

Not only did the participants know about different things, they also had different personal interests and views about the way things are in the division and why. In the

individual round where two of the participants were allowed to discuss but did not have to come to an agreement on capability level, this was nicely illustrated in the following dialogue:

Dialogue 21

> Staff Employee 2: I answered level three here. That we do have, that there are many systems now, computer information systems but also all the reporting that we get, which is specific in relation to how we are doing. [LESAT practice I.E.1 level 3: "Systems and policies have been defined, rationalized, and standardized to support the enterprise vision."]
>
> -
>
> Union Representative 1: Yes, but then there is also, I am sitting in a project now about resource management, and what tools they have for managing finances. It is very vague. For example with respect to increasing outpatient treatments, what is to be counted, what should we measure? There are so many discussions about what is what. Who is included, who is excluded, what counts and what doesn't. We have a lot of systems yes, but that some of them are in conflict with each other now at the start? I sure think so. [LESAT practice I.E.1 level 1: "*Systems and policies are in conflict with each other* and with desired enterprise behaviors."]
>
> -
>
> Staff Employee 2: Yes then you have to put down level 1. And I give a higher score than you do, because I am sitting in the staff. And we just get the numbers, and think: oh, that's how it looks, some big changes have happened over there, over there they have adjusted themselves, and you can see it month for month. But then you are saying now that the understanding of what the parameters imply differs.

7.2.2.3 PARTICIPANTS DISTORTED CONTENT OF THE LESAT TO BECOME RECOGNIZABLE

The guiding force of the participant's efforts to find recognizable parts of the LESAT seemed to lead to the participants sometimes reading what they expected to read. The participants often twisted formulations and meanings in the LESAT to become something more easily recognizable. Sometimes words were simply misunderstood, but often it became evident that words, sentences, capability levels and even the meaning of practices acquired a new meaning based on what the participants expected it to be about. This even happened during word-for-word reading of the tool, often observed as the participants read aloud to each other. In the following dialogue the participants are discussing practice I.A.1, which overall is about *plans* regarding *transformation* being considered in the enterprise's *strategic plans*. Going into the dialogue below, the participants had interpreted the practice to be about the degree of *implementation* of transformation initiatives. This was what they expected to see, and in fact did see, when reading the capability level descriptions out loud:

Dialogue 22

[Talking about "the Project", a large project from 2011-2013 that resulted in "Master Plans". The "Master Plans" are being followed today, as they are still implementing the change initiatives determined in the Project.]

Participant 4: We have Master Plans that we, considering where we were in the Project, where we are now. And then I am thinking level three. That we have not come all the way out, integrated all the way out in the departments with all that. I don't think all level four leaders and employees are familiar with all the Master Plans. So it is not integrated well enough I think. [LESAT practice I.A.1 level 3: "Enterprise transformation plans are formulated, but not *integrated into the strategic plan*."]

-

(...)

-

Participant 4: I think it's that sentence about it not being integrated everywhere, and with everywhere I am thinking the line employees, that makes me land on level three.

In this example, the participant is reading from the LESAT and filling in with his own words, and "integrated into the *strategic plan*" is swopped with "integrated into the *departments*". At this point the participants thought the practice was about implementation of changes, and not integrating change- and strategy-plans. Integrating changes into the departments fit much better with the interpretation the participants were expecting to see than "integrated into strategic plan". As this previous example illustrates, the text the participants perceived to read became decoupled from the actual text written in the tool.

7.2.3 SUMMARY

The participants were unable to comprehend the content of the original English version of the LESAT because words and phrases were foreign and unknown. The English language was however not the only language barrier to be crossed. The LESAT language is very different to the division language, and the LESAT language was experienced as very foreign, complex and hard to employ to the participants. They struggled to identify division-specific definitions of concepts introduced in the LESAT and many abstract foreign concepts became ambiguous to the participants.

The participants' process of making sense of the practices was characterized by a selective focus guided by what sparked recognition to the participants. Attention was especially focused on the capability level descriptions, and often on sentences and phrases taken out of their context. The understanding of the overall abstract meaning of the practices was rarely given attention. The overall meaning of the practices with respect to the division was fragmented, and only a small subset of potential division-specific definitions of the concepts in the practices were considered when answering the practices. Instead of reading the tool and discussing the overall meaning of the practice until the abstract meaning and a division specific interpretation became clear to the participants, they often picked up the pieces of a practice that were recognizable. These pieces would then guide the discussion about the division, which often ended up being about established issues and topics that the participants were used to discussing. The guiding force of what the participants recognized made the

141

topics of discussions highly dependent on the participants and sometimes led the participants to distort the content by reading what they expected to read.

7.3 WORKING AS A GROUP TO BUILD PREVAILING UNDERSTANDINGS

The participants worked as a group to collectively interpret the practices and come to a decision about the current and desired state of the division with respect to the practices.

7.3.1 PARTICIPANTS SHARED VIEWS AND BUILT ON EACH OTHERS' CONTRIBUTIONS

During conversation, participants shared their view on interpretations of a practice or concept, and the state of the division with respect to these interpretations. They often made their understanding known through comments like, "this is what I am thinking", to form the basis for their contributions. Interactions with the tool played a central role in this process. The participants had frequent and sporadic interactions with the tool. The interactions included reading aloud a sentence, a capability level description or a word, referring to a capability level score or mentioning a task implied in the tool to move the discussion along. Below is a representative process around a practice they found relatively easy to discuss.

Dialogue 23

>Participant 2: I think this, I mean, we have a completely different economic climate now than we did before. It makes people very focused on their own unit. We are very afraid of ending up with red numbers on unit leader level, they are very focused on money. Which is good, but then we also become much less generous with each other.
>
>-
>
>Participant 1: I agree with Participant 2, this is about that. And I also think it is about the size of the division.
>
>-
>
>[...]
>
>-
>
>Participant 5: I am thinking like what you are saying, I remember a time when we had more money, and things were less tight, I remember discussions in the

leadership group about "ok, you can take those and we can take some of yours".

(Practice I.B.1)

When a participant stated his view or suggestion, other participants would agree to this statement and try to build on it; or criticize it and try to change it. In this way, the participants together seemed to determine some kind of a shared understanding that they could build on and change. Together the participants determined a form of agreement, or a shared acceptance of "this is what we are talking about" or "this is how we view the organization". They did not at all times understand or agree with each other, but they always worked to build on each others contributions and reach some kind of an accepted understanding. We term this accepted, common understanding the prevailing understanding. Prevailing understandings could exist for several issues regarding the same practice, such as how to interpret the practice and capability level descriptions to the division, and how to view the relevant aspects of the division and their relationship to the interpreted practice.

Dialogue 24

> [Arguing that level 5 in practice II.B.1 is too hard to reach for the division] Participant 1: We need to use some tempo, no, time to be able to change. We can't, it is kind of restricted how flexible we can be, and how responsive we can be to "changes in the market place". I see "changes in the market place" as, that has to be changes in the patient population, new illnesses coming in that we need to [snaps his fingers]. So to a certain extent we need to be able to do that too, but the organization can't change very fast I think.

The prevailing understanding became something that could be referred to and used in future contributions to the discussion. For example, it could be called "what Participant E argued for," or "what we talked about earlier". Once a prevailing understanding was determined it became the valid base for arguments when deciding the desired or current capability level of the division for a practice. On several occasions, the participants would refer to something that had been agreed on earlier. They would make it known that they chose to see it in the way that they had agreed

143

on, and then let it form a basis for an argument for determining a current or desired capability level for the division.

Dialogue 25

> Participant 5: And now I was thinking about what we were discussing earlier, about including the patient. Then you get a completely different feedback.
>
> -
>
> Participant 1: User competence, I guess we should be at level five then.
> (Practice II.C.1)

7.3.1.1 INTERACTION WITH THE LESAT KICK-STARTED DISCUSSION

The participants often interacted with the tool to change the focus of discussion, move the process along or kick-start conversation when discussion stagnated. If the conversation had stagnated, any referrals to the tool could work as a kick-starter of further conversation. In the first example below Participant 5 prompts for a decision on what capability level to set, changing the subject and moving the process along.

Dialogue 26

> Participant 2: Nobody wants to be presented with fifteen different medication options.
>
> -
>
> Participant 1: Some people might. [Silence for half a minute]
>
> -
>
> Participant 5: Where should we land?
>
> -
>
> Participant 2: I think, I am a fan of five. [Conversation starts up again]
> (Practice II.C.1)

Here, we see how prompting for a capability level decision restarts the discussion. Referral to a capability level or reading aloud from some sentence that caught a participant's attention could also kick-start the conversation even though no other explanation or argument was provided. Even if it was just a way of filling the silence, it seemed to often spark recognition for the other participants and in a way rebooted the conversation. However, even though the participants frequently used interaction

with the tool to prompt conversation, it should be noted that we as facilitators had to prompt for capability levels on several occasions. In this sense we used the tool ourselves to help facilitate the process. This would either be because they were drifting off-topic and did not redirect themselves to setting a capability level, or because they spent too much time on one practice and we deemed it appropriate to move onto the next practice. For almost every practice, we were the ones to ask the participants to go to the next practice. After having set current and desired capability levels, the participants often kept talking about the topic they were on, and even if the conversation went quiet it seemed to become our role to move them to the next practice.

7.3.1.2 PARTICIPANTS COMMUNICATED THEIR POSITION WITH CAPABILITY LEVEL PREFERENCES

Participants' positions often became associated to capability level preferences, and the participants often shared their opinion about which capability level was appropriate as a way of communicating their position. Sharing capability level preferences was sometimes, and sometimes not, accompanied with arguments for the position. Sometimes participants would just share their preference, focusing attention on that capability level. The other participants would then read the mentioned capability level, and argue the position or agree. Other times participants would check their understanding of someone else's comments by inferring the capability level they thought the other person wanted.

Dialogue 27

> [Discussing level of resource sharing and cooperation]
> Department Head 1: I think we have come a long way, and I think it has something to do with us becoming one department. That when Early Intervention has a neuropsychologist, and another acute post needs one and uses the one at Early Intervention, I think we have more of that now.
> -
> Staff Employee 2: But that is because, because your department has changed and become one.
> -
> Department Head 1: Yes

145

-

Staff Employee 2: But what about using people from other departments?

-

Department Head 1: No, I don't think we are good enough there. I think we are still struggling with things in the department, and we don't cross the threshold to the next, and ask about.

-

Staff Employee 2: Do you think we are at level two? That we are there now? Or?

-

Department Head 1: Yes, now yes, I think that may be right.

(Practice II.A.1)

In this way, the capability level descriptions, and sometimes just the capability level number, became communicative tools the participants used to communicate their positions, focus the group's attention and check their understanding of others' opinions.

7.3.1.3 DESCRIPTIONS OF PRACTICES GUIDED DISCUSSION BUT DID NOT PREVENT DECOUPLING OF PREVAILING UNDERSTANDING AND CONTENT OF THE TOOL

The content, or text, in the LESAT guided the content of the discussions. Before every new practice was discussed, the participants took time to read through it first. Starting on a new practice sparked completely new conversations based on the content of the tool. Throughout the discussions, participants would read from parts of the practice, point out different elements, read aloud (sometimes in their own words) and share what capability levels they thought were appropriate. This initial reading and continuous interaction with the tool kept the discussions focused on the topics presented in the LESAT.

Pointing to elements in a practice or reading aloud was a common way for the participants to contribute to the interpretation process. This would focus the participants' attention on an aspect of the practice, and at the same time serve as an argument for or against a capability level depending on if it sounded right for the division.

Dialogue 28

> Participant 5: Yes. No. Well, with upstream stakeholders I am thinking not internally, but maybe a municipality for example. Who can be both before and after. But, "the divisions' elements informally incorporate the municipality's knowledge and capability. Communication lines are established to allow exchange of relevant." That's how I was thinking.
> [LESAT practice II.D.1 level 3: " Enterprise elements informally incorporate upstream stakeholders' knowledge and capabilities. Communication lines are established to allow exchange of relevant information."]
> (Practice II.D.1)

As shown in chapter 7.1 however, the contents in the LESAT dictated the content of the discussions to a surprisingly small extent. The participants' selective and recognition-dependent focus often led to the contents of the tool becoming distorted. The prevailing understanding developed for a practice could end up surprisingly decoupled from the actual text in the LESAT. When this happened, interaction with the tool did not necessarily steer the discussion back to be in line with the text in front of them. As the prevailing understanding strayed from the descriptions in the tool, the descriptions were twisted and read in a way that fit with the prevailing understanding rather than the actual text. This means that even though the participants interacted with and moved down the tool, the prevailing understandings could be highly decoupled from the content in the LESAT.

7.3.2 Having a Prevailing Understanding did Not Imply Consensus

There was not always one clear prevailing understanding for every practice that was collectively built. In other words, there was not always one prevailing understanding that was in the end agreed upon by all participants, and which was then used to determine capability levels scores for that practice. Rather, the group could come to a decision about what understanding was the prevailing one even if one or more participants had a significantly different view. The participants had different roles in the division, different understandings of the organization and different information. All this led to different perspectives on the interpretations of the practices, what they

thought was relevant in the organization, what they thought was important, and what they thought was good for the division. This was often, but not always, reflected in different opinions about what capability levels should be put down.

We could not observe one rule for how understandings were built and how they became the prevailing ones. Sometimes everyone seemed to initially share the same understanding of a practice and got the same associations to the divisions from it. It then became evident by the way they spoke that a prevailing understanding was established almost immediately. The participants often built on each others' proposals, building a prevailing understanding by including more information, making others' proposals more clear and finding examples from the division that fit. At other times the participants would argue for different views long and hard, and what became the accepted basis for further discussion and determining a capability level was determined by which view the majority of the participants ended up adhering to. The prevailing understanding could also be set or changed by an unopposed statement, or sometimes a statement would be seemingly unopposed but in reality ignored as the other participants would keep using an earlier version. These processes of building on each other's understandings, and debates and negotiations took place for many related topics. The participants could find a prevailing understanding for how to interpret the practice to the organization through building on each other's contributions, but negotiate when establishing what aspects of the division should be given the most weight during the assessment.

So, although one understanding typically became the prevailing one and the one used to establish a capability level, this did not mean that all the participants were in agreement. Several opposing understandings could exist for one practice. Following is an example of a process where participants argue for different interpretations of practice I.A.1. As mentioned earlier, during this practice the participants were torn between seeing it either as a question of the degree of successful implementation of changes (Participant 5 has this interpretation), or if the change- and strategy-plans are coordinated with each other (Participant 1 has this interpretation).

Dialogue 29

>Participant 1: And that it shouldn't be that someone sits on the top and makes a strategy, while others are pursuing improvement work that is never reflected in a strategy.
>
>-
>
>Participant 5: I was thinking that, like in the Project, it was deeply rooted in the entire organization. And I think that when you are able to do that, involving people and have them participate all the way from the start, then changes start being implemented very early.
>
>-
>
>Participant 1: Yes. Well. What level do we land on?
>
>(Practice I.A.1)

In the dialogue above you can see how the earlier prevailing understanding, the implementation interpretation, is unchanged by the new suggestion from Participant 1 as the other participants keep using the previous understanding. For this practice the participants never came to a point where they all acknowledged each others' view, integrated them or agreed on one, but rather moved on to determining a capability level largely based on the implementation view. The participants would often remain in disagreement on for example how to see the division, what level of importance things had, what was the right thing for the division, and sometimes how to interpret the practice.

Even when there was explicit agreement on using one understanding there could still exist alternatives that were not established as worse, just different, but they chose to go for one to be able to set a capability level. The following dialogue shows such parallel understandings explicitly stated as collaboration with upstream stakeholders is discussed:

Dialogue 30

>Participant 3: Isn't upstream stakeholder. If we think of a patient, and if you are to see the patient you have to kind of look at what they have been a part of

before, membership to a family, an earlier history. That is, knowledge of the life outside the hospital? Is that what is thought?

-

Participant 5: It will perhaps be a different assessment, if you look at this with respect to the patients' and families' experience and interests, or if you look at a system like a municipality or the labor and welfare system. OK so if we try to see what level we are, if you look at level two it says "in an informal way". It is not just informally the way we operate, taking in these upstream interests, what do you think?

-

Participant 3: some places it is and some not.

-

Participant 5: If you look at it my way, with respect to public entities, then we have, or try to have, systems for information sharing and collaboration. But if you look at patient history?

-

Participant 3: I have seen cases where the patient's and family's experience was that what has happened the last ten years was not considered at all. Where the only thing considered was what was here and now. And I am thinking, we have information ten years back, hospitalizations, outpatient visits, all the data is accessible. But many don't even look at it! Perhaps people are oriented towards the here and now, and don't see all the things we think we do, or should.

-

Participant 3: Ok, but then I understand it in your way, that's fine. Then we set level two.

(Practice II.D)

The participants did not always collectively accept one prevailing understanding, which then became the basis for determining a capability level. The determined capability levels were therefore not always based on all the participant's views, as only one capability level for the desired and one for the current state of the division was to be set for each practice. Deciding on capability levels could often be a process

150

of negotiation, as in the following dialogue. This dialogue is a compressed version of the one found in the chapter below, dialogue 36.

Dialogue 31

> Participant 3: No I was thinking in the lines of, is it really necessary that all the supporting systems. Maybe we should establish the ones we have already instead of taking it even further? That maybe it is enough to get what we have in order. That I don't necessarily think that we have to go any further in. That it could be possible to stay on level two.
>
> -
>
> Participant 2: No I want to be at four.
>
> -
>
> Participant 5: Me too, I want to be at four.
>
> -
>
> Participant 2: (Laughing) This is a majority decision.
>
> -
>
> [...]
>
> -
>
> Participant 3: It's not a problem for me if it lands on four. I assume I don't get any punishment.
>
> (Practice III.A.1)

This dialogue illustrates how of the participants could determine a capability level without being in agreement in their view of the division. Participant 3 has a different underlying view of the issue of measurement in the division than the other participants. He argues for remaining on capability level 2, implying he does not want to increase the amount of measurement and quantitative focus in the division. The other participants do not share his view, and argue for a different vision of the division's future. This culminates in the participants disagreeing on what should be the desired capability level for practice III.A.1. The disagreement remains within the group, and a negotiation process where Participant 3 gives up determines the capability level.

7.3.3 Experience from Operational Position in Division Gave Legitimacy

The participants' statements seemed to carry somewhat different weights. It seemed that having experience from the operational part of the division made the participants' words carry extra weight. This was especially evident for the difference between the participants on the division staff and the department head. The staff members often directly asked what the department head or, "you guys who know the division better" (meaning the department head, union representative and staff member who had recently been a department head) thought about things such as in the dialogue below.

Dialogue 32

> Staff Employee 2: I put down level four because I think that. But. I'm thinking, I am not in a department so.
> -
> Department Head 1: No, well, the structural or organizational changes are [gives his view on the matter...]

7.3.4 Summary

Through discussion the participants informed each other's views of the organization and the practices, debated and negotiated these same views and established an accepted common understanding. This understanding is termed a prevailing understanding. We found that once a prevailing understanding became established in the group, it could be referred to and used in future contributions to the discussion. Throughout the process of using the LESAT the participants interacted quite often with the tool. We observed that they would prompt for a capability level decision or refer to some content in the tool to kick start conversation if the discussion stagnated. We also employed such kick-starters as facilitators at times, to bring the conversation forward. Furthermore, we found that much of the discussion was focused on the tool, where the participants read aloud, pointed to different elements, or shared capability levels they thought were appropriate. However, due to the selective focus and the focus on recognition, these frequent interactions with the tool did not necessarily mean that the participants ended up having the discussion intended by the author. However, as the participants did not always agree on how to view the organization, and we found that even though one understanding typically became the prevailing one

to be used to set down the capability level, this did not necessarily mean that the participants had reached a consensus. A few times (twice clearly enough to observe easily), the group could not establish one prevailing understanding, and the results of the assessment became the result of an explicit negotiation. In this process of establishing a common understanding, we notice that the opinions of the participants with experience from the operational part of the division carried extra weight.

7.4 DYSFUNCTIONAL USE OF CAPABILITY LEVEL SCORES (NUMBERS)

The focus on the task of determining capability levels for a practice, hereby referred to as setting scores, was large. The significance given to the scores was high, but the scores did not necessarily represent the content in the tool or the meaning originally attributed to them.

7.4.1 CENTRAL ROLE OF DETERMINING SCORES CONTRIBUTES TO DECOUPLING

At the start of the individual pilot rounds the participants were asked to consider what the practice meant in their division, and then determine capability level scores for the division. They were also asked to think of change initiatives that would take the division from the current to the desired capability level.

The participants' initial approach to using the tool surprised us. The initial reaction in the individual pilot rounds was to as quickly as possible mark down the capability levels. The participants spent only a minute or two on the first practices, jotting down their capability level scores and moving on to read the next practice. A good illustration of how they viewed the tool is the comment from Participant 5 after seeing our surprised faces when we saw that he was already on practice II.C.1.

Dialogue 33

     Participant 5: [After a long silence where the participants have quickly determined capability level scores for 4 practices in a row without asking questions]
     Yes, well. You can't think too much. You just have to take the intuitive, what you perceive it to be.
     (Individual pilot round)

Anyone who has answered a job satisfaction questionnaire with answer alternatives or other questionnaires with answer alternatives will recognize this way of thinking. Many such questionnaires typically include instructions with some statement about how you should answer, often something along the lines of "don't spend too much time on each question and if you are in doubt, follow your intuition and choose the first answer that came to mind". The way the participants intuitively reacted to the LESAT speaks to that the questionnaire-like design of the tool resulted in questionnaire-like associations. With these associations, the design of the tool encourages participants to leave the job of understanding the content of the tool to their intuition.

Realizing that the participants had treated the LESAT like a questionnaire to be answered on the basis of intuition rather than a deep understanding of the content, we intervened. In the group assessment, we informed the participants that if anything was unclear, they should work on the interpretation until it made sense either by discussing it or asking us if necessary. They were asked to first reflect on what the practice meant in the division before making that the basis for setting scores. However, the participants still often went on to determine scores before they felt that the overall meaning of the practice, the meaning of the concepts in the practices or even individual capability level descriptions were understood. Sometimes they cut discussions about the interpretations of the practices short without concluding them, in favor for a discussion about what scores to set. In dialogue 29, two competing prevailing understandings were being discussed when the discussion was cut short by a suggestion to set a capability level. The continuation of that discussion follows:

Dialogue 34

> Participant 1: And that it shouldn't be that someone sits on the top and makes a strategy, while others are pursuing improvement work that is never reflected in a strategy.
>
> -
>
> Participant 5: I was thinking that, like in the Project, it was deeply rooted in the entire organization. And I think that when you are able to do that,

involving people and have them participate all the way from the start, then changes start being implemented very early.

-

Participant 1: Yes. Well. What level do we land on?

-

[...]

Participant 2: I think we should try to get to five. We should have something ideal.

-

Participant 5: I agree.

-

Participant 4: Yes, we should reach for something high.

(Practice I.A.1)

In the discussion that followed the prompt for setting a score, the participants left the discussion about how to see the practice. The final score decision was not based on either of the interpretations, but rather a discussion about wanting something ideal. This is an illustrative example of an important point. The participants did not have to understand the descriptions in the capability levels, or even agree on their interpretations, to be able to set a score. As Union Representative 1 said, "a number is a number." It can be discussed without having meaning beyond its numerical value. In some cases the scores were discussed as being high or low, and not as representing the prevailing understanding of the capability level. We term these discussions about scores as being decoupled from the prevailing understanding. More often, the capability levels were discussed and used without representing the description found in the tool for that capability level, and we call these discussions decoupled from the content in the tool.

We explicitly asked them to discuss the practices and their interpretations before determining capability levels. However, the implicit instructions in the LESAT's questionnaire-like design, the instructions in the LESAT, and the main task we gave them was still to reach a decision for the capability level scores. The only necessary fill-out encouraged by the design in the LESAT is to draw a circle around the C in the

capability level box you think should be the current capability level, and a circle around the D for the desired capability level. There are small boxes for comments about evidence for current capability level and opportunities for desired, but these were however not mandatory. This means that the visible, documented results you are left with after having used the tool is a set of circles, or a list of scores for each practice. A set of scores is therefore the final package of results the design of the LESAT encourages the participants to create. How these were to be determined was not specified further than us urging them at the start of the day to try to understand the practices before setting a score. The content of the LESAT was found to be very hard for the participants to interpret, and they did not always reach an agreement. Given that their main task was to decide on a set of capability level scores, it made sense for the participants to move on to that task even though they knew they were on shaky interpretive ground. The resulting decoupling of the capability levels from the descriptions in the tool and prevailing understandings can therefore be seen as the participants just doing what they were asked to do to the best of their ability.

7.4.2 DESIRED CAPABILITY LEVEL SCORES AND CHANGE INITIATIVES WAS USUALLY DECOUPLED FROM CONTENT IN CAPABILITY LEVEL DESCRIPTIONS

Discussions about the desired capability level scores were usually highly decoupled from the content in the capability level descriptions, much more so than for the current capability level scores. The discussions about the practices almost always started with respect to the current state. An understanding of the practice with respect to the division would take form and a capability level for the current state would be determined. When determining the desired capability level, the participants did not repeat the process of reading the practice, mapping the capability level descriptions to the division and then make a decision for the desired state. Instead, the participants would often start with the prevailing understanding developed for the current capability level, and expect the higher capability levels to represent better versions of this understanding. When the potential desired capability level descriptions were read, their content was often subject to distortions. The content of the desired capability level descriptions was distorted to become in line with what the participants expected the higher capability levels of the practice to be. This is illustrated in the following dialogue where the participants had set capability level five as the desired score, but

156

changed their mind and set a lower score when we later prompted them to read it thoroughly. We see that they read only parts of the practice, and made up their own mind about what the content probably was, effectively distorting the content of the capability level description.

Dialogue 35

[Discussing current level for practice III.B.I, after we intervened and said that they could consider their procedures when it says, "standardized processes"]
Participant 4: Yes, we review those.

-

Participant 1: Yes!

-

Participant 5: We adjust those.

-

Participant 4: Yes. And now there is a large project going on to try to get the most uniform [Norwegian: enhetlig] procedures on all levels. Less on the lower levels.

-

Participant 1: So are we at four and are going to five? Or?

-

[All participants utter their agreement]

-

[We ask them to read level five to get inspiration for change initiatives]

-

Participant 1: Hm. But on the IHT level, the focus is on standardization. And then I think that the flexibility automatically goes down a little.

-

Participant 4: It is a little contradictory that sentence. A little Winnie the Pooh. [Winnie the Pooh is a much used expression in Norway for someone wanting everything and not making a choice. The sentence referred to in the LESAT reads, "Extended enterprise interface processes have been standardized while allowing for flexibility in innovation in support of local needs."]

-

Participant 5: And extended enterprise. Then we are talking outside the division. It could be both IHT and Health South-East. If we think both procedures and the Ledelsens Gjennomgåelse.

-

Participant 3: Maybe we should stay on four?

-

Participant 1: (Laughing) I think we might want to stay on four there, yes. (Practice III.B.1)

The change initiatives were also rarely coupled to the descriptions of the capability levels they had determined they wanted to reach. Rather, they were coupled to the prevailing understanding, and when this was decoupled from the descriptions of the practice, so were the change initiatives. In the dialogue above, we observed such a decoupling, and asked them to read the desired capability level they had chosen to see if the capability level description inspired further thoughts on how to get there. Instead of taking inspiration from this capability level description however, the participants discovered that the description of the capability level they had chosen was not in line with what they wanted to work for.

7.4.3 CAPABILITY LEVELS HAVE NORMATIVE POWER

Establishing an understanding of a practice, and then expecting the higher capability levels to be more optimal versions of this understanding speaks to the participant's expectation of higher capability levels being better. In the following situation, one participant disagreed with a practice in the LESAT, and did not want to set a high desired score because he felt the descriptions of the higher capability levels were wrong for the division.

Dialogue 36
[Reading practice II.A.1] [...]
Participant 5: We don't measure on a local level, it is measured, perhaps locally and, both places [referring to description of level 1]

-

Participant 3: We have to be at three? With all the data that we, that we have, don't you think?

[...]

-

Participant 2: And you can get numbers for most of what you need to do things.

-

Participant 1: You can?

-

Participant 2: We get, well numbers are presented in heaps and loads these days.

-

Participant 5: I put four during the individual round.

-

Participant 2: In three years? Yes I agree.

-

Participant 3: I think I was thinking, I put down two for both current and desired.

-

Participant 2: (Laughing) You don't think there should be any production then?

-

Participant 3: No I was thinking in the lines of, is it really necessary that all the supporting systems. Maybe we should establish the ones we have already instead of taking it even further? That maybe it is enough to get what we have in order. That I don't necessarily think that we have to go any further in. That it could be possible to stay on level two.

-

Participant 2: No I want to be at four.

-

Participant 5: Me too, I want to be at four.

-

Participant 2: (Laughing) This is a majority decision.

-

Participant 3: And they have outsourced many of these services. They have outsourced a lot. [Here an association between measurement and outsourcing leads to a gradual distortion of the practice from being about measurement to being about outsourcing.]

-

Participant 1: Yes we have.

-

Participant 3: And is that the way we want to go?

-

Participant 1: No.

-

Participant 3: So that's what I was thinking.

-

[… They discuss for some time how the services they have outsourced are very bureaucratic and cumbersome.]

-

Participant 1: So the conclusion then is that we should be at level four. Then we get a "seamless information exchange" with services that are outsourced. Because like you said, having to send an email to change a light bulb is not very seamless.

[The actual text in level 4 is, "Performance measurement system scope is expanded to integrate with non-traditional measures of value creation (e.g., intellectual capital, balanced scorecard, etc.)." The statement "seamless information exchange" is found in level 5, in the following context, "*Performance measurement systems* provide seamless information exchange across the extended enterprise (...)". This is a major distortion of the text.]

-

Participant 3: No, it's not a problem for me if it lands on four. I assume I don't get any punishment.

-

Participant 4: Yes we should be a little visionary here.

-

Participant 3: Yes but I am just as visionary I think! But I am thinking that we could.

-

Participant 4: Yes I understand, but I think we are a bit further.

-

Participant 3: I guess.

(Practice III.A.1)

Instead of seeing the preference for a low score as criticism of the tool's implied direction of change, the other participants viewed this preference for a low desired score as a less visionary view of the future. This, together with the tendency to distort the content of higher capability levels to fit with participants' idea of what would be the best for the issue they are considering, speaks to the normative power of having scores from low to high. Setting low scores is interpreted as being less optimistic or visionary, and the high capability level descriptions are interpreted as ideal solutions to the issues the participants believe the practice considers.

The dialogue above is interesting because it includes evidence of the participants first changing the meaning of the practice, and then selectively reading a potential desired capability level to fit with this meaning. Then it shows negotiation of the results because Participant 3 has underlying assumptions about measurement in the division that differ from the rest, and lastly how proposing a low desired score is interpreted as little visionary. There was a clear decoupling of the discussions about what capability levels to set and the actual description of the capability levels found in the tool. Further, proposing a desired capability level was often used as a communicative tool for implying how important an issue was or how realistic it is to solve it, rather than being the result of a consideration of the text in that capability level description.

7.4.4 SCORES ARE NOT CHANGED AFTER THEY HAVE BEEN SET

Settling on a score had a surprisingly definite effect. Once it was set, it was rarely changed. On several occasions the participants kept discussing a practice after a current or desired capability level was determined. When the prevailing understanding of the meaning of the capability levels changed during these discussions, the

161

participants did not once go back and change their capability level decision in light of the new understanding. For one of the practices they even changed their interpretation of the content of the practice between deciding the current and desired state, but did not, as illustrated in the dialogue below, go back and update the current state. As a result, the number set for the practice could be decoupled from the prevailing understanding that the participants had after concluding the discussion about the practice. In the following dialogue, practice II.C.1 is first discussed as being about user feedback, and the use of feedback forms and patient feedback after appointments to improve the service. The current capability level was set to two based on this understanding. The understanding was then changed to be about the use of user competence during patient treatment. Inclusion of user competence in the treatment process was an issue the participants had discussed for an earlier practice, and established to be very important. The change of the prevailing understanding from user feedback to user competence happens between setting current and desired capability levels, but the current level is not revisited in light of the new understanding.

Dialogue 37

> Participant 5: During the individual round, we were thinking about the user feedback survey and feedback from the patient, right?
>
> -
>
> Participant 3: Yes.
>
> -
>
> Participant 5: If we swopped. Swopped customer with patient.
>
> -
>
> Participant 1: I would say, I think we are at level two, because we don't do much continuous, collection of downstream, eh. [Reading from LESAT level 3: " Feedback is continuously collected from customers and other downstream stakeholders."]
>
> -
>
> Participant 5: And now I was thinking about what we were discussing earlier, about including the patient. Then you get a completely different feedback.
>
> -

162

Participant 1: User competence, I guess we should be at level five then. The goal. But I think, maybe that's a little hairy?

-

Participant 5: Mmm

-

Participant 2: Yes we should be at five. The law. The law says we have to be at five.

-

[The rest of the discussion is about whether they should make the desired level 4 or 5 using the new interpretation]

(Practice II.C.1)

It became evident that determining a capability level was a task that was at times completed without making sure that it fit the capability level descriptions, or even that the determined capability levels fit with the most updated prevailing understanding. This means that just as there was sometimes a decoupling of the prevailing understanding and the content of the tool, there was also a decoupling of the capability levels set and the prevailing understanding as the understanding could change while the capability level scores remained as they were.

7.4.5 NUMBERS DECOUPLED FROM PREVAILING UNDERSTANDING WERE STILL USED TO REPRESENT IT

Regardless of the at times loose coupling between the prevailing understanding and the capability level set, once a number was set it became a representation of a prevailing understanding. Numbers were often referred to at a later time to represent an agreement that was made. Using capability levels in such a way was common, even if the underlying understanding had been forgotten or changed. The discussion leading up to numbers determined at one point was often very disconnected from the discussions where those numbers were used later. We did not notice this before listening closely to what was discussed as the scores were set, noting it down, and then comparing it to the use of the numbers later. We took no notice of this during the pilot and felt that this was a completely natural thing to do. When looking closely at the participants' process later, it was surprising to see how often earlier established

numbers were used as representations for decisions almost completely disconnected from what the number was actually based on.

Dialogue 38

[Continuation of previous dialogue 34. They are now discussing what the desired capability level should be. The prevailing understanding is that a high level means a high level of inclusion of user competence in the patient treatment.]

Participant 5: So where should we land?

-

Participant 2: I think, I am a fan of five.

-

Participant 4: In three years?

-

Participant 2: (Laughs)

-

Participant 3: Yes in three years, I was thinking that we can't get to five in three years.

-

Participant 2: So you guys want four? Three?

-

Participant 3: No I put down four in the individual round. No! I put three! [During the individual round, participant 3 had an understanding where the practice was about the use of user feedback forms and after-appointment feedback sessions. He put the desired level on 3 because there is a long way to go from the user feedback system they have today to a system where all departments have continuous patient feedback after appointments and use this to improve. However, he still shares this level as a representation of what he thinks, even if the discussion is a completely different one than what he based this answer on.]

-

Participant 5: We were pessimistic last time! I also had three, but I am with you on four. [The same comment applies here. There is some recognition of a

different view forming the basis for the capability level from last time, but this is explained as pessimism rather than a different interpretation of the practice.] (Practice II.C.1)

An additional dialogue illustrating such use of numbers is presented below.

Dialogue 39

[We prompt them to determine a current level]

Participant E: I certainly landed on the wrong level last time because I answered an English questionnaire. I am certainly not on three as it says in my earlier answer!

-

Participant X1: I would say two.

-

Participant E: Yes I think we are further than one at least, because I think the part about leaders understanding, that is at least. One thing is to understand, another is to practice it.

-

Participant X1: Yes.

-

Participant X3: But what about in three years? Desired?

-

Participant E: I think I answered four last time.

-

Participant X2: Yes five seems unrealistic in our world really.

-

Participant E: Yes, it will take some effort to get to four as well.

-

Participant X1: Maybe we should expect it though? (...)

(Practice III.A.1)

Participant E was very aware that his previous answers from the individual assessment were based on a process where he did not feel he understood the practices

165

well enough. He was therefore very restrictive with referring to capability levels determined during the individual assessment, as seen in the start of the dialogue. However, neither he, we nor the other participants reacted when he did use an earlier determined capability level to support the choice of a capability level. This speaks to how easy it is to refer to numbers even when they have no clear meaning attached.

7.4.6 SUMMARY

The participants had a very high focus on capability level scores. When determining desired capability level scores, the participants would use a prevailing understanding build during discussions concerning the current capability level score. They would then expect the higher capability levels to represent better versions of this understanding rather than critically reading the descriptions in the higher capability levels. Also, the change initiatives they came up with were disconnected from the descriptions. The capability levels had a normative effect, as the higher level was automatically perceived as being better, and not wanting a higher capability level was perceived as being less visionary. Furthermore, once the number was set, it was rarely changed. Even in cases where the interpretation of the practice was changed after setting a capability level, they would not reassess it. This resulted, on occasion, in a decoupling between the prevailing understanding of the practice and the number set. Regardless of this, numbers were often referred to at a later time to represent an opinion on the practice. We observed several incidents where the participants would refer to numbers they had set at an earlier time, and this was accepted as a contribution to the discussion without requiring the inclusion of the thoughts that the number was based on. Due to this, numbers that were set on the basis of a different interpretation than the one currently employed in a discussion was perceived as valid input.

7.5 Participant's Reflections about the pilot

The participants gave direct feedback about the tool throughout the pilot and in the follow-up interviews. These were meta-comments about the process or the tool, which were not a part of the actual pilot process, but comments about it. The main bulk of comments were made during the follow-up interviews, as we asked the participants to reflect on the process. However, we also received comments during the pilot, as the participants would break up the discussion to for example voice frustration and criticize the language or the applicability of the tool for use in the division. To better round out the presentation of our findings, we here present a representative set of the explicit feedback comments made by the participants about the process and the tool.

7.5.1 Participants Learned From Each Other

The participants found the most interesting and valuable part of the pilot to be the mixed group of participants and the opportunity to discuss and listen to the others' concerns and learn from each other. As seen from the comments below, several explicit comments were made about this.

Dialogue 40

> Staff Employee 1: Sitting down together has an educational effect. We don't take the time to reflect in the daily routines. It is valuable to sit down together.
> (Follow-up interview)

Dialogue 41

> Staff Employee 1: Being able to sit down and take a bigger look at things, that is luxury. And it is done far too little. One tries to minimize the amount of meetings. And especially for our division, when there are meetings people have to come from far away. You can't just talk together and say, can you come to a meeting tomorrow at two? You have to make it known a week in advance. And people have to travel and sleep over at hotels to come to meetings. So it is not very flexible, with respect to just saying hey, lets take a day to brainstorm.
> (Follow-up interview)

167

Dialogue 42

> Union Representative 1: When I am presented with a reality as seen from the top of the division, it does not necessarily coincide with the image I have further down the hierarchy. And that was interesting! To hear good discussions here.
> (Follow-up interview)

Dialogue 43

> Staff Employee 2: I thought it was cool with Union Representative 1. But we did get a little, on some of the topics we were a little, you know, on different, we kind of had different realities. And that was very good!
> (Follow-up interview)

As can be seen from these comments, the participants valued the chance to sit down and talk to each other, as this gave them a chance to reflect. It was also noted how this is not done very often in the division, as the great geographical distances makes this challenging. Furthermore, the inclusion of the union representative was found to contribute interesting insights, and both Staff Employee 2 and Union Representative 1 commented on how interesting it was to be presented with another perspective on the organization.

Throughout the pilot we observed several situations that support these comments and speak for the participants having learned from each other. The participants shared information from different places in the organization such as the lack of formally educated employees, which not all the participants were aware of. They also shared information such as good ideas that were successfully implemented in parts of the division, for example how some departments gathered feedback from patients at once after an appointment and adjusted their services continually. There were several instances during the pilot rounds where participants were presented with new information as a result of participants sharing their insights and experience from the division.

Dialogue 44

> Participant 5: [Referring to the individual pilot round] It really surprised me, when you said there were so many unskilled [Norwegian: ufaglært] employees in the division.
>
> (Group pilot round)

Dialogue 45

> Department Head 1: And then you have, like BUP [a category of departments: mental health clinics for children and adolescents], they do things in a way so that if their patients are admitted with us, the have a very different.
>
> -
>
> Staff Employee 2: Yes, they follow, the outpatient clinics always follow the patient even if they are admitted to an overnight unit [Norwegian: døgnenhet].
>
> -
>
> Staff Employee 1: Do they really?
>
> (Group pilot round)

7.5.2 THE OVERALL DESIGN OF THE LESAT WAS ACCEPTABLE

No criticism was directed at the overall design of the LESAT with its list of practices with capability level descriptions and boxes for current and desired capability levels. While being critical to aspects of the content, the design was found to be acceptable. Furthermore, Department Head 1 was familiar with use of tools with variants of this design, and noted that gap analysis is not a new concept to the division.

Dialogue 46

> Department Head 1: But when you implement ROP-guidelines for example, you use a gap analysis. A lot of this is the emperor's new clothes, it is just new words for the same thing.

Dialogue 47

> Union Representative 1: But there is nothing wrong with the whole, finding out where we are and where we are going.
>
> (Follow-up interview)

169

The task of setting a score for where you are and when you want to go is a central aspect of the design of the LESAT and found to be known and acceptable.

7.5.3 Too Much Energy was Spent Trying to Comprehend and Interpret the Language

In the follow-up interviews, when asked to talk about the whole process and their experience of the tool, one of the main topics represented was how challenging it had been to understand and interpret the language in the LESAT. This is illustrated by quotes from the participants below.

Dialogue 48

>Staff Employee 1: I think you spend too much energy trying to kind of translate it in your head all the time, to comprehend it. Because the entire time I am sitting and reading this thing, I think enterprise, hospital trust. Enterprise, hospital trust. [Recall that "enterprise" in the pilot meant the division, this comment illustrates how confusing this was.] You have to do that all the time, so that you don't concentrate too much on what this really is, what the question is.
>(Follow-up interview)

Dialogue 49

>Staff Employee 4: Very many of the words here [pointing to the LESAT] are never used in the public health sector. Even if it is translated to Norwegian, we still have to translate it again. Try finding words that are used here, to get it into the context.
>(Follow-up interview)

Dialogue 50

>Staff Employee 2: It was very demanding just understanding the questions.
>[...] I feel it is very hard to translate it to the health language.
>(Follow-up interview)

Dialogue 51

> Department Head 1: It is a language that I, I don't know if I get it all the time. I think there were words and expressions and the sentences were, that made me think it was hard to see what they really wanted.
> (Follow-up interview)

As can be seen from the comment made by Staff Employee 1 above, he experienced that he had to concentrate too much on trying to understand the content. He also felt that too much energy was spent trying to translate words and expressions in his head. This is backed by the other participants, who also commented on the difficulty of understanding what was really meant in the practices and how challenging it was to employ the language in the division.

7.5.4 Participants Reacted to the Values Reflected in the LESAT Language

In addition to vocabulary being unknown or from a different context and therefore tough to interpret and use for the division, there were reactions to the values implied in the language of the LESAT. This concern was voiced explicitly both during the pilot and in the follow-up interviews. In the pilot, the reactions were often in the form of a participant reacting to words or phrases that they did not feel comfortable using to describe and discuss the division. These were terms such as "customer", "business model" and "production".

However, these reactions were often disconnected from the participants' ability to understand the practice and interpret the concepts in the division context. This means that disliking the language was not the same as not being able to interpret and use it.

Dialogue 52

[Reading practice I.B.1]

-

Department Head 1: First of all I think that this is a foreign language to the public health service. It is a very, very cold and cynical language. That we immediately don't feel at home in. And that does something to the way we relate to it. Now,

everyone knows that we are getting closer to a business model in the public health service. We are using words today that we would not have in our mouth a few years ago. And some of us who work in health services want to still stand for those values. And that means a language that is much more value loaded. That does not mean that we don't want to work efficiently, give good services and have good quality and such, but there is something with that language. (...) We want to have a language that communicates values.

-

Staff Employee 2: I agree. [...]

-

Department Head 1: It's not that we don't think the patient is a customer, but it is about not taking away the language we actually have. Sick people are not first of all customers, they are actually patients.

(Practice I.B.1)

After reading practice I.B.1 the participants reacted to the language, voiced their frustrations, before going on to interpret and employ the capability level descriptions and practice relatively easily.

In the follow-up interviews it became clear that there was a clear personal difference between the participants in their reactions to the values in the language. One of the participants reacted mostly with curiosity to the foreign language and did not give it much attention in the follow-up interview. Also, although the remaining four of the participants agreed that the language should be more in line with the values of the division, two of the participants were mostly concerned with the language being hard to understand. The language in the LESAT however actively provoked the last two participants. As illustrated below, Department Head 1 and Union Representative 1 perceive there to be a clear link between how people talk about things and how they behave, and they found the language in the LESAT to be irreconcilable with the mindset and values that they need to have in their work.

Dialogue 53

Department Head 1: We actually work with people who are incredibly sick. To be able to work with these people you have to be able to let yourself be touched without becoming completely drained. There is a balance to be found between how much you should let yourself be affected and how much distance you have to have. And I think a language like this [points to the LESAT] to a certain extent alienates us from this issue. I mean we create a very large distance between our self and the patient. And we should have a distance, but we are completely dependent on being able to understand what it is people are actually talking about. And then you have to in some way or another let yourself be affected by what they say. And I don't think this language is any good with respect to, of course they are customers, but we are in the health sector and they are first and foremost patients. They are here because they are patients. [...] So when we talk to patients and, I use the word patient, they are patients when they are here, we have a different approach, perspective, than we do if we think about them as customers. So I find it very hard to relate to this, and I don't know if I want to either. I don't know if I want to be a part of this. We turn them into a product rather than people of flesh and blood.
 (Follow-up interview)

Dialogue 54

Union Representative 1: I think when you put in terms like these, for example when we are supposed to help people that are sick and call it a form of production, then I think we loose the values. Then the value-concept becomes different to me. Then we are producing, and that is not what we are supposed to do. We are not delivering a "service" either, I don't think I can express it any better. When you use words like that, there are some financial terms that turn us people into a form of producing a car or something, and we become a number. And that, in many ways we probably have to become numbers, but I think we should be very careful with what words we choose to use. We should not use words like that.
(Follow-up interview)

173

As we can see, they are both aware of what language they use, and what values or attitudes are reflected in it. They are also concerned with how the way they talk about the patients affects how they see them and approach them, and Department Head 1 related how this could be used as a tool to maintain a balance between understanding the patient and keeping a necessary emotional distance. Furthermore, Union Representative 1 argues that they are doing production, and if they say that they are and use terms from economics, then humans become form production or a number.

However, as noted, one of the participants was mainly curious about the new language. Staff Employee 1, when asked, commented that he failed to understand why this was seen as such an issue.

Dialogue 55

> Staff Employee 1: No, well, I can probably. I am struggling to see that this language is about values. It's just a tool. For me, this does not imply any choice of values. I just can't see it. It is a tool to measure production, or something that is going on. Some service or another. [...] One can't feel threatened by some new words that come wandering into your work place. (Follow-up interview)

As can be seen from the quote, he did not think that the language in the LESAT had much to do with the values in the division, and felt that the perceived threat from introducing words from another context was somewhat exaggerated. However, two participants reacted very strongly to this issue, and two participants stated their agreement. We therefore consider this issue an important concern, even if all participants did not back it.

7.5.5 NEGATIVE TO INCREASED USE OF NUMBERS AND MEASUREMENT

Two of the participants were very concerned about the increasing use of counting and measurement in the division. Although they were not wholly negative to being evaluated or measured, they are very skeptical to the effect of the increasing focus on numbers. As can be seen from the comments included below, they were both concerned with how the focus on numbers means less focus on quality.

Dialogue 56

> Department Head 1: I absolutely think we need to evaluate ourselves. And I think we have to evaluate ourselves with respect to quality, not just numbers. And I know we have to be evaluated with respect to, that we have to use numbers. [...] But numbers for the sake of numbers, it has no value if you are not able to see what is behind the numbers. I think my concern is that I am afraid we start discussing so much numbers, and then forget that those numbers without quality or specialization [Norwegian: fag], well. We work in a very profession-based organization. And my experience is simply that we become alienated. That we become more obsessed with numbers than the quality of our work.
>
> (Follow-up interview)

Dialogue 57

> Union Representative 1: Not everything is very measurable, quality for example. We told the leadership during the last meeting that numbers are all we talk about. Numbers are the topic in ninety percent of our meetings. How do we manage this because the budget is smaller, the demands are this and that, this here does not look good and we have to save, we have to come with suggestions for changes. Mostly we just focus on what worries. Sometimes I think we should lift our gaze and look at what we are succeeding at. I think it does something to the mood when we have that pressure all the time. [...] Everything is to be counted! And that's how it has become. That is how the hospital is run, new public management way of thinking.
>
> (Follow-up interview)

In addition, Union Representative 1 notes how the increased focus on measurement results in a lot of work, but not necessarily much value. In the quote below, he relates how it is hard to find good, reliable indicators for the things they are trying to measure.

175

Dialogue 58

    Union Representative 1: It can easily become quite subjective in large organizations like ours. But it is also hard to measure. For example, when we measure activity, how many patients we have, it really tells us nothing. Because patients can be very different, different in how demanding they are. Six patients can be much more demanding for a unit than ten other ones. (Follow-up interview)

Dialogue 59

    Union Representative 1: And often I think it is a very complicated way of doing it. It's not that hard. You have lots of people who are set to work making tables, and, and making circles. And making codes, and in one way it looks incredibly interesting with all that. But when you bring it down to having X people at work and X patients, then perhaps it is not necessary to do it in that way. It is amazing what people can do, can be employed to do. Without necessarily discovering all that much. But this here [points to the LESAT], its a little different I guess, its not so much about numbers I guess. (Follow-up interview)

In addition to concerns about increased use of numbers and measurements, Department Head 1 relates that the employees often do not know what the numbers are used for and as a result, they feel like they were not trusted to do their work.

Dialogue 60

    Department Head 1: I think I spend a lot of my time explaining to unit leaders and employees this here about the numbers. What it is we are using them for. People feel, there was one who came to me today, and said that she felt so devaluated. It is almost as if someone thinks we are not doing anything. And well, that does not necessarily have anything to do with tools, but how we communicate it. (Follow-up interview)

7.5.6 BUSY LEADERS WANT MORE TIME, NOT MORE TOOLS

Throughout the pilot, the participants made several comments about how many new systems have been introduced into the division lately, and how much time is spent in meetings, seminars and documenting work. These comments were typically made when the participants were asked about how the LESAT would be received if it were adopted by the division management, or to discussions about how the LESAT could be used by the division.

Dialogue 61

> Staff Employee 1: It is really becoming a lot now. An employee told me that he spent one and a half day a week on documentation. That is probably a little exaggerated, but there is a lot of stuff that has to be followed up, checked, updated and aggregated. And a lot of it has been introduced in a short amount of time.
> (Follow-up interview)

Dialogue 62

> Union Representative 1: What I am usually told is that the more tasks leaders are given, the more you are taken out of the daily work. There are leader gatherings, meetings, you are pulled out all the time. So the presence at post [leaders' presence in the unit are leading] is becoming less and less. Meetings and leader gatherings have become very, and they are usually far away too, it is experienced by unit leaders as nice, but they wished they had more time to actually be a leader.
> (Follow-up interview)

In the pilot, Union Representative 1 also noted that there is a lot of resistance to new things in the division now after so many things has been introduced lately.

>Union Representative 1: A lot has been introduced the last years. If you now also introduce something like this [LESAT] then I think, even if it is good, I think you would have many out there who would go, "now it is enough!"
>(Follow-up interview)

7.5.7 THE DIVISION ALREADY HAS SIMILAR TOOLS

The division already has tools that bear resemblance to the LESAT. This became apparent to us through comments scattered throughout in the pilot and follow-up interviews.

Dialogue 64

>Staff Employee 2: It is a little like the Ledelsens Gjennomgåelse, just a different tool.
>(Individual pilot round)

As can be seen from the explaining comment made by Staff Employee 4 below, Ledelsens Gjennomgåelse is a self-assessment of the activity of the organization. It covers some of the same topics as LESAT, and is a tool (set of documents) that is to be filled out and sent upwards to the next hierarchical level where they make an aggregated version that they send upwards and so on.

Dialogue 65

>Staff Employee 4: In the Ledelsens Gjennomgåelse, that we do twice a year, we go through in effect everything we do in the division. Not necessarily numbers and such, but how we, formal competence/education, quality, treatment, organizational challenges, yes pretty much everything that affects our activity is included in it.
>(Follow-up interview)

Also in the follow-up interview, Staff Employee 2 again commented on the similarity between Ledelsens Gjennomgåelse and LESAT. He proposed that perhaps if the

LESAT were to be introduced, it might replace some parts of Ledelsens Gjennomgåelse.

Dialogue 66

> Staff Employee 2: [Answering the question, "could LESAT with division language be useful to you?"] Yes, well, perhaps. Then you have to think, where are you going to use it? And with respect to all the other things we have. Because we do a thing called Ledelsens Gjennomgåelse regularly in the division. This tool looks a bit like that. Maybe one could replace some of that? Because I think if you come with another tool now, people would. People feel there is enough of that. So I think it would have to replace something else.
> (Follow-up interview)

This is supported by Staff Employee 4, who also argued that you cannot have two tools with the same responsibilities.

Dialogue 67

> Staff Employee 4: Well, actually I guess we do have a tool that is adapted to the division already. No we don't need any more tools. We can't have different tools answering the same things.
> (Follow-up interview)

The notion that they already have tools that have some of the same responsibilities as the LESAT is supported by Department Head 1 and Staff Employee 2, when they comment that there already exist tools and guidelines, that are available to them, but not used. These include national guidelines and treatment lines (Norwegian: behandlingslinje).

Dialogue 68

> Department Head 1: I don't think there is a lack of procedures and guidelines in the hospital.
> (Follow-up interview)

Dialogue 69

> Staff Employee 2: We could use, there are so many, we have national guidelines for courses of patient treatment [Norwegian: pasientforløp]. And we have those, what are they called? The patient treatment lines [Norwegian: behandlingslinjer]. We could always evaluate ourselves against those.
> (Group pilot round, practice II.A.1)

7.5.8 SUMMARY

Throughout the pilot, and in the follow-up interviews the participants made some explicit feedback comments on how they perceived the LESAT. The positive feedback was mainly centered on how they appreciated a chance to talk to the other participants. Furthermore, they seemed to be quite positive towards the overall design of LESAT, noting that this form was known to them, and that it made sense to use this form of analysis. The main concerns voiced by the participants revolved around the language, the increased use of documentation and focus on numbers, that leaders were busy and would resist the introduction of more tools, and that it overlaps with already existing tools in the organization. Their reactions to the language was mainly concerned with how it is difficult to understand and use. However, some participants raised concerns about how the language did not reflect the values of the organization.

8 EVALUATION

In the following chapter the LESAT is evaluated with respect to the Psychiatric Health Services Division at IHT. The evaluation is done with respect to the predetermined criteria developed in chapter 3 and concerns from the participants that are not covered by these criteria. A final judgment is then made on the basis of the criteria-based evaluation and these concerns.

We established that when evaluating an SAT, this must be done with respect to the organization, the users of the tool and stakeholders of the self-assessment. The organization is the division, and the participants in the pilot rounds represent the users. The use of the LESAT may initiate large-scale changes, and the stakeholders of the division are therefore the stakeholders of the self-assessment as well. The division's main stakeholders were identified in chapter 5. The participants of the pilot rounds do not only represent likely users of the tool, but also represent important stakeholder groups and their concerns. Although not all stakeholder groups were represented, a union representative, operational leadership, and staff from the division leadership were included to tap into these important stakeholder group concerns. The stakeholder groups that were not represented through the participants were considered on a theoretical basis.

The evaluation draws on the findings from chapter 7 as well as the descriptions of the division and the LESAT in chapters 5 and 6. To provide a form of overview over the findings and their use in this chapter, the figure 16 illustrates the mapping of findings into the criteria.

Challenges for the Participants Interpreting the LESAT

| Foreign Tongue (English) Made the Participants Unable to Use the Tool |
| Language and Concepts in The LESAT were Hard to Employ |

Participants' Interpretation Process

| Participants had Selective Focus |
| Participants were Guided by what they Recognized |

Working as a Group to Build Prevailing Understandings

| Participants Shared Views and Built on Each Others' Contributions |
| Having a Prevailing Understanding did Not Imply Consensus |
| Experience from Operational Position in the Division Gave Legitimacy |

Dysfunctional Use of Capability Level Scores (Numbers)

| Central Role of Determining Scores Contributes to Decoupling |
| Desired Capability Level Scores and Change Initiatives were Usually Decoupled from Content in Capability Level Descriptions |
| Capability Levels Have Normative Power |
| Scores are not Changed After they Have Been Set |
| Numbers Decoupled from Prevailing Understanding were still used to Represent it |

Participant's Reflections about the pilot

| Participants Learned From Each Other |
| The Overall Design of the LESAT was Acceptable |
| Too Much Energy was Spent Trying to Comprehend and Interpret the Language |
| Participants Reacted to the Values Reflected in the LESAT Language |
| Negative to Increased Use of Numbers and Measurement |
| Busy Leaders Want More Time, Not More Tools |
| The Division Already has Similar Tools |

Comprehensibility

Usefulness

Effort

FIGURE 16: MAPPING OF FINDINGS TO
CRITERIA

182

8.1 COMPREHENSIBILITY

From chapter 3 we established that in order for the tool to be comprehensible, the structure of the tool and how to use it should be intuitive. Further, the explanations provided needs to be clear, simple and exhaustive. This will be discussed in chapter 8.1.1. The language used in the tool needs to match the user's competence level and the concepts presented in the tool should be clearly expressed and not ambiguous. This is discussed in chapter 8.1.2.

8.1.1 OVERALL DESIGN OF THE LESAT

The design of the LESAT closely resembles a classic questionnaire and this resemblance makes the design recognizable and intuitive to use. The participants were accepting of the overall design and the idea that you should establish the current and desired state. The structure of the tool and how to use it is found to be intuitive. With respect to the explanations provided, the task of matching the organization with the capability levels that best describes the current state and desired state was easily grasped by the participants. The department head was even familiar with this form of analysis, meaning that it is not a new concept to the division. Furthermore, there were no questions about the explanations, including any questions about aspect such us how the time horizon of three years were to be employed. The participants understood the task at hand, and we take this to indicate that explanations provided were clear, simple and exhaustive.

8.1.2 REQUIRED LANGUAGE COMPETENCE LEVEL

The language used in the tool did not match the user's competence level. Firstly, the LESAT is in English and this made the participants unable to use the tool. If the LESAT were to be left in its original language, the tool would be completely incomprehensible for the identified potential users. Secondly, even when translated to Norwegian, the language and concepts in the LESAT were still difficult to use and experienced as foreign. The language and concepts in the LESAT were found to be hard to employ, and foreign and abstract concepts became ambiguous. How problematic this was differed between the capability level descriptions and the practice titles, where the latter was found to be more problematic than the former.

Ideally the titles should help users understand what the practices overall are about. However, the participants found these titles very hard to employ and therefore selectively focused on the capability levels. As the LESAT is to be judged with respect to likely users of the tool in the division if it was adopted, and our participants represent such a likely group, the titles in the LESAT are judged to be almost completely incomprehensible. The participants' understanding of the overall meaning of the practice in the division was often fragmented and decoupled from the contents in the tool. The capability level descriptions increased the clarity of the practices, allowing users to compare the division with each description to find a best match. However, although the participants found the capability level descriptions easier to relate to than the surrounding content, they still found the language and concepts hard to employ. We conclude that the language in the tool does not match the competence level of its users, the concepts in the tool were not expressed clearly, and the concepts became highly ambiguous to the users.

8.2 USEFULNESS

To fulfill the usefulness criteria, the presentation of the concepts in the tool should support a reflection process where the participants become more sophisticated by learning from the theory presented in the tool, which will be discussed in chapter 8.2.1. The tool should also support a reflection process where participants become more informed and understanding of others' views, which is discussed in chapter 8.2.2. In chapter 8.2.3 we consider if the tool enables participants to take action. It is here relevant to judge if the tool serves as a boundary object, enabling collective action without necessitating consensus. Further, in chapter 8.2.4 we discuss if the concepts chosen for assessment are applicable and key to understanding the functioning of the organization, and if the theory coded into the tool is suited to explain the underlying structures of the division. Many of the issues highlighted in this chapter overlap with considerations of potential conflicts with important stakeholder concerns. These will therefore be discussed when relevant in this chapter. Lastly, chapter 8.2.5 considers the remaining issues of conflicts with stakeholder concerns, discussing if the recommendations of the tool and influence of the discourse embedded in the tool is in line with important stakeholder values and interests.

8.2.1. LEARNING FROM THE THEORY IN THE LESAT

The presentation of the concepts in the tool should support a reflection process where the participants become more sophisticated by learning from the theory presented in the tool. It was very hard for the participants to learn from the theory coded into the LESAT. For the most part the practices in the LESAT sparked continuations of discussions that were already ongoing in the division, instead of new discussions guided by the theory introduced in the tool through the practice descriptions.

The participants focused their attention selectively on what they recognized, and because of the foreign language and concepts it was hard for the participants to recognize and therefore employ the contents of the LESAT. The selective focus on recognizable topics steered the discussion away from the contents in the tool towards topics they were familiar with instead. The presentation of the concepts in the tool was therefore too incomprehensible to support users in their reflection about the theory in the tool.

In addition, the LESAT's design and instructions encouraged the participants to discuss what scores to set even when they felt that they had not comprehended the content of the tool. Rather than discussing the overall meaning of a practice in abstract or division-specific terms, or even the meaning of individual capability level descriptions until they were made clear, the participants went on to determine scores even when they felt they were on shaky interpretive ground. Scores were not changed after they had been set even if the interpretation of the practice changed, and scores were used to represent opinions even when the interpretation of that capability level had been forgotten. The design of the LESAT did therefore not encourage reflection about the theory in the tool, but rather discussions about numbers. This resulted in the discussions about numbers becoming decoupled from what those numbers actually meant. The participants had discussions about numbers that were removed from the understanding of the capability level descriptions intended by the author (content of the tool), and even the meaning attributed to the scores by the participants. When discussions become decoupled from the descriptions in the tool, the theory coded into the tool was effectively overlooked. The presentation of the concepts in a manner so

focused on numbers therefore does not support the participants in reflecting on and learning from the theory in the tool.

The quick change of focus from interpretation to scoring may also have been a result of our facilitation of the process. We told the participants at the start of the process that they should try to understand the practices before making final score decisions. However, a handful of times when the discussion went quiet for a long period of time and we were unsure about what score they had landed on, we asked them if they had determined a score so that they could move on to the next practice or start the discussion again. We never explicitly gave them the option of saying that a practice was too unclear or unfitting to use and move on without setting a score, because we wanted to see if they would react in such a way on their own. As the leaders of the process we could have opened for other alternatives, given explanations without being asked and asked them to read the practices again when we observed that the discussion became decoupled from the content of the tool. However, because we did not want to influence their process, we did not. This may have strengthened the participants' perception of the scores being the most important result, rather than reflection and learning from the tool's content.

We considered many of the practices in the LESAT and their implications for the division to be important topics that could have been useful for the participants to discuss. However, the theory in the tool had surprisingly little influence on the discussions, and we did not see that the participants' views became significantly more sophisticated as a result of learning from the theory in the LESAT.

8.2.2 PARTICIPANTS MORE INFORMED AND UNDERSTANDING OF EACH OTHER'S VIEWS

To fulfill the usefulness criteria, we found that the presentation of the concepts in the tool should support a reflection process where the participants become more informed and understanding of others' views. Learning from each other was in fact the most positive result experienced by the participants from the pilot rounds. The feedback after the pilot round was clear. For the participants, the most valuable part of the pilot rounds was the opportunity to sit together with people with different information and

186

views on the organization and discuss important issues. Even though the discussions they had were not new to them, they appreciated the opportunity to take the time to discuss these important issues in depth together.

Based on our data it is not possible to attribute this knowledge and perspective sharing to the tool. However, the tool did not seem to hinder these discussions. We also observed that the participants interacted with the LESAT while building prevailing understandings. For example, sharing capability level preferences was used to communicate positions, spurring investigative discussions into each others' views. The positive feedback was given with respect to the design of the tool and that they could sit together and discuss important topics. This speaks to the presentation of concepts (practices) using capability level descriptions being positive, or certainly not negative, for the participants' ability to share knowledge and perspectives.

8.2.3 ENABLING PARTICIPANTS TO TAKE ACTION

Answering if using the LESAT enabled participants to take action requires more than a pilot where they use the tool. For example, the potential long term effects of increased sophistication of the participants' views from their collective learning experience cannot be judged on the basis of the pilot. Further, the participants were not given authority or official responsibility to act on the results of the pilot, as it was first and foremost a research project and not part of the participants' official tasks. However, we observed that the capability level descriptions did not influence the generated change initiatives, that a normative force was embedded in the best practice design of the tool, and that the tool did not function as a boundary object.

The descriptions of the desired capability levels and of those between the current and desired had little influence on the change initiatives generated by the participants. Vinge and Knudsen (2002) note that a prerequisite for changing how the work is done in health organizations, is that change initiatives need to be discussed and formulated in a language that reflects the actual work being done. The language used when discussing change initiatives should reflect the language and realities of the environment where the change is to take place (chapter 5.3.2). The language in the LESAT is foreign to the division and does not reflect the division language or realities

(the LESAT language not reflecting the division's realities is further examined in chapter 8.2.3). The intuitive idea behind having descriptions of capability levels is that participants can use the capability level descriptions to envision what the division could look like at a higher capability level, and through that image see what is missing and therefore needs to be done. In this way the descriptions of the best practices could enable participants to take action by showing what action to take. However, the participants' change initiatives were based on the prevailing understanding of the practice, which was typically decoupled from the contents in the tool and represented familiar issues in the division. In effect, it would not be too harsh to claim that the participants would have come up with the same change initiatives if they had been simply asked to discuss a predefined list of important topics in the division and to come up with change initiatives based on this discussion. We therefore attribute little of the change initiative generation to the tool itself, but rather to the task given to them of discussing issues and coming up with change initiatives.

The design of the LESAT had a normative effect on the participants. After determining a current state, the participants' default reaction was to suggest some higher desired capability level, sometimes even without having read the higher capability levels before the suggestion was made (see chapter 7.4.3). Going against this trend was labeled non-visionary rather than constructively critical to the practice. Whether the normative effect increased the participants' feeling of urgency is however unclear. There was a clear feeling during the discussions of "we have to do something", but there was not a clear relationship between this feeling of urgency and the use of the tool or the content of the tool. The urgency was most often based on issues that the participants already felt strongly about. In this sense, setting scores and the freedom to interpret the tool so that the content represents issues the participants feel strongly about can be seen as a way for participants to communicate their interests and increase the collective feeling of urgency on these issues. The task of setting current and desired scores may therefore increase urgency and enable participants to get collective support for taking action on issues they are passionate about. Even though decoupled from the contents in the tool, this can be seen as enabling action.

We found that the LESAT did not serve as a boundary object to our participants. To recap, boundary objects facilitate the actors in the process of building a common representation of the different actors' knowledge and interests, enabling collective action (Kaplan & Jarzabkowski, 2006, p. 13). When participants have a range of different interpretive frameworks and interests, Kaplan and Jarzabkowski (2006) argue that it can be very useful for participants to be allowed to modify the boundary object so that it can situate the different interests and perspectives. Interpreting the LESAT is such a process of modification, where the participants interpret the meaning to situate their perspectives and interests. Allowing a tool to be actively under construction helps the participants construct a common representation of the different participants' knowledge and interests that enables collective action (Kaplan & Jarzabkowski, 2006). This happens through a process of tacking back-and-forth (Star, 2010) from each individual's or local group's understanding and reasoning, and the common representation that is being built.

Our participants made up a heterogeneous group, where each individual had their own view of the division. Ideally, all participants would understand the LESAT and all the capability level descriptions in their own way (local representations). Through discussion the participants would tack back-and-forth, explaining their interpretations of the LESAT and establishing a common understanding of what the different elements of the LESAT represents in the division (global representation). Together the participants would model the division in LESAT terms and concepts, based on their views of the organization. They would not all have to agree with everything and could have a local representation that was different, but they would understand, contribute to and agree to use the global representation in the further discussion. This global representation can be paralleled with the prevailing understanding. However, the participants struggled to understand the contents of tool, finding the language and concepts hard employ. Thus, the content in the tool did not satisfy the participants' information requirements well enough for them to understand it and build their local understandings. Furthermore, we found that even though a prevailing understanding was established, the participants often remained in disagreement about how to interpret the LESAT or did not even understand each other's interpretations. It follows that they were not able to establish a common, global understanding of the

189

content, and as such, the commonly interpreted LESAT did not serve as the global representation it should have. The process of building a prevailing understanding was often less prioritized than agreeing on a score, scores that were more or less decoupled from the contents of the tool and the prevailing understanding. Consequently, the LESAT did not serve as a boundary object enabling collective action. Its failure rests on the participants struggle to comprehend the content and focus on capability level scores.

8.2.4 CONCEPTS AND THEORY IN THE LESAT

To fulfill the usefulness criteria, the concepts chosen for assessment need to be applicable and key to understanding the functioning of the organization. The theory coded into the tool should be suited to explain the underlying structures of the organization. In broad terms, this means that the LESAT should introduce concepts that are applicable to the division and that are important for the division to function well. Also, moving towards the best practices, represented by the fifth capability level of each practice, should be useful for improving the division. Lastly, the theory in the tool should match the realities of the division and support the users in their understanding of it.

The participants did not fully grasp the overall meaning of all the practices, and we observed several misinterpretations where the prevailing understanding became decoupled from the contents of the tool. Further, only eight practices were covered during the group pilot round, and though a few more were covered in the individual pilot rounds only the first practice in each sub-section was covered. This means that the empirical foundation for evaluating the underlying theory in the LESAT is limited. The applicability and importance of the concepts, and the theory's match with the realities of the division, is therefore considered on a predominantly theoretical foundation.

8.2.4.1 OVERALL TOPICS

The feedback from the participants about the overall topics covered in the pilot was that they were important issues to discuss for the division. The LESAT is divided into three sections, "transformation/leadership", "lifecycle processes" and "enabling infrastructure". Each section contains a set of practices associated with the processes

of that section. For these overall concepts in the LESAT to be applicable, it should be possible to describe the division as a set of processes categorized in leadership, lifecycle and enabling infrastructure processes. This is possible, as stated by Department Head 2 when asked if it seemed reasonable to use these categories to model the division:

> Department Head 2: Sure, I think you can. We have core processes like patient treatment and care, a whole lot of support activities like IT and HR, and of course we have leadership functions.

Further, most of the practices in the LESAT seem sensible to use in the division and are about very non-industry specific topics, for example planning for the future, aligning systems, incentives and policies, and communication internally and externally. However, we have identified what we judge to be significant mismatches between the theory in the tool and the realities of the division. The normative force of the best practices, and potential influence of discourse embedded in the tool, call for a thorough investigation of such mismatches. The identified issues are presented in the remaining parts of chapter 8.2.4.

8.2.4.2 THE CONCEPT OF TRANSFORMATION AND ITS UNDERLYING THEORY

The most central concept in the LESAT is the "transformation" concept. This concept is one of the main aspects that sets the LESAT apart from similar tools. The tool is created to support transformation efforts, which are large-scale change efforts that include large parts of the organization. Following the LESAT is supposed to help organizations become more flexible and ready to transform. The concept of large-scale changes is very applicable to the division. Being able to change, including large-scale changes, is very important in the division as campaigns, new political directions, budget cuts and a constant pressure to improve necessitates being able to change.

The applicability and salience of the concept does however not entail that all the theory about transformation introduced in the tool is suited to explain the underlying structures of the division. We have identified potential conflicts between the transformation theory introduced in the tool, and theory about change management in Norway and success factors identified by the participants for large-scale change

initiatives in the division (see chapter 5.7). The conflict is between the need for considering employee and patient participation in division change processes, versus a purely management-driven conceptualization of change efforts in the LESAT.

In the division, leaders must walk a fine line between undisputable, top-down decisions, and involving stakeholders in the process. Through their unions or associations, employees and patients expect to be heard and informed in all major decisions affecting them. They do not have the formal power to make decisions, but they have a right to be heard and included (see chapter 5.5). The practices and sections in the LESAT do not include considerations of balancing involvement with top-down decision-making.

The discourse embedded in the LESAT reflects the underlying theory, the Enterprise Transformation Paradigm. The best practices in Section I, "Transformation/ Leadership", reflect the implementation process of the Enterprise Transformation Paradigm. The discourse of Section I is characterized by a strong focus on the central role of senior management and the enterprise leadership. Throughout this section, "transforming" the organization is talked about as a responsibility that rests on the management alone, especially senior leadership. The enterprise's leadership is talked about in a way that makes them seem to have the ability to change fundamental aspects of the organization. The tool's discourse further implies that the organization's leaders are the legitimate actors to decide how the organization should change. This is illustrated nicely in a quote from section I.B,
"It is imperative that the enterprise leadership understands and buys into enterprise thinking because they will be required to create a vision for doing business, behaving, and seeing value in fundamentally different ways" (Nightingale et al., 2012, p. 9).

The LESAT legitimizes the leadership's position as change agents, therein their right and power to make changes, by explicitly specifying the leadership's role in several of the practices as well as emphasizing their central role several times.
When analyzing a discourse, Trowler (2001, p. 186) points out that what is absent from a text is often at least as important as what is present. In the LESAT, the enterprise leadership is explicitly mentioned on several occasions, and their role as

change agents is firmly established. By the power of their position as formal leaders they are legitimized to act as initiators of enterprise-wide change. No other groups, such as other occupational groups within the organization, are legitimized in this way through the LESAT's discourse. The need for other change agents is recognized in practice I.E.4 (see figure 18), but there is no mention of who this should be. They are spoken of as "key people" who are identified and equipped with decision-making power and appropriate skills needed to support the cause of transformation. The wording does not open for these change agents to be critical of the transformation initiated, rather they are to be trained in the cause and given responsibility to work for it. This is not a description that opens for change agents being union or patient association representatives. The change agents here are people recruited to the cause and could absolutely be useful in a change initiative in the division, but this practice cannot be read as covering the issue of involvement and co-determination in the sense described in chapter 5.5.

I.E.4	**Empower Change Agents** *Enable key people to inspire and enact change*	Change agents are sporadically distributed but do not have change authority.		There is formal identification of change agents, along with role definition, delegation of authority, definition of roles, and provision of training/education for all change agents.		Appropriately skilled change agents are assigned to key areas with the authority to effect changes.		Change becomes self-generating, initiated by employees as well as change agents.		Change agents are providing a critical resource of enterprise knowledge, skill and experience in transforming the extended enterprise.	
		C	D	C	D	C	D	C	D	C	D
	Indicators (Examples)	• Change agents have been designated and empowered. • Change agents operate throughout all areas and cross-transfer transformation implementation experience. • Process for developing transformation process owners and other change agents has been established.									

FIGURE 18: PRACTICE I.E.4 (LESAT VERSION 2.0, 2012)

I.E.7	**Empower Employees** *Decision-making at lowest possible level*	Centralized decision-making occurs in a hierarchical structure with limited delegation of authority.		Appropriate structure and training is being put in place to enable empowerment.		Organizational environment and management system supports limited decision-making at point of use.		Decision processes are continually refined to promote increased accountability and ownership at point of use.		Decision-making across the extended enterprise is delegated to the point of use.	
		C	D	C	D	C	D	C	D	C	D
	Indicators (Examples)	• Managers and supervisors serve as mentors and educators, promoting lower level decision-making. • The extent and types of empowerment are tailored to match the environment and people empowered. • Empowerment enables swift and effective decision-making closest to the point of use.									

FIGURE 17: PRACTICE I.E.7 (LESAT VERSION 2.0, 2012)

194

The only practice mentioning something close to the topic of employee participation is in practice I.E.7 (see figure 17). In this practice the organization is recommended to foster "empowerment" by delegating decision-making to the closest point of use. Empowerment is only presented as being concerned with delegating the right to make decisions to the point of use, so that decisions can be made without for example asking for a supervisor's consent. There is no mention of including employee and union representatives in decision-making processes beyond their immediate area of responsibility, for example at the enterprise-level. The view of "empowerment" in the LESAT is in line with a view described by Greenberg (1975), (Wilkinson, 1998) and (Wilkinson & Fay, 2011) as a view of employee participation focused on performance. Here employees are mainly invited to participate to issues relating to their immediate conditions at work, such as work schedules and pace, and participation is seen as a part of the management agenda (Greenberg, 1975; Wilkinson, 1998; Wilkinson & Fay, 2011). Wilkinson (1998) and Wilkinson and Fay (2011) describe this view of employee participation as exclusively aimed at enhancing employee contribution to the organization and it is ultimately introduced to improve efficiency. "Empowerment" in the LESAT has a narrow scope that does not cover employees' expectation of taking part in decisions regarding their work on a higher level of the organization. This is far from the division's system of co-determination (see chapter 5.5), which goes beyond allowing employees to make their own decisions when performing their own tasks. Through co-determination employees are also included in decisions regarding for example what new technology should be employed (for example if the division should introduce the LESAT), changes to the organizational structure and the way the work is organized. There is no trace of such a concept in the LESAT.

Although focus on the role of formal leaders, use of "change agents" and "empowering" employees are important issues for the division to consider when making changes, a very central element is lost. The co-determination system including both employee and patient representatives is deeply rooted in the in the Norwegian specialist health care services. One could argue that the central focus on stakeholders in the LESAT can be made to include focus on employees and patients. However, no focus is explicitly established on the inclusion of these considerations.

195

The co-determination system is an important underlying structure of the division, and the theory in the LESAT does not help users to understand it. Not only does the LESAT not give insight into an important issue for the division, but the tool also presents itself as a list of the most important things to consider during transformation efforts. As co-determination is excluded from this list, the LESAT could cause its users to become less attuned to the importance of co-determination for the success of change efforts. This could lead to prioritization and focus away from an important issue, tipping the balance between top-down decisions and involvement, and could therefore have a negative impact on the divisions' change processes. Taking this argument one step further in the direction of discourse as a socially constructive force, the discourse embedded in the LESAT may influence the way the users see change processes and the division. The discourse could legitimize the more top-down change approach implied in the LESAT and influence its users to rather think of "empowerment" than co-determination. The discourse in the LESAT could influence users to think of it as natural that those involved in the process, or "change agents", are supposed to be fighters for the cause, and not constructive and at times critical counterparts in decision processes. This in turn could undermine the culture of involvement, which is dependent on mutual respect and the parties' acknowledgement of each other's legitimacy. This would certainly be in conflict with the interests of important stakeholders, most prominently patients and employees and their associations. Even without the discourse changing peoples' underlying assumptions and views, having a checklist for large change initiatives in the division that does not even mention the issue of involvement or employee participation, is judged to be a problem for the usefulness of the tool in the division.

8.2.4.3 FORMAL SYSTEMS, QUANTIFICATION AND MEASUREMENT

The topics covered by the practice and the descriptions in the LESAT imply that quantification and measurement are key success factors for an enterprise. In the discourse of the LESAT, topics such as information gathering are often coupled with words such as quantification or measurement. For example, practice I.G.4 is about following up the implementation of change efforts. A very central focus is the quantification and aggregation of information gathered, see the description of capability level 2 for example, "Process is under development to permit tracking and quantification of progress of the detailed implementation. Data from some projects is

being reviewed" (LESAT Version 2.0, 2012, p. 18). No mention is given to other ways of gathering information about implementation, such as interviews or meetings with the people at lower levels who are experiencing and implementing the changes.

As seen in practice III.A.2, I.E.3 and III.A.1 (see figures 19, 20 and 21), the LESAT recommends a tight system of formal steering devices such as incentive systems based on performance measures and data-driven performance management.

Capability Levels

EP #	ENTERPRISE PRACTICES	Level 1	Level 2	Level 3	Level 4	Level 5
III.A.1	**Enterprise Performance Measurement System Supports Enterprise Transformation** *Transformation requires appropriate measurement*	Performance (e.g., financial, productivity, deliveries, innovation, etc.) is measured at the local rather than enterprise level. Measures are subjective in nature and data integrity is low.	Initial efforts are under way to adapt or modify performance measurement systems to compensate for the inadequacies of the scope or scale of the existing system. Data are objective.	Performance measurement system provides data to support and enable transformation at the enterprise level.	Performance measurement system scope is expanded to integrate with non-traditional measures of value creation (e.g., intellectual capital, balanced scorecard, etc.).	Performance measurement systems provide seamless information exchange across the extended enterprise and emphasize value creation for all stakeholders. Frameworks exist for assessing the performance of the enterprise, and metrics are continuously refreshed.
		C D	C D	C D	C D	C D

FIGURE 19: PRACTICE III.A.1 (LESAT VERSION 2.0, 2012)

Capability Levels

EP #	ENTERPRISE PRACTICES	Level 1	Level 2	Level 3	Level 4	Level 5
I.E.3	**Align Incentives** *Reward the behavior you want*	There is sporadic use of incentives, and awareness that some incentives elicit localized optimization and harm interactions across functional boundaries.	Parts of the enterprise have implemented incentives that reward and encourage achieving enterprise goals by working across boundaries.	Executive compensation and employee incentives are linked directly to attainment of enterprise objectives.	Incentive systems successfully contribute to achievement and sustainability of enterprise objectives.	Enterprise incentives are deployed, with measurable success across the extended enterprise.
		C D	C D	C D	C D	C D
	Indicators (Examples)	• Incentives include a balance of monetary rewards, non-monetary rewards and recognition to encourage transformation activity. • Incentives are based on performance measures that encourage transformation activity. • Incentives encourage local improvements that will benefit multiple processes and/or value steam performance.				

FIGURE 20: PRACTICE I.E.3 (LESAT VERSION 2.0, 2012)

Capability Levels

EP #	ENTERPRISE PRACTICES	Level 1	Level 2	Level 3	Level 4	Level 5
III.A.2	**Enterprise Stakeholders Pull Required Metrics** *Data on demand*	Lagging performance measures are reported through regularly scheduled standardized reports. What is shared may not be relevant or actionable. Specific requests for measures require extraordinary (often manual) effort.	Internal users actively provide traditional performance information to assist users in planning and programming activities. Emphasis is on metrics that indicate progress or activities (i.e., project status, number of initiatives, etc.) but ignore outcomes.	Internal users are able to directly access and use performance information to make trade-off decisions. There is a blend of progress and outcome measures.	Internal users are able to pull performance and other value creation information to support decision analysis in the format desired. External partners have access to the necessary metrics to support continuous improvement. Emphasis is on outcome metrics (productivity, cost reduction, etc.) rather than progress metrics.	Stakeholders across the extended enterprise generate and share timely enterprise performance data. Data reflect extended enterprise results.
		C D	C D	C D	C D	C D
	Indicators (Examples)	• Financial and performance measurement data can be accessed as needed in user-defined format. • Financial information can be extrapolated to forecast outcomes. • Enterprise performance measurement system provides up to date information on request and constantly refreshes information needs.				

FIGURE 21: PRACTICE III.A.2 (LESAT VERSION 2.0, 2012)

The amount of attention given to measurement, data collection and quantification presents an image where more documentation and use of numbers to aid decisions and to control the organization is the answer. Informal systems are recommended replaced with formal systems. Implementing formal incentive systems such as linking executive compensation to indicators for enterprise goal achievement (Practice I.E.3, figure 20) is presented as how employee behavior is controlled and changed.

Focus on formal steering mechanisms is further illustrated in practice I.E.6 (see figure 23), considering the communication in the organization. A striking feature of this practice is that the only communication mentioned is vertical communication. The focus is on leaders' communication with their subordinates, establishing a "two-way communication" where "leaders are accessible and visible". The informal structures that are considered an important feature of the Norwegian specialist health care services necessitates communication horizontally as well as vertically, since not all real influence goes through the vertical command chain. These structures are completely neglected in this top-down, bottom-up depiction of communication. The discourse here may therefore influence the organization to consider this as covering the need for improving communication. This would not be a useful input to the organization, as it could lessen the focus on improving horizontal communication.

FIGURE 23: PRACTICE I.E.6 (LESAT VERSION 2.0, 2012)

EP #	ENTERPRISE PRACTICES	Capability Levels				
		Level 1	Level 2	Level 3	Level 4	Level 5
I.E.6	**Establish Open and Timely Communications** *Right information at right time*	Communication is largely top-down, limited, and lagging. C | D	Basic communication mechanisms are employed but are not uniform; communication strategy is under development. C | D	Enterprise leaders are accessible and visible, developing two-way communications in open, concise, and timely manner. C | D	Communication processes are undergoing continuous refinement and information is exchanged or can be pulled as required. C | D	Comprehensive system of two-way communication is employed throughout the extended enterprise. C | D
	Indicators (Examples)	• Open and timely communications exist among stakeholders, i.e., regular meetings with employees, newsletters, etc. • Technology has been leveraged to speed communications flow and accessibility while filtering unnecessary communications. • Employee input is valued and plays a key part in decision-making.				

FIGURE 23: PRACTICE I.E.6 (LESAT VERSION 2.0, 2012)

FIGURE 22: PRACTICE I.E.8 (LESAT VERSION 2.0, 2012)

I.E.8	**Encourage Innovation** *From risk aversion to risk rewarding*	Innovation initiatives are sporadic and *ad hoc*; security, stability, and risk aversion drive most decision-making. C | D	Initial efforts are under way to develop systems, processes, and procedures for fostering innovation. C | D	Innovation initiatives are under way in selected areas; measures for assessing impact are in use. C | D	Innovation initiatives are flourishing across the enterprise; prudent risk taking is encouraged and rewarded. C | D	A comprehensive innovation program is implemented and positive results recognized across the extended enterprise. C | D
	Indicators (Examples)	• The review process for suggestions has been streamlined and gives clear visibility of the progress of each suggestion. • Suggestion programs have been properly incentivized to give recognition to originators of innovative ideas.				

FIGURE 22: PRACTICE I.E.8 (LESAT VERSION 2.0, 2012)

When describing the characteristics of Norwegian hospitals in chapter 4.4, we found that a combination of high levels of trust and informal structures are important elements of the Norwegian specialist health care services. Christensen et al. (2007, p. 117) found that this trust-based feature is a strong overall feature of the system, also covering the leadership's relationship to the professions. He argues that the high level of mutual trust contributes to goal acceptance, reducing the need for formal, data-driven steering devices and thereby keeping down the technical complexity and monetary costs of such systems. He also found that it supports the leaders' high levels of autonomy throughout the organization and that traditionally, regulation through trust has been an important feature of the medical profession (Christensen et al., 2007, p. 117).

The discourse and practices in the LESAT has exclusive focus on formal and non-trust based steering devices, which stands in contrast to the trust-based systems in place. A case can be made for substituting or supplementing trust based systems with data-driven, formal steering devices. For example, the professions - especially the doctors - are the ones who drive costs by controlling what the patients receive of medication and services (Sveri, 2004). Introducing formal steering devices may be a way to gain control of the costs. However, by only focusing on these systems, the LESAT leaves out a very important aspect of the division's underlying structure in its underlying model of enterprises. For example, following the recommendations of the LESAT in practice I.E.3 level 3 could entail linking compensation of the department heads to indicators developed for division goals. This could for example be done in the form of giving department heads bonuses based on how many patients are treated in their department. In the Norwegian specialist health care service as it is today this seems simply absurd, and we expect that many department heads would react to this as an insult and a declaration of distrust, as management would with such a move imply that they thought the department heads needed to be controlled with monetary incentives to do their job.

Not only may following the LESAT cause users to neglect the informal systems, but formal and data-driven steering devices also stand in contrast to and would replace informal, trust-based systems. Gathering large amounts of data and making it

accessible for "decision-making support" substitutes a trust-based decision-making system. In their research on Norwegian hospital reforms, Christensen et al. (2007, p. 117) found that the introduction of more formal, external and non-trust-based steering tools challenge the trust-based regime. Performance measurement systems may be experienced as a reduction of trust, where the leader's trust in their employees to do their work is replaced by a system that quantifies and controls your actions. This effect is supported by the comment made by Department Head 1 (see chapter 7.5.5). An increased amount of documentation and incentive systems (such as activity based financing) have been introduced the last years in the division. Department Head 1 said that unit leaders experience the increased amount of documentation and quantification of their work as devaluating, and as a sign that they are not trusted to do their work. Threats to trust experienced between the parties in the division is a serious matter. Union Representative 1 illustrated that effective solutions to tough issues and conflicts can be achieved though informal conflict resolution when trust between parties is cultivated (see chapter 5.4.1).

In addition to reducing trust and shifting power in the organization, formal data-driven steering devices have high technical complexity and monetary costs. Not only is it costly to acquire, maintain and adjust such systems, but there is also a significant cost associated to having people to handle the data and having employees spend valuable time on filling the systems with data. The time and effort spent on documentation was a major concern voiced by our participants during the pilot and follow-up interviews (see chapter 7.5.6). During the group pilot round a participant reacted very negatively to the phrase "Performance measurement system scope is expanded to integrate with non-traditional measures" (LESAT Version 2.0, 2012, p. 34). This is a recommendation found in capability level 4 of practice III.A.1, a practice that was covered during the pilot rounds. When discussing the desired capability level for this practice, one participant went so far as to propose a low desired score because he did not think the higher capability levels were the right way to go for the division. Our participants were concerned with too much time spent on documentation and too much focus on numbers at the expense of focus on quality and professional specialization (Norwegian: fag). The effect and quality of treatment are very hard to measure, and the participants were concerned about a decoupling of

discussions about numbers and what was behind those numbers. As representatives of important stakeholder groups, we take our participants' concerns as indicative of important stakeholder concerns in the division.

Furthermore, the LESAT itself may contribute to an increased use of numbers in the division. During the pilot rounds the participants had a central focus on numbers, or scores for the practices, and these discussions were often decoupled from the content of the tool. At times, the discussions about scores also became decoupled from the prevailing understanding of the practice, meaning that the number chosen for capability levels, in effect had no common meaning to the participants. The participants themselves commented on the increasing use of numbers in the division, and warned against that this increase might lead to less focus on quality and professional specialization (Norwegian: fag). They note that numbers must carry meaning. Otherwise, numbers are just numbers, and do not add any value. However, as numbers became decoupled from the content of the tool or the prevailing understanding, the scores they set for the practices in the LESAT became just so - only numbers. As our participants where very aware of potential effects of the increased use of numbers and the potential for decoupling, we do not see this as a consequence of their preference towards numbers, but rather as a potential effect of using the tool. Therefore, in addition to having the content that encourages an increase use of quantification, the tool itself may become an additional system for quantification and use of numbers decoupled from their meaning.

The discourse and recommendations in the LESAT is here judged to be in conflict with important stakeholder concerns. This includes leaders, employees and their representatives who today have influence through informal and trust-based systems that may lose influence and control over their own work situation with more formal, quantified data-driven steering devices. At best the theory in the LESAT does not support users' understanding of the underlying structures of the division. Taken further, the LESAT could influence users through its recommendations and discourse, to see steering mechanisms as something that should be left to formal data-driven systems. This can harm the division through reducing the trust between parties and

substituting informal and trust based steering systems with resource demanding data-driven systems.

8.2.4.4 ASSUMPTIONS ABOUT THE ENTERPRISE'S ROLE IN ITS VALUE CREATION PROCESS

Use of the concepts upstream and downstream stakeholder in the LESAT brings with it assumptions about the enterprise's role in its value creation process. Using these concepts necessitates viewing the enterprise's role as a central entity that the value stream goes through. Stabell and Fjeldstad (1998, p. 413) note that Michael Porter's value chain framework from 1985 is the "accepted language for representing and analyzing the logic of firm-level value creation". They further state that it appears to be well suited for traditional manufacturing firms. The value chain is a sequential representation of the activities in the organization, where the value creation logic is to transform inputs into outputs (Stabell & Fjeldstad, 1998). The concepts of upstream and downstream stakeholders are meaningful and often used in conjunction with this framework. Stabell and Fjeldstad (1998) argue that the primary activity typology found in the value chain is less suited to analyze activities in many service industries. For hospitals, they propose an alternative value configuration, termed a value shop (Stabell & Fjeldstad, 1998, p. 414). This is a value configuration where the value creation logic is concerned with solving customer problems, and it is modeled as a cyclic chain of activities. Using this value configuration, the concepts upstream and downstream stakeholder become much less meaningful.

Furthermore, the value chain is not in line with how the employees in the division model its role in the patient treatment process. They view their role in patient treatment as being one of several entities surrounding the patient, who all contribute to provide the necessary care. The other entities that provide care to the patient contribute both before, during and after the division provides its services. The centrality of the patient in this view is illustrated by the discourse in the division, with commonly used phrases such as "the patient is in the center" and "the patient's own themselves". The concepts upstream and downstream stakeholder were found to be difficult to use for the participants, as the other entities taking part in the patient treatment do not strictly contribute before or after the patient receives treatment in

division. For example, municipalities both provide primary health services before and after the patient has received specialist healthcare in the division.

Consequently, the practices II.B.1, II.C.1 and II.D.1 in the LESAT, where upstream and downstream stakeholders are central concepts, became difficult to understand. These practices are concerned with communication, inclusion and coordination of organizations and groups that the organization depends on for value creation, which are valuable topics of discussion. For example, it would make sense to talk about communication, inclusion and coordination of municipalities, doctors, patients, families, and other entities in the specialist health service system and the elements internal to the division. However, the distinction made between upstream downstream stakeholders did not make sense in the division, and the participants were rendered confused and unable to tap into these ideas. These concepts do not contribute to the participants' understanding the functioning of the organization. Furthermore, using such concepts means employing a new discourse where patients are given a less central role. Considering the potentially socially constructive force of discourse, such a change in discourse could bring the employees to consider the needs of the division to be more central than the needs of the patient. This would go against the concerns of patients, an important stakeholder group.

8.2.4.4 PRIVATE MARKET AND MANUFACTURING INDUSTRY ASSUMPTIONS AND DISCOURSE

A quick glance at the discourse in the LESAT immediately brings to mind images of production industry enterprises operating in a private market. Assumptions better suited for private market and manufacturing firms than for public Norwegian hospitals can be found throughout the LESAT.

One example of an entire practice that is unfit for the division is practice I.E.8 (see figure 22). This practice reflects the LESAT's orientation towards manufacturing enterprises. The practice is called "encourage innovation", and encouraging risk-taking is a central focus. It is possible to interpret this practice in a way that does not conflict with important stakeholder values. For example, encouraging risk could mean allocating resources to projects testing new treatment methods that may fail, but can also contribute to life-saving treatments for patients. However, a general

encouragement of risk-taking may conflict with the interests of important stakeholders such as patients, their associations and external contributors to patient care. Manufacturing enterprises can take calculated risks that may result in poor quality for a time, which can result in reduced sales and perhaps be a blow to the enterprise's reputation. This would still end up as an economic consequence, perhaps by reducing the enterprise's competitive position. The same logic does not apply in Norwegian public hospitals, as taking calculated risks that put the "production" quality at risk for the sake of innovation puts people's health at risk. Such risk-taking would conflict with the concerns of most of the important stakeholders, especially patients and relatives, the society and patient associations. With the explicit focus on risk taking without a specification of what kind of risks are encouraged, we consider this practice to be in conflict with stakeholder concerns.

Early in the pilot rounds our participants started labeling the language in the LESAT as having an "industry" and "market language". The choice of vocabulary is a central part of the presentation of the underlying theory. The theory is discursively mediated, as the words in the tool are what paints images of the content in the users' minds (see chapter 2.2.3). The discourse in the LESAT is characterized by a vocabulary oriented towards private market and manufacturing industry with words such as "customer", "market", "competitive advantage", "growth" and more. Practices II.C and II.D for example use the concepts "customer", "marketplace", "development process", "products", "product usage data", "downstream customer value", "customer's business solution", "manufacturing" and "product and process design". To grasp the entirety of the implications of these practices for the division, one has to envision the division as a competitive organization with customers and products. Modeling the division in such a way is not judged to be useful to users trying to understand the underlying structures of the division. For example, the term "customers" brings to mind a consumer who chooses their offerings, and who's wallet size, judgment of quality and needs determine what their money is spent on. Patients in the public specialist health care services on the other hand do not have much of a choice of where to go for treatment. They do not pay for the services, meaning that more patients getting treated and more offerings provided cost more but does not generate any revenue from "customers" who are willing to pay (see chapter 5.2.1). The

meaning of "market" and "competitive advantage" in the division therefore becomes absurd in the context given in the LESAT. The division has a responsibility, not a market potential, and trying to model the division with such terms becomes misleading rather than helpful for uncovering the division's underlying structures.

Further, drawing on discourse analysis we understand the discourse embedded in the tool as having potential to influence the construction of knowledge and meaning in the organization. This means that not only does the private market and manufacturing oriented discourse reduce the applicability of concepts in the LESAT, but it may also influence the way in which the organization sees itself and its mission. Our participants were very concerned with the "market" and "industry" language in the LESAT and the values that are communicated through it (see chapter 7.5.4). Some participants were provoked by the language used, feeling appalled at the thought of talking about treating and caring for sick people with "cold and cynical" words like production. Speaking of the constructive effects of discourse in their own words, they argued that if you see a patient as a customer, you will end up taking a different approach to the issues you are attending to.

An important perspective from the participants is that the language used by employees in the division serves as a tool for a balancing empathy and an emotional distance to their work (see chapter 7.5.4). Working with tragic situations and people in need of help demand an ability to understand them and reach out a helping hand, while at the same time keep an emotional distance and not let everything in. This is a challenging mental exercise demanding certain ways of thinking and communicating with co-workers about their work. Trowler (2001, p. 186) argues that a discourse provides the language resources you need to think about and describe alternatives. The language employees can draw on to describe and think about their work therefore becomes an important resource for health care employees. The participants argued that the language used in the division should be "value based", implying the values in the division such as the division's responsibility to care for and treat sick people. They argued using the LESAT's manufacturing-oriented language with terms such as "production" and "customer" to talk about the division's work created too much of an emotional distance, tipping the balance too far in the favor of distance. In this sense

207

the discourse in the LESAT goes against employee values in a very direct way. Considering the potential constructive effects of discourse further, increasing the distance between patients and division employees implies that patients are met with less empathy. This goes against patients' concerns of being treated with understanding and being offered a helping hand beyond the "production goals" of the enterprise.

8.2.5 FAIRNESS

The self-assessment should satisfy the most important stakeholders' most important concerns. For the tool, this means that the recommendations of the tool and influence of the discourse embedded in the tool should be in line with important stakeholder values and interests. Judgments about conflicts with important stakeholder concerns have been made throughout this chapter where relevant. This section includes further fairness concerns that have not fallen under previously discussed topics.

The discourse embedded in the tool may have a significant impact on the self-assessment process. This potential impact is increased by the normative power that the capability levels were found to have, as users believed the higher levels to describe better practices. The discourse of the LESAT is characterized by a relatively strong management jargon, frequently using concepts from management theory such as "stakeholder", "core-competencies" and "enterprise". Klev and Levin (2012, p. 110) argue that even though different parties are represented in a decision-making process, this will not guarantee their influence because the participants will have different prerequisites for getting their interests across. This means that even if the committee members of the self-assessment are chosen so that the most important stakeholders are represented, this does not guarantee that their interests and values will be fairly represented in the process. Stein Bråten (as cited by Klev & Levin, 2012, p. 110) argues that participants make sense of information through different ways of reasoning, and that the models used to interpret situations are based on the person's experience and education. When a given way of making sense of information, or "model", is considered the legitimate one, those actors who are competent in this model will have a better ability to get their arguments across. Stein Bråten (as cited by Klev & Levin, 2012, p. 110) found that in discussions, "model strong" actors are able to use information to make their own arguments stronger and

clearer, while the "model weak" actors find it much harder to use information to produce arguments to support their interests. This is highly related to the relationship between discourse and power described in chapter 2.2.3, where the discourse affects what ways of relating to concepts become "ruled in" or "ruled out". Those participants who are best able to connect their interests to what is ruled in are the ones who will be able to generate the arguments that serve their interests. By choosing to use the LESAT to guide discussion, the discourse embedded in the tool is legitimized and sets the standard for what is ruled in or out. The strong management jargon found to dominate the LESAT's discourse may therefore legitimize the models of those with management experience and education. Those who master the management jargon will more quickly understand what is being discussed and how their interests can be protected, as well as save effort on comprehension that can be spent developing arguments. In this way, the management jargon influences the power distribution by legitimizing the models of those with management backgrounds. However, this impact on the power distribution may be mitigated in the division. We observed that legitimacy was attributed to those with experience in the division. Therefore, employees who are not familiar with the management jargon may still attain power through their knowledge and experience. Even with this effect however, the management jargon in the tool still makes the LESAT harder to employ for those without management education background. The tool's influence on the power distribution is judged to not be in line with important stakeholder interests, as important stakeholders without management education background may experience that the power distribution is shifted to their disadvantage.

8.3 EFFORT

In this section we will evaluate the LESAT with respect to the effort criteria. In chapter 8.3.1 we discuss whether the level of effort needed by the self-evaluation committee to discuss and respond to the tool is so high that it negatively affects the participants' active participation and attitudes. This is important to consider because the quality of the results is highly dependent on the participants. Chapter 3.3.2 discusses the resources used in the assessment compared to the benefits gained.

8.3.1 EFFECT OF HIGH EFFORT ON ATTITUDES AND ACTIVE PARTICIPATION

The LESAT was very hard for our participants to employ. Formulations became ambiguous or difficult to understand, and most of their effort was spent on trying to define what was asked in the practice. This meant that considerable time was spent thinking about what the practices meant, leaving less time, energy and focus to discussing the topic raised in the practice. Being repeatedly interrupted because of a need to understand words or expressions that could have been written in a language more appropriate for the division was experienced as frustrating. The participants commented that this task took up too much energy and were not satisfied with the effort required to use the tool.

8.3.2 RESOURCES SPENT ON ASSESSMENT VERSUS BENEFITS

We cannot say at this time how much the pilot affected the future actions of the participants. However, we saw that although the topics considered in the LESAT were useful to discuss, they did not discuss any topics that they had not discussed before. Also, while the participants learned from each other's insights, they did not learn from the tool. Furthermore, the ideas for change initiatives that were presented during the process could not be attributed to the LESAT. Moreover, in the seven hours that was spent on the group assessment, the participants only made it through eight practices. The extensiveness of the task of conducting the entire LESAT (43 practices) in this same fashion would therefore be daunting. To add to this, much of the time was spent interpreting the practices, in an attempt to deal with the foreign, complex and abstract language in the LESAT. This is an activity that does not add much value to the division. Additionally, we have found that the participants reduced the complexity of the task of understanding the practices and deciding on capability levels by employing selective focus. This means that the effort spent on interpreting the content of the LESAT only included deliberation on fractions of the practices. This makes the time spent on interpretation even less valuable. Following, the benefits gained from using the LESAT without adapting it further to the division would not outweigh the effort of using it.

8.4 Concerns not Covered by Predetermined Criteria

As can be seen from figure 13, two of our findings in chapter 7 cannot be mapped into the predetermined criteria of comprehensibility, usefulness and effort. One of these findings is that busy leaders want more time, not more tools. The leaders in the organization already have so much documentation and meetings to attend to that have been introduced in recent time, making them negative towards the introduction of more tools. The second finding is that they already have similar tools. The participants noted that the LESAT is similar to a tool that they already have, called Ledelsens Gjennomgåelse. They also noted that they already have more guidelines than they can follow. As a consequence of these findings, we add to our evaluation that if the LESAT is introduced into the division without altering the existing systems and tools in use, or altering and perhaps piecing up the LESAT to fit into these systems, it will create double work for leaders that are already pressed for time.

8.5 Final Evaluation

In this final evaluation the findings and judgments from above are drawn together. A concluding summary is presented for each of the predetermined criteria before a final evaluation is presented lastly.

8.5.1 Comprehensibility

The structure of the tool and how to use it was intuitive, and explanations provided are found to be clear, simple and exhaustive. However, the language does not match the users' competence level. In its original English version, the LESAT is almost completely incomprehensible to the likely users of the tool in the division. When translated to Norwegian the tool became more comprehensible, and the participants could use it to guide and contribute to their discussion to some extent. However, the language was still experienced as very foreign and hard to employ by the participants, and central concepts in the tool became ambiguous when interpreted to the division. The LESAT, even in Norwegian, is far below an acceptable comprehensibility level for use in the division.

8.5.2 USEFULNESS

The presentation of concepts in the LESAT did not support a process where the participants learned from the theory in the tool. Low comprehensibility and central focus given to the task of determining scores in the LESAT led to decoupling of participants' discussions from the content of the tool. Participants did however become more informed and understanding of each other's views, but this was mostly a result of the process, rather than a result of using the LESAT. There was no indication of the LESAT enabling the participants to take action any more than what could be expected from the same participants discussing the division without the LESAT. The desired capability levels had little influence on the generation of change initiatives, and the LESAT did not serve as a boundary object. The capability level design of the LESAT had a normative effect, but the decoupling of discussion from the contents of the tool mostly connected this effect to issues participants were already passionate about. The LESAT's three overall topics of transformation and leadership, lifecycle processes and enabling infrastructure are applicable to the division. However, the concept of transformation, assumptions about the existence of private market and manufacturing concepts and mechanisms in the enterprise, and the assumptions about the enterprise's role in its value creating process are found to be ill suited to the division. Also, the conceptualization of steering in the LESAT is characterized by an exclusive focus on formal systems, quantification and measurement, which effectively overlooks important structures in the division. The severity of these conflicts is significant. The LESAT is therefore not suited to explain the underlying structures of the division, and following its recommendations may be harmful to the division and go against important stakeholder concerns. Furthermore, the management jargon in the LESAT has potential to influence the power distribution in the assessment process itself by making it easier to employ for those with management education background.

8.5.3 EFFORT

The effort needed to understand the contents in the tool was too high, and the amount of time spent on each practice was not outweighed by the benefits. The participants' reflection process was repeatedly interrupted by discussions about the meaning of words and concepts and this was experienced as frustrating. The time spent on understanding the content caused the participants to only cover eight practices. Additionally, this interpretation process did not lead to the participants learning from the tool, making this time spent on interpretation of little value.

8.5.4 FINAL EVALUATION

The participants were unable to construe a common image of the organization with helpful theoretical influence from the LESAT. The process was hindered by the very low comprehensibility of the tool, effectively making the participants unable to use it and the effort required high. The participants spent most of the time trying to interpret the content, discussing topics they were already familiar with without influence from the theory in the tool, and discussing what scores to set even though the scores often became devoid of meaning beyond a numerical value. Additionally, the underlying theory of the tool is inappropriate for use in the division regardless of the participants' ability to comprehend and learn from the LESAT. The recommendations and discourse in the LESAT neglect and threaten positive aspects of the division such as the trust-based and informal systems, existing power relationships, the co-determination system and the existing value-based discourse in the division. With respect to the division, the LESAT therefore is judged to be neither comprehensible, useful, or worth the effort. The LESAT is also found to overlap with existing systems and tools. Managers in the division are pressed for time, meaning that their time is a very valuable resource. In conclusion, the LESAT is judged to be both inappropriate for the division and not worth prioritizing over existing systems and management tasks.

9 DISCUSSION

In this chapter we discuss three central issues related to our findings, then research limitations and lastly managerial and research implications. The first central issue, discussed in chapter 9.1, is that not all of the findings from our pilot project mapped into the predetermined criteria developed from theory. The next two chapters are discussions where we explore underlying reasons for why the LESAT is evaluated to be inappropriate for the division. The problems identified are argued to occur because the overall design of the LESAT is based on a misconception of SATs serving as measuring devices, and that the content of the tool is based on managerialist assumptions that do not apply for the Norwegian specialist health care services. Following this discussion, chapter 9.4 covers research limitations, and chapter 9.5 and 9.6 cover managerial and research implications respectively.

9.1 REMARKS ABOUT CRITERIA

In order to let our research be guided by our participants' concerns, in line with an action research strategy, our approach to data collection and presentation was mainly an inductive one (see chapter 4.2). This was combined with a more deductive approach to casting the final judgment on the tool, in order to ensure that we considered all the most important questions. The latter entailed using a structured approach, following detailed criteria for SATs. The low comprehensibility of the tool hindered the participants from tapping into the content of the tool, and resulted in the pilot rounds only covering eight out of 43 practices. Following the usefulness criterion, which includes consideration of underlying theory, structures and discourse, we therefore considered aspects of the tool that would not have been encountered in a completely inductive approach that followed participants' concerns. The underlying theory of the LESAT was found to be in significant conflict with the underlying structures of the division, and a deductive approach was therefore fruitful. Our predetermined criteria served us well as a framework for evaluation, guiding us to reveal aspects that we otherwise would not have discovered.

When mapping our findings into the criteria for the evaluation it became apparent that not all the findings corresponded to a criterion. The findings that could not be mapped

into the criteria have the commonality that they were both concerned with the current state of the organization. When expanding on the criteria of Rebolloso et al. (2002) we used criteria for self-assessments as a foundation, and then considered what aspects of these criteria could be attributed to the tool. However, in doing so we did not consider how introducing a tool to an organization also depends on the needs of the organization, not just whether the tool has a potential have a positive impact on the organization. The criteria developed by Rebolloso et al. (2002) are very attuned to the organization and its benefits from using the tool, but have no explicit comment on the opportunity cost of the tool. No attention is given to considering other potential tools or systems to provide the same benefit, or towards the potential value of the time spent using the tool and alternative uses of this time. Additionally, there is no reference to exploring the need for the tool in the organization. A tool can be very beneficial for the organization, but still not be needed because they already have systems providing the same benefits. Therefore, as Rebolloso et al.'s (2002) criteria contain no reference to opportunity cost of introducing the tool or the need for the tool in the organization, and our theoretical exercise of considering criteria for self-assessments did not alert us to this need, our resulting criteria ended up having no consideration of these matters. Our initial inductive approach to the data was designed to capture such unanticipated findings, and this was an important finding for our final evaluation. The developed criteria from chapter 3 were well equipped to explore the potential of the LESAT in the division, but before a final evaluation could be done, we needed to consider the division's initial needs and systems. There is no indication of this finding being unique to the division. It is reasonable that if an organization is considering introducing an SAT, they must both evaluate the potential of the tool, which can be done using our three criteria, and consider the organization's needs and existing systems.

9.2 MANAGERIALISM IN THE LESAT

We found the recommendations and discourse in the LESAT to neglect and threaten the trust-based and informal systems in the division, the informal power and legitimacy of the professionals, the co-determination system and the existing value-based discourse in the division. These were threatened by the LESAT's recommendation of increased focus on measurement and formal steering devices, the

presentation of management as the only legitimate and natural change agents, presentation of participation as a part of a management agenda to increase efficiency, and the use of a discourse with concepts such as "core-competencies", "market", "competition" and "stakeholder". These aspects of the LESAT bear a strong resemblance to what has been expressed as "managerialism". Trowler (2001, p. 185) explains managerialism as "providing a guide and justification for behavior oriented towards efficiency and economy, market responsiveness and the control of employee behavior towards these ends by managers." In addition, we find the following description of managerialism useful:

> (. . .) The world should be a place where objectives are clear, where staff are highly motivated to achieve them, where close attention is given to monetary costs, where bureaucracy and red tape are eliminated. If one asks how this is to be achieved the managerialist answer is, overwhelmingly, through the introduction of good management practices, which are assumed to be found at the highest pitch and most widely distributed in the private sector. (Pollitt (1993) as cited by Traynor (1996, p. 317))

Managerialist research is argued to unreflectively assume the managerial perspective to be the natural and exclusive starting point for analysis of organizational issues, marginalizing or overlooking alternative approaches (Keenoy, 2009, p. 464). By doing so, managerialist theories and discourse aim to answer the question of what can be done with management to solve a problem, not what can be done with the organization.

An exclusive management perspective is not the natural starting point when trying to understand the division. The division is a Norwegian public, professional bureaucracy with a mandate to care for and treat sick people. Analyzing the division with management as the natural and exclusive starting point neglects important underlying structures of the division related to these characteristics. As a professional bureaucracy, those providing the medical services, the professionals, have a high degree of informal power and legitimacy. An exclusive focus on management neglects this very strong and powerful group of professionals that does not take the

management's word as law. As a professional bureaucracy with a mandate to care and treat sick people, it is the health workers' professional and personal competences that take center stage. It is not first and foremost the management's ability to rethink new and smarter ways for how the organization can deliver value that will ensure the fulfillment of the organization's mission. If you see the organization as an instrument that the management has to optimize and steer in the direction they know to be best, ensuring formal steering devices that management controls and a constant feed of objective data makes sense. As the management is responsible for making the decisions and keeping the organization on the right course, they also have to be as fully informed as possible and have an organization that is highly responsive to their commands. However, by only casting light on the need for data-collection and formal steering devices, the informal power of professionals and the informal and trust-based systems that work parallel to such formal systems are neglected. These are central aspects of the division associated to it being a public Norwegian professional bureaucracy. The assumption of management being the natural and exclusive analytic starting point also neglects considerations of other alternative decision makers and contributors to decision-making processes. Seeing managers as the sole responsible actors in steering and change processes makes employees and patients unnatural actors in the scene of enterprise-level decision-making. The co-determination system is effectively ignored, a system that is considered a central aspect of the Norwegian public health services.

A language with terms such as "core-competencies", "market", "competition" and "stakeholder", what we have termed management jargon and the participants called technical, industrial market language, can also be described as managerialist discourse (Gilbert, 2005; Trowler, 2001). We argued that choosing a tool to guide a process legitimizes that tool's discourse, and that those users who are most competent in that discourse will have a better chance at getting their views across. Using a language easily recognizable to those with management education background then makes sense if the assumption is that these are the people who make the decisions and who's views should be represented the strongest. This is not the case in the division or in most entities in the public Norwegian specialist health care services, as most leaders have professional backgrounds. Using this language further assumes that this is the

language that should be used to discuss and think about the organization and change initiatives. This is not an assumption that holds for the division, in fact it does not hold for specialist health services in all of Scandinavia. Vinge and Knudsen (2002) found that the specialist health care services in Scandinavia have very similar characteristics, and that these characteristics are very special for the specialist health care services. The language used in this context is very specialized and reflects the infrastructure of the work and the health mandate. They found that in order to make changes in this context, the changes must be talked about in this language. The managerialist discourse in the LESAT is far from the language employed in the division, or the "health language" as our participants called it. Our participants ignored, commented on or struggled with almost all the central concepts in the discourse of the LESAT, and were completely unable to use it to discuss change initiatives. This lack of comprehension was the most easily observed problem resulting from assumptions embedded in the tool being in conflict with the division.

The managerialist language was also found to be a threat to the division's own discourse. We found that the language used in the division and about the work performed there should reflect the values of the health workers. Their language is a tool that is used to balance the need for empathy with the need for an emotional distance, a fine balance that can be disrupted in favor of emotional distance if the language in the LESAT is employed. With a managerial assumption, the management and associated management theory is the natural starting point to understand and speak of organizations. This neglects the health language and the need for this language in health organizations, needed both for the ability to describe change initiatives in a language that reflects the infrastructure of the work, and as a tool for handling the emotional balance that is essential for health workers to do a good job. This motivates the need for exercising caution and performing evaluations when considering introducing tools with managerialist discourse into this context. It should certainly be taken as a warning about the LESAT specifically to other entities in the Scandinavian specialist health services, as these have much more in common with the division's discourse than that of the LESAT.

The topic of introducing managerialism into the public health sector is subject to an ongoing debate. In his study of managerialism in the Australian public health care sector, Germov (2005) found that the use of managerial strategies in professional practice places constraints upon professional autonomy. In the British health care sector, Gilbert (2005) found that the introduction of managerialism has established a culture of contracts and targets and that the welfare-customer is now considered as a consumer. With this he argues that when managerialist discourse was introduced into the public health sector in the UK, this actually changed the view of the welfare-customer. Furthermore, he notes that by using a managerialist discourse, activities of care are objectified and go from being socially intimate activities to a product that can be managed. Trowler (2001) looks into the effect of introducing managerialism in the higher education sector in the UK and notes the absence of a discourse considering students as "real, diverse, responsible individuals who might react in different ways to the policies proposed" (Trowler, 2001, p. 187). Instead, a totalizing category is used by referring to them as "the market" (Trowler, 2001, p. 187). The effect of theories and practices from the private sector in the Norwegian health care services specifically have been warned against by authors such as Lægreid et al. (2005), Christensen et al. (2007) and Byrkjeflot (2004).

Arguments have also been made for the need to strengthen the role of management in the public health sector. Proponents of managerialism challenge professional opinions, and argue that they represent the interests of professions rather than service-users (Gilbert, 2005). The professionals will have strong values linked to their profession, and these may not coincide with the needs of all patients. For instance, Hall (2005) has found that physicians mainly value saving a patient's life, not the improvement of the patient's quality of life. She argues that a consequence of this is that areas such as preventive health, care of mental illness, and care of chronically ill or dying patients have traditionally not held us much attraction for physicians and therefore merited little attention. As a consequence, a management perspective could contribute to the public health sector by providing a means of managing and controlling the professionals. However, the main argument we have made against introducing managerialism in the Norwegian specialist health care services is that employing an exclusive management perspective leads to neglect of important and

positive aspects that are already in place. This is not an argument against that some practices derived from a management-oriented perspective can be useful. It is the neglect of perspectives other than the management-oriented one that results in the conflicts between the managerial discourse and practices, and the Norwegian specialist health care services.

The managerialist assumption that management is the natural and exclusive staring point of analysis fits with the aspects of the LESAT that were considered to conflict with the division. An exclusive management perspective is not the natural starting point when trying to understand the division, and such an approach will lead to the neglecting and threatening of the aspects of the division identified in the evaluation. The embedded managerialist assumption in the LESAT may therefore be the underlying reason for the LESAT to be so much in conflict with the needs of the division. The aspects of the division that are neglected and threatened by the LESAT are associated to the division being a Norwegian public professional bureaucracy with a health mandate. These are characteristics that are shared by the entire Norwegian specialist health care services. We therefore have reason to think that the LESAT would be ill suited for any organization in the Norwegian specialist health care services. It is also reasonable to expect other tools and theory that is founded on managerialism to be in conflict with the underlying structures of the Norwegian specialist health care services. By identifying pitfalls in the LESAT that can be attributed to managerialism, we feed into the ongoing debate about introducing managerialism into the Norwegian specialist health care services.

9.3 SAT Measurement Misconceptions

We have argued that the role of an SAT is to facilitate the process of construing an organization, meaning creating an image of it, which in turn serves as a foundation for organizational change (see chapter 2). We argued that the main potential of SATs is that they can facilitate a construal of an image of the organization through reflection, where participants draw on their knowledge and experience. Through introducing topics for discussion and theory that the participants can employ to understand their organization better, SATs can facilitate a construal process that results in the participants construing a more informed an image of the organization to serve as a

basis for action. However, when using the LESAT, much of the participants' time and focus was spent on discussions about what sores to set for the capability levels. These discussions about scores did not contribute to the participant's process of construing their image of the organization. On the contrary, these scores were often discussed and used without being coupled to any other meaning than being "high" or "low", and were often the topic of discussion instead of reflection about the organization or the contents of the tool. This focus on numbers is attributed to the design of the tool. The tool is designed like a questionnaire and gives the impression that the choice of capability levels for the organization is the central task, rather than the central task being reflecting and learning.

There is therefore a breach between the potential benefits of using an SAT to facilitate a self-assessment process and the reactions that are encouraged by the choice of design for the LESAT. This motivates a consideration of why the LESAT is designed the way it is. A reasonable explanation is that an assumption of the tool's creators is that the LESAT is in effect a measuring instrument. Evidence of this being the assumption held by the creators of the tool is abundant. Searching the LAI Enterprise SAT (LESAT) Version 2.0 Facilitator's Guide (Nightingale et al., 2012) for the word "measurement" gives 29 hits. The author's introduction to the tool is as follows:

> [The LESAT] includes leading indicators associated with organizational excellence; the indicators are based on principles developed through academic research and field experience in enterprise transformation (Nightingale, 2009). The tool is designed to both measure the current state and envision a future state, which allows users to assess and prioritize gaps between the current state and a desired future state. (Nightingale et al., 2012, p. 9)

If the tool *measured* some aspect of the organization (maturity of organization's capabilities in this case), the main results would be the numerical output of the process. The main result from the exercise would not be a construed image of the organization. The numerical output would be *measures* that would say something true about the organization. A design choice with central focus on establishing these numerical outputs makes sense if this is the underlying assumption of the creators.

The organization could use these numerical results to learn about the organization, and use them to base more informed decisions on. With the assumption that the tool is a measuring instrument, a design where the most important task is perceived to be the task of setting scores would be appropriate because the scores themselves would be the important result.

Consequently, the number-focused design of the LESAT is attributed to the authors' misconception that the tool has the ability to measure abstract concepts such as "the maturity" of the organization. Note that we do not argue against measuring in an organization, there are many situations where measurement would be a better choice than using an SAT, where valid measures are needed rather than an image construal process based on reflection. However, in chapter 2 we established that SATs, performed by participants within the organization and with the high level "indicators" (capability level descriptions) typically found in these tools, do not fulfill the quality criteria required of measurements. One could argue that these tools are measuring devices that produce invalid measures that the organization could use for some other purpose than learning about the concepts they claim to measure, but such a purpose is not considered here. SATs do not fulfill the quality criteria to produce measures that are anywhere near valid measures of the concepts they claim to measure, and we use the term "quantification of thoughts" to describe what these tools call "measurement".

This is very much the case for the LESAT, and the use of the LESAT by our participants illustrate well how the measurement assumption is a misconception. As a *self*-assessment, the numerical results of SATs are highly sensitive to participant's biases and interests. When using the LESAT, one participant for example interpreted a practice very selectively to be about user competence, a topic he was passionate about. He then proposed a very low score for the current capability level to drive home the point that the division was far too inconsiderate of user-competence. In capability maturity models such as the LESAT, the part of the process most related to a measuring mechanism is the part where participants read the "indicators" (capability level descriptions), and try to match their organization to the most fitting category. This process is highly participant dependent. When using the LESAT, our participants were guided by what they recognized. What they thought of and considered relevant

in the organization, how they interpreted the practices and capability level descriptions and how they weighted importance was very dependent on the individual. We assume that this applies to all SATs with abstract content that has to be interpreted by the users to acquire organization-specific meaning. This means that the reliability of these "indicators" is very low. This is the part of the process where the quantification takes place, which implies that the reliability of the numerical output from the process is very low. When the interpretation of the content is dependent on who reads it the measures not only become unreliable, but trying to understand the results later by reading the capability levels that were chosen becomes futile. This was illustrated by our participants as even the participant who had decided on a score would forget how he had interpreted that capability level when using it later. When participants went back to read capability levels that they had set at an earlier point, they often interpreted their own earlier intentions to be completely different from what they had been thinking when setting the score. This means that the numerical output from the process tells you close to nothing about the state of the organization, because you do not know how the "indicators" were interpreted or the participant's motivation for setting the scores. The result of this measurement misconception is a design in which generation of numerical outputs is given priority, but as these tools do not measure (at least do this very poorly), these prioritized numerical results become close to valueless.

This assumption of measurement combined with a number-focused design is not unique to the LESAT. The Maturity Model Grid design (like the LESAT and designs in chapter 2.3.3 and 2.3.4) has become a popular way of codifying and communicating best practices for a wide range of activities (Fraser et al., 2002, p. 244) (see chapter 2.3). These have the score-focused design that encouraged our participants to focus on scores. Most of these are presented as tools for external evaluators. However, the line between external and internal assessment is often blurred, and many such tools encourage organizations to use them as SATs for organizational development. Furthermore, it is not uncommon for SATs to be uncritically presented as measurement instruments, as the LESAT is. Bhasin (2011) for example presents an SAT for "measuring how lean" an organization is. A quick Internet search for "SAT measure" reveals a range of commercial and free SATs that

claim to measure things like how lean, mature, agile and innovative your organization is (see Jim Collin's "good to great" SAT (Collins, 2006), Ronald Mascitelli's book with embedded SAT (Mascitelli, 2007)). Through our work we have also come across three large Norwegian production firms that have adopted or developed SATs that are presented as measuring instruments with focus on creating, collecting and analyzing numerical outputs.

This uncritical presentation and design of SATs as measuring instruments speaks to a widespread misconception of what SATs are and how they work. In chapter 2 we established that for SATs to provide trustworthy measures, this requires fulfillment of quality criteria that none of the SATs we have come across satisfy. This is supported by our empirical study of one of these tools in use. The numerical output from these SATs are therefore results of a process better described as participants quantifying their thoughts than measurement. Not only does the numerical output from these tools not represent valid measures, but as illustrated by our empirical research, the centrality of numerical outputs may draw attention away from reflecting on the content in the tool and the organization. The centrality of numerical outputs may therefore have a very negative impact on SAT's potential for providing value to the organization by facilitating a reflection-based construal process of a more sophisticated and potentially collective image of the organization.

9.4 LIMITATIONS

To critically consider our work, and hopefully feed into future research, we note several limitations in our research. First, we underestimated the extensiveness of the process of interpreting and scoring each practice. It was a very time consuming activity, and this meant that we were able go through less practices than expected. Ideally, more practices should have been tested. However, we were not able to get access to the participants for more time than what was spent on the pilot, and we wanted to give them enough time to interpret each practices properly and not rush the process. Second, additional pilot rounds to test even further adjustments to the tool would have been favorable. By doing so, we would have been able to get closer to the source of what can be applicable in the organization, and what is not a good fit. However, with the limited time available for this research, this was not a possibility

and these considerations had to be done on a predominantly theoretical foundation. Third, our participants represented the most likely users of the tool in the division, and this did not include a person with management education background. The problem of comprehending the LESAT was more substantial than anticipated, and a person with management education background may have been able to tap into the content of the LESAT more successfully than our participants. Again, more pilot rounds would have been favorable, as we then could have included someone familiar with this language to see the effect of this on the process.

9.5 MANAGERIAL IMPLICATIONS

Even though many SATs are presented as measuring instruments, they are not and we call for issuing caution when using these tools. This is an important implication for organizations using, or considering introducing, SATs. As these tools are presented as measuring instruments, both in the way they are sold by their authors and by their design, it is easy to mistake the numerical outputs from these tools as good measures of concepts the organizations seeks information about. However, the numerical outputs from these tools are not valid measures, and should therefore not be used as a basis for decision-making. Also, they should certainly not be used as a basis for benchmarking, not against other organizations or even internally over time to benchmark the organization against itself. We do not claim that these numerical outputs cannot be used at all, there may for example be motivational value in scoring low, but they should not be used as measures. If an organization indeed seeks to measure something there are many good ways of doing this, for example using external audits or developing indicators for the concept in question. The SATs available today in academia and commercially that claim to let the organization measure itself, do however not present such a measurement solution.

When considering employing an SAT, it is important to keep in mind that the underlying theory of the tool may not be applicable to your organization. A potential conflict between the tool and the underlying structures of the organization may however not be obvious. We therefore urge organizations to critically evaluate the SATs they consider introducing. It seems reasonable that this should apply for all management tools, as it is in power of being a tool that SATs communicate theory.

225

The results from our evaluation of the LESAT to the division illustrates how assumptions from the authors of the tool is embedded into the tool. Through its discourse, content and design, the tool communicates a representation of the organization, recommendations, values and norms. If assumptions and values embedded in the tool do not fit your organization, introducing the tool can have a potentially harmful effect. Analyzing the discourse, underlying theory and design of the tool are therefore important tasks organizations should attend to when considering introducing a management tool.

9.6 RESEARCH IMPLICATIONS

Our research has implications for the design of SATs. Conceptualization of SATs as measuring instruments leads to designs that undermine a good facilitation of construal processes by leading attention away from reflecting and learning and towards creating numerical outputs. We call for a reconceptualization of SATs as facilitators of construal processes, rather than measurement instruments. Such a reconceptualization opens for and encourages new design choices for such tools, as ensuring the creation of numerical outputs is not a part of the design consideration. This does not imply that the popular designs of such tools today are valueless. Our empirical and theoretical critique of SATs' focus on numbers should for example not be taken as arguments against the design choice of having capability levels with descriptions. It is only meant as a critique of the focus on numerical outputs, for example the task of circling chosen capability levels, the room made for documenting these scores in the tool, and the impression that these scores are made available for collection. Without such a focus these designs can be fruitful. We attributed most our participants' lack of learning from the LESAT to issues of comprehensibility. There is still a reasonable theoretical possibility that with capability level descriptions that its users easily recognize and find easy to employ, that the discussion will not become decoupled from the contents in the tool. It is not hard to imagine a tool where the descriptions are clear and unambiguous, for example tailored to the individual organization using the organization's language and referring to concepts the users are familiar with. In such a tool, the descriptions could perhaps paint a picture in the users' mind that is an organization-specific image of the implications of the tool's underlying theory. The theory in the tool may be new, and the users may not find it easy to see its

implications for their organization. However, with capability level descriptions that are based on theory but described in organization-specific terms, this may help the users in envisioning the implications of the theory without even having to take the mental path up to the theory and back down to what it implies for the organization. Furthermore, a conceptualization of SATs as construal process facilitators opens for SAT designs such as the "bare minimum" design from chapter 2, or something else entirely. We encourage researchers and other SAT developers to design these tools with a central aim of facilitating good construal processes where participants learn from each other and from the tool, resulting in users acquiring a better understanding of their organization and its opportunities for change.

Our participants selectively focused on what they recognized, and this was very few of the concepts in the LESAT. This implies that what is very clear and specific to one person, the author of the tool for example, may be meaningless strings of words to others. For future research and development of SATs, consideration of what would be appropriate language in these tools would be very useful. We expect that the tolerable level of abstraction and references to theory will vary greatly between potential user groups, for example between professionals and academics. A challenge for future research is to find out what language, design and perhaps preparatory introduction to theory is needed for common user groups to employ all the concepts and theory introduced in an SAT.

10 CONCLUSION

SATs facilitate users in building an image of their organization, which can serve as a basis for changing the organization. In order to fulfill this potential, SATs need to be comprehensible to the users, be useful in the process of building an image of the organization, and worth the effort required to use the tool.

The LAI Enterprise SAT (LESAT) is evaluated as inappropriate for use in the Psychiatric Health Services Division at Innlandet Hospital Trust. The concepts introduced in the LESAT do not reflect the underlying structures of the division, and introduction of the LESAT into this context is considered potentially harmful. The sources of conflict between the LESAT and the division also apply to similar organizations. The LESAT is judged inappropriate for use in all entities in the Norwegian specialist health care services, and probably inappropriate for all Scandinavian specialist health care service entities. Our research demonstrates the potential severity of conflicts between the underlying assumptions of SATs and organizations that consider adopting them. We therefore issue a warning and call for thorough evaluation of SATs for the specific context they are to be employed.

Just like many other available SATs from academic and commercial sources, the LESAT is designed and presented as a measuring instrument. However, these SATs do not possess the necessary attributes to facilitate a measurement process that results in anything close to valid measures. Because of a widespread misconception of SATs serving as organizational measuring instruments, we call for a reconceptualization of SATs in the academic and commercial community. We urge developers and users of SATs to conceptualize these tools as facilitators of a process where users learn from each other and from the tool, resulting in a better understanding of their organization and its opportunities for change.

Bhasin, S. (2011). Measuring the Leanness of an Organisation. *Journal of Lean Six Sigma, 2*(1), 55-74.

Bhaskar, R. (1989). *Reclaiming Reality: A Critical Introduction to Contemporary Philosofy*. London: Verso.

Bhaskar, R. (1998). *The Possibility of Naturalism - A Philosophical Critique of the Contemporary Human Sciences* (3 ed.).

Bhaskar, R., & Lawson, T. (1998). Critical Realism. In M. Archer, R. Bhaskar, A. Collier, T. Lawson & A. Norrie (Eds.), *Critical Realism: Interventions*. New York, NY: Routledge.

Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out: Classification and its Consequences*. Cambridge, Massachusetts: The MIT Press.

Bryman, A. (2012). *Social Research Methods* (4 ed.). New York: Oxford University Press.

Byrkjeflot, H. (2004). The making of a health care state? An analysis of the recent hospital reform in Norway. In A. a. Grønlie (Ed.), *Rokkan Rapport*.

Chen, H.-T. (1990). *Theory-Driven Evaluations*. Newbury Park, California: SAGE Publications, Inc.

Christensen, T., Lægreid, P., & Stigen, I. M. (2007). Performance Management and Public Sector Reform: The Norwegian Hosptial Reform. *International Public Management Journal, 9*(2), 113-139.

Collins, J. (Producer). (2006, June 1, 2015). Good to Great Diagnostic Tool. [Tool] Retrieved from http://www.jimcollins.com/tools/diagnostic-tool.pdf

Corbin, J., & Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (3 ed.). Thousand Oaks, California: Sage Publications, Inc.

Departementenes sikkerhets- og serviceorganisasjon. (2014). Regjeringen.no. Retrieved December 10, 2014, from regjeringen.no

Donaldson, S., & Grand-Vallone, E. (2002). Understanding Self-Report Bias in Organizational Behaviour Research. *Journal of Business and Psychology, 17*(2), 245-260.

Dunphy, D. (1996). Organizational change in corporate settings. *Human Relations, 49*(5), 541-552.

Edgren, L., & Roald, J. (2001). Employee Experience of Structural Change in two Norwegian Hospitals. *International Journal of Health Planning and Management, 16*(1), 311-324.

Fairclough, N. (2010). *Critical Discourse Analysis - the Critical Study of Language* (2 ed.). United Kingdom: Pearson Education Ltd

Fraser, P., Moultrie, J., & Gregory, M. (2002). *The use of maturity models/grids as a tool in assessing product development capability.* Paper presented at the Engineering Management Conference, 2002. IEMC'02. 2002 IEEE International.

Germov, J. (2005). Managerialism in the Australian public health sector: towards the hyper-rationalisation of professional bureaucracies. *Sociology of Health & Illness, 27*(6), 738–758.

Gilbert, T. P. (2005). Trust and Managerialism: Exploring Discourses of Care. *Journal of Advanced Nursing, 52*(4), 454–463.

Grant, D., Iedema, R., & Oswick, C. (2009). Discourse and Critical Management Studies. In M. Alvesson, T. Bridgman & H. Willmott (Eds.), *The Oxford Handbook of Critical Management Studies* (pp. 213-231). New York: Oxford University Press Inc.

Greenberg, E. S. (1975). The Consequences of Worker Participation: A Clarification of the Theoretical Literature. *Social Science Quarterly, 56*(2), 191-209.

Greenwood, D. J., & Levin, M. (2007). *Introduction to Action Research: Social Research for Social Change*. Thousand Oaks, California: Sage Publications, Inc. .

Guba, E. G., & Lincoln, Y. S. (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Newbury Park, CA: Sage Publications.

Guba, E. G., & Lincoln, Y. S. (2001). Guidelines and checklist for constructivist (aka fourth generation) evaluation.   Retrieved April, 20, 2015

Hall, P. (2005). Interprofessional Teamwork: Professional Cultures as Barriers. *Journal of Interprofessional Care, 19*, 188-196.

Haukås, T. H. (2014). Styreleder Marthe Styve Holte trekker seg [Press release]. Retrieved from http://www.helse-midt.no/no/Media/Nyhetsarkiv/Nyheter/Arkiv-2014/Styreleder-Marthe-Styve-Holte-trekker-seg/130328/#.VE5sbYtws6Q

Helseforetaksloven. (2001). *Lov om helseforetak m.m. .* Retrieved from www.lovdata.no.

Helsepersonelloven. (1999). *Lov om helsepersonell m.v.*: Retrieved from lovdata.no.

Jacobsen, D. I. (2012). *Organisasjonsendringer og Endringsledelse* (2 ed.). Bergen: Fagbokforlaget Vigmostad & Bjørke AS.

Jacobsen, D. I., & Thorsvik, J. (2007). *Hvordan Organisasjoner Fungerer* (3 ed.). Bergen: Fagbokforlaget Vigmostad & Bjørke AS.

Johansen, J. R. (2006). Health Systems in Transition: Norway. Copenhagen.

Kaplan, S., & Jarzabkowski, P. (2006). *Using Strategy Tools in Practice: How Tools Mediate Strategizing and Organizing*: AIM Research.

Keenoy, T. (2009). Human Resource Management. In M. Alvesson, T. Bridgman & H. Willmott (Eds.), *The Oxford Handbook of Critical Management Studies* (pp. 454-472). New York: Oxford University Press Inc.

Klev, R., & Levin, M. (2012). Participative Transformation: The Role of Technology and the Political Economy *Participative Transformation: Learning and Development in Practicing Change* (pp. 49-132). England: Gower Publishing Limited.

Kotter, J. P. (1996). *Leading Change*. Boston, Massachusetts: Harvard Business Review Press.

Lean Advancement Initiative at MIT. LESAT Template Presentation for Enterprise Leadership.   Retrieved December 10., 2014, from http://dspace.mit.edu/handle/1721.1/84695

LESAT Version 2.0. (2012). LAI Enterprise Self-Assessment Tool. In D. Nightingale, L. Abdimomunova, T. Shields, L. N. Perkins, J. Srinivasan & R. Valerdi (Eds.). Massachusetts: Massachusetts Institute of Technology.

Levin, M. (2012). Academic Integrity in Action Research. *Action Research, 10*(2), 133–149.

Levin, M., Nilssen, T., Ravn, E. J., & Øyum, L. (2012). *Demokrati i arbeidslivet. Den norske samarbeidsmodellen som konkurransefortrinn*. Bergen: Fagbokforlaget Vigmostad & Bjørke AS.

Lægreid, P., Opedal, S., & Stigen, I. M. (2005). The Norwegian Hospital Reform: Balancing Political Control and Enterprise Autonomy. *Journal of Health Politics, Policy and Law, 30*(6), 1027-1064.

Mascitelli, R. (2007). *The Lean Product Development Guidebook: Everything Your Design Team Needs to Improve Efficiency and Slash Time-to-Market*: Technology Perspectives.

Mintzberg, H. (1983). *Structure in fives: Designing effective organizations*. Englewood Cliffs, New Jersey, USA: Prentice-Hall, Inc.

Mitchell, R. K., Agle, B. R., & Wood, D. J. (1997). Toward a Theory of Stakeholder Identification and Salience: Devining the Principle of Who and What Really Counts. *The Academy of Management Review, 22*(4), 853-886.

Mize, J. H., & Nightingale, D. (2002). Development of a Lean Enterprise Transformation Maturity Model. *Journal of Information-Knowledge-Systems Management, 3*(1), 15-30.

Nightingale, D. (2009). *Principles of Enterprise Systems*. Paper presented at the Second International Symposium on Engineering Systems, Cambridge, Massechusetts, USA.

Nightingale, D., Abdimomunova, L., Shields, T., Perkins, L. N., Srinivasan, J., & Valerdi, R. (2012). LAI Enterprise Self-Assessment Tool (LESAT) Version 2.0 Facilitator's Guide. Massechusetts: Massachusetts Institute of Technology.

Nightingale, D., & Srinivasan, J. (2011). *Beyond the Lean Revolution*. New York: AMACOM American Management Association.

NTB. (2014, November 14). Toppsjefen i Helse Midt-Norge går av. *Hegnar.no*. Retrieved from http://www.hegnar.no/okonomi/artikkel521816.ece

Oswick, C., & Robertson, M. (2009). Boundary Objects Reconsidered: from Bridges and Anchors to Barricades and Mazes. *Journal of Change Management, 9*(2), 179-193. doi: 10.1080/14697010902879137

Oxford Dictionaries. (2014). Definition of "Evaluate". Retrieved December 10, 2014, from http://www.oxforddictionaries.com/definition/english/evaluate

Oxford Dictionaries. (2015). Measure - Definition. Retrieved May 28, 2015, from http://www.oxforddictionaries.com/definition/english/assess

Pasient- og brukerrettighetsloven. (1999). *Lov om pasient- og brukerrettigheter*. Retrieved from lovdata.no.

Ponzi, L., & Koenig, M. (2002). Knowledge management: another management fad. *Information Research, 8*(1), 8-1.

Psykisk helsevernloven. (1999). *Lov om etablering og gjennomføring av psykisk helsevern*. Retrieved from www.lovdata.no.

Rebolloso, E., Fernández-Ramírez, B., & Cantón Andrés, P. (2008). Responsibility of Educational Institutions for Strategic Change. *Journal of MultiDisciplinary Evaluation, 5*(10), 5-20.

Rebolloso, E., Fernández-Ramírez, B., & Cantón Andrés, P. (2009). Quality Criteria for Self-Evaluation in Higher Education. *Journal of MultiDisciplinary Evaluation, 6*(11), 16-31.

Rebolloso, E., Fernández-Ramirez, B., Canton, P., & Pozo, C. (2002). Metaeveluation of a Total Quality Management Evaluation System. *Psychology in Spain, 6*(1), 12-25.

Scriven, M. (1969). An Introduction to Meta-Evaluation *Educational Product Report* (Vol. 2, pp. 36-38).

SNL AS. (n.d). Store Medisinske Leksikon Retrieved November 22, 2014, from https://sml.snl.no/distriktspsykiatrisk_senter

Spesialisthelsetjenesteloven. (1999). *Lov om spesialisthelsetjenesten m.m.*: Retrieved from lovdata.no.

Stabell, C. B., & Fjeldstad, Ø. D. (1998). Configuring Value for Competitive Advantage: On Chains, Shops and Networks. *Strategic Management Journal, 19*, 413-437.

Star, S. L. (2010). This is not a boundary object: reflections on the origin of a concept. *Science, Technology and Human Values, 35*(5), 601-617.

Star, S. L., & Grisemer, J. R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkley's Museum of Vertebrate Zoology. *Social Studies of Science, 19*(3), 387-420.

Stufflebeam, D. L. (2011). Meta-Evaluation. *Journal of MultiDisciplinary Evaluation, Volume 7*(15), 99-158.

Sveri, T. (2004). Strukturer og reformer. *En kvali*.

Sykehuset Innlandet. (2014). Om Sykehuset Innlandet.   Retrieved December 10., 2014, from http://www.sykehuset-innlandet.no/omoss_/om_sykehuset_innlandet_

Tjora, A. (2012). *Kvalitative forskningsmetoder i praksis (English: Qualitative Research Methods in Practice)* (2 ed.). Oslo, Norway: Gyldendal Akademisk.

Traynor, M. (1996). A Literary Approach to Managerial Discourse after NPM Reforms. *Sociology of Health & Illness, 18*(3), 315-340.

Trowler, P. (2001). Captured by the Discourse?

The Socially Constitutive Power of New Higher Education Discourse in the UK. *Organization, 8*(2), 183–201.

Vinge, S. (2005). Arbejdsorganisering og organisationsforståelse i sygehussektoren. In Å. Johansen & E. Døving (Eds.), *Organisasjonsteori på Norsk* (pp. 95-113). Bergen: Fagbokforlaget.

Vinge, S., & Knudsen, M. (2002). Infrastruktur og Reformer i Sygehusvæsenet. *Dansk Tidsskrift for Sundhedsvæsen, 18*(1), 268-279.

Wilkinson, A. (1998). Empowerment: theory and practice. *Personell Review, 27*(1), 40-56.

Wilkinson, A., & Fay, C. (2011). New Times for Employee Voice. *Human Resouce Management, 50*(1), 65-74.

Womack, J., Jones, D., & Roos, D. (1990). *The Machine that Changed the World*. New York: Macmillan Publishing Company.