The assessment gap:

The case of Det Norske Veritas (Maritime)

Karen Klyve Sunde

Master thesis for the Master of Pedagogic degree in Didactic, organization and leadership, Institute for Educational Research

UNIVERSITY OF OSLO
January 20, 2006
Acknowledgements

First of all I would like to thank my supervisor at the University of Oslo, Terje Grønning, for giving me great supervision and guidance during my thesis.

I am grateful to DNV Learning who hosted me during the interviews and helped me with practical information. I would like to thank especially DNV Learning for their openness and for letting me use their facilities.

Tor Endre Gustavsen, my supervisor at DNV Learning deserves special thanks for fruitful discussions and friendliness, along with George McHenry, Jens Martinius Nilsen, Ann Margit Doublou and Ole Erik Eikenes. I would like to thank all the respondents for taking their time at DNV. Hopefully, all the participants will read my thesis.

DNV Maritime department has given me the opportunity to collect information and given me access to key informants and interviewees. Maritime Learning department has been open and willing to share information with me, and especially Stein Gaarder deserves thanks for being so friendly.

Oslo, January 17, 2005.
Karen Klyve Sunde
Abstract

The research project identified the gap between the existing situation and the desired situation when it comes to how Det Norske Veritas Maritime (DNV M) is using evaluation strategies on competence development for the employees. DNV is an international foundation and is a provider of services for risk management. I have focused on DNV M department which is specialists in technology and business risks consulting for customers in the maritime industry. The employees are basically engineers who are going through continuous competence development to be able to perform services of risk management. However, there is a lack of focus on evaluation of the learning processes and outcomes of the training programs. The thesis will highlight a part of this knowledge gap by first describing the existing competence development procedures offered at DNV M for the employees. I will then present how training courses are evaluated; to identify which stage DNV M is in today according to Kirkpatrick’s four level evaluation strategy. I am focusing on an “assessment of the evaluation procedures” mainly using four evaluation strategies designed by one of the main proponents when it comes to evaluation, Donald Kirkpatrick. Kirkpatrick’s four level evaluation model is basically based on evaluating training courses:

1) Measuring reaction: it measures the customer’s satisfaction after a training program.
2) Measuring learning: “the extent to which participants change attitudes, improve knowledge, and/or increase skills as a result of attending the program” (Kirkpatrick 1998 p.20).
3) Measuring behaviour: the extent to which change in behaviour has occurred because the participants attended the training program.
4) Measuring result: The final results can include increased production, improved quality, decreased costs, reduced frequency and/or severity of accidents, increased sales, reduced turnover and higher profits (Kirkpatrick 1998 p.23).

The next step is to describe the desired situation of the training programs and evaluation methods:

1) A desired situation according to ten interviewees/ employees at DNV,
2) A desired situation according to DNV Learning and DNV Maritime Learning,
3) A desired situation according to evaluation literature and organizational learning literature.

The evaluation literature will be linked to- and analyzed in view of organizational learning literature. I have chosen an empirical approach framed in an eclectic theoretical framework
within the thesis, where I am exemplifying with DNV as a case study. A critical theory perspective will guide me. Kirkpatrick’s four evaluation strategies will be highlighted through the thesis as a framework to understand and measure the evaluation situation. I will raise the evaluation discussion through an analytical understanding of evaluation as a tool of organizational learning. The empirical data is analyzed through theoretical literature, both evaluation literature and organizational learning literature. I have done this through three main research activities:

1) A literature review of evaluation and organizational learning;
2) A comprehensive review and analysis of DNV written documents about competence development strategies and evaluation procedures;
3) Interviews of ten key persons who work in- or with evaluation at DNV Maritime.

If the competence development procedures are not satisfactory and are not being evaluated, this could have serious implications on the outcome of performance and quality of risk management. The training programs should continuously be evaluated and reconsidered as a quality and safety routine. This thesis highlights the discussion about how it is possible to use evaluation as a tool in organizational learning.

Some of the findings consist a tendency of level 1 (reaction) evaluation, and in some training courses a level 2 (learning) evaluation. It has indeed been accomplished a few projects where a four- and five level evaluation were used, but otherwise there are no formal and continuous procedures of level 2, 3 (behaviour) or 4 (result) evaluation. The discussion of “Return on Investment” (ROI) is also present. Even though level 1 evaluation is almost present in all the training courses, the procedures are different, and the focus on evaluation varies. The visions of DNV as a holder of the best competence in their field and how DNV M document their competence in reality can indicate a gap in two ways:

1) What the different levels (HR- and training managers, line managers and employees) at DNV M state and what happens in reality;
2) The existing situation of evaluation at DNV M and how evaluation literature (i.e. Kirkpatrick) and describes evaluation and how organizational learning literature describes a learning organization.

DNV shows a strong resistance against testing and ranking their employees. DNV M lives by evaluating their customers, but inside the organization, there is a lack of systematic approach
towards assessment of evaluation procedures. By questioning operating norms, a learning organization approach can be possible. The need to have a dialogue, reflect and raise questions within the organization will be important. To improve communication awareness and share visions and goals in the different organizational levels will be important for DNV M.
List of tables and figures:

Figure 1: “Picturing the thesis” 15
Figure 2: “Kirkpatrick’s four level evaluation model” 20
Figure 3: “Business Areas in DNV” 37
Figure 4: “DNV Maritime” 38
Figure 5: “International perspectives for measure training and ROI” 61

Table 1: “DNV Interviewees” 48
Table of content:

Aknowledgement 2

Abstract 3

List of tables and figures 6

Table of content 7

1 Introduction 9
    1.1 Introduction and contextualization 9
    1.2 Purpose and objectives for the thesis 10
    1.3 Limitations 11
    1.4 Significance of the project and its potential impacts 12
    1.5 Clarification of concepts 13

2 Theoretical perspectives 15
    2.1 Evaluation theory 15
       2.1.1 Approach 1: Four different dimensions of evaluation 16
       2.1.2 Approach 2: Top-down and bottom-up perspective of evaluation 17
       2.1.3 Approach 3: A more practical view of evaluation 17
    2.2 Kirkpatrick and his model for evaluating training 19
       2.2.1 Measuring Reaction 20
       2.2.2 Measuring Learning 21
       2.2.3 Measuring Behaviour 22
       2.2.4 Measuring Result 23
       2.2.5 Comments about Kirkpatrick’s model 24
       2.2.6 Phillips and Return on Investment (ROI) 27
    2.3 Organizational learning theory 29
       2.3.1 Argyris and Schôn 29
       2.3.2 Nonaka and Takeuchi 31
       2.3.3 Senge 33
    2.4 Organizational learning and evaluation 35

3 Det Norske Veritas (DNV) 37
    3.1 DNV Maritime 38
       3.1.1. DNV Maritime, Training and Impact 39

4 Methodology and Analytical Tools 43
    4.1 Methodological basis 43
    4.2 Methodology 44
       4.2.1 Case Study 45
       4.2.2 Interviewing 47
       4.2.3 Conclusions 49

5 Analysis 50
    5.1 Individual level and Kirkpatrick’s evaluation model 50
       5.1.1 Measuring Reaction 51
       5.1.2 Measuring Learning 53
1 INTRODUCTION.

1.1 Introduction and contextualization.

Companies today are facing increased complexity and uncertainty, making it difficult to manage technical and business risks. Tolerance of failure to manage risk from society and stakeholders is low. In turn, regulators and authorities are developing stricter requirements. Although safety is more emphasized than ever before, boat-ship-oil platform, accidents (i.e. Alexander Kielland, Estonia, Scandinavian Star) still occur. How different risk companies handle accidents, document and argue for their competence, has raised my curiosity of evaluation of competence development in risk management organizations. To meet the criteria of competition and complexity, competence development and training is necessary. Although training is becoming a more focused and developed area in many organizations, formal procedures and experience of managing training and training results can vary. Evaluation of training programs and processes are in many cases not present.

Evaluation of competence development is a critical activity. Without systematic evaluation, there is no basis to evaluate the effects of training procedures and how they can improve. The evaluation activity is often neglected in many organizations. Lai (2004) argues that possible reasons for this can be time limitations, lack of economic resources and lack of evaluation competence. I think it is very important to emphasize results of the competence development. Some organizations may have a lack of formal goals and visions when it comes to competence development. Another reason for neglecting evaluation can be the lack of experience of evaluation and competence (e.g. methodological evaluation competence). The scepticism of mapping and clarifying the results can also be present. Evaluation is often seen as a final step in a competence development procedure. It should rather be seen as an integrated activity in the process of competence development (Lai 2004).

After spending 6 months in Canada, at the University of Alberta, I noticed an enormous focus on assessment and evaluation in schools, universities and organizations. Norway does not have the same presence or emphasis on assessment and evaluation in schools or organizations as Canada. Focus on assessment and evaluation at the University of Oslo was not as present as in Canada. I found this interesting, and therefore I decided to write my thesis on evaluation. The idea of using evaluation processes and evaluation results as a tool in
organizational learning in risk management organizations made me decide to write about evaluation I stayed 6 weeks in one of Norway’s biggest companies, The Norwegian Veritas (DNV), as a part of the practicum period at the University of Oslo. DNV is a leading international organization and an independent foundation established in 1864. DNV is a provider of services for risk management. I was working in DNV Learning department (DNV L) making a global tailor made training course for the Certification department. I recognized the lack of focus on both process and result evaluation when it comes to evaluation of the employees.

I decided to contact DNV L when I started writing the thesis. They approved my desire for writing the thesis using DNV Maritime (DNV M) as a practical example in my discussion about evaluation as a tool for organizational learning. DNV is a complex international organization, and evaluation of competence development can be an important part of the organizational learning. Because the DNV is a provider of risk management, the quality of its work has implications for many people, i.e. safety on children’s playgrounds, tanker ship safety, transportation safety etc. Knowledge of different evaluation methods will therefore be decisive in the effort to evaluate the outcome of competence development procedures of the employees. After reading and reflecting on evaluation literature, I noticed that research on evaluation was not as present as I assumed. The evaluation literature I read did not focus a lot on how to use evaluation as a tool in organizational learning. Writing about evaluation and link this to organizational learning in this thesis can contribute to broadening the understanding of evaluation literature. Using evaluation theory in a case study (DNV M), can give indications on how evaluation theory can be seen in practice. The results can also be seen as a contribution to DNV M because I am reflecting and discussing their daily work routines when it comes to evaluation and organizational learning.

1.2 Purpose and Objectives for the thesis.

I have chosen an empirical approach framed in an eclectic theoretical framework within the thesis. I am exemplifying with DNV M as a case study to test Kirkpatrick’s four levels of evaluation and the eventual relationship between evaluation and organizational learning. Kirkpatrick’s model is a global well known and widely used model. I will link this model to organizational learning literature. DNV has to be up to date at all times when it comes to

---

1 For further details see chapter 3
knowledge, quality, safety and research, because they have responsibility for human life and security. Continuous competence development is a key factor. The employees at DNV are provided training, courses and further education at work. But how is this evaluated?

Some fundamental problems arise when it comes to evaluation (concept validity, criteria validity etc), and this will be addressed in my research.

My **main research objective** is: to make an assessment of how evaluation is practised in the case of DNV.

This is guided by the research objectives outlined below:

1) To describe the existing competence development procedures offered at DNV M for the employees and how the training courses are evaluated today.
2) To identify which stage DNV M is in today according to Kirkpatrick’s evaluation strategy today.
3) To describe the desired situation of the training programs and evaluation methods:
   a) A desired situation according to ten interviewees/employees at DNV,
   b) A desired situation according to DNV Learning and DNV Maritime Learning,
   c) A desired situation according to evaluation literature.
4) To identify barriers of how and if DNV M can reach all four levels of Kirkpatrick’s evaluation strategies in a desired situation.
5) To assess this according to evaluation theory and link this to organizational learning theory.

**1.3 Limitations**

I will not evaluate the training programs or different kinds of competence development procedures or learning strategies. I have chosen to analyse Kirkpatrick’s four levels to training courses offered at DNV M. I am focusing as such gap between the existing and desired situation when it comes to evaluation in DNV M. I will reflect on possible hindrances and barriers in reaching a desired situation of evaluation, but this thesis is not a recipe on how DNV M should implement and use Kirkpatrick’s model.
1.4 Significance of the Project and Its Potential Impacts

DNV L department and I wanted to clarify and map the gap when it comes to the existing situation and the desired situation of evaluation procedures (in comparison of Kirkpatrick’s evaluation levels) used at DNV M. The findings will have significance for the corporate community as they identify the different gaps viewed above, where there is insufficient focus on evaluation of competence development. Identifying the implications of competence development and evaluation can have a social significance in light of the fact that greater proportions of Norwegians (and other inhabitants of countries where DNV is present) are currently using different tools and objects that have been certified by DNV. The society must therefore trust that DNV is providing safety.

Kirkpatrick does not stress how evaluation can be seen in organizational learning. The thesis can be intellectually significant because the evaluation literature will be discussed and linked to organizational learning literature. As mentioned, this can contribute to a broader understanding of the field of evaluation. It can also result in new insight and usefulness for DNV with respect to evaluation. The potential impact of the project includes the ability to focus and inform about the Kirkpatrick’s evaluation methods in all countries where DNV is present. The observation of why knowledge firms, like DNV, do not have assessment of their evaluation methods can be seen in different perspectives. This thesis is not a causality analysis. It does however contain possible reasons for the situation, such as unawareness, time pressure, beliefs, lack of resources etc.

The theoretical framework will be highlighted first. I will present three different approaches of evaluation before I focus on Kirkpatrick’s four level model. An insight in Return on Investment (ROI) will also be highlighted here. The organizational learning literature will be presented by contributions from Argyris & Schön, Nonaka & Takeuchi and Senge. DNV and DNV M structure will be presented with a focus on why competence and evaluation is important for DNV M. Methodology and analytical tools will then be introduced. I will map Kirkpatrick’s four levels in how DNV M is evaluating their employees individually in training courses today.

After I conducted all the interviews, I realized that all my reflections around evaluation could be seen in a wider context. I could not seem to find the answers I was looking for when I was
mapping DNV M’s training courses to Kirkpatrick’s four levels. There was something missing to get a broader picture of the existing evaluation situation. I therefore decided to lift the discussion about evaluation on an individual level to an organizational level. I am therefore linking Kirkpatrick to Argyris & Schôn’s conception about espoused theory- and theory- in- use, and single loop- and double loop learning to get a broader understand of the existing- and desired situation at DNV M. Nonaka and Takeuchi’s conceptions about tacit and explicit knowledge and middle-up- down management is used in a further analysis together with Kirkpatrick. Senge’s five disciplines will be presented in a further discussion about Kirkpatrick’s evaluation model. Seeing the field of evaluation and Kirkpatrick’s contribution together with Argyris & Schôn, Nonaka & Takeuchi and Senge give a framework for further discussion and interpretation in the thesis. At the end of the thesis, I will present reflections and alternatives to how DNV M can implement the four level evaluation model or if the evaluation model should be implemented at all.

1.5 Clarification of concepts.

The competence concept is from the Latin term competentia, and it involves having sufficient knowledge, being able to evaluate and have strength to perform and obtain desired results (Lai 2004). There are several definitions of competence according to Lai, but I will use Lai’s definition:

“Competence is the gathered knowledge, abilities, skills and attitudes that makes it possible to perform actual functions and tasks in relation to the defined demands and goals” (Lai 2004 p.48).

Formal competence can be seen in documentation from i.e. educational systems or approved certification systems. Real- competence is an expression for the gathered knowledge a person has acquired through formal education, work experience, private life and family life. This contains both formal and informal competence. Basic competence is the essential knowledge and skills that can be present in the organization and through organizations (Lai 2004 p.55- 58).

Scriven defines evaluation as:

“the determination of the worth or value of something- in this case, of educational and social programs, policies and personnel- judged according to appropriate criteria explicated and justified” (Scriven 1991).
I think his definition shows that the concept of evaluation can be seen in different approaches. A definition of evaluation for learning is from Preskill and Torres (1999):

“…..is a means for fostering individual and team learning about complex organizational issues. Evaluation inquiry for organizational learning and changes is more a means to an end; it is more than developing skills that result in increased competence or improvement profits. A significant consequence of evaluative inquiry is the fostering of relationships among organizations members and the diffusion of their learning throughout the organization; it serves as a transfer-of-knowledge process. To that end, evaluative inquiry provides an avenue for individuals as well as the organizations ongoing growth and development” (Preskill and Torres 1999 p.18)

Formative evaluation improves the quality of the training during the design, development and implementation stages by carrying out a subject matter expert review, a user review, or a pilot test. When the design of the training program is near completion, both subject matters experts and users provide feedback to further refine the training. Summative evaluation determines the impact on individual and organizational performance during and after the training through direct observation, surveys of training stakeholders, measurement of performance indicators (such as quality, productivity, and customer satisfaction), and/or measurement of a business outcome (such as numbers of sales, customer retention, profitability and market share) (Hale 2002).

When it comes to evaluation, sorting out hard and soft data will have a decisive impact on the result. Hard data are easy to measure and easy to convert to monetary values. It achieves high credibility with management. It represents output, quality, costs, and time of work related processes. Soft skills represent usually subjective, sometimes difficult to measure, and often difficult to convert to monetary values, and are behaviourally oriented (Phillips 2003).
2 THEORETICAL PERSPECTIVES.

In this chapter I will give an introduction to evaluation theory and organizational learning theory. In the evaluation theory section, contributions from three different approaches to evaluation will be presented before I specifically focus on Kirkpatrick’s four evaluation levels. In the organizational learning section, a discussion of Argyris & Schôn’s conception about single and double loop learning; Nonaka & Takeuchi’s conceptions about tacit- and explicit knowledge and middle-up-down management; Senge’s conception about a learning organization will be presented. Figure 1 illustrates how I can analyze and picture the situation in my thesis:

![Diagram of Evaluation and Organizational Learning](image)

**Figure 1:** "Picturing the thesis"

Source: Author’s research.

2.1 Evaluation Theory.

I will give an introduction into three different approaches of evaluation, before I present Kirkpatrick’s evaluation model. It is important to understand the numerous perspectives in the field of evaluation in order to be able to get a comprehensive view and choose a practical evaluation method. Dahler- Larsen (2001) presents “some mysteries about evaluation”: it
contains in example the idea of a common understanding that everybody knows that they should perform evaluation, but not knows why, and what it should be used to. Evaluation can also be seen as performed without any kind of formal criteria, and says nothing about the quality of the results. I think that some of the mysteries give a fruitful picture of mapping DNV and evaluation, and I will discuss this further in my analytical part later in the thesis.

2.1.1 Approach 1: Four different dimensions of evaluation.

Dahler- Larsen and Krogstrup (2001) emphasize four different approaches in the discussion about evaluation:

1) *The knowledge dimension*: there is a great discussion about what kind of scientific demands evaluation methods should have. The scientific approach to evaluation is combined with hermeneutical, phenomenological and post-modern ideas, and thereby the question about validity is present here.

2) *The value dimension*: the question about how the evaluator can represent the partners or his own values in the evaluation procedures is present here. The demands on evaluation procedures are more focused, because the evaluator has to choose and argue for the value dimensions he/she uses, and for his/her own biases in the procedure.

3) *The applicability - dimension*: it is not just the evaluation results that have been given paid attention but also the evaluation process. The expansive use of evaluation can be seen in many different areas: in society, policy, organizations etc.

4) *The evaluation*: today’s understanding of evaluation is numerous. There is no longer an understanding of evaluation based on experimental programs. The use of evaluation in all areas has raised the question about how and why evaluation should be performed.

As one can see here, the field of evaluation is controversial because of the different viewpoints, and the evaluation literature can therefore be understood as fluctuating. To acquire more understanding of evaluation, it can be interesting to get an insight of a top-down and bottom up perspective of evaluation.
2.1.2 Approach 2. Top-down and bottom up perspective of evaluation.

A top-down perspective and bottom up perspective of evaluation is commonly discussed in the evaluation literature. Top down evaluation is used as an instrument to assess decisions where employees are “objects” and the learning process is top-down controlled. The participants do not influence the knowledge or development. Bottom-up evaluation focuses on the actor’s values, motivation and experiences. The participants influence the knowledge-development and can influence the learning process (Lindøe 2003 p.56).

Generally, the traditional use of evaluation has had a more top-down focus, but during the last decade one can see a change in this tradition. Karlsson (2003) focuses on democratic evaluation to give different groups and participants in evaluation a voice in the evaluation process. Gray (1998) emphasizes how organizations can perform evaluations without having a top-down focus on the evaluation process. Evaluation as an effective tool can be used to enhance the organizational learning. Self assessment can also be used in organizations as an alternative to formal evaluation procedures. The top-down and bottom up perspective of evaluation give an overview, but it does not provide an insight in the different practical procedures of evaluation. I will therefore present some different practical views before I emphasize Kirkpatrick’s four evaluation levels.

2.1.3 Approach 3: A more practical view of evaluation.

Sverdrup (2002) presents four different approaches when it comes to evaluation:

1) The decision-making approach
   Goals and results are based on goal-remedy rationality. The main activity is to measure the effect from a program compared with the goals is the main activity. The results can contribute to changes in the program or the goals.

2) The user-orientated approach.
   This is a more bottom-up approach where processes, quality and clients are more focused on comparison than within the decision making approach. It is more subjective, descriptive and explaining. It is the target group expectations, needs and goals that are the main focus.

3) The process orientated approach/ formative evaluation.
Insight, understanding and learning on how one can improve process efficiency in a training program are the main points in this approach. Qualitative methods are often used here to improve the learning process. Reflection, dialogue and self-assessment are commonly used methods.

4) The goal-free evaluation

The process is an activity, and it is difficult to measure or identify the goals. The arguments for this approach are that the individuals should gain an overall picture of the evaluation and not just focus on the goals. The evaluator can also enter the program or evaluation without being informed of the purpose of the program, and he/she has to find out what it is.

Stocker and Van der Heijden’s (2001) article *Competence Development and Appraisal in Organizations* suggests that there are three ways in which “self- and other” ratings are fruitful:

1) The first concerns communication about the results of “self-and other” ratings and proposes that regularly feedback could be a start;
2) Both employees and the supervisors should be mutually engaged in the development of each other’s competencies;
3) The use of a personal development contract between the employee and supervisor can be a fruitful contribution for competence development.

Kearns (2005) presents ideas and steps for how it is possible to link evaluation to organizational learning. This will be highlighted later in the thesis. Even though Kearns contribution can seem quite abstract, it is though different from my contribution, because I have used DNV M as a practical example in a case study to reflect in how it is possible to link organizational learning and evaluation.

The literature review shows different approaches of evaluation, and one can picture the field of evaluation to be quite numerous. During the last decade, the focus and understanding of evaluation has expanded, and therefore many evaluation methods have been developed. A turn from a more top-down focused evaluation to more bottom-up evaluation has been present the last decade. On the other hand, I think the literature review shows that there is a gap in all the evaluation literature about how to be critical and reflective about practical evaluation methods and how one can link this to organizational learning. Vedung (2003) recommends a
combination of several models, because every model provides partial perspectives and answers only. It can be difficult to choose and perform evaluation because of the numerous techniques, and the lack of research and research results of the different techniques. The processes and effects of evaluation depend on the view the evaluator has. The evaluator must be aware of that his/her reflections, reactions and evaluation are always more or less influenced by one’s pre-understanding.

I have chosen to focus on Kirkpatrick’s model of evaluating training and use it with a critical view, because it is a well known and used model for many decades. Kirkpatrick can be seen as a representative for the applicability dimension (viewed above) because he pays attention to the evaluation process and result. His perspective can also be understood as a representative of a more bottom-up perspective, because the four levels emphasize involvement, dialogue and influence from the individual in the learning process. The evaluation focuses on the actor’s values, motivation and experiences (Kirkpatrick 1998). I will now present Kirkpatrick’s four level evaluation model.

2.2. Donald Kirkpatrick and his four level model for evaluating training.

Kirkpatrick published articles on the levels of evaluation in 1959 in American Society for Training Directors (ASTD)'s Journal. During the past decades, the articles have become a classic model of evaluation used by training professionals around the world. Kirkpatrick has written several books on evaluation and he is also a well-known speaker and consultant in the training field. Kirkpatrick states that his model is simple and practical. A lot of trainers want a model they can understand and use, therefore has this model has been used widely (Phillips 2003). There are three main reasons for doing evaluation according to Kirkpatrick: first, you have to justify the existence of the training department by showing how it contributes to the organization objectives and goals. Second, you have to decide whether to continue or discontinue training programs. Third, you can use evaluation to gain information on how to improve future training programs (Kirkpatrick 1998 p.16). Kirkpatrick states that many organizations forget the evaluation part, and therefore according to him, he offers a simple and easy evaluation model that includes four levels:
2.2.1. Measuring reaction.

The first level is the reaction phase. It measures the customer’s satisfaction after a training program. Many use “happy sheets” as feedback forms; verbal reaction; post-training surveys; questionnaires to conduct information about how the participants reacted on the training program. This is important because it can give you valuable feedback about the training, and it tells trainees that the trainers are there to help them do their job better and that they need feedback to determine how effective they are. The reaction sheets can provide quantitative information that can be given to managers and others concerned about the program and give information that can be used to establish standards of performance for future programs. This can be seen as quick and very easy to obtain, and it is not expensive to gather information or to analyse it. Questions to be proposed profoundly at this level can be:

- Did the trainees like and enjoy the training?
- Did the trainees consider the training relevant?
- Did they like then venue, the style, timing, domestics etc?

Kirkpatrick provides guidelines for evaluating reaction:

a) Determine what you want to find out.
b) Design a form that will quantify reactions. The organization should establish standards of reaction sheets for future evaluations.
c) Encourage written comments and suggestions.
d) Get 100 % immediate response.
e) Get honest responses.

f) Develop acceptable standards. The trainer can use the ratings (i.e. a scale from 1-5) to establish standards of acceptable performance, and analyze results from i.e. past programs.

g) Measure reactions against standards, and take appropriate action.

h) Communicate reactions as appropriate (Kirkpatrick 1998 p.26-36).

2.2.2 Measuring learning.

The second level is the learning phase which the author defines as “the extent to which participants change attitudes, improve knowledge, and/or increase skills as a result of attending the program” (Kirkpatrick 1998 p.20). One or more of these changes must take place if a change in behaviour is to occur. Evaluation tools and methods used at this level are typically assessments or tests before and after training, interview or observation. The trainer must submit reliable, clear scoring and measurement assessment. It can be quite simple to arrange; it is clear-cut for quantifiable skills for example technical skills, but it can be difficult to measure complex learning like attitudinal development. Kirkpatrick suggests certain questions to be profound at this level:

- Did the trainers learn what was intended to be taught?
- Did the trainee experience what was intended for them to experience?
- What is the extent of advancement or change in the trainees after the training, in the direction or areas that was intended?

Kirkpatrick provides guidelines for how you can evaluate learning after an accomplished training program:

a) Use a control group if practical because it can provide better evidence that change has taken place. Any differences between the control group and experimental group can be explained by the learning result because of the training program.

b) Use a paper and pencil test to measure knowledge and attitude.

c) Use a performance test to measure skills

d) Get a 100% response to be able to analyze the results statistically.

e) Use the results of the evaluation to take appropriate action.” If the learner hasn’t learned, the teacher hasn’t taught” (Kirkpatrick 1998 p.40-50).
2.2.3. Measuring behaviour.

The third level is how to evaluate behaviour, and the author defines that as “the extent to which change in behaviour has occurred because the participants attended the training program” (Kirkpatrick 1998 p.20). It is important that the person who accomplishes the program must have a desire to change and that the person must know what to do and how to do it. The participant must work in the right climate and he or she must be rewarded for changing. Evaluation tools or methods used at this level can be observation and interview over time to measure the sustainability of change. 360-degree feedback can also be used as an effective method, and it does not need to be used before training, because the participants can make a judgment about change after training, and this can be analyzed for groups of respondents and trainees. (Phillips 2003). The most preferred assessment situation is to design tasks around relevant performance scenarios and specific key performance indicators or criteria. Electronic and online assessment can be difficult to use at this level, because the results of the assessments can often be too subjective. Kirkpatrick states that you can use self-assessment if you use carefully designed criteria and measurements (Kirkpatrick 1998). To measure behaviour change is difficult, because it is less easy to quantify and interpret that reaction and learning evaluation. The evaluator must cooperate with line-managers to be able to reach the employees (the participants). Certain questions can bring in ideas and guidelines for gathering information;

- Did the trainees put their learning into effect when back on the job?
- Were the relevant skills and knowledge used?
- Was there noticeable and measurable change in the activity and performance of the trainees when back in their roles?
- Was the change in behaviour and new level of knowledge sustained?
- Would the trainee be able to transfer their learning to another person?
- Is the trainee aware of their change in behaviour, knowledge and skill level?

Kirkpatrick offers guidelines for evaluating behaviour:

a) Use a control group if practical.
b) Evaluate both before and after the program if it is possible and practical.
c) Survey and/or interview one or more of following: trainees, their immediate supervisors, their subordinates, and others who often observe their behaviour.
d) Get a 100% response or a sampling if the organization has resources to interview or sample all the participants.

e) Repeat the evaluation at appropriate times. There is no right answer to what “appropriate time” is. It can be 2-6 months after a training program.

f) Consider cost versus benefits (Kirkpatrick 1998 p. 48-56).

You can gather baseline measures before people attend the training program, just like giving pre-tests in order to make level 2 evaluation meaningful. Questionnaires (or telephone interviews) can help reduce the cost of level 3 evaluation. In the information-gathering process here, you need to take into account possible factors in the work environment that could prevent the application of newly learned knowledge and skills.

2.2.4. Measuring result.

The fourth level is how to evaluate the results. The final results can include increased production, improved quality, decreased costs, reduced frequency and/or severity of accidents, increased sales, reduced turnover and higher profits (Kirkpatrick 1998 p.23). Kirkpatrick states that one should be aware of the fact that many organizations already have their own measurement systems, and therefore it can be a challenge for the evaluator to identify and relate to the trainee’s input and influence. At the beginning of a training program, the course instructor should identify and agree accountability and relevance, so everybody understands what is to be measured. It can be difficult to measure final results for programs on such topics as leadership, communication, motivation, time management, empowerment etc. These are the kind of conceptions one can refer to as soft skills, because they are more difficult to describe. It can be easier to measure results individually, and it is often difficult to generalize the results to the whole organization. Further external factors will influence both individual and organizational performance. Therefore it is important to take into consideration that the results can be biased. Guidelines for evaluating results can be almost as same as guidelines for evaluating behaviour and they have some of the same strengths and difficulties:

a) Use a control group if practical.

b) Allow time for results to be achieved. Some skills need time to internalize.

c) Measure both before and after program if practical.

d) Repeat the measurement at appropriate times.

e) Consider costs versus benefits. It can be less costly to evaluate result than behaviour.

f) Be satisfied with evidence if proof is not possible. (Kirkpatrick 1998 p.61-70).
2.2.5 Comments about Kirkpatrick’s evaluation model

As reviewed above, there are several approaches to evaluation. Kirkpatrick’s model is meant to uncover if learning is present or not of the participants of a training course or competence development program. Further, his model is widely used and positively experienced in many countries as a contribution to the field of evaluation (Bramley 1996, Taschereau 1998). Phillips also states that: “The framework developed by Kirkpatrick has been the most widely used approach in organizations” (Phillips 1997 p.44). Taschereau states that Kirkpatrick’s model can be seen as a contribution to the field of evaluation:

“A number of methods can be and are used in evaluation synthesis, ranging from pooling data from multiple programs, to analyzing effect sizes across programs in a meta-analysis, and finally to use I qualitative methods. These methods attempt to overcome the limitations often imposed by assessing individual programs by using data and/or results from a number of often different, but similar evaluations. In the specific area of evaluating training programs, Kirkpatrick (1975) describes four evaluation steps that provide much greater detail in assessing the results of these programs including their impact” (Taschereau 1998 p.3).

I will use Kirkpatrick’s approach to evaluation as a tool in my thesis. As mentioned, it is important to understand Kirkpatrick’s evaluation model as one contribution to evaluation. Even though research on evaluation varies, some recent research about evaluation has shown that there are multiplicities of methods available. Although researchers have had difficulty determining the exact number of evaluation and assessment methods it has been shown that the appropriate evaluation depends on the context. Alliger & Janak’s (1998) article Kirkpatrick’s levels of training criteria: thirty years later stress the validity of Kirkpatrick’s evaluation model in three assumptions of the model (1) the levels are arranged in ascending order of information provided; (2) the levels are causally linked; (3) the levels are positively correlated. Levels of training evaluation are however problematic. Kirkpatrick’s model may never have been meant to be more than a first, global heuristic for training evaluation but as such it has done well. There are several other models of training evaluation criteria; most of them are very similar to Kirkpatrick’s. Alliger & Janak (1998) stress that each of three assumptions is apparently implicit in the general understanding of Kirkpatrick’s model, and when using his model, one should take into consideration the critique of the model.
Kearns’ (2005) article *From Return on Investment to Added Value Evaluation: The Foundation for Organizational Learning* also raises the question about whether Kirkpatrick’s model offers a sufficiently solid and robust foundation for developing both general theory and practical methodology to ensure that the effect of Human Resource Development (HRD) learning investment by organizations is maximised. Further, he distinguishes between two types of value statement: (1) basic value, and (2) added value. Basic value is about consistently maintaining the status quo, which is necessary, but it will not contribute to progress, innovation, or continuous improvement. Added value is when the organization improves over and above its normal level of operation. The evaluation process itself has a profound effect on both organizational learning and operational management beyond training alone. The author presents an evaluation model that almost looks identical to Kirkpatrick’s. It contains baseline- pre-training, post-training, checking reactions, and learning, transfer/application, re-measuring the performance of each trainee and comparing with the pre- measurement and calculating the return on investment.

Holton (1996) gives Kirkpatrick’s evaluation model a critical view, and his remarks should be taken into consideration when testing the evaluation model in DNV M. Holton’s article “The Flawed Four-Level Evaluation Model” (1996) stresses that Kirkpatrick’s model is not a model, but a taxonomy for training and the results of training.

“One shortcoming of taxonomies is that they do not fully identify all constructions underlying the phenomena of interest, thus making validation impossible. Attempts to test causal assumptions within a taxonomy are futile because, by definition taxonomies classify rather than define causal constructions” (Holton 1996 p.6).

Some of the weak points of Kirkpatrick’s four level evaluation model can be seen if the model is used as a diagnostic tool:

“Consider the case where performance change or positive results are not found to occur. The only conclusion possible using data within the four-level model is that something is wrong with the training program. However, if the many intervening variables that remain unmeasured are considered, it is quite possible that the training program is well designed and that the
problem lies outside the classroom with some elements of the organization, job, or individual” (Holton 1996 p.8).

It does not seem to me that Kirkpatrick places emphasis on this approach to the problem in his evaluation model. Kirkpatrick focuses on the positive reaction from the participants after completed training program, because he states further that learning can be present here. I think this is a controversial statement, because there is no evidence for positive learning even though the results are positive. Holton (1996) raises critique of Kirkpatrick’s model and he states that change in individual performance can be as a result of the learning being applied on the job not because of a training course. Holton says that Kirkpatrick does not focus enough on this aspect, and therefore he excludes other variables when considering why the participants react as they do, and one can not exclude other variables that may influence reactions and learning outcomes. On the other hand, Holton states that the four-level evaluation model is quite plain and elegant, and its contribution to the evaluation field in HRD can be seen as positive. “If HRD is to continue to grow as a profession, an evaluation model grounded in research is necessary” (Holton 1996 p.6).

As a response to the critique, Kirkpatrick states that:

“Personally, I don’t care whether my work is called a model or taxonomy as long as it helps to clarify the meaning of evaluation in simple terms and offers guidelines and suggestions on how to accomplish an evaluation. (…) I only hope that my model- or taxonomy- continues to be a help to HRD professionals who are more interested in practical ideas than scholarly research” (Holton 1996 p.24).

Kirkpatrick further states that “Progress in evaluation of training will result if all of us will freely exchange information on objectives, methods, and criteria” (Kirkpatrick 1960 p.17).

I think Holton and Kirkpatrick have a pragmatic view of evaluation, even though Holton stresses more the research side of evaluation. The importance of being able to perform evaluation in the practical field is therefore one of Kirkpatrick’s building stones. Kirkpatrick’s model is not perfect, but it can be seen and used as a framework that includes categories on how to perform evaluation.

References to training evaluation typically centre on the four-level evaluation framework.
There are still issues to be addressed like how to measure the learning outcome and the return on investment (Kearns 1999). A relevant discussion during last decade has been on the benefits and return on investments for training programs. Jack Phillips adds a fifth level to Kirkpatrick’s model, and raises the question about how and why the return on investments should be focused on. I will therefore present a short introduction to return on investment because it gives a broader understanding of the evaluation field.

2.2.6 Phillips and Return on Investment (ROI).

Phillips emphasizes the first four evaluation levels in a similar way that Kirkpatrick does it, and he adds the fifth level which contains comparison on the monetary values of the results with the costs for the program. The values for return on investment are usually quite large, in the range of 25 to 500 %, which illustrates the potential impact of successful programs (Phillips 2003). Phillips emphasizes that the best approach to ROI is to base the calculations strictly on business impact data obtained after the program has been implemented. Business performance measures (level 4) are easily converted to a monetary value, and this is necessary for an ROI calculation. Phillips states that ROI is not for every training program, because ROI is more appropriate for those training programs that have a long life cycle and that are very important to the organization in meeting its operating goals.

As one can see in this section, Phillips raises more questions about return on investment than Kirkpatrick does. Are Kirkpatrick’s four evaluation levels not enough to evaluate a training program? Or is it possible to recognize Phillips fifth level in Kirkpatrick’s fourth level? Kirkpatrick states that many training professionals mistakenly call his level 4 “ROI”, but Kirkpatrick labels it “results”. Kirkpatrick stresses that ROI can have three major shortfalls;

1) they usually do not capture all of a companies strategic objectives
2) ROI is a snapshot in time that tells you where you have been; it has no ability to predict where you will go
3) Since ROI is a lagging indicator, it is not a good diagnostic tool.
   (Kirkpatrick 1998 p.88).

I will raise the question related to this ROI method briefly later in my thesis: Is the ROI method present at DNV?
I have now given an introduction to several approaches to evaluation. Evaluation can be a difficult and interesting field, because of the numerous methods available. Kirkpatrick (and Phillips) present practical evaluation methods. A literature review shows that there is a gap in evaluation literature about how to be critical and reflective about practical evaluation methods and how one can link this to organizational learning. As reviewed, Vedung (2003) recommends a combination of several models, because every model provides partial perspectives and answers only. Since Kirkpatrick’s model holds one perspective, it can be important to broaden this perspective and try to map how it is possible to use this model by linking it to organizational learning literature. As reviewed in chapter 1, I will discuss how-and if evaluation can be used as a tool in organizational learning. Argyris and Schõn’s conception of single and double loop learning will be highlighted on how evaluation can be seen as an attempt to increase or decrease organizational learning. Kirkpatrick does not stress how it is possible to use explicit and tacit knowledge in his model, and therefore I have chosen to link his model to Nonaka and Takeuchi’s conception about tacit- and explicit knowledge, and middle-up down management. To understand what it takes for an organization to implement a four level evaluation model, Senge can be seen as a contributor with his five disciplines. Even though he does not stress evaluation, I think his five disciplines show a framework for discussing organizational learning and evaluation. In the next section, I will therefore present an introduction to three approaches when it comes to the field of organizational learning.
2.3 Organizational learning theory.

In this section, I will present an introduction to true key contributions to organizational theory, namely Argyris & Schôn, Senge’s, and Nonaka & Takeuchi’s and their contributions to the field of organizational learning. I will use their conceptions and theories to understand, picture and highlight the existing situation and the desired situation (from a theoretical perspective) at DNV when it comes to DNV as a learning organization. As mentioned, the analytical part in the thesis will have two different levels: individual and organizational. I will analyze Kirkpatrick’s evaluation model and the individual in DNV M. This will be followed by a discussion of Argyris & Schôn’s conception about single and double loop learning, Nonaka & Takeuchi’s conception about sharing tacit and explicit knowledge and Senge’s conception about a learning organization.

2.3.1 Argyris and Schôn.

In most of the organizational learning literature, learning is referred to as gaining new knowledge and change in behaviour. Argyris and Schôn’s understanding of learning as a detection and correction of errors; either a product (something learned) or the process that yields such a product (Argyris and Schôn 1996).

Argyris and Schôn state that individual learning is an assumption for organizational learning together with the organizational behaviour, action, intelligence and memory. Organizational learning can be defined as organizational improvements of work tasks over time. The authors understand of the concept of organizational learning that includes the organizational members act upon their theories of actions. Further, Argyris & Schôn differ between two types of theories of action: theory- in-use and espoused theory and it provides a broad background for understanding organization development and how to enhance organizational learning. Espoused theory can be seen as the organization’s visions and goals. The theory-in-use is how it really is in reality. The gap between espoused theory and theory-in-use can be big. The conception about single- loop learning rests in an ability to detect and correct error in the relation to give a set of operating norms (Argyris and Schôn 1996). The status quo traps many organizations. Double loop learning depends on being able to take a “double look” at the situation by questioning the relevance of operating norms and challenges the basic paradigms. One has to review, question and challenge the basic paradigms and operating norms. A lot of
organizations believe that they are performing organizational learning (and double loop learning), even though they are performing single-loop learning. A top-down approach to management, especially one focusing on control through clearly defined targets, encourages single-loop learning and discourages the double-loop thinking that is so important for an organization to evolve (Morgan 1998 p.86).

The authors claim that many individuals and organizations tend to have trained incapacity, and it makes it difficult to create double-loop learning. They call this Model 1 theory-in-use and it is the general action theory that people are using. People’s behaviour and action can be in non-agreement with the theory-in-use that they have expressed earlier. Argyris and Schôn state that individuals can show four signs at the Model 1 theory in use:

1) To keep the control
2) To win as much as possible, and loose as little as possible
3) To oppress negative feelings
4) To be as rational as possible, oppress feelings


The typical model 1 theory in use makes double loop learning impossible, because people are not able to take a double look at the situation and question the reality and behaviour. Argyris and Schon state that the Model 2 theory in use is a more effective theory, because it challenges to create offensive learning between the individuals and the organization. Three signs are representative here:

1) valid information
2) free and informed choices
3) Internal engagement, commitments, follow-ups and accomplishments.

It is not possible for a human being to act in accordance to model 2 without help, because they do not know how to do it. Therefore, the management has to implement Model 2 and “de-learn” Model 1 behaviour. Bateson (1972) defines deuterolearning as a second order learning, or learning how to learn. Argyris and Schôn describe this in term of a shift from O-1 to O-2 learning systems. Organizational deuterolearning is critically dependent on individual deuterolearning, which they describe in terms of a shift from Model 1 to Model 2 theories-in-use (Argyris & Schôn 1992 p.29).
Argyris and Schõn`s theories and conceptions can be critiqued as not being scientific, and based on observations and anecdotes. An organization in the real life is seldom as dualistic as they present it, and their argumentation can also be seen as circular. I also question the linear idea about identifiable assumptions for action.

Espoused theory and theory in use can be seen as containing tacit and explicit knowledge. I have chosen to use Nonaka and Takeuchi`s conception about tacit and explicit knowledge to broaden the understanding of Argyris and Schõn`s conception about espoused theory and theory in use, and single and double loop learning. Middle-up down management can be seen as one way of stimulate single and double loop learning.

2.3.2 Nonaka and Takeuchi.

Nonaka and Takeuchi`s contribution to research is mainly the generic model for organizational knowledge building and this model can be seen in a wide context. They present further knowledge building processes for how the learning processes can be present and grow. In a knowledge spiral, an interaction between tacit and explicit knowledge remains. There are two types of knowledge: 1) explicit knowledge: contained in manuals and procedures, often expressed in words and numbers. This knowledge can be readily transmitted across individuals formally and systematically; 2) Tacit knowledge is learned by experience, and communicated indirectly, through metaphors and analogies. Tacit knowledge can be difficult to communicate or share, because it is deeply rooted in the individuals. Nonaka argues that focusing on tacit knowledge as well as explicit knowledge is the key to success (Nonaka 1995). Nonaka emphasises the differences- but also the link between individual knowledge and social knowledge. Individual knowledge contains individual actions, communication, observations etc, and in an organization explicit and tacit knowledge can allow for effective individual action. Organizational members share the social knowledge. Based on individual experiences of shared organizational events, social knowledge allows organizational members to share rules in the form of practices. Social knowledge can be both explicit and tacit (Nonaka 2000).

According to Nonaka, socialization is the process of sharing the tacit knowledge, of individuals. Sharing the same experience through joint activities such as being together, spending time can contain tacit knowledge. Externalization requires the articulation of tacit
knowledge and its translation into forms that can be understood by others, like explicit knowledge. *Combination* involves the conversion of explicit knowledge into more complex sets of explicit knowledge; knowledge is exchanged and combined through communication networks. *Internalization* means the conversion of newly created explicit knowledge of individuals. Learning by doing and training are important to embody explicit knowledge. Thus, on-the-job training, simulations or experiments are used to induce internalization of new knowledge. Knowledge that has been created is shared throughout an organization (Nonaka 2000).

Front-line employees, middle managers and top management all play a part in creating knowledge. The creation of new knowledge is the product of dynamic interaction among the three kinds of players. Front-line employees are immersed in the day-to-day details of particular technologies, products or markets. Middle managers (often team leaders or line managers) serve as a bridge between the visionary "ideals" of the top and the often chaotic "reality" of those on the front line of business. In other words, if top management's role is to create a grand theory, middle managers create more concrete concepts that front-line employees can understand. They may convert the tacit knowledge of both frontline employees and top management, make it explicit, and incorporate it into new technologies, products or systems. In this sense, middle managers are the true knowledge engineers of what Nonaka and Takeuchi call "the knowledge-creating company" (Nonaka 2000).

Engstrøm (1999) criticizes Nonaka and Takeuchi because they ignore the small cycles of continuous team-based development. Engstrøm states that their knowledge model (the spiral of tacit to explicit knowledge), is deterministic and has little empirical evidence because every situation is contextual and different, therefore one can not match the knowledge spiral in every situation. However, I find Nonaka and Takeuchi’s conceptions about tacit and explicit knowledge interesting when discussing how one can use and share learning and evaluation in DNV M. The individual will hold a lot of tacit and explicit knowledge in a knowledge organization like DNV. The challenge of how DNV M share and build knowledge can be seen as contribution to my discussion later in the thesis about tacit and explicit knowledge. Certain factors should be present to create a knowledge creating organization, and sharing tacit and explicit knowledge can be seen as one factor. By linking this to Senge’s five disciplines, the idea of a learning organization can broaden the understanding and discussion to how DNV M can share tacit and explicit knowledge, and stimulate of doing double loop learning.
2.3.3 Peter Senge.

According to Peter Senge, learning organizations are:

“…organizations where people continually expand their capacity to create the results they truly desire, where new and expansive patterns of thinking are nurtured, where collective aspiration is set free, and where people are continually learning to see the whole together” Senge (1990 p.3)

The basic rationale for such organizations is that in situations of rapid change only those that are flexible, adaptive and productive will excel. For this to happen, organizations need to “discover how to tap people’s commitment and capacity to learn at all levels” (Senge 1990 p. 4). For a ‘learning organization it is not enough to survive. “Survival learning” or what is more often termed “adaptive learning” is important and necessary. But for a learning organization, “adaptive learning” must be joined by “generative learning”, learning that enhances our capacity to create (Senge 1990). Argyris and Schön argue that the organization should be controlled by problem solving, but Senge states that it should be controlled by creative tensions. Senge states that a goal for a learning organization is to minimize the tension between the visionary and the reality. To create common visions and develop a joint picture of the organizations mission and goals is very important for an organization. The five disciplines that Peter Senge identifies are said to be converging in innovating learning organizations. A central thought in these disciplines are that the employees have to use critical reflection, open dialogue, be open about their value systems and challenge other value systems critically. The five disciplines are:

1) **Systems thinking**: this is an integrated element and a bridge builder in the organization, so that its members can understand and integrate the system in the whole organization.

2) **Personal mastery**: it is the discipline of continually clarifying and deepening our personal vision, of focusing our energies, of developing patience, and of seeing reality objectively (Senge 1990 p.7). People with a high level of personal mastery live in a continual learning mode.

3) **Mental models**: These are deeply ingrained assumptions, generalizations, or even pictures and images that influence how we understand the world and how we take action (Senge 1990 p.8).
4) **Building shared vision:** it is the capacity to hold a shared picture of the future that the organization seeks to create, and it has to share and foster a commitment and enrolment.

5) **Team learning:** the processes of aligning and developing the capacities of a team to create the results its members truly desire (Senge 1990 p.236). Personal mastery and shared visions need to be seen and acted with other people in team learning (Senge 1990).

All the disciplines are, in this way, “concerned with a shift of mind from seeing parts to seeing the whole, from seeing people as helpless reactors to seeing them as active participants in shaping their reality, from reacting to the present to creating the future” (Senge 1990). Senge argues that learning organizations require a new view of leadership, and the leaders should be designers, stewards and teachers. They are responsible for building organizations where people continually expand their capabilities to understand complexity, clarify vision, and improve shared mental models – that is they are responsible for learning (Senge 1990).

Senge’s work is not seen as part of the academic literature of organizational learning, because he basically writes for practicing and aspiring managers and leaders. He promotes a holistic vision of organizations. Some of the critique is that it is difficult to find organizations that match Senge’s idea about a learning organization. However, I would like to use Senge’s idea about a learning organization and try to map how DNV M can be seen in the five disciplines. This can be seen as a contribution to the discussion about how one can use evaluation as a tool in organizational learning. But is it possible to link evaluation literature to organizational learning literature? This leads me to the next section of discussing and linking the contribution to organizational learning literature, reviewed above, to evaluation literature.
2.4 Organizational learning and evaluation.

Argyris & Schon present an action-related theory about learning at different levels. Senge’s theory about the “learning organization” and Nonaka and Takeuchi’s “knowledge-building organizations” can be seen as further contributions to the concept of learning organizations. They present two different reactions to Argyris & Schön’s “Organizational development” paradigm. Senge can be seen as a continuous development of Argyris and Schön’s ideas, but Nonaka and Takeuchi represent a reaction against the organizational development paradigm, which is influenced by western ideas: Nonaka says that to create knowledge in an organization, depends on the ability to create new competence and spread the knowledge in the organization and articulate it through new products, systems or services (Nordhaug 2002).

I think Senge’s five disciplines together with Argyris & Schön’s idea about learning organizations and Nonaka & Takeuchi’s conceptions about tacit and explicit knowledge give a broader framework for understanding DNV as an organization. Nonaka and Takeuchi’s concept about the externalization phase has something in common with double-loop learning (Argyris & Schön), and generative learning (Senge). Nonaka and Takeuchi do not see this phase as problematic but rather as an integrated and unproblematic part of the organizational learning process. A generative learning process seeks to identify tensions between visions and reality. This learning process can be seen as related to Argyris and Schön’s concept about double-loop learning (Nordhaug 2002). Nonaka and Takeuchi state that creating double loop learning is difficult for organizations to implement on their own.

To create knowledge, the best management style is neither top-down nor bottom-up, but rather what Nonaka and Takeuchi call "middle-up-down". The model shows as mentioned that the middle leaders operate as a “bridge builder” between the visionary and the operative (Nonaka 1995). This resembles Mackay’s (1964) view of sharing “routine decision making”. Decentralized decision making allows for more flexibility (Morgan 1998). Senge’s theory about shared leadership can be seen as link to Nonaka and Takeuchi’s understanding of leadership in some areas, but Nonaka and Takeuchi focus on the middle leader as a knowledge engineer and a bridge builder in the organization. Senge does not emphasise this.

Even though Kirkpatrick does not stress which level in the organization that is responsible for the evaluation the idea of shared leadership or middle-up-down management can have an important role in evaluating employees. Using evaluation as a tool to stimulate from single loop learning to double loop learning can also contribute to focus on sharing tacit and explicit
knowledge, and reduce the gap between espoused theory and theory in use. This will be further discussed in the analytical chapter.

Before I present the methodological and analytical chapter I would like to give an introduction to DNV and DNV M to create an understanding of the organization. A short presentation of competence development procedures offered at DNV M will also be present to understand how evaluation can be performed and in what kind of procedures. As mentioned, I have chosen to focus on training courses/programs, but I think it is fruitful to present all the competence development procedures to receive a broader understanding of what is offered for the employees at DNV M.
3 DNV

DNV\(^2\) is a leading international organization and an independent foundation established in 1864. DNV is a provider of services for risk management in about 100 countries and has about 5,800 employees. DNV’s objectives are safeguarding life, property, and the environment and is a leading international provider of services for managing risk. DNV’s customers include ship owners and management companies, charterers, yards, designers and manufacturers, financiers, law firms, insurance and brokers, authorities and administrations, maritime organizations, navy, ports and terminals, training providers and crewing agents.

DNV has four main Business Areas (BA):

\[\text{Figure 3: Business Areas in DNV.} \]
\[\text{Source: www.dnv.com.}\]

1) DNV Certification is working with measure management systems against national and international best practices. It provides formal documentation that your product has been tested to comply with applicable codes, standards and/or directives. Risk Based Certification is a trademark of Det Norske Veritas AS.

2) DNV Maritime is one of the world's leading classification societies, and helps the maritime industry manage risk in all phases of the ships life, through ship classification, statutory certification, fuel testing and a range of technical, business risk, financial and competency related services.

3) DNV Consulting offers cross-disciplinary competence within management and technology, whether it concerns oil and gas installations, transportation systems, process facilities, ICT, food safety or patient safety.

\(^2\) The information in this section is, unless otherwise noted, based on www.dnv.com retrieved in December 2005
4) DNV Technology Services Combining cutting-edge technology knowledge and industry experience, DNV Technology Services is a preferred provider of independent services for the oil, gas, process and energy industries.

3.1 DNV Maritime

**Figure 4: DNV Maritime**

Source: www.dnv.com

DNV M contains several departments:

*Ship Classification:* it entails verification against a set of requirements during design, construction and operation of ships and offshore units. Statutory Certification

DNV operate a worldwide network of survey stations and is authorized by more than 120 flag administrations to carry out surveys and, in most cases, issue statutory certificates on their behalf. *DNV Petroleum Services* (DNVPS) is the foremost name in marine fuel management, commanding over 75 per cent world-wide market share for contracted Fuel Quality Testing services. *Maritime Solutions* specializes in technology and business risk consulting for customers in the maritime industry. *DNV Sea Skill* offers standards, tests and certification services related to competence development.

As one can see here, DNV’s main business area is to perform and deliver competence to

---

3 [www.inside.dnv.com](http://www.inside.dnv.com)
customers. The employees are basically engineers with a high level of specific competence. To understand how the competence inside DNV M is maintained and continuously improved, I will in the next section present shortly how DNV M is accomplishing and offering competence development procedures for their employees.

3.1.1 DNV M, Training and Impact.

Person B (competence manager) said that 3% of the budget is set for training, and this is based on seven organized training days for each employee over 230 working days a year. The new employees accomplish 10 training days a year. All M Regions have Training Responsibles and M Learning is a staff unit in Maritime Technology and Production (MTP) responsible for support to all training activities. The training managers have psychological, pedagogical and engineer background. M Learning evaluates training courses, trainers, tutors, competence, individual – and organizational performance and return on investment. M Learning is collaborating with DNV Learning, which is serving all DNV Business Areas. They support M Learning with both course development and operation. DNV Learning helps DNV employees and divisions to perform daily work by providing training and learning support. M Learning requires all divisions and employees to have a Task Plan (a scheme to compare the current performance situation with expected tasks – and identify the needs for change) a Competence Plan (the employees are compared with competencies, basically VQSC elements-Veritas qualification scheme- class) and a Training Plan (this is based on the results from the Task and Competence Plans, and you can tailor made courses, coaching etc). VQSC is a tool to secure planned quality and uniformity of work for all personnel in Maritime involved in Classification work. The scheme is based on practical and theoretical training and includes registration and documentation of competence. Local managers have the responsibility of ensuring that each employee has the necessary training, experience and competence before being sent on own responsibility on any particular assignment.

Several competence strategies are offered at DNV:

1) Training courses: development of training courses go through a formal revision once a year from regional and central needs, prioritizing and budgeting. 1/5 of the competence development program at DNV contains training courses. 4/5 is set for on-the-job/practical training and supervision. Some training courses are mandatory for all employees, i.e. course nr.2001 “Introduction to DNV” and nr.2002 “DNV Services”.
Some courses are mandatory for specific VQSC competence elements (i.e. course 2115 Certification of Materials & Components) and others are recommended for Continuous Professional Development (i.e. course nr.1358 Intercultural Communication). The Maritime department performs around 32 courses (2005) as classroom courses, and 15 web based courses. DNV Learning arranges about 10 courses for DNV M. The training plan for 2005 and 2006 will be adjusted for some courses.

2) *Seminars and foras:* a) *Experience Exchange Seminars* is based on the idea of sharing and reflecting about learning and work experience in a group, both among surveyors and across regions. b) *Customer Reality Seminars* for Maritime personnel in general to learn about the customers business and perception of DNV services. c) *Technical forum:* Regular technical seminars arranged by MTP 5 times a year for all MI employees at Høvik. D) The Project Manager’s Forum: it is a face to face arranged seminar for all Research &Development project managers to learn more about their role and how to succeed in it.

3) *The Web Course Reader* gives employees either access to a "light" version of a web course or only to the course material in a PDF format. Only by personally enrolling in the courses will you get access to expert guidance on its subject matter. And only by successfully completing the course, i.e. passing its final test, will you get formal credit for your work. This may be of special interest for those who just need update on the subject.

4) *On-the-job training and supervision:* As mentioned, 4/5 of the competence development contains practical training and supervision and guidance at work. This is a continuous procedure with focus on dialogue-communication and active situated learning in certain contexts, without any form of testing.

5) *Managing Individual Performance (MIP):* This is like an appraisal interview performed as a dialogue once or twice a year between a leader and an employee about the employees work, expectations, performance, goals and competence development. The MIP process shall be used to manage the individual performance of all employees within DNV. The leader and employee shall complete the annual MIP process. This consists of the following parts: The goals shall be quantifiable whenever possible and include a goal statement, measure and target. If it is not possible to quantify a particular goal, the result should be verifiable, with a specific and clear description of what the output, product, modified behaviour etc will be and how its occurrence and quality will be evaluated. The result of the dialogue shall be a documented agreement
between the leader and employee. Effective coaching motivates the employee and gives him/her a clear picture of progress vs. goals and targets, meaning that the result of the subsequent assessment process comes as no surprise. If the performance is not meeting expectations, the leader and employee shall discuss the situation and may define additional actions or steps to close the gap. Coaching should be done continuously and can be done relatively informally. In addition, a formal progress review based on individual goals shall be scheduled at least once during the year. The assessment process shall begin with the employee compiling his/her overall year end (self assessment). The leader shall base his/her assessment on the input from the employee and other relevant sources. The assessment consists of three parts:

1) The leader shall evaluate the results vs. each of the individual goals and determine a rating (MIP Form).
2) The leader shall assess the performance of the job against normal job expectations defined in job descriptions.
3) The leader shall consider other significant contributions beyond the normal job scope (big unexpected projects, deliveries etc). (www.inside.dnv.com)

The assessment of the employee shall be discussed with him/her in a meeting. The next step for the leader is to determine an overall rating of the employee based on the goals assessment and the evaluation of the job. This final rating shall be communicated with the employee as soon as possible.

Why is evaluation important for DNV M? This chapter shows that DNV and DNV M offer a lot of training programs and competence development procedures. DNV M lives by/of competence and they invest enormous amounts in competence development. The need for evaluating an effect will therefore be present. DNV M is continuously evaluating their customers with a high professional level. But the question is: are they evaluating themselves? Does it exist a gap between the documents and what happens in reality? After my practical period in DNV Learning I got, as mentioned, interested in evaluation, because I could notice a lack of evaluation focus. One of my concerns is how come DNV (and DNV M) evaluate their customers with a high professional level, when evaluating their own employees is not as present as evaluating externally. This will be discussed further in the analytical chapter.
Writing the thesis about DNV M, I needed theoretical perspectives and analytical tools to be able to receive a broader understanding of DNV M as an organization and the field of evaluation. In the next chapter, I will therefore present how I have conducted the study, and argue for the choices I have made gathering the data and writing the thesis. I will consider the organizational theoretical understanding that I have used in this thesis, because it has influenced the choices I made during the thesis.
4 METHODOLOGY AND ANALYTICAL TOOLS.

I will now present the methodological basis and the concrete methods I have used. My methodological basis is in so called critical theory, and I have made use of a case study approach including interviews as the main method.

4.1 Methodological basis.

I have chosen a critical theory perspective for this thesis. I will use Denzin and Lincoln definition of a criticalist:

“as a researcher or theorist who attempts to use her or his own work as a form of social or cultural criticism and who accepts certain basic assumptions: that all thought is fundamentally mediated by power relations that are social and historically constituted…….; that facts can never be isolated from some form of ideological inscription; that the relationship between signifier and signified is never stable or fixed and is often mediated by the social relations of capitalist production and consumption;… that language is central to the formation of subjectivity……” (Denzin & Lincoln 1994 p.139)

Critical research can be understood in the context of the empowerment of individuals. Inquiry that aspires to the name critical must be connected to an attempt to confront the injustice of a particular society or sphere within the society. The researcher has to be aware of his/her own subjective, intersubjective and normative reference claims.

According to Denzin & Lincoln (1994) there consists among critical researchers a firm recognition that ideologies are not simply imaginary mental relations that individuals and groups live out relative to their existence. It can also seem like inscribed in the materiality of social and institutional practice. The authors emphasize that the aim of inquiry in a critical theorist perspective is the critique and transformation of the social, political, culture, economic, ethnic, and gender structures. These structures can constrain and exploit humankind, by engagement in confrontation or even conflict. The inquirer can be seen as an instigator or facilitators, implying that the inquirer understands a priori what transformations are needed. Knowledge does not accumulate in an absolute sense. Because the nature of knowledge can be seen as series of structural/historical insights that will be transformed as
time passes. The appropriate criteria for judging the goodness or quality of an inquiry can be seen as historical situatedness of the inquiry (i.e., that it takes account of the social, political, cultural and economic antecedents of the inquiry), the extent to which the inquiry acts to erode ignorance and misapprehensions, and the extent to which it provides a stimulation to action, that is, to the transformation of the existing structure (Denzin & Lincoln 1994 p.113-115).

Thus, a critical theorist perspective can be used as a framework in several areas, I have chosen to use DNV M as a case study to critically picture and investigate their existing situation of performing evaluation. I will try to use my thesis as a form of confronting and questioning the existing norms. At DNV, the contextualized knowledge is based on historical insights that have been transformed as time has passed. To avoid accidents, DNV’s criteria for judging quality is based on critical inquiry.

It is important to have in mind that facts can never be isolated from some form of ideological inscription, and therefore I can not interpret the interview answers as objective statements. As mentioned above, the language is central to the formation of subjectivity (Denzin & Lincoln 1994). Since I have been working at DNV before, I have to be aware of my own subjective and intersubjective references.

4.2 Methodology.

The data will primarily be qualitative where contextual factors are described with a case study research. The researcher has to be conscious and critical in relation to choices and decisions he or she makes, and how her or his position in the inquiry may influence on the research results. The credibility of the research-results depends on the way the basis that the knowledge rests upon is made explicit. This involves accounting for the approaches chosen during gathering of data, how the analysis is conducted and how the results are interpreted (Thagaard 1998). My methods will be: 1) Singular case study, and 2) Interviews.

Qualitative research involves the collection of large amounts of information within a natural setting and requires that humans are the main research “instrument.” It typically involves the research methods of observation, interview, documents collection and analysis (Wellington 2000). Though Creswell (2003) stresses that the researcher using a mixed method-
qualitative and quantitative methods, best provides an understanding of a research problem, I have studied DNV M as a case study with a qualitative approach and focusing on interviews. It gives me an in-depth understanding of how DNV M employees understand the field of training and evaluation. A case study is defined by interest in individual cases, not by the methods of inquiry used (Denzin & Lincoln 1994 p.236).

4.2.1. Case study

One definition of case study can be:

“An empirical inquiry that investigates a contemporary phenomenon within real-life context, especially when the boundaries between phenomenon and context are not clearly evident. The case study inquiry copes with the technically distinctive situation in which there will be many more variables of interest than data points, and as one result. It relies on multiple sources of evidence, with data needing to converge in a triangulating fashion, and as another result., benefits from prior development of theoretical propositions to guide data collection and analysis” (Yin 2003 p.13-14).

Yin emphasises that when using a case study as a research method, is when

“A “how” and “why” question is being asked about a contemporary set of events, over which the investigator has little or no control” (Yin 2003 p.9).

I think this statement match my thoughts and curiosity on how DNV M is performing evaluation and why there is a lack of focus on evaluation. According to Denzin & Lincoln (1994), an *instrumental case study* can be seen as a case which is examining to provide insight into an issue or refinement of theory. The case plays a secondary role, facilitating the researchers understanding of something else. The case if often looked at in depth, its contexts scrutinized, its ordinary activities detailed, but because it helps the researcher to pursue the external interests (Denzin & Lincoln 1994 p.237). I think the description of an instrumental case study fits to my research. Using DNV as a case study gives me an opportunity to, according to Yin:

1) Explain the presumed causal links in real-life interventions that are too complex for experimental strategies;
2) To describe an intervention and the real life context in which it occurred;

3) To illustrate certain topics within an evaluation- in a descriptive mode.

(Yin 2003 p.15).

Kvale emphasizes the importance of a continuous validity discussion though the thesis regarding the researcher’s role, the research design, the data material, interpretation and theory. In a qualitative study, it’s more difficult for another researcher to collect the same material. Individual and contextual factors may change over time. Since my thesis has a qualitative approach, I have chosen Kvale’s (1997) reconstructions of concepts in qualitative research. Validity, reliability and generalization are conceptions that are originally descended from a positivist research paradigm. 4 I have tried to cover two steps in relation to emphasize construct validity: I have used multiple sources of evidence (interview, document analysis, content analysis etc), established a chain of evidence and I have had key informants at DNV Learning to review drafts of the case study report during the composition of the thesis. I will therefore argue that I have established correct operational measures for the concept being studied (Yin 2003 p.34). When it comes to external validity, I have tried to test Kirkpatrick’s evaluation model at DNV M and therefore I can state that I am using theory in a single- case study. I have chosen to not generalize results to all DNV divisions, but mainly focus on DNV M. This thesis represents a microscopic selection of employees at DNV M. Hopefully, the result will contribute to new theoretical and practical reflections. I will describe all levels of my study carefully to be able to strengthen the reliability of my research so that another researcher could do the same project and get almost the same results. I have not emphasized to generalize my case study to other organizations as DNV, even though I can expect the readers of this thesis to comprehend their interpretations and draw conclusions of their own.

There has been a lot of discussion about case study as a research method, and Yin summarizes these prejudices against using the case study: there has been a lack of rigor of case study research and a concern that case study research provides a small basis for scientific generalization. Critics claim that using a case study takes too long, and result in massive, unreadable documents (Yin 2003 p.12). On the other hand, from a case report, one can learn both propositional and experiential knowledge. By refining theory and suggesting

---

4 Sometimes it can be difficult to translate Norwegian words and conceptions into English. I will therefore refer to this section when I discuss reliability and validity through my thesis: I have chosen to use Kvale’s understanding of these conceptions in a different form, and he states that reliability can be understood as “reliable” (pålitelighet), and validity as “valid” (gyldighet).
complexities for further investigation, a case study as a method can be fruitful (Denzin & Lincoln 1994).

4.2.2 Interviewing.

Kvale sees the interview as a conversation, an art of asking questions and listening. He states further that the interview is not a neutral tool. The interviewer creates the reality of the interview situation, and the interview produces situated understanding grounded in specific interactional episodes. The purpose of the interview, according to Kvale (1996 p.6) is:

“To obtain descriptions of the life world of the interviewee with respect to interpreting the meaning of the described phenomena”.

I will have semi-structured interviewing with ten key informants representing the DNV M. I will use some open ended questions, to seek opinions, to invite the interviewees to express views and attitudes or to encourage prediction or sheer speculation. A genuinely open question will invite opinions or views without either leading or prompting (Wellington 2000 p. 79). The purpose for these interviews was to:

1) To elicit in-depth descriptive accounts from the top leaders and the middle leaders about how evaluation is accomplished in DNV M.
2) To identify connections between an individual and organizational understanding at DNV M in relation to how evaluation are accomplished today and how the desired situation looks like.

Key informants at all “levels” can be valuable in establishing different perspectives and also in creating some kind of “in-house triangulation” (Wellington 2000 p.73). I selected ten employees from DNV M:
Table nr.1: DNV interviewees.

<table>
<thead>
<tr>
<th>HR- or training managers:</th>
<th>Position</th>
<th>Length of employment in DNV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person A</td>
<td>HR Manager in Maritime/ Head of Section.</td>
<td>Over 10 years.</td>
</tr>
<tr>
<td>Person B</td>
<td>Principal Research Engineer working in DNV Maritime in Maritime Technology and Production Centre (MTP).</td>
<td>Over 25 years.</td>
</tr>
<tr>
<td>Person C</td>
<td>VQSC manager/Maritime Manager.</td>
<td>Over 25 years.</td>
</tr>
<tr>
<td>Person D</td>
<td>Consultant/Training Counsellor at DNV Learning.</td>
<td>Under 5 years.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Line- managers:</th>
<th>Position</th>
<th>Length of employment in DNV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person E</td>
<td>Head of Section in MTP.</td>
<td>Over 25 years</td>
</tr>
<tr>
<td>Person F</td>
<td>Head of Section in MTP.</td>
<td>Over 10 years</td>
</tr>
<tr>
<td>Person G</td>
<td>Head of Section in MTP</td>
<td>Over 5 years</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Employees:</th>
<th>Position</th>
<th>Length of employment in DNV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person H</td>
<td>has attended the international trainee program and is now working for MTP.</td>
<td>Under 5 years</td>
</tr>
<tr>
<td>Person I</td>
<td>has attended the international trainee program and is now working for MTP.</td>
<td>Under 5 years</td>
</tr>
<tr>
<td>Person J</td>
<td>has attended the international trainee program and is now working for M Solution.</td>
<td>Under 5 years</td>
</tr>
</tbody>
</table>

Source: Author’s research.

Reading and analyzing, literature, theory and DNV material have provided a rich data material and contributed to different perspectives during the study. I made three semi-structured interview guides for this qualitative study; one for the top management, one for the middle leaders and one for the “normal” employees (Appendix 1, 2, 3). The reason for doing this was to map the line between the top management and the employees. I had to analyze and interpret the answers in light of the interview guide so that the answers given may be seen in relation to what is being asked. There are so many factors that may influence during the interview process, and it is therefore important to be aware of that as an interviewer and to note that interview guides are not able to grasp all of the biases. None of these interviews have been tape-recorded. By not tape-recording the interview, it is possible to loose or forget some material after writing. But taping the discussion can also be understood and interpreted as a way of leaking information outside DNV.
4.2.3 Conclusion.

In this chapter I have presented and argued for my theoretical framework and the analytical tools I have used. A critical theory perspective has guided me through collecting the data and writing the thesis. In the next section, I will analyze how Kirkpatrick’s evaluation levels can be present at DNV M in light of my theoretical framework and analytical tools. First, I would like to give a practical example to show how one can use Kirkpatrick’s four evaluation strategies. Then I will proceed by examining DNV M as a case of how they measure reaction, learning, behaviour and results according to Kirkpatrick’s evaluation model.
5 ANALYSIS

An example of how to use Kirkpatrick’s four evaluation levels:
Imagine a small risk management company in Norway with 30 employees, where 20 employees are engineers. Ten of the employees have accomplished a new training program, learning a new technique of certificate tank ships. The other ten employees are performing as usual. To evaluate this training program, the manager has chosen Kirkpatrick’s four evaluation levels. The manager hands out reaction sheets after the training course (level 1), and the results from the reaction sheets can inform the manager and the training instructor about how relevant participants thought the training was; how engaged the trainees felt by training; in which area the trainees thought information was missing (Kirkpatrick 1998). The manager wants to measure if the participants have learned something (Level 2). He can therefore e.g. measure the trainees’ performance before and after training (pre-test and post-test), and see if the result is different. The manager wants to measure if the training had a positive influence on job performance (Level 3). He can interview, observe, use questionnaires etc the employees. He needs to take into account that possible factors in the job environment can prevent the application of newly learned knowledge. To measure the result (Level 4), the manager can count e.g. how many tank ships Jon has certified (before the training program, and after the training program), and see how this will influence reduced costs, improved quality, increased sales and increased profitability. Using experimental and control groups, gives the possibility to recognise that other variables could have contributed to the result. One can have evidence that training influenced the result, but one cannot state positively that training was the only cause.

Next, I will present how the individual employee at DNV M can be seen in Kirkpatrick’s four levels of evaluation: reaction, learning, behaviour and result. For each part, I will first present shortly each level, before I analyze and link the existing situation at DNV M to Kirkpatrick levels. I will use statements from the interviews and analyze them in accordance to the four evaluation levels.

5.1 The individual and Kirkpatrick’s evaluation model.

“Training is more than courses, and Learning is more than Training”
(Person B- competence manager)
5.1.1 Measuring Reaction.

The reaction level measures the customer’s satisfaction after a training program. As reviewed in chapter 2, many use “happy sheets” as feedback forms, or verbal reaction, post-training surveys or questionnaires to conduct information about how the participants reacted on the training program. This is important because it can give you valuable feedback about the training, and it tells trainees that the trainers are there to help them doing their job better and that they need feedback to determine how effective they are. The reaction sheets can provide quantitative information that you can give to managers and others concerned about the program. Reaction sheets can provide trainers with quantitative information that can be used to establish standards of performance for future programs. This can be seen as quick and very easy to obtain, and it is not expensive to gather information or to analyse it (Kirkpatrick 1998).

After reflecting and considering the results from the interviews and reading and analyzing DNV material, I saw that the reaction phase was present in almost all the different training courses: technical and “soft-skills”. Almost all the courses complete a “happy sheet” as a feedback form handed out from the trainer to the participants. Several training managers said that they used “happy sheets” because they are quick and easy to obtain, and not expensive to gather. It can be done immediately when the training ends. In technical courses, focus on the reaction phase was improved a lot during the last years. In soft-skills training courses, there were always handed out a reaction sheet, and often followed by a verbal discussion/feedback. According to person B (a competence manager): “we always emphasize participants’ reaction after attending a course, and this is important for further development of a course”. In addition, a judgment of the trainer, the technical/professional content is also present”. M Learning and DNV Learning want to map the “successful courses”, and design and implement courses that have the successful factors in every course.

One of the employees at DNV L said that DNV L always send a “Quest Feedback Evaluation Form” to the participants after every training course that DNV L is accomplishing, and they calculate the average (on a scale from 1-6) of the feedback. If the trainers feel a need to discuss the results, DNV Learning offers guidance. Not all the trainers receive the results. If the instructors want to see the result of the reaction phase, they should be shown to them. Members of the training departments or human resource department should have access to the results (Kirkpatrick 1996).
I noticed after interviewing the line managers (head of section of different technical departments), that in some technical courses the importance attached to the reaction phase was not present. As person D (line manager) stated: “sometimes it is good enough that people are present at the courses, and we have to envisage that the participants are pleased with the course and that they have learned something. Sometimes, I use reaction sheets. Often, we don’t have time to evaluate at all”. Kirkpatrick emphasizes the importance of getting a reaction, and try to get a positive reaction. The future of a program is dependent on a positive reaction. If the participants react negatively, they will probably not be motivated to learn (Kirkpatrick 1998). According to Phillips, level 1 evaluation is important as a customer’s satisfaction measure, a favourable reaction does not ensure that the participants have learned new skills or knowledge (Phillips 2003 p.34). The statement from Person D can indicate that even though reaction sheets are suppose to be easy to obtain, the interest of focusing on reaction sheets vary.

I can suspect a need of continuous reaction-evaluation from the participants of the trainers, especially the technical trainers. All the employees completed the trainee program were of a similar opinion, as in this example: “Sometimes, the technical course instructors use poor pedagogical methods in the courses, and we desire a reaction evaluation of the instructors- not just the content, after completed a training course”. Person J mentioned that even though the course participants completed a reaction sheet with valuable information on how to improve the course, the trainer continued the same procedures. Person J therefore questioned the use of reaction sheets. Training managers from M Learning and DNV Learning said that they are focusing a lot of the reaction score, and even though the score could differ with some decimals, it was important to work and improve the result score.

How come the communication and evaluation understanding is so divided and unclear in DNV M? Why did I get so many different answers about measuring reaction phase? How come some of the trainers did not know that DNV Learning send out quest-back forms? How can DNV M become better in measuring reaction? I can suspect a desire from the interviewees that even though they have some different thoughts about training and evaluation, they all have the same goal: to be able to evaluate with easy and valid evaluation procedures. When I asked the interviewees about which factors that could be present in a successful reaction-evaluation situation, they all said: all training courses should have a reaction sheet
followed by verbal feedback and dialogue from the participants”. On the basis of the different statements, it is possible to suspect that there is a gap between

a) The existing situation and the desired situation of using reaction sheets in DNV M.

b) The evaluation literature highlights that the reaction phase can be easy and quick to obtain, but this varies at DNV M.

Since some of the training courses have different reaction sheets and procedures, there might be a lack of agreement of the importance of having common formal procedures. It can therefore be necessary to establish a common routine for all the training courses, and an agreement of how DNV M will consider the reaction results. This will be further discussed in chapter six.

5.1.2 Measuring Learning.

First of all, one should stress: How does Kirkpatrick define learning? There are many different definitions of learning, but there is an agreement that learning is related to an actual and potential change in behaviour (Lai 2004). The whole learning debate is complex, and I have chosen to use Kirkpatrick’s definition of learning: “to the extent to which participants change attitudes, improve knowledge, and/or increase skill as a result of attending program” (Kirkpatrick 1998 p.20). One or more of these changes must take place if a change in behaviour is to occur. In a large organization, like DNV M, there is possible to use a control group if one makes sure that the groups are equal in all significant characteristics. If not, the comparisons are not valid. For instance, one department at DNV M can conduct a training program for 10 employees, while the other 10 employees are continuing as usual. The pre-test and post-test results for both groups can give valuable information about the learning result.

Increased knowledge (for instance from a technical training course), can be easy to measure by a test related to the content of the program before (pre-test) and after training (post test). If the training course is building on already existing knowledge, performing pre-test and post-test can be possible (Kirkpatrick 1998). To measure skills, a performance test can be necessary, for instance how to test non-destructive material (NDT). At DNV M, some of the most important technical courses have a test at the end of the course, and each individual needed 70% to pass. The participants need a minimum understanding of the content. To
register the course on the VQSC platform, the participants need to pass. If they do not pass, the line managers offer post-supervision to make sure that the employee has acquired the new information. The participants could even attend the course again if desired. Person E (line manager) stated: “we have to rely on our employees, and envisage that they are paying attention and acquiring new knowledge. Often, we don’t have time or resources to check everything after a course”. In the technical courses, pre-tests, interviews and observations were hardly present. In some of the soft skills courses, some testing during – and after the course were present, but there was no formal procedures of any course, it depended on the individual course. It can seem like the line managers and some training managers blame external factors like the lack of time, resources etc, and not them selves.

As Kirkpatrick states that it is often easier to measure quantifiable skills such as for instance technical skills, but it can be difficult to measure complex learning like attitudinal development. None of the interviewees mentioned anything about performing pre-testing or using control groups in training programs. Therefore the possibility of measuring the gap between pre-post testing is not present. Some of the line managers said that it could be interesting to perform a pre-testing, but they did not have the resources to do it. Person B (training manager) said that: “we have to emphasize the post--testing, because it gives important information about the learning outcome and as an evaluation of the course. But we have to be careful and not make it too academic, and the questions should contain the most important knowledge”.

The line managers agreed on trying to measuring knowledge after attending a course, but they were more sceptical of ranking and testing the participants. They stated further that they did not always know how they were supposed to measure learning. I will discuss this later in the analysis. Kirkpatrick says that it is relatively simple to set up assessments and tests before and after training, but more investment and thought is required for reaction evaluation (Kirkpatrick 1998). It is possible to use tests, skill practices, role plays, simulations. Person A (HR-manager) emphasized especially the need of measuring learning after each course: “DNV culture is very special because we trust our employees, but on the other hand, I see the need of developing more formal tests after attending training courses. We can not afford to fail our knowledge, because it can have fatal consequences for testing boats etc”. 
The evaluation culture at DNV M contains more over informal evaluation perspective and procedures. Person C (training manager) said that 1/5 of competence development procedures are courses, and 4/5 is set for on-the-job training and supervision. He stated further: “on-the-job training is a continuous informal procedure where supervision and coaching between a leader/tutor and an employee is a daily routine. We don’t have a lot of formal material that explains objectives, procedures and implementation of on-the-job training”. The intention of on-the-job training is that employees can have a possibility to learn, experience and discuss during performance of the job, and reflect on their immediate reaction on their experience. The content from the training courses are also expected to be discussed in on-the-job training.

But the supervision, coaching and dialogue after completed training courses etc were often neglected. As person H (employee) stated: “I am quite happy with the training courses, but I could wish for more dialogue and coaching after having completed a training program to plan further work tasks and development... We received reaction sheets almost every time in the end of a training course, but it was never mentioned anything about it again. After completing the reaction sheet, the course was over”. Person I said: “I wish I could have more supervision from my tutor because then it would be easier to maintain and reflect on the newly required information from training courses. My tutor is often travelling, so I can not talk to him face to face”.

Evaluating learning can be seen as a varied procedure. It does not exists any formal documents of how to test learning after completing a training course at DNV M. As the reaction phase; measuring reaction and learning is dependent on the individual course. The interview answers indicate that evaluate learning is not as common as using reaction sheets. On the basis of the different statements, it is possible to suspect that there is a gap between

a) The existing situation and the desired situation of how to measure learning at DNV M.

b) The evaluation literature highlights that measuring learning should be done after completed training course, but this varies at DNV M.

Is it possible to implement formal test procedures in an organization like DNV M, where the culture is so strong with a perception of non-testing? I noticed that the training managers saw the need and importance of more testing in all the training courses. It can therefore be
necessary to establish better formal routines for supervision and follow-up after completed training programs. The question of who is responsible for the formal procedures of measuring learning and analyze the results is a challenge for DNV M. This will be further discussed in chapter six.

5.1.3 Measuring Behaviour.

The third level is how to evaluate behaviour and the author defines it as “the extent to which change in behaviour has occurred because the participants attended the training program” (Kirkpatrick 1998 p.20). It is important that the person who accomplishes the program must have a desire to change and that the person must know what to do and how to do it. The participant must work in the right climate and he or she must be rewarded for changing. Evaluation tools or methods used at this level can be observation and interview over time to measure the sustainability of change. 360-degree feedback can also be used as an effective method, and it does not need to be used before training, because the participants can make a judgment as to change after training, and this can be analyzed for groups of respondents and trainees (Phillips 2003). The most preferred assessment situation is to design tasks around relevant performance scenarios and specific key performance indicators or criteria. Electronic and online assessment can be difficult to use at this level, because the results of the assessments can often be too subjective. Kirkpatrick (1998) states that you can use self-assessment if you use a carefully designed criteria and measurements. To measure behaviour changes are difficult, because it is less easy to quantify and interpret that reaction and learning evaluation. The evaluator must cooperate with line-managers to be able to reach the employees (the participants).

Behaviour evaluation is the extent to which the participants applied the learning and changed the behaviour, and this can be immediately and several months after the training, depending on the situation. Kirkpatrick says that evaluation of implementation and application is an extremely important assessment. There is little point in a good reaction and good increase in capability if nothing changes back in the job; therefore evaluation in this area is vital. Behaviour change evaluation is possible given good support and involvement from line-managers or trainees, so it is helpful to involve them form the start, to identify benefits for them, which links to the level 4 evaluation. Management and analysis of ongoing subtle
assessment are difficult, and virtually impossible without a well-designed system from the beginning.

Person C (competence manager) said that: “To develop simple procedures and evaluation is needed at DNV M. Many employees travel on business, and they can receive other inputs from doing their job, communicating with other equals etc, and this can result in a wryness of evaluating behaviour in relation to the training course”.

Person B (competence manager) said that: “behaviour evaluation is only project-based, and not a formal implemented action. The evaluation methods at DNV M are therefore developed, because of the single projects using a four- and even five level evaluation”. He referred to one example: One NDT-course in Korea was evaluated with a survey 3 months after a completed training course. The employees still considered the course (content and learning outcome) as positive. Their line-managers said that they noticed a difference in how their employees communicated with the customers. The course participants still wanted more training. The conclusion of this evaluation project identified a need to implement level 3 evaluation in more training courses, but the situation is still status quo. How come that it is not happening? One possible explanation can be that the evaluation demands more resources. Or is this an excuse for something else? Several people at M Learning and DNV Learning emphasize 360-degree feedback as a useful method in Kirkpatrick’s third level. The leaders, training managers, colleagues, customers etc are indicators that can be a part of the 360 degree evaluation. Assessment can be designed around relevant scenarios. Person F (line-manager) said: “we don’t have any formal procedures for evaluating behaviour change; we just “feel” when something is right or wrong. If some customers, colleagues or others give negative feedback, of course we take action. We have to rely on our employees that they are doing their job.”

In combination with training courses, the employees are supposed to have continuously dialogue and feedback from the line-manager. According to the DNV-web site, the Measure Individual Performance (MIP) is supposed to be performed twice a year. Person G (employee) said that in his division, they had MIP 3-4 times a year, and the feedback with the line-manager was excellent. The employees had an opportunity to receive feedback- both technical and personal, discuss training courses and further needs of competence development. Person

---

5 www.inside.dnv.com
H (employee) said that in his division, the MIP was once a year and that it was a lack of possibility to have a continuous dialogue and receive feedback from the line-manager. Person I (employee) had not accomplished a MIP this year (2005) at all. This can indicate a gap between the practical performances of MIP between the different divisions. Even though MIP is a formal procedure, its practice varies. The formal procedures are not always in agreement with the reality. Kirkpatrick stresses that the person who accomplishes the program must work in the right climate and he/she must be rewarded for changing.

As mentioned, it can be difficult to measure behaviour, and the evaluator must be able to reach the employee and cooperate with line-managers. Person D (a line manager) said: “Things are working ok, and if there is a gap or error, we will correct it as soon as possible, and try to prevent it from happening again. Formal evaluation procedures will probably be a work-load; we don’t have time or competence to do it…we can not act like strict teachers here, people are professional enough at DNV to make sure their competence is excellent.”

On the basis of the different statements, it is possible to suspect that there is a gap between

a) The existing situation and the desired situation of how to measure behaviour at DNV M.

b) The evaluation literature highlights that measuring can be done after completed training course, but this is not a formal common procedure at DNV M.

How can an evaluator perform a level 3 evaluation when the line-managers do not see the importance of doing so? In this section, it occurs that measuring behaviour is not present formal procedure beside a few projects performed. The MIP can be seen as one contribution to evaluate behaviour (also from training courses), but the practice of MIP also varies. Kirkpatrick says that measuring result can often be easier than measuring behaviour because it demands less resource.

5.1.4 Measuring Result.

The fourth level is how to evaluate the result. The final results can include increased production, improved quality, decreased costs, reduced frequency and/or severity of accidents, increased sales, reduced turnover and higher profits (Kirkpatrick 1998 p.23). Kirkpatrick states that one should be aware of the fact that many organizations already have their own
measurement systems, and therefore it can be a challenge for the evaluator to identify and relate to the trainee’s output and influence. At the beginning of a training program, the course instructor should identify and agree on accountability and relevance, so everybody understands what is to be measured. It can be difficult to measure final results for programs on such topics as leadership, communication, motivation, time management, empowerment etc. These are the kind of conceptions one can refer to as soft skills, because they are more difficult to describe. Further will external factors influence both individual and organizational performance. Therefore it is important to take into consideration that the results can be biased.

It can be important to define the business effect criteria before the training is designed, if not it is difficult to perform a credible level 4 evaluation. Kearns states that no business can claim real organizational effect unless it ultimately results in a financial benefit that would not have been achieved otherwise (Kearns 2005). Level 4 is a great challenge to training professionals. As mentioned, result evaluation is the effect on the business or environment resulting from the improved performance of the trainee, and it is possible that many of these measures are already in place via normal management systems and reporting. (E.g. return on investment, staff turnover, numbers of complaints etc).

It can be easier to measure result or map the cost of in some training programs, i.e. programs with the aim of reducing accidents, increase sales etc. Training programs with the aim of i.e. improving communication can be more difficult to measure because so many factors can influence the result. Therefore, the evaluator must end up with intangible results.

It can be challenging to identify and relate the trainee’s output and influence. To measure result individually is not very difficult; across the whole organization it becomes more challenging. Because of the reliance on the line-management, and the frequency and scale of changing structures, responsibilities and roles, which complicates the process of attributing clear accountability. External factors can also influence the business performance. (Kirkpatrick 1998)

Person C said that there was only one project performed on level 4 evaluation at DNV M with a survey and in-depth interview. M Learning used the ASTD’s Training Scorecard as a further development of Kirkpatrick’s 4 levels to 6 levels. Return on Investment was also used, and the courses were ranked. M Learning has compared themselves with the ASTD’S industry
report. DNV M has a quite high budget when it comes to “Investment Leaders” in ASTD, 56 hours per person per year in MTP. DNV M uses a smaller amount on formal training courses (16 hours in 2003).

Person G (line-manager) said that his department wants to map and use formal procedures of result evaluation, but they don’t know how to do it or who can be responsible of it. He wants more formal and automatic procedures of result evaluation. Person E (line-managers) stated that the communication between the employees and the line managers (together with MIP, on the job-training etc) were enough indicators for evaluating the employees performance. If something went wrong, the line manager was informed. He did not see a need to have formal evaluation procedures to evaluate the result from training programs. He saw formal evaluation procedures as more work demands together with the lack of time. Person B (competence manager) stressed the difficulties of measuring the different results with common procedures. The employees are performing varied tasks in over 100 countries, and the need to take local considerations is also very important. Even though DMV M (Høvik) creates common evaluation procedures, it can be difficult to implement it in all the regions, because of local procedures, local resistance and cultural differences. On the other hand, person B also stated that the importance of creating and having the same tools to measure and evaluate results, are essential for all the DNV departments.

On the basis of the different statements, it is possible to suspect that there is a gap between

a) The existing situation and the desired situation of how to measure results at DNV M.

b) The evaluation literature highlights that measuring results can be done in several ways inside the organization, but I think the formal and common procedures at DNV M varies a lot.

It can be seen as the best criteria of effect are results. An economist will probably want to know the return on investment. By adding a fifth level to Kirkpatrick’s model, Phillips (2003) has reinforced the importance of defining ultimate value in monetary terms (Kearns 2005). What is the ultimate level of evaluation? Is it return on investment (ROI)?
5.1.5 DNV M and ROI.

Many organizations do not perform evaluation in four levels. How can one add a fifth level? Maybe the need of improving the evaluation procedures at a lower level is more present, than implement ROI? Is ROI non-existing in DNV M? It can be interesting to picture the situation from an international perspective. This table shows how companies measure training and ROI today according to Bersin & Associates (2003) from a research of 6000 organizations:

![Information Routinely Tracked](image)

Figure nr.5: International perspectives for measure training and ROI.

This table shows that only 11% measure business impact and job impact. This can indicate that to measure business impact and job impact is challenging for many companies. Phillips stresses that all four levels should be evaluated before performing ROI. Many organizations want to use ROI, but they do not have the tools or knowledge to perform ROI (Bersin & Associates 2003).

I have recognized at DNV M that almost all the training programs are evaluated at level 1, and some courses at level 2. Maritime Learning and DNV Learning update the computer systems of how many participants have enrolled and completed training courses. Thus, there are no indicators from the interview answers that DNV M is evaluating at a higher level. Only
a few projects have completed a fifth level evaluation. As reviewed, Person B (training manager) referred to a few projects in Korea completing five levels of evaluation. This can indicate that ROI is not present in daily routines. The training managers at DNV M want to use and implement a fifth level, but they can all see the limitations and difficulties of doing so. The training managers also highlighted their leaders’ desire of performing return on investment, but resources to perform return on investment are not given. Although an economical perspective of evaluation can eager the focus on evaluation from the top management. As reviewed, according to Phillips, the need to evaluate in all the four levels before evaluating at a fifth level should be present; the return on investment can not be seen in level 4. Kirkpatrick on the other hand states that it is possible to evaluate a fifth level (ROI) in Kirkpatrick’s level 4 (result). The final results (level 4) can include, as mentioned, increased production, improved quality, decreased costs, reduced frequency and/or severity of accidents, increased sales, reduced turnover and higher profits. Therefore, Kirkpatrick argues that return on investment is already present in his level 4, and a need for another level is not present (Kirkpatrick 1998).

5.1.6 Conclusions.

I have now tried to map which level of Kirkpatrick’s evaluation strategies DNV M uses to evaluate their employees. As reviewed, I can see a tendency of level 1 evaluation, and in some training courses- level 2 evaluation. Chevalier (2004) stresses that level 1 and level 2 evaluation can lead to a false sense of security; there may be no relationship between how participants feel about the training and improved individual or organizational performance. M Learning has indeed accomplished a few projects where they used all the four levels (and five), but otherwise there are no formal and continuous procedures of level 2, 3 or 4 evaluation. Even though level 1 evaluation is almost present in all the training courses, the procedures are different, and the focus on evaluation is fluctuating. How come DNV M (as a strong competence organization) does not evaluate at a higher level than 1- and in some cases level 2? One can therefore suspect that there is a lack of ROI evaluation because it is only being performed as single projects, not implemented procedures.

To understand and reflect of the field of evaluation, several other aspects are coming up. As mentioned, some of the interviews were interested in what they called “DNV Culture”, and I think this matches Argyris and Schôn’s conception about espoused theory and theory-in-use.
The interviewees referred to DNV Culture as “the way things are”. When it comes to changing and developing new ideas and visions, how can one link this to what the employees refer to as DNV culture? Is evaluation used as correcting errors, or questioning the whole competence development strategies? All the interviewees involved in HR or in creating the pedagogical aspects of training courses, said that the need of questioning the competence development procedures and change some of the visions (and values) of the evaluation focus was absolute necessary. But how can this change if the top management do not share the same opinion? This can also be seen in relation to espoused theory and theory in use. I will discuss this later in section 5.3. The values often contribute to set the priorities, and as one can see at DNV, evaluation is not a priority from the top-management. The norms inform how the employees should act to be a part of the organization. Person G (line manager) stated that: “why should we evaluate something that works ok? We follow routines here, as we have always done, and I do not see a need to change that”. All the ten interviewees told me about the “strong DNV Culture”; how they emphasized the autonomy and that DNV has a “non-testing” culture on their own employees. The organizational culture is not static and can change. All the employees at DNV have their own personality, their own visions, goals, knowledge etc which influence the culture. The leaders can change the organizational culture by a long commitment and engagement to map and clarify the values and create a debate about the values in the organization.

The four level evaluation model is basically meant for evaluating individuals. As reviewed, the four-five level evaluation methods are used in single projects. Maybe the implementation or problems using evaluation procedures is not the real case. M Learning and DNV Learning know how to use the evaluation methods, but how come it is not commonly used? Maybe there are other factors at DNV M which hinders using evaluation methods? I think Kirkpatrick’s evaluation strategies should be seen in relation to other theory about organizations and organizational learning. This can maybe indicate why evaluation methods are not commonly used at DNV M. In the next sections, I will raise Kirkpatrick’s evaluation methods in an organizational learning aspect, to give a broader perspective of DNV M.
5.2 Evaluation and organizational learning.

“A person who only has a hammer as a tool, will define all problems as nails to strike” (Abraham Maslow)

Argyris and Schön’s conception of single and double loop learning will be highlighted and linked to how evaluation can be seen as an attempt to stimulate double loop learning. As reviewed, Kirkpatrick does not stress how it is possible to use explicit and tacit knowledge in his model, and therefore I have chosen to link his model to Nonaka and Takeuchi’s conceptions about tacit- and explicit knowledge, and middle-up down management. To understand what it takes for an organization to implement a four level evaluation model, Senge can be seen as a contributor with his five disciplines. Even though he does not stress evaluation, I think his five disciplines show a framework for discussing organizational learning and evaluation. In the next sections, I will therefore present the three contributions of organizational learning and link this to Kirkpatrick’s evaluation model.

5.2.1 Kirkpatrick’s evaluation model and single- and double loop learning.

I would like to challenge Kirkpatrick’s evaluation model and see it together with Argyris and Schön’s theories, because I think it can contribute to expand the understanding of the four-level evaluation model. As reviewed in chapter 2, espoused theory can be seen as the organization visions and goals. The theory-in-use is how it really is in reality. Single- loop learning rests in an ability to detect and correct error in the relation to give a set of operating norms. Double loop learning depends on being able to take a “double look” at the situation by questioning the relevance of operating norms and challenge the basic paradigms. One has to review, questioning and challenging the basic paradigms and operating norms.

I noticed in some of the interview answers that some of the line managers and trainers had to “rely on their employees that they have learned something”. How does Argyris and Schön’s idea about learning match Kirkpatrick’s definition of learning at the secondary level of evaluation? Argyris and Schön’s state that learning can be seen as a detection and correction of errors; a process and a product, what we have learned and how we learned it (Argyris and Schön 1996).
Kirkpatrick states that the learning phase is “the extent to which participants change attitudes, improve knowledge, and/or increase skill as a result of attending program” (Kirkpatrick 1998 p.20). One or more of these changes must take place if a change in behaviour is to occur. I think the Argyris and Schö̈n and Kirkpatrick’s understanding and definition of learning differ in some sense. In one way, I think Kirkpatrick’s conception of learning seems more progressive, in other words, it involves a change to something new— not just a correction or detection of errors. On the other hand, Kirkpatrick’s statement about learning is not an evidence that change in behaviour necessarily means questioning and challenging existing norms. The two definitions reviewed above can seem to be in agreement to one another, if one sees learning as a product and a process. As mentioned, Argyris & Schö̈n state that individual learning is an assumption for organizational learning together with the organizational behaviour, action, intelligence and memory. Kirkpatrick does not focus further on how individual learning and evaluation of individuals can be seen as a part of organizational learning process. Therefore, it can be fruitful to discuss how evaluation can contribute to map and identify a gap between espoused theory and theory-in-use and single- and double loop learning. I do not understand Kirkpatrick’s model as stimulating to double loop learning, because he does not stress how to use evaluation results to questioning existing courses or procedures.

I noticed a gap between the theory-in-use and espoused theory when I was interviewing. Persons E, F and G stated: “We know that M Learning emphasizes the importance of evaluation, but we do not always have time to evaluate. The training courses are working ok, and since there are no indicators of errors or bad results from the courses, we do not see the importance of evaluate”. The top management at DNV does not, as mentioned, attach importance to evaluation as much as M Learning does. M Learning’s visions are therefore to attach importance to evaluation, but the line managers’ goals and performance do not match M Learning’s goals and visions. What happens in reality? The interviewees claimed that they were all engaged in evaluation. When I asked more specific about the evaluation procedures and their visions and goals about evaluation, the answers varied. The line managers were satisfied with the existing norms, and did not see a need to change. After analyzing the answers of the interviewees, I noticed that some of the trainers and line- managers showed several signs of Model 1 theory in use: they wanted to keep control, win as much as possible and loose as less as possible. To be as rational as possible and oppress negative feelings were also present.
After reflecting and analyzing the interviews, I saw a tendency of correcting errors (single loop learning). Statements from line managers like: “Things are working ok, and if there is a gap or error, we will correct it as soon as possible, and try to prevent it from happening again”, can be seen as a representative for single loop learning. When I asked about how they can use and focus more on evaluation, many of the interviewees said that lack of time and resources were present. From the line-managers, I did not recognize the desire to review, question or challenge the existing situation of evaluation, because they said that they did not have time. One of the recurrent comments was: “why should we evaluate something that already works ok? If one of my employees has completed a training course, I have to count on that that the course was ok and that they have learned something”. Statements from training managers like: “we have to develop and focus more on evaluation procedures, and question existing norms”, can be seen as a representative for double loop learning. How can the line-managers be sure of that their employees have learned something?

The typical model 1 theory in use makes double loop learning impossible, because people are not able to take a double look at the situation and questioning the reality and behaviour. Argyris and Schõn state that the Model 2 theory in use is a more effective theory, because it challenges to create offensive learning between the individuals and the organization. It is not possible for a human being to act in accordance to model 2 without help, because they do not know how to do it. Therefore, DNV management has to implement Model 2 and “de-learn” Model 1 behaviour.

I think Kirkpatrick’s contribution to evaluation can also be seen as an attempt to focus on tacit and explicit knowledge. By evaluating individuals, one can map their explicit and tacit knowledge in level 2, 3 and 4, with different methods, e.g. interviews, exams, observations, dialogues etc. All individuals, according to Argyris and Schõn, are continually striving to complete his/her theory-in-use in re-describing himself/herself in relation to others in the organization. Individual inquiry leads to both new understanding and action but remains outside the stream of distintively organizational activity and produces no change in the theory-in-use (Argyris and Schõn 1996 p.17). I think Argyris and Schõn’s focus on organizational inquiry can be seen as one way to evaluate: interpretations of past experiences of success or failure, interference of causal connections between actions and outcomes and their implications for future action; analysis of the potentials and limits of alternative
organizational strategies, structures, techniques, information system, or incentive systems (Argyris and Schön 1996 p.17).

A top-down approach to management, especially one focusing on control through clearly defined targets, encourages single-loop learning and discourages the double-loop thinking that is so important for an organization to evolve (Morgan 1998 p.86). If DNV management visions and goals are not clear enough about how they want to perform and focus on competence development, learning and evaluation, there will be a gap between espoused theory and theory-in-use. How can the individuals at DNV create deuteron-learning when the management does not focus enough of double loop learning and Model 2 behaviour? Is it possible to use evaluation results as one possibility to encourage double loop learning? These questions raise an interesting debate about organizational learning. Nonaka and Takeuchi’s conception about sharing tacit and explicit knowledge might to broaden the understanding on how it is possible to stimulate evaluation and double loop learning. I want to lift the discussion and focus on how Kirkpatrick’s evaluation model can be seen together with Nonaka and Takeuchi’s theories about tacit and explicit knowledge and middle-up down management and Argyris and Schön’s contribution as reviewed above. This can contribute to further discussion about evaluation as a tool in organizational learning.

5.2.2 Kirkpatrick’s evaluation model, and tacit and explicit knowledge and middle up down management.

Training courses can be seen as one possibility to share explicit and tacit knowledge by performing different learning activities. Even though Kirkpatrick does not explicitly stress how to share explicit and tacit knowledge (the way Nonaka and Takeuchi do) it can be interesting to discuss how competence development and evaluation can be seen in Nonaka and Takeuchi’s conceptions. According to Nonaka, an organization can not create knowledge without individuals. The organization creates a context where the individuals can create knowledge. The average age at DNV is 41, 2. 7 of the interviewees have been working in DNV M over 25 years. This can indicate that they hold a lot of tacit and explicit knowledge. The new employees do not have that much experience. The challenge and need to share information- both tacit and explicit knowledge, is very important for an organization like DNV.
Individual knowledge can be reinforced through interaction between experience and rationality. To share tacit knowledge is one possible way for individuals to build and create knowledge. As reviewed, Experience Exchange Seminars is based on the idea of sharing and reflecting about learning and work experience in groups, both among surveyors and across regions. On-the-job training, supervision and MIP can also be seen as a possibility to share tacit and explicit knowledge in dialogues and active situated learning. Person H (employee) said that: “The idea of supervision and MIP is great, but the tutors often do not have time to communicate with us. I have not had MIP at all this year, and it is suppose to be performed at least twice a year”. Person I (employee) said: “I have had MIP twice this year, and in my department the focus on evaluation of employees is high. Although I could wish for more supervision with my tutor. He has been working at DNV M for 25 years, and I have probably a lot to learn from him”. Learning from each other and experience learning is very important to enhance and expand the knowledge about DNV’s work. The whole learning cycle of socialization, externalization, combination and internalization can therefore be seen as an important contribution to understand how knowledge and competence development can be seen at DNV.

At DNV M, there is relative autonomy in several levels: individual, groups and organizational. When the individual is acting autonomously, the organization can increase the ability to introduce new opportunities. Autonomy gives the human being a possibility to create and absorb knowledge. Chaos and fluctuation can create new forms of interaction between individuals and their environment. The individuals can therefore question the value of their actions, and this can lead to reconstruction of action and routines (Nonaka 2000). This is similar to Argyris and Schôn’s correcting of errors (single-loop learning). After analyzing the answers from the interviews, I could see that the individual at DNV M has an exceptional autonomy status. How can one evaluate more if the autonomy is very strong and many line managers do not see a need for more evaluation? Why should one use a lot of resources on letting the employees share knowledge when each individual in the same group/division have different desires and needs of it? Person G (line manager) said: “In my division, I have to count on that people take responsibility and know their stuff. The autonomy is very strong here at DNV M. I am getting feedback from others if something is wrong. I could wish for more evaluation, but I don’t have the time.” It can be interesting to picture how the line managers understand evaluation and sharing knowledge. It is possible to see the line
managers as middle leaders; Nonaka and Takeuchi´s concept about middle up-down management will therefore be presented in the next section.

As reviewed in chapter 2, middle managers (often team leaders or line managers) serve as a bridge between the visionary "ideals" of the top management and the often chaotic "reality" of those on the front line of business. In other words, if top management's role is to create a grand theory, middle managers create more concrete concepts that front-line employees can understand. The middle managers may convert the tacit knowledge of both frontline employees and top management, make it explicit, and incorporate it into new technologies, products or systems. In this sense, middle managers are the true knowledge engineers of what Nonaka and Takeuchi call "the knowledge-creating company" (Nonaka 2000). One can argue that the line-managers at DNV M can be seen as middle managers because they are acting as a bridge-builder between visions/goals and the chaotic reality. They are trying to transfer the goals and vision of DNV into the real life. One of my concerns during the interviews was: how can the middle managers act as a bridge between the frontline employees and the top management, when the goals and thoughts about competence development and evaluation vary?

Even between the line managers it exist different thoughts and desires for competence development and evaluation. Again, I suspect another indication that maybe implementing and using evaluation methods, for instance Kirkpatrick model, is not the real problem. It can seem like the variation of visions, goals and practice at DNV M can make it difficult to share the same opinion about how to evaluate competence development, for instance in training courses.

You can not afford to fail in an organization like DNV, because the consequences can be too fatal of loosing human lives. Thus, visions and goals about competence development and evaluation procedures should be consistent and be as much as possible in accordance with the reality. How can DNV M create a common procedure and vision about evaluation? Should they implement Kirkpatrick’s evaluation model in all the departments and make sure it is being used continuously? After discussing Kirkpatrick, Argyris and Schôn, Nonaka and Takeuchi, I am noticing a lack of a general understanding of what it takes for an organization to implement evaluation procedures, and how it can be done. Therefore, I will try to highlight these questions in a further discussion, and refer to Peter Senge and his conception about a
learning organization. It is possible that his five disciplines can show how evaluation can be seen as a part of a learning organization.

5.2.3 Kirkpatrick’s four evaluation levels and Senge’s five disciplines.

“The person(s) who has the gold makes the rules”.

According to Senge it is not enough for a learning organization to simply survive. “Survival learning” or what is more often termed “adaptive learning” is important and necessary. But for a learning organization, “adaptive learning” must be joined by “generative learning”, learning that enhances our capacity to create (Senge 1990). Argyris and Schôn argue that the organization should be controlled by problem solving, but Senge states that it should be controlled by creative tensions. Senge stresses that a goal for a learning organization is to minimize the tension between the visionary and the reality. To create common visions and develop a joint picture of the organizations mission and goals are very important for an organization. I understand that top-managers at DNV M share a common picture of DNV as “the best knowledge organization in their field”. The visions of continuous competence development for their employees is also present. But it exists different visions and goals about evaluation.

How can DNV as a learning organization minimize the tension between the visionary and the reality? The five disciplines that Peter Senge identifies are said to be converging to innovating learning organizations. A central thought in these disciplines are that the employees have to use critical reflection, open dialogue, be open about their value systems and challenge other value systems critically;

1) Systems thinking: this is an integrated element and a bridge builder in the organization, so that its members can understand and reflect how their own action can influence the whole system like DNV.

2) Personal mastery: all the employees at DNV M should have a desire of focusing on their personal visions, focus on their energies, developing patience and seeing reality objectively. Senge states as mentioned that people with a high level of personal mastery live in a continual learning mode. The new employees that I was interviewing at DNV M, all said that they could wish for more evaluation from their tutor or line manager in several levels. More
supervision, guidance and dialogue after completed training programs (especially training courses) together with on-the-job performance were desired. The employees also advertised for more supervision for developing personal vision together with DNV M for further career development.

3) **Mental models**: The different interview answers can be seen as a representative of different mental models. The top-management, the line-managers and employees all have unlike pictures and images that influence how they understand DNV and how they take action. Mental models can hinder learning. If the mental models are different, how can one expect a whole division to share same mental models or shared vision? Senge (1994) states that the management is central for creating and testing existing mental models, to identify and reject defensive routines, to clarify espoused theory and theory-in-use etc. If one can map the existing deep-rooted mental models, challenge them and create a debate in the whole organization, the individuals can be aware of their own mental models and be open for influence from others.

4) **Building shared vision**: DNV M has a shared vision about focusing on competence development. A shared vision about evaluation is not as present. One possibility of building a shared vision about evaluation can come from the top management. If the top management does not stress evaluation, how can one expect line managers, training managers etc to attach importance to evaluation? If the top-management contains almost only engineers, how can one expect that they emphasize a lot on more pedagogical sides, like learning and evaluation? This can indicate a gap between the DNV’s vision as a top knowledge organization and how the reality is: 1) the lack of systematic evaluation systems, 2) presence of view of systematic evaluation not needed. To create common visions in an organization can contribute to stimulate learning through the actor’s loyalty and will to change. To create a bridge between the individual and organizational shared visions are therefore a desired situation (Nordhaug 1990 p.132).

5) **Team learning**: Smaller departments at DNV M can be seen as groups working with common goals and vision. Personal mastery and shared visions need to be seen and acted with other people in team learning. Individual trust in a group contains shared values and shared goals or vision. Creating a dialogue in the group will be essential in a team learning process. If one refers to Nonaka and Takeuchi’s conceptions about tacit and explicit
knowledge in this discipline, one can see the possibility of creating a process of sharing tacit- and explicit knowledge in a team/group. (Senge 1990).

All the disciplines are, in this way, “concerned with a shift of mind from seeing parts to seeing whole, from seeing people as helpless reactors to seeing them as active participants in shaping their reality, from reacting to the present to creating the future” (Senge 1990). Senge argues that learning organizations require a new view of leadership, and the leaders should be designers, stewards and teachers. They are responsible for building organizations were people continually expand their capabilities to understand complexity, clarify vision, and improve shared mental models – that is they are responsible for learning (Senge 1990). Senge’s five disciplines should be seen together with the use of critical reflection and open dialogue. The employees must be open about their value systems and challenge other value systems critically. It is possible to argue that if Senge’s disciplines are implemented together with critical reflection, stimulus to double loop learning can occur. With shared visions and mental models about how to use and perform evaluation, I think the five disciplines can be seen as a framework for evaluation as a tool in organizational learning.

5.2.4 Conclusion.

As reviewed, analyzing DNV M to Kirkpatrick’s levels can indicate that maybe it is not performing evaluation that is really the hindrance. M Learning has indeed completed projects performing a four and five level evaluation. The answers to why DNV M is not performing a four level evaluation routinely can be seen in other aspects. Even though it exist a desire from the top management to perform return on investment, it is not being focused on or performed. The idea of lifting the discussion to an organizational level can contribute to broaden the understanding of evaluation. Argyris and Schôn’s conception of single and double loop learning can be linked to DNV M by seeing their daily work routines as more single loop learning. A gap between espoused theory and theory in use can also be present at DNV M. Employees in several levels in the organization have different thoughts and visions about how to practice the top managements “espoused theory” (DNV’s visions and goals). The theory-in-use can in some cases indicate a gap between the espoused theory and theory-in-use. Nonaka and Takeuchi’s concept of sharing explicit and tacit knowledge can also highlight the field of practical evaluation methods. It is important to have in mind when evaluating, the different values of tacit and explicit knowledge. To share knowledge between experienced employees and newly hired employees (in training courses and supervision) can be a
challenge at DNV M. Senge’s five disciplines can be seen as an idea of creating and sharing visions, goals and mental models to be able to be a learning organization. The challenge for DNV M can be to share visions and goals about evaluation. In the next chapter, I would like to offer some further interpretations of the situation at DNV.
6 FURTHER INTERPRETATIONS.

In 1984, the sociologist Charles Perrow (1984) published a study of risk- and accidents. He claims that many of the already developed technological systems also contain room for accidents to happen. He called them “normal” accidents. It is difficult when the technology is complex and can affect the different parts in unexpected ways, and when the connection is contemporarily hidden for the operator. Perrow states that these kinds of technologies can only be steered by a decentralized organization. But if the technology is closely connected, sequences are following each other without buffers or stops. This can only be steered by a centralized organization. Complex technological systems must according to Perrow, be managed by a centralized and a decentralized organization, but he states further that this is not possible, and therefore accidents happen. Those industries that Perrow is focusing on are dependent on trust to be able to function (Lindøe 2003). On the one hand, DNV M can be seen as a hierarchical organization. On the other hand, since the autonomy at DNV M is so strong, it can indicate that the organization is more decentralized. As mentioned, DNV’s visions are safeguarding life, property, and the environment as a provider of services for managing risk. I think evaluation on several levels is central here. One of my concerns around this area is why the training managers and HR- managers claim that evaluation is so important, but the top management does not seem to stress evaluation. Often only reaction sheets of completed training courses are accomplished, but where is the focus on the actual learning, behaviour and result? A critical goal of training is to improve human performance. Is one supposed to always measure whether or not course participants have learned anything? It can be several reasons why programs are not scored: lack of infrastructure, the program did not have the budget, inability to decide what to measure, or lack of time. This can be business limitations. There is also some confusion about whether assessments can adequately measure actual learning and job performance. In the next chapter, I will basically focus on three areas:

1) The lack of formal procedures,
2) DNV- “the non-testing culture”
3) The gap between theory-in use and espoused theory
6.1 The lack of formal procedures:

According to Phillips (2003) measuring training can be easy to conduct and is widely used. More general business measures are confounded by too many other factors; it is and would be impossible to identify the training impact. Person G (line manager) stated: “To identify the business impacts of training would have been interesting. I think it is difficult, and I do not know how to use the evaluation methods”. Many organizations consider the business impact of training to be the most valuable information to obtain, but very few firms are actually measuring it. Companies can have a lack of tools and processes to make such measurements easy. In most organizations, there is a tremendous opportunity to more effectively measure training. Person A, B, C and D all stated that DNV aims at perform a Return on Investment. One of my concerns here is that: how can an organization strive to reach level 5, when level (sometimes 2), 3 and 4 are not regularly present? As reviewed in chapter 2, both Kirkpatrick and Phillips state that it is necessary to complete all the four levels before one can reach and measure level 5. Kirkpatrick also stresses the need of some formal procedures in all four levels. Analysis of level 1 shows that almost every trainer hand out reaction sheets to the participants after completed training course. The reaction sheets however are quite different, and there are no formal procedures for how they should look like and what they should contain.

I can suspect a desire from the interviewees that even though they have some different thoughts about training and evaluation, they all share the same goal: to be able to evaluate with easy and valid evaluation procedures. When I asked the interviewees about which factors that could be present in a successful reaction-evaluation situation, they all said: “all training courses should have a common reaction sheet followed by verbal feedback and dialogue from the participants”. As reviewed in chapter 5, the training managers saw the need and importance of more testing in all the training courses. In the technical courses, pre-tests, interviews and observations were hardly present. In a few soft skills courses, some pre-test and post-test were present, but there are no formal common procedures in any course on how to evaluate. It depends on the individual course. Person E (line manager) said: “After the employees have completed a training course, it is their own responsibility to maintain and discuss the internalized information with other people to maintain the information from the course”. One of the other line-managers (Person F) said: “we do not have any formal procedures for evaluating behaviour change; we just “feel” when something is right or
wrong… Formal evaluation procedures will probably be a work-load; we do not have time or competence to do it…” One of the line managers (person G) said that his department wants to map and use formal procedures of result evaluation, but they do not know how to do it. I think all of these statements can indicate the lack of formal procedures.

One example of formal evaluation procedures is the Measuring Individual Performance MIP. (For further details see chapter 3). In combination with training courses, the employees are supposed to have continuously dialogue and feedback from the line-manager. According to DNV-web site, the MIP is supposed to be performed twice a year. As reviewed, Person H (employee) said that in his division, they had MIP 3-4 times a year, and the feedback with the line-manager was excellent. Person I (employee) said he did not accomplished a MIP during 2005. This can indicate a lack of agreement on how formal evaluation procedures should be performed.

I can suspect that there are a number of different thoughts about implementation and use of more formal evaluation procedures. Especially the training managers want to have more formal routines. Some of the line managers do not see the need to implement evaluation procedures. Since the practice of evaluation differs in all the departments, there might be a lack of agreement of the importance of having common formal procedures. It can therefore be necessary to establish a common routine for all the training courses. An agreement of how DNV M will consider the evaluation results will be fruitful. It is possible to draw a conclusion that reflection on evaluation results should be taken into consideration together with the trainers and the HR- managers. To establish better formal routines for supervision and follow-up after completed training programs can have a successful outcome. The question of who is responsible for the formal procedures of measuring learning and analyze the results is a challenge for DNV M and M Learning.

DNV M live by evaluate and rank their customers, but I noticed a resistance of ranking the employees internally after training programs. In the next section, I will therefore present my reflections to why DNV M is sceptical of ranking their employees.
6.2 DNV- “The non-testing culture”.

As mentioned, DNV and DNV M live by testing and evaluating their customers. I have noticed a resistance of regarding testing the employees at DNV M. How come DNV tests external customers, but they have a resistance of testing them selves? Is it possible to implement formal test procedures in an organization like DNV M, where the culture is so strong with a perception of non-testing? Kirkpatrick says that it is relatively simple to set up assessments and tests before and after training, but more investment and thought is required for evaluation. The tests do not need to be published (Kirkpatrick 1998). Management and analysis of ongoing subtle assessment are difficult, and virtually impossible without a well-designed system from the beginning. The results from the evaluation can be used in different ways. If the results are being ranked, it is not necessary to publish the results if the participants have a resistance against that. The trainer can discuss the training results with each and one of the course participants (Phillips 2003).

I noticed that the training managers saw the need and importance of more testing in all the training courses. Training managers, HR- managers and line managers stated that it is important that DNV M do not test and rank their employees because they want to maintain the individual freedom and rely on their employees. Person B (competence manager) said that: “we have to emphasize the post-testing, because it gives important information about the learning outcome and as an evaluation of the course. But we have to be careful and not make it too academic, and the questions should contain the most important knowledge”.

The line managers agreed on trying to measure knowledge after completing a course, but they were more sceptical of ranking and test the participants. According to Kirkpatrick, it is possible to use tests, skill practices, role plays, simulations etc. The line managers stated further that they did not always know how they were supposed to measure learning. Person D (competence manager) emphasized especially the need of measuring learning after each course: “DNV culture is very special because we trust our employees, but on the other hand, I see the need of developing more formal tests after attending training courses. We can not afford to fail our knowledge, because it can have fatal consequences for testing boats etc.” Person C (competence manager) said that it is possible to rank the employees in the VQSC elements. He saw a need to rank the employees. But he noticed a resistance from all the departments, because none of the line managers want to rank their employees. He stated
further that it can be fruitful for DNV M to be able to map the “experts” in DNV M, and therefore a ranking system can be an indicator to map the experts.

It is possible to argue that some evaluation methods are used at DNV M, but evaluation demands more resources. As reviewed above, simple procedures and evaluation are therefore needed. Several people at M Learning and DNV Learning emphasized 360-degree feedback as a useful method in Kirkpatrick’s third and fourth level. The leaders, training managers, colleagues, customers etc are elements that can be a part of the 360 degree evaluation. Assessment can be designed around relevant scenarios. If DNV M has a strong non-testing culture, it is possible to argue that it is a challenge to implement more evaluation procedures.

The training managers had a resistance against ranking the course participants, because they said that the “DNV culture” is so strong and they all share the same thought that everybody are equal. The interviewees explained “DNV Culture” as what I will refer to as Argyris and Schön concept about theory-in-use: the reality at DNV shows a strong resistance against testing and ranking their own employees. Testing and ranking external customers are on the other hand accepted. DNV’s objectives and vision of having the best competence level on their employees can be seen in a conflict of testing and evaluating the employees to make sure the employees hold the competence. Statements like: “we can not act like strict teachers here, people are professional at DNV to make sure their competence is excellent” (Person E and F), can be seen as a gap between the visions of DNV as a holder of the best competence in their field and how DNV M document their competence in reality This leads me to further interpretation in the next chapter about the gap between theory-in-use and espoused theory.

6.3 The gap between theory in use and espoused theory

Weick & Sutcliffe (2001) have summarized 15 years of research about robust organizations and claim that they have five general signs:

1. They are constantly concerned by risks and failures.
2. They are open about accidents and they are reporting systematically about all risk-situations and they avoid simplifying and generalize about the learning process. Nothing is taken for granted.
3. They have a continuous attention and focus on existing operations.
The tolerance of doing mistakes is low, and they have a continuous focus on preventing work.

They have respect for – and give power to make decisions to persons with the right competence instead of those in formal positions (Lindøe 2003 p.142).

DNV can be seen as a representative here, because they show several signs of what Weick & Sutcliffe call a robust organization. I think especially nr.4 and nr.5 are present at DNV M. Organizations like this must have an ideology where security is the main vision. To be able to build a system based on trust, they must have access to resources that can be used to improve security. When it comes to certain activities in some organizations, like DNV, the consequences of failing can be so big that you cannot afford to learn from experience. The failing results can be too fatal, i.e. tank ship accident and people's death because of technical defects. Therefore, the importance of developing systems that are robust and enable to avoid accidents are crucial. It will require extra demands for resources to reduce risks and accidents, both when it comes to technologically solutions and human resources. The top management must focus on prevention and evaluation of all situations and operations in the company. The employees can experience work demands that are not appropriate with evaluation and security procedures, and falling into the trap of taking short cuts here can be dangerous. The balance between the ethical demands and the economical demands can be a challenge to retain security (Lindøe 2003 p.145).

The HR-managers stated in the interviews that DNV M has “the best competence level” in their field. It is possible to argue that to be able to state that DNV M has the best competence level, there is a need to know how and why they have the best competence level. Evaluation of their employees is therefore necessary. The top management at DNV M does not focus enough on evaluating the employees according to the interviewees, but M Learning and DNV Learning stress the need of focusing more on evaluation. The organization’s espoused theory can be seen in conflict with the theory in use.

As reviewed in 5.3, the gap between the theory-in-use and espoused theory became clear when I was interviewing. M Learning’s desires and visions do not match the top management espoused theory. Person E, F and G (line-managers) stated: “We know that M Learning emphasizes the importance of evaluation, but we do not always have time to evaluate. The training courses are working ok, and since there are no indicators of errors or bad results from
the courses, we do not see the importance of evaluate”. M Learning’s visions are therefore to attach importance to evaluation, but the trainers or line manager’s practice is not in accordance with M Learning’s visions. But what happens in reality? The interviewees claimed that they were all engaged in evaluation. When I asked more specific about the evaluation procedures and their visions and goals about evaluation, the answers varied. As reviewed in chapter 5, the evaluation practice is different in the division.

A top-down approach to management, especially focusing on control through clearly defined targets, encourages single-loop learning and discourages the double-loop thinking that is so important for an organization to evolve (Morgan 1998 p.86). Several signs of Model 1 can be seen in at DNV M, especially with the line managers; 1) To keep the control, 2) To win as much as possible and loose as little as possible, 3) To be as rational as possible, and oppress feelings (Argyris 1992 p.90). This makes it difficult to stimulate double loop learning, because people are not able to take a double look at the situation and question the reality and behaviour. It is not possible for a human being to act in accordance to model 2 without help, because they don’t know how to do it. Therefore, the management has to implement Model 2 and “de-learn” Model 1 behaviour. Organizational deuterolearning is critically dependent on individual deuterolearning, which Argyris and Schön describe in terms of a shift from Model 1 to Model 2 theories-in-use (Argyris & Schön 1996 p.29). It is possible to state that it is not the use of evaluation procedures or not that is the problem, but how the top management, HR-managers, training managers and line managers communicate. M Learning and DNV Learning know how to use the methods based on earlier projects. But again: why is nothing happening? I am referring to Senge’s idea of creating shared vision and goals for the organization. If DNV M is going to compete and challenge their competitors, DNV M should be able to document and evaluate their employee’s competence. Reflection, dialogue and better communication skills among different levels in organizations can be necessary. This resembles Dale (1999) and his so called K3 level (Competence level) about the importance and necessity of reflections and dialogue in a daily work routine. When the top management does not focus on evaluation, how can one expect the line managers to see the need of evaluation? This can indicate a gap between the DNV’s vision as a top knowledge

---

6 Dale’s third competence level- K3 stresses the need to reflect and questioning existing values, theories etc. Routine based actions can hindrance reflection (Dale 1999)
organization and how the reality is: 1) the lack of systematic evaluation systems, 2) presence of view of systematic evaluation as not needed.

I have presented some interpretations about DNV and their possible lack of formal procedures, and DNV as the non-testing culture. This can make sense in some ways because the gap between theory-in-use and espoused theory is present. I can therefore notice a gap in two ways:

1) What the different levels (HR- and training managers, line managers and employees) at DNV M state and what happens in reality.
2) The existing situation of evaluation at DNV M and how evaluation literature (for instance Kirkpatrick) describes evaluation.

Is it possible for DNV M to use evaluation as a tool in organizational learning? Kearns (2005) presents four steps to how an organization can change toward an organizational learning approach to evaluation:

1) Picture the status quo of evaluation
2) Create awareness of a need to change and the exploration of a new approach to evaluation
3) Transitioning to an organizational learning approach. There is a need for continued understanding and dialog about how existing tools, methods and frameworks for learning- centered evaluation operate in practice. Adoption and implementation of an organizational learning approach.
   - guidelines developed for using evaluation as a tool for the organizational effectiveness;
   - increased professional development opportunities
   - organizational leaders who embrace the role and importance of evaluation in strategic decision making, and are available to participate in inquiry efforts;
   - Providing various kinds of learning opportunities to help employees participate successfully on inquiry processes (e.g., team building, conflict management, deliberation); and
   - Operate from a system orientation with the infrastructure for cross-functional learning.
4) predominance and refinement of the approach (Kearns 2005 p.138).
It is possible to understand Kearns as contributor to combine evaluation and organizational learning. Even though Kearns contribution can seem quite abstract, it is though different from my contribution, because I have used DNV M as a practical example in a case study to reflect in how it is possible to link organizational learning and evaluation. As reviewed, employees at DNV M have quite different understanding of the existing situation of evaluation. It can therefore be challenging to create awareness of a need to change and the exploration of a new approach to evaluation, according to Kearns four steps. Being able to transitioning to an organizational learning approach at DNV M will contain a need for a common understanding and dialog about how existing tools, methods and frameworks in evaluation in practice will also be a challenge.
7 CONCLUSIONS.

Although training is becoming a more focused and developed area in many organizations, the focus of evaluation processes often disappear. In this thesis, I have tried to picture a knowledge intensive organization with risk control as its business, guided by a critical theory perspective. My main research objective was to make an assessment of how evaluation is practised in the case of DNV M to Kirkpatrick’s four level evaluation model. I have tried to picture the status quo of evaluation by interviewing ten key persons working at DNV M. As mentioned, it is possible to argue that DNV M evaluate training courses in level 1 and sometimes in level 2, according to Kirkpatrick's four evaluation levels. DNV M Learning has indeed accomplished a few projects where they used all the four levels (and ROI), but there are no formal and continuous procedures of level 2, 3 or 4 evaluation outside single projects. Even though level 1 evaluation is almost present in all the training courses, the procedures are different, and the focus on evaluation is fluctuating. It is therefore possible to argue that DNV M does not practice a complete four level evaluation according to Kirkpatrick. Even though ROI is performed on a few projects, there are no formal or common procedures of performing ROI.

However, it is important to question Kirkpatrick’s model as well: are Kirkpatrick's views applicable and, if so, should they be implemented under all circumstances? Or, is it possible to avoid some levels under some circumstances? It may not be practical to implement all the four levels in all circumstances. Chevalier (2004) stresses that by documenting the impact that training has made, the training function will stand better against budget cuts and more credible when resources are requested. By performing evaluation systematically, information on continuous improvements can be developed. However, he also states that it may not be desirable, practical or necessary to evaluate in four levels. The training management should choose the level that will produce the information needed to evaluate the training activity.

Analyzing DNV M to Kirkpatrick’s levels can indicate that maybe it is not performing evaluation that is really the hindrance. Since Kirkpatrick’s model holds one perspective; I have tried to broaden this perspective and map how it is possible to use this model by linking it to organizational learning literature. DNV M Learning sees the need to change and create new approaches to evaluation. The need to clarify and picture the existing use of evaluation tools in all the competence development procedures offered can be present. This can lead to
questioning the existing operating evaluation tools. According to Kearns, adoption and implementation of an organizational learning approach is then possible. As reviewed, by questioning operating visions, norms and procedures, this can according to Argyris and Schön double loop learning. When the top management do not focus enough of double loop learning and Model 2 behaviour, the possibility to encourage and create deutro- learning will be a challenge. In some evaluation theory it is possible to argue that it is possible to use evaluation results as one possibility to encourage double loop learning. The challenge inside DNV M is whether evaluation can be used as a tool in organizational learning when the visions, goals vary inside the organization when it comes to evaluation. A top-down approach to management, especially one focusing on control through clearly defined targets, encourages single-loop learning and discourages the double-loop thinking that is so important for an organization to evolve. An ongoing inquiry of evaluation, providing time for reflection and dialog among evaluators, program staff, and organizational leaders should be present at DNV M.

The challenge and need to share information, both tacit and explicit knowledge, is very important for an organization like DNV. Nonaka and Takeuchi’s conception about sharing tacit and explicit knowledge might broaden the understanding of how it is possible to stimulate evaluation and double loop learning. Many employees at DNV M hold a very specialized knowledge- both tacit and explicit. Basically all the line managers have been working at DNV for over 10 years. Thus, on- the- job training, Experience Exchange Seminars, MIP, etc can be seen as a possibility to share tacit and explicit knowledge in dialogues and active situated learning. Nonaka and Takeuchi’s theory about middle-up-down management can also be seen as one way to share tacit and explicit knowledge, if the line managers serve as a bridge between the visionary "ideals" of the top and the often chaotic "reality" of those on the front line of business. But communication between top management, line managers and employees seem chaotic at times at DNV M. One can therefore question the possibility of focusing and facilitating knowledge sharing in DNV M existing situation. It is possible that Senge’s five disciplines can show how evaluation can be seen as a part of a learning organization. The effort to minimize the tension between the visionary and the reality, can comprise can be to create common visions and develop a joint picture of the organizations mission and goals. I have noticed a desire from the training managers to initiate evaluation work, but the difficulty of getting top management support is present. The training managers or HR managers can have difficulties of implementing and influencing the organizational
culture if it is not a desire or demand from the top management. Line managers can have limited experience in performing evaluation on their employees, or they can see more evaluation demands as a workload or nonessential. If line managers do not see a need for more evaluation, the challenges for HR- and training managers increase. Since the visions and goals are very different in all the levels at DNV M, one can questioning how it is possible for DNV M to minimize the tension between the visionary and the reality. If a general understanding of the need to focus more on evaluation is present from the top management, it is possible to implement Kirkpatrick’s four evaluation levels. Integrating the evaluation function and evaluator role within DNV M will be a challenge. The question of who is responsible for implementing more evaluation and how the results of the evaluation should be taken into consideration is also a challenge for DNV M. A central thought in Senge’s five disciplines is that the employees have to use critical reflection, open dialogue, and be open about their value systems as well as challenge other value systems critically.

DNV M lives by evaluating their customers, but inside the organization, there is a lack of systematic approach towards assessment of evaluation procedures. This comprises shortcomings relative to theoretical approaches, which highlight the importance of performing assessment. DNV shows a strong resistance against testing and ranking their employees. DNV M lives by evaluating their customers, but inside the organization, there is a lack of systematic approach towards assessment of evaluation procedures. The visions of DNV as a holder of the best competence in their field and how DNV M document their competence in reality, can indicate a gap in two ways:

1) What the different levels (HR- and training managers, line managers and employees) at DNV M state and what happens in reality.

2) The existing situation of evaluation at DNV M and how evaluation literature describes evaluation and how organizational learning literature describes a learning organization.

A need to clarify and picture the existing use of evaluation tools in all the competence development procedures offered can be present. This can lead to questioning the existing operating evaluation tools. In an organization like DNV M evaluation can play an expanded and more productive role. The future emphasizes a learning approach to evaluation that is contextually sensitive. It is important, however, that the top management, HR- managers,
training managers and employees have an understanding of the evaluation process and use the evaluation findings. The eager to demand outcomes within short period of time can be present. The need to have a dialogue, reflect and raise questions within DNV M will be important. As mentioned, this can lead to double loop learning and even deutrolearning. It might be possible to use evaluation as a tool in organizational learning in the future at DNV M; the need to improve communication awareness and share visions and goals in the different organizational levels will be important for DNV M.
References


INTERNET ADDRESSES:

www.bersin.com

www.dnv.com

Appendix 1

Abstract for all the interview guides: I am taking a master of education and organizational learning at the University of Oslo, and I am writing my thesis this semester. In my practicum period at the University, I stayed 6 weeks in DNV Learning, making a global tailor made training course for the Certification department. I recognize the lack of focus on both process and result evaluation when it comes to evaluation of the employees and training courses many departments. I will therefore create a project for DNV Learning department and ‘assess the evaluation methods’ used on competence development for DNV Maritime department. The proposed research project will identify the gap between the existing situation and the desired situation when it comes to how DNV Maritime is using evaluation strategies on competence development and training courses for their employees. The project will fill part of this knowledge gap by focusing on an “assessment of the evaluation procedures”, mainly using four evaluation strategies authored by one of the most famous professors when it comes to evaluation, Donald Kirkpatrick. I will therefore interview ten people at DNV Maritime to map the situation. I would like to ask you in this interview about your experience and thoughts when it comes to training and evaluation, so you can help me picturing the situation for my thesis.

INTERVIEW GUIDE. For the Top Management.

1.0 QUESTIONS ABOUT THE JOB.
   1.1. How long have you been working for the DNV?
   1.2. What is your present job classification?

2.0 TRAINING
   2.1 How many percentages of the budget are set for training?
   2.2 What kind of training activities (i.e. training courses, supervision, on-the-job training etc) do you have in your division, and what is the reason for choosing these activities?
   2.3 How do you adjust training activities?
   2.3 Can you name three factors that you think contribute to successful competence development training program (training courses, on-the-job performance etc)?
3 EVALUATION

3.1 How do you arrange evaluation of training courses, supervision and on-the-job performance in your division?
3.2 Do you think it's valuable to focus more on evaluation?
3.3 How can DNV become better when it comes to using evaluation strategies? What kinds of factors have to be present if you want to change the situation?
3.4 Do you see any hindrances (either on an organizational level or an individual level) of focusing more on evaluation?
3.5 Think of an experience you had where you felt success from a process or a result of evaluation. What kinds of factors were present at this situation? How did you and your colleges create this situation?

4.0 DNV AS A LEARNING ORGANIZATION

4.1 Can you explain what you mean with the concept of a “learning organization”, and which factors do you think can create a learning organization?
4.2 Where do you place DNV in the field of a learning organization?
4.3 How do you think DNV can reach a better approach of organizational learning?
4.4 How will you place evaluation when it comes to the concept of a learning organization?
Appendix 2

Interview guide for the middle-leaders.

1.0 QUESTIONS ABOUT THE JOB.
   1.3. How long have you been working for the DNV?
   1.4. What is your present job classification?

2.0 TRAINING
   2.1 What kind of training activities (i.e. training courses, supervision, on-the-job training etc) do you have in your division, and what is the reason for choosing these activities?
   2.2 How do you adjust training activities?
   2.3 Are you in charge of or involved in any specific training programs today (training courses, supervision, on-the-job performance etc)? If yes,
      2.3.1 Can you explain your existing role and work tasks with this training program?
      2.3.2 How can this program become better?

3.0 EVALUATION
   3.1 How do you arrange evaluation of competence development (training courses, supervision and on-the-job performance) in your division?
   3.2 Are you involved in evaluation procedures today, and if yes, are you using any specific evaluation models or theory in your division?
   3.3 Do you think its valuable (on an organizational and individual level), to focus more on evaluation?
   3.4 How can DNV become better when it comes to using evaluation strategies? What kinds of factors have to be present if you want to change the situation?
   3.5 Do you see any hindrances (either on an organizational level or an individual level) on focusing more on evaluation?
   3.6 Think of an experience you had where you felt success from a process or a result of evaluation. What kinds of factors were present at this situation? How did you and your colleges create this situation?
Appendix 3

INTERVIEW GUIDE. For the employees.

1.0 QUESTIONS ABOUT THE JOB.
1.5. How long have you been working for the DNV?
1.6. What is your present job classification?

2.0 TRAINING
2.1 What kind of training have you yourself completed at the DNV?
2.2 How will you describe your experience and learning from the training?
2.3 Can you name three factors that you think contribute to successful training courses?
2.4 How can the courses you have completed become better?
2.5 Do you have any thoughts when it comes to further needs of competence development for your self (training courses, supervision, on-the-job performance etc)?

3.0 EVALUATION
3.1 How were the training courses you participated in evaluated, and how were you evaluated?
3.2 What do you think of these evaluation methods?
3.3 How are you being evaluated in your job today?
3.4 How would you like to be evaluated in your job?
3.5 Think of an experience you had where you felt success from a completed training course and evaluation of the training course. What kinds of factors were present at this situation?