



CHALMERS
UNIVERSITY OF TECHNOLOGY



How do organizations evaluate their training in 2022?

Studying the prevalence of Kirkpatrick's four levels of training evaluation

Master's thesis in Learning and Leadership

Jack Ahlkvist & David Larsson

DEPARTMENT OF COMMUNICATION AND LEARNING IN SCIENCE

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022
www.chalmers.se

MASTER'S THESIS 2022

How do organizations evaluate their training in 2022?

Studying the prevalence of Kirkpatrick's four levels of training
evaluation

Jack Ahlkvist & David Larsson



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Communication and Learning in Science
Division of Education Engineering Research
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022

How do organizations evaluate their training in 2022?
Studying the prevalence of Kirkpatrick's four levels of training evaluation
Jack Ahlkvist & David Larsson

© Jack Ahlkvist & David Larsson, 2022.

Supervisor: Michael O'Connell, Communication and Learning in Science
Examiner: Philip Gerlee, Mathematical Sciences

Master's Thesis 2022
Department of Communication and Learning in Science
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Kirkpatrick's four levels of evaluation – reaction, learning, behavior, results –
illustrated. Credit: Max Egnell.

Typeset in L^AT_EX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2022

How do organizations evaluate their training in 2022?
Studying the prevalence of Kirkpatrick's four levels of training evaluation
Jack Ahlkvist & David Larsson
Department of Communication and Learning in Science
Chalmers University of Technology

Abstract

Learning technology has been on the rise in recent years, a rise that was further accelerated by the covid-19 pandemic. This has led to unprecedented growth of educational technology companies who focus on digital learning solutions. Implementing digital learning in an effective way is no trivial task and the need to evaluate training programs to find out what does and does not work is imperative. A popular model for evaluating training is presented by Kirkpatrick. With his model as a base, this thesis investigates to what extent learning professionals actually evaluate their training programs.

Knowly is one of many companies focused on helping other organizations with structuring their digital training programs by using their platform. By analyzing course data from the Knowly platform, the goal is to find out if there are factors that affect the habits of evaluation in courses. Studies investigating evaluation habits of learning professionals are a limited quantity and most studies that exist are based on surveys rather than analyzing individual courses which gives this thesis a unique approach. The results mainly confirm that customer satisfaction is the most common way of evaluating training and that usage of the Knowly platform did not seem to impact the likelihood of evaluating training programs. The results suggest that further research into the motivation behind evaluating training is needed.

Keywords: Kirkpatrick, evaluation, transfer of training, learning technology, learning

Contents

1	Introduction	1
1.1	Aim and purpose	2
1.2	Ethical aspects	2
1.3	Delimitations	3
1.4	Research questions	3
2	Theory	5
2.1	The Kirkpatrick Model	5
2.1.1	Level 1: Reaction	6
2.1.2	Level 2: Learning	7
2.1.3	Level 3: Behavior/Transfer of training	7
2.1.4	Level 4: Results	8
3	Methods	9
3.1	Data from the Knowly platform	9
3.1.1	Organizational factors	9
3.1.2	Course specific factors	11
3.1.3	Criteria for evaluation habits	12
3.1.4	Data analysis	12
3.2	Data from the survey	13
4	Results	15
4.1	Statistical analysis	16
4.1.1	”Evaluation”	16
4.1.2	”Two or three”	18
4.2	Survey	18
4.2.1	Analysis of qualitative survey data	19
5	Discussion	21
5.1	Interpretation of the results	22
5.1.1	Organization and course sample	22
5.1.2	The Knowly platform	22
5.1.3	Survey respondents	23
5.1.4	Comparing the data from Knowly with the survey	23
5.1.5	Choice of methods	24
5.2	Implications of the results	24
5.2.1	Implications for learning professionals	24

Contents

5.2.2	Implications for Knowly	25
5.2.3	Implications for further research	26
	Bibliography	29
	References	29
A	Appendix 1 - Survey	I

1

Introduction

Learning technology is on the rise. In 2019 the world saw all time high investments in learning technology development (Adkins, 2020a) and with the Covid-19 pandemic, digital learning was no longer just a possibility, but a necessity. With continued growth projected throughout the next decade (Adkins, 2020b), the utilization of digital solutions for learning, creating new behavior, and saving time, will see unprecedented highs (Lund et al., 2021). The supply of digital solutions for learning and education is vast, and many tools enable the use of learning models that previously have been complicated and time consuming. With innovation at hand, learning professionals have an opportunity to develop their training programs and learning initiatives in new exciting ways. However, having a lot of possibilities and little time, knowing what to prioritize can be a tough challenge even for the most experienced professionals.

Knowly is a Gothenburg-based company that offers a flexible platform which helps learning professionals to create learning journeys. Learning journey is a term used to describe a sequence of interactions with the aim of optimizing the learning experience. The knowly platform is flexible and enables its users to design their courses in a way that maximizes the effect of the training. The users also have freedom to prioritize as they wish to fulfill their goals and expectations.

Tools provided on the Knowly platform enables the users to gather data for training evaluation. The learning professionals have the freedom of choosing what data to gather for their evaluation. Since the desired effect of training is a matter of discussion, there exists several models that aim to explain how to evaluate for different results. One widely adapted model for training evaluation was proposed by Kirkpatrick (Kirkpatrick, 1959; Kirkpatrick & Kirkpatrick, 2006), who suggested that evaluation could be divided into four levels, where the levels represent four separate approaches to measuring the effect of training.

According to prior research, learning professionals are experiencing a growing pressure in presenting their results. Simultaneously, their knowledge on how to evaluate training and in particular more advanced evaluation, is generally low (Kennedy, Chyung, Winiecki, & Brinkerhoff, 2013). Knowing what kind of results are evaluated on the Knowly platform can provide Knowly with insight into customers and potential customers goals. Identifying patterns regarding the habits of evaluation predicted by factors such as organization field and course subject could give Knowly the opportunity to help their customers reach their respective goals.

1.1 Aim and purpose

The main goal of this thesis is to investigate to what extent organizations gather data to evaluate their training. Both data gathering for evaluation using the features of the Knowly platform and use of evaluation methods outside the platform will be considered. To put the evaluation habits of the different organizations in context, factors describing the organization and factors describing the courses will be gathered as well. The data set created from the courses in Knowly will then be analyzed with the goal of finding correlations between different types of organizations, courses and evaluation habits.

One reason for conducting this study is to find if evaluation habits have changed with the rise of learning technology. E-learning has proven useful in the context of organizational learning (Falconer, 2006; Wild, Griggs, & Downing, 2002) and tools like the ones provided on the Knowly platform gives the course designer new opportunities compared to classic classroom education. Automating tasks like sending emails to training participants enables organizations to put more time into designing the learning experience. This in turn gives room for tailoring the training to the course designers respective goals. However, it is not established how to approach the course design to get the best results. To be able to develop the course design towards gaining the most effect, evaluating the effect of different approaches could be of great interest.

Most studies on evaluation in organizational learning is based on surveys and general adaptation in contrast to observing each individual course (Kennedy et al., 2013; Blanchard, Thacker, & Way, 2000; Kirkpatrick & Kirkpatrick, 2006). Analyzing a set of courses on a digital platform could give insight into if there is a discrepancy between the accounts of the organizations and the evaluation habits observed in the courses. It could also give insight in to how tools that aim to help with evaluation affect our habits of evaluating training.

1.2 Ethical aspects

Regarding the research subject itself there are no ethical factors to consider. There are however some to consider regarding the data collection. A non-disclosure agreement have been signed by both authors with respect to the information available on the Knowly platform. All organizations whose courses are analyzed have given either verbal or written consent for it to be part of the study. It has also been made clear that only the content they themselves have uploaded to the Knowly platform is studied and that responses to their evaluations as well as user information are not part of the data set gathered for the thesis.

1.3 Delimitations

The study will focus on determining to what extent and which type of data organizations gather to evaluate their training. The study does not seek to answer why organizations do or do not evaluate their training. If the data is gathered to evaluate, this is treated as if the data is evaluated upon as well.

The data gathering will be limited to what can be gathered from the Knowly platform as well as a survey issued to organizations analyzed in the study. Further, the selection of organizations subject to the study will be limited to currently paying customers of Knowly, as of February 2022, that gives consent that information about their courses are used in the study.

1.4 Research questions

The issue investigated by this thesis is the habits of evaluating training in organizations. The questions to answer are the following:

- How do organizations evaluate their training?
- What factors affect how organizations evaluate their training?

2

Theory

In a renowned meta study of training evaluation concepts, training evaluation is defined as "the measurement of a training program's success or failure with regard to content and design, changes in learners, and organizational payoffs" (Alvarez, Salas, & Garofano, 2004). Training and education has been a part of society for a long time, but according to the same meta study, the evaluation of said training has garnered the most attention from the 21st century forward. This could be one of the reasons why scholars find that evaluation methods adapted for training are often insufficient, or missing (Eseryel, 2002).

Evaluating different approaches to course design is a powerful tool for any learning professional that seeks to increase the effect of their training (Kirkpatrick & Kirkpatrick, 2006). However, what results should be evaluated in training programs is a more complex topic, and a lot of models and ideas to tackle the problem have been proposed. The general idea of most models is to establish what training should seek to affect, and the question is where the focus of the training designer should be when stipulating their goals. When defining the goals that learning professionals in organizations should aim for, many models propose that the ultimate goals of learning should be the goals of the stakeholders involved (Phillips, 1996; Carnevale & Schulz, 1990; Kaufman & Keller, 1994). Others seek to broaden the ideas of evaluation by creating a more complex model than Kirkpatrick (for example, see Thalheimer, 2018).

One of the oldest and most adapted models for evaluating training was proposed by Kirkpatrick (Kirkpatrick, 1959), and many models since are based on his ideas (Alvarez et al., 2004). For the sake of this thesis we are using the Kirkpatrick model as a base for the study, not because it is the most usable or effective model, but because it is arguably the most comprehensive and most adapted model in the field (Kirkpatrick & Kirkpatrick, 2006).

2.1 The Kirkpatrick Model

The Kirkpatrick model was defined by Kirkpatrick in a series of papers regarding evaluation of training published in 1950s (Kirkpatrick, 1959) and is the most-used training evaluation model in the world (Kirkpatrick & Kirkpatrick, 2006). Since then the model has been criticized, and new versions have been proposed but to this day its level-terminology is widely used as a pragmatic way of discussing training

evaluation. The model is divided into four levels:

1. Reaction
2. Learning
3. Behavior (Transfer)
4. Results

Apart from dividing evaluation into different levels, the model comes with a set of assumptions, which is the subject of most of the criticism (Alliger & Janak, 1989; Blanchard et al., 2000; Holton III, 1996). The assumptions can be simplified into two different statements:

1. Each level is more informative than the previous level
2. The results measured on each level is caused by the previous level

The assumptions are controversial, in particular the second one. This is addressed in the "New World Kirkpatrick Model" presented by the Kirkpatrick foundation shortly before the passing of Kirkpatrick Sr. For the purpose of this study, the New World-model will be referenced to as the Kirkpatrick model, however the assumptions and the criticism will be addressed where it is due.

2.1.1 Level 1: Reaction

The reaction level is referred to by Kirkpatrick as the customer satisfaction level. This level aims to evaluate training on the basis of the participants feelings and impressions. This is commonly gathered through surveys asking the participants what they thought about the course (Kirkpatrick & Kirkpatrick, 2006).

The reaction level is the most criticized level and some argue that including this level in the model was a mistake (Holton III, 1996). Studies show that this level has little to no causal relation to the other levels, and therefore it disproves the second assumption (Alliger & Janak, 1989; Blanchard et al., 2000; Holton III, 1996). Despite the criticism, studies show that the reaction level is the most adapted level by learning professionals (Kirkpatrick & Kirkpatrick, 2006; Blanchard et al., 2000).

According to economists and management theorists, customer satisfaction should be seen as intellectual property, and customer relations are critical to a business success rate (McCull-Kennedy & Schneider, 2000). Even though not all training is sold to a paying customer, it is most likely of interest to any training supplier that their training participants are happy. According to neuroscientists, the environment in which the learning takes place in combination with the attitude of those learning has a significant effect on the learning (Hinton & Fischer, 2010). Since both the training provider and those learning can benefit from a positive experience, there are arguments that support evaluating on level one, even though it is criticized.

2.1.2 Level 2: Learning

Evaluating learning is commonly done through tests, written assignments and presentations. This aim is to evaluate the course participants understanding of concepts, theory and understanding surrounding the subjects covered (Kirkpatrick & Kirkpatrick, 2006). Some theorists argue that achieving a high level of learning is crucial to enable transfer of training, the third level of Kirkpatrick's model (Thalheimer, 2018). It is also shown that the act of testing a participant, in contrast to instructing, enables further understanding of concepts (Whitten II & Bjork, 1977).

In terms of organizational learning, some sources say that the learning level is the least adopted (Blanchard et al., 2000). It also seems that the learning level, when talking about the Kirkpatrick model, is the least studied. There are sources that point to a causal relationship between learning and transfer (Gessler, 2009; Alliger & Janak, 1989), but because of flaws in the model the relationship is hard to establish with certainty (Holton III, 1996).

Evaluation of learning is often seen as the same thing as assessment of learning (Kizlik, 2012). Maybe even more so in organizations, as assessment is a concept mostly adapted in education. Where evaluation is determining if a training participant reaches goals, assessment is monitoring the progress towards said goals along the way (Kizlik, 2012). Assessment is important because research suggests that students need tests to activate their learning potential (Bjork, n.d.; Scanlan, 2012). Of course, not all assessment leads to more learning, but if we can effectively assess training participants, we not only have data to evaluate our training in the future, but also give training participants extra motivation.

2.1.3 Level 3: Behavior/Transfer of training

Transfer of training refers to the "extent of which trainees effectively use the knowledge, skills and attitudes they have acquired in the training context in the work context" (Weinbauer-Heidel & Ibeschitz-Manderbach, 2018). This concept is what Kirkpatrick is addressing on his "Behavior"-level. A course after which almost none of the participants change their behavior is considered to have a low amount of transfer of training. A course with a high level of transfer of training will result in most participants applying their new knowledge to their work.

Evaluating behavior is a lot more time consuming than the first two levels since you need to let some time pass from the training to let the behavioral change manifest (Kirkpatrick & Kirkpatrick, 2006). Clear goals for what behavior the training seeks to change is also preferable (Weinbauer-Heidel & Ibeschitz-Manderbach, 2018). Behavior is generally evaluated through observations, interviews, surveys or both. Of these, the only observable method on the Knowly platform are the surveys.

Transfer researchers claim that about 10 to 30% of training is actually used at the workplace (Weinbauer-Heidel & Ibeschitz-Manderbach, 2018) and they propose that

the remainder of the time spent in training is in most cases useless since it does not affect our performance at the job. This discrepancy in the time and money spent in training versus the effect it has on performance is by researchers called the "transfer problem" (Baldwin & Ford, 1988). To be able to tackle the problem, evaluating the transfer of training is crucial.

2.1.4 Level 4: Results

The fourth level of evaluation, "Results", focuses on the organizational results that can be affected by the training (Kirkpatrick & Kirkpatrick, 2006). To evaluate on the fourth level, some factor that the training aims to affect has to be isolated and measured before and after training. The factors affected can be anything that is of interest for the organization, from less injuries to higher revenue.

Kirkpatrick's assumptions imply that level four is the most informative measure of training effect, and experts argue that the ultimate measure of successful training should be acquired revenue, even though others suggest that it is more complicated (Alliger & Janak, 1989; McEvoy & Buller, 1990). Some popular refined versions of the Kirkpatrick model suggest to add levels that specify the outcome. Levels added are for example societal outcomes (Kaufman & Keller, 1994) and return of investment (Phillips, 1996; Carnevale & Schulz, 1990).

3

Methods

To be able to answer the research questions, data was gathered on the habits of evaluation in organizations. The data gathering for this thesis was divided into two separate data sets. One set was gathered directly from observations on the Knowly platform and the other from survey results. The Knowly platform was the main source of data for the thesis, but since not all organizations using the Knowly platform for their training use it for evaluation, the survey acts as a complementary data source to get a more complete view of the organizations evaluation habits. After the data was collected from the Knowly platform it was analyzed with statistical tools to determine if any significant differences in the habits of evaluation between different categories could be found.

3.1 Data from the Knowly platform

The data gathering from the Knowly platform consisted of manually analyzing all the courses created by the participating organizations. Each course was categorized based on factors describing the organizations themselves, the course subject, design choices and at what levels evaluation was performed. The data was documented in a spreadsheet to be used for data analysis. The factors that could be interpreted as binary were documented as 1s and 0s for the data analysis. For the purpose of accuracy, only courses that had evidence that they had actually been completed or were currently in progress during the data gathering were analyzed. Duplicates of the same course have also only been counted once unless there were changes in the course material between the courses. In total 25 organizations were represented with a total of 166 courses.

3.1.1 Organizational factors

To find out what organizational factors might affect the evaluation process the organizations must first be categorized. The two factors that will be analyzed are size of the organization and their line of business. Information about these factors were collected from various online databases, such as www.allabolag.se, www.ekonomifakta.se, the organizations page on www.linkedin.com and the organizations respective websites.

The categories of size used are micro, small, medium and large based on the categories used by the *Organisation for Economic Co-operation and Development* (n.d.).

3. Methods

The categories are only based on the number of people the organization employs. Other classifications of organizational size might use the company total capital as well, but this information might be considered sensitive and was thus disregarded for the purpose of this study. The division of the categories as well as the amount of organizations and courses in each category can be found in table 3.1.

Category	Number of employees	Number of organizations	Number of courses
Micro	9 or fewer	6	19
Small	10-49	4	7
Medium	50-249	3	34
Large	250 or more	12	106

Table 3.1: Amount of organizations partaking in the study categorized by size, the definition of each category, and the number of courses in each category.

For line of business the organizations were categorized based on the their SNI-code (Statistikmyndigheten, n.d.) and then further generalized by the authors⁷. The categories defined are as follows:

- Business service - Organizations selling services such as audits or organizational consultants.
- Private service - Organizations selling services to private individuals, for example pharmacies or dentists.
- Retail - Organizations that sell products to private individuals or other organizations.
- State - Municipalities and counties as well as interest groups between these entities.
- Technology - Organizations that develop technological innovations.
- Training - Organizations that create training programs for other organizations.

The amount of organizations and courses in each category can be seen in table 3.2.

Line of business	Number of organizations	Number of courses
Business Service	2	53
Private Service	2	6
Retail	4	33
State	2	15
Technology	4	25
Training	11	34

Table 3.2: Amount of organizations partaking in the study categorized by line of business as well as the number of courses in each category.

3.1.2 Course specific factors

Course specific factors that have been taken into consideration are the subjects of the courses as well as course design choices. For the purpose of analyzing how the subject of the courses affect the habits of evaluating the training, the subject matter of the courses had to be categorized.

To categorize the course subjects the content of each individual course had to be analyzed. These categories were formulated during the data gathering process according to observations. The final categories formulated from the data set and are defined as follows:

- Sales - courses that teach general sales strategy or product knowledge.
- Onboarding - courses that introduce employees to new working conditions or to the company as a whole.
- Skills - courses that teach new skills, for example *spreadsheets, it systems* or *law*.
- Leadership & Coaching - courses that teach strategies for self leadership, leading others or coaching.

In table 3.3 the amount of courses categorized by course subject can be found.

Course subject	Amount of courses
Leadership & Coaching	73
Sales	38
Onboarding	11
Skills	44

Table 3.3: Amount of courses categorized by course subject.

Beyond analyzing to what extent the individual courses were evaluated some other factors were also taken into account regarding general course design. These design choices were also analyzed to see if they had an impact on course evaluation habits. The four factors that were investigated were defined as follows:

- Formative knowledge checks - Courses that use activities that could be used to evaluate learning but no data was gathered for evaluation.
- Setting goals - Courses that establish goals for what the participants should have learned by the end of the course which could be used as a basis for further evaluation.
- Gathering participant expectation - Courses that gather participant expectations at the beginning of the course.
- Self-reflection - Courses that use self-reflection activities.

3.1.3 Criteria for evaluation habits

In order to determine if a course meets the criteria to be considered to have evaluated at a certain level, conditions must be standardized. The conditions formulated have mainly been based on Kirkpatrick's model (Kirkpatrick & Kirkpatrick, 2006).

For a course to be considered to have evaluated on level one it must include some sort of activity that gathers data about the participants' reaction to the course. This is typically done through surveys on the Knowly platform, but evaluation on other platforms have also been counted if there was a clear link to the survey on the Knowly platform.

For a course to be considered to have evaluated at the second level there must be some sort of knowledge check that gathers the answers that the participants submit. This is typically done through quizzes on the Knowly platform, but if it was clearly stated that a knowledge test would be performed on another platform or at a scheduled event, it has also been counted to have performed evaluation at the second level.

For a course to be considered to have evaluated at level three the course had to include a clear evaluation method at least three months after the course's initial conclusion. For a course to be counted there had to be evidence that there were evaluation data gathered at this point and thus surveys are again the main way that has been considered for the evaluation to count.

For level four evaluation the factors evaluated cover a very large spectrum. Many of these factors need tools and data that is not available on the Knowly platform. There certainly are scenarios where level four evaluation would be possible on the Knowly platform, but since there are so many situations where it is not possible, information about evaluations on level four was not gathered from the Knowly platform.

3.1.4 Data analysis

To see how different categories affect the habits of evaluation, the categories were used to predict the frequency of evaluation found in the courses. The analysis was made out of comparisons between categories and their relative effect on the frequency of evaluation. Evaluation is either present or not present in the data set, so the frequency of evaluation was defined as a percentage of courses in that category where evaluation was found. The percentage was then tested against the null hypothesis with a one-sided student's t-test and the result was presented in the form of a p-value. The percentage for one particular average was called the intercept when used in the comparison with others.

After the intercept was calculated it was compared with all other categories. When compared, the other categories were calculated in relation to the intercept, so that their relative percentage could be tested against the null hypothesis with a 2-sided

students t-test. This result was also presented as a p-value. When testing for significance in both the intercept and the relative average, the cutoff for the p-value was 0.05 or 5%, where all lower values were treated as significant.

3.2 Data from the survey

Not all organizations evaluate their training through Knowly, and therefore a survey was sent out to participating organizations with a set of questions regarding their habits of evaluation. The main purpose of the survey was to get a more complete view of the organizations' evaluation habits. In the survey the respondents were asked to report if they evaluate the courses they had on the Knowly platform outside of Knowly for each of the four levels. They were also given an opportunity to elaborate on their evaluation habit on each level. The comments was later thematically analyzed as suggested by Metodpraktikan (Esaiasson, Gilljam, Oscarsson, & Wängnerud, 2007).

The questions were accompanied with information that explained each of Kirkpatrick's levels of evaluation, from the authors point of view. The respondents were asked to use the definition provided even if they were already were familiar with the evaluation framework. This was done in an effort to normalize the bias in the interpretation of the evaluation framework.

Since the survey was designed by the authors based on survey theory (Esaiasson et al., 2007) the survey design was subject to a pretest. During the pretest five respondents, of which two were surveyed by phone and three filled in the digital survey, had the possibility to issue feedback on questions and design. After reviewing the feedback, no significant changes were made between the pretest and the final survey design. The full survey can be found in appendix A.

4

Results

The results are divided into three sections, first, the findings on the evaluation habits from Knowly is presented in a bar chart, here is also the findings on different design choices presented. Second, the statistical analysis of the data from Knowly is presented, and last the results from the survey is presented.

The following figure shows the share of courses from Knowly that evaluated on the first three levels respectively. The percentage is out of the total 166 courses analyzed. It was found that 40.4% performed evaluation on at least one of first three Kirkpatrick levels.

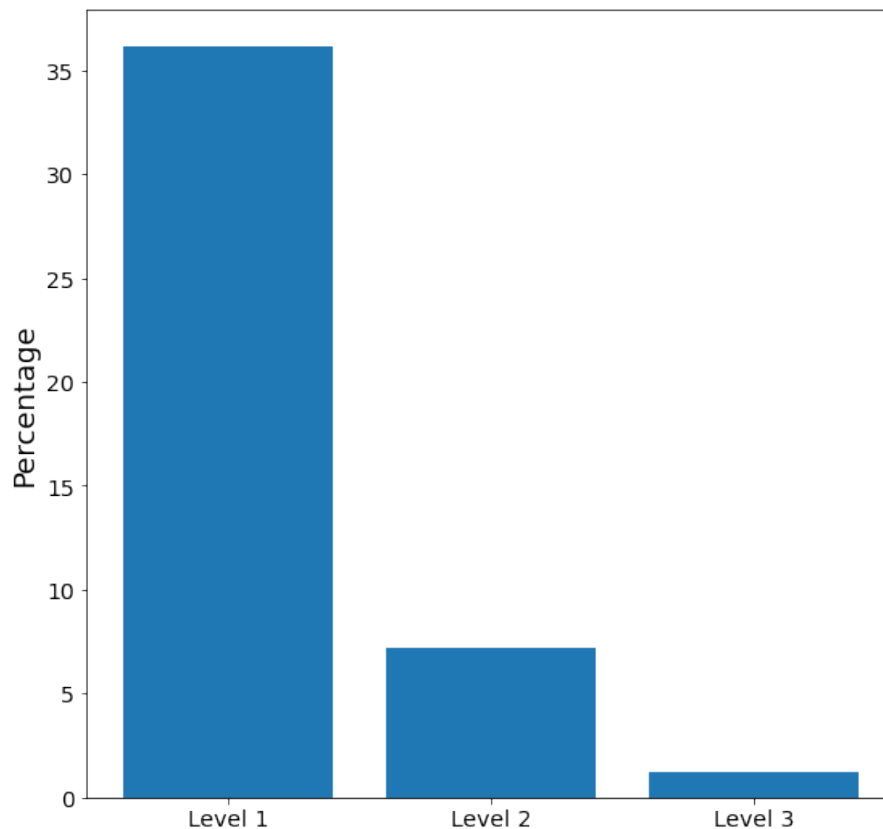


Figure 4.1: The percentage of all 166 courses analyzed that perform evaluation at Levels 1, 2 and 3.

Of the 166 courses that were analysed for the thesis 49 of them used formative exercises. The amount of courses that set up goals that could be used to evaluated

at the third level was 113. Finally, 89 of the courses used or encouraged self-reflection exercises and 87 gathered participants expectations ahead of the training.

4.1 Statistical analysis

The next step was analyzing the results to compare the effect of different categories on the frequency of evaluation. While analyzing the results, it was found that evaluations at the second and third level were too rare to make any statistically significant observations. For this reason the results have been combined into two separate categories: "Evaluation" and "Two or three".

"Evaluation" refers to courses that evaluate at any of the first three levels. Any course that has any type of evaluation present in Knowly is included in this category. "Two or three" refers to courses that evaluate at the second or third level. This is the same as "Evaluation" but without those that *only* evaluate on the first level. This is to show if different factors and categories can affect different types of evaluation. Significant results have been sought for all classifications within the same category, but only correlations that were significant enough, that is with a p-value lower than 0.05, are presented in the tables below. Prediction of evaluation on only level one was not included because it is implied by the relation between "Evaluation" and "Two or three".

Each table consists of three columns. The first column describes what factor is investigated, with the category that had the most significant intercept value as the first entry in the column. The intercept is compared to the category listed below. The first entry in the second column lists the percentage of courses in that category that performed the sought type of evaluation. The second percentage shows how many more or less (+ or -) percentage points the second factor is to evaluate compared to the first factor. The third column displays the statistical significance of the value of the intercept and the value of the difference from the intercept respectively.

4.1.1 "Evaluation"

In 4.1 we are comparing categories of organization field and their effect on evaluating. The most significant intercept value was Business service which is compared to Technology organizations. The intercept value of 45.28% indicate that an average of around 45.28% of courses under organizations in the field business service show some sort of evaluation in Knowly. The -25.28% indicates that it is around 25.28% percentage points less for Technology organizations courses in Knowly to have any type of evaluation.

In the third column we can see that the p-value approaches zero for the intercept value and that p-value is 0.033 for the difference to the technology organizations. Note that we should not read this as public service companies having an average of 45.28% of courses evaluated, due to sample sizes and variance, but we could interpret

that we see a significant trend that technology companies show a lower frequency of evaluating in Knowly.

Organization field	%	p-value
Intercept: Business service	45.28	$< 10^{-9}$
Technology	- 25.28	0.033

Table 4.1: The difference in evaluation frequency between courses made by public service organizations compared to technology organizations.

Table 4.2 shows the difference between courses that do and do not use self reflection activities. Courses that use self reflections seem to have a higher frequency of evaluation than courses that do not.

Course design category	%	p-value
Intercept: No self reflection	28.57	$< 10^{-6}$
With self reflection	+ 21.99	0.004

Table 4.2: The difference in evaluation frequency between courses that do and do not use self reflection exercises.

Table 4.3 show the frequency of evaluation for courses that do and do not gather participants expectations at the beginning of the course. Courses that gather expectations seems to have a higher frequency of evaluation.

Course design category	%	p-value
Intercept: No expectations	24.05	$< 10^{-5}$
With expectations	+ 31.12	$< 10^{-4}$

Table 4.3: The difference in evaluation frequency between courses that do and do not gather participant expectations.

Table 4.4 show the impact that using formative knowledge exercises have on the likelihood of evaluating. Courses using formative tests seem to be more likely to evaluate than those who do not.

Course design category	%	p-value
Intercept: No formative tests	31.62	$< 10^{-10}$
With formative tests	+ 29.60	$< 10^{-3}$

Table 4.4: The difference in evaluation frequency between courses that do and do not use formative knowledge tests.

Table 4.5 shows that courses that include goals focused on behaviour seem to have a higher frequency of evaluation than those courses that do not.

Course design category	%	p-value
Intercept: No goals	28.30	$< 10^{-4}$
With goals	+ 17.72	0.030

Table 4.5: The difference in evaluation frequency between courses that do and do not set goals focused on behaviours.

4.1.2 "Two or three"

The following tables show the categories that seem to have significant impact on evaluation frequency at the second level, third level or both. Table 4.6 shows that medium sized organizations seem to be more likely to evaluate at level two, three or both than large organizations.

Organization size	%	p-value
Intercept: Large	5.66	0.037
Medium	+ 11.99	0.029

Table 4.6: Comparison between large and medium organizations evaluation habits at the second or third Kirkpatrick level.

Table 4.7 shows that evaluation at level two, level three or both seem to be more frequent in onboarding courses than courses in leadership & coaching.

Course subject	%	p-value
Intercept: Leadership & Coaching	6.85	0.031
Onboarding	+ 29.51	0.001

Table 4.7: Comparison between leadership & coaching courses' and onboarding courses' evaluation habits at the second or third Kirkpatrick level.

4.2 Survey

For the survey there were three possible answers for whether they evaluate on each of the four levels respectively: "Yes", "No" and "I don't know". For the respondents that answered "I don't know" they have been categorized as "Yes" or "No" if their comments could be clearly interpreted as such. If no such interpretation was possible they have instead been excluded from the data set for the purpose of calculating the percentages. A total of 17 organizations answered the survey. Figure 4.2 shows the percentage of the respondents that report that they evaluate on each level respectively. Using the previously defined "Two or three" category it was also found that 70.6% answered that they perform evaluation on at least one of the second and third level.

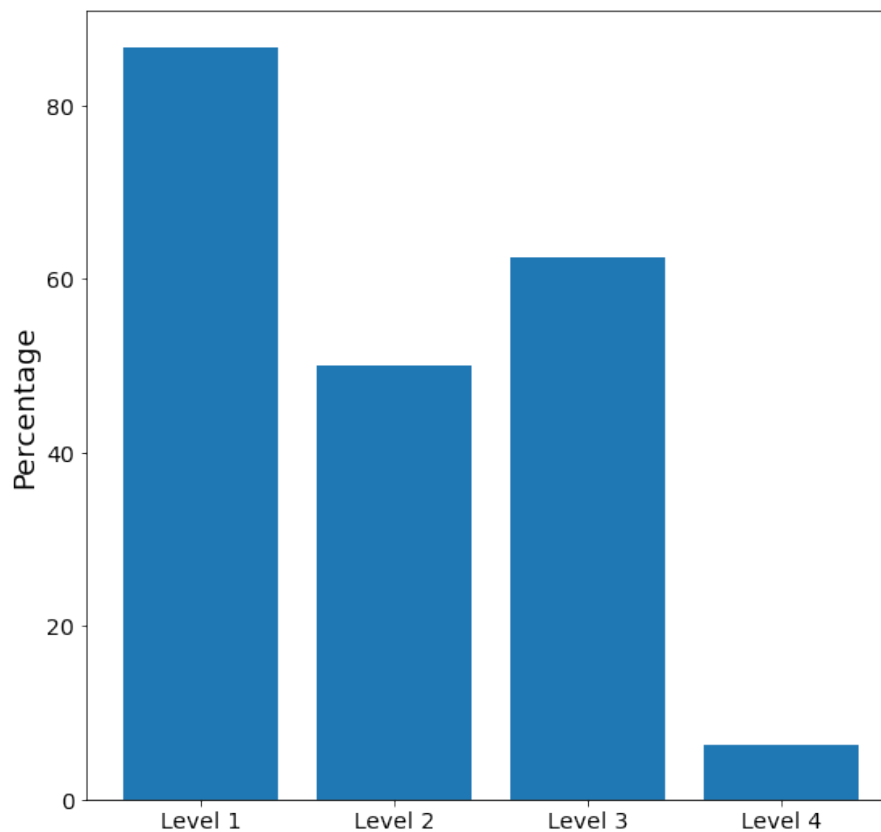


Figure 4.2: The percentage of participating organizations that evaluate at each of the four levels as found from the survey.

4.2.1 Analysis of qualitative survey data

For each of the questions the respondents were given the option to elaborate on their evaluation habits. Following a thematic analysis, these are the trends that could be found based on their answers.

- Nine respondents chose to elaborate further on level one. The trend that could be found in the answers were that many organizations evaluate level one with written surveys on other platforms than Knowly.
- Ten organizations shared their thoughts on evaluating on level two. The answers did not show any particularly strong trends on how they evaluated.
- Twelve respondents elaborated on their habits of evaluating on level three. No apparent unified trend could be drawn from the responses, but the general way of evaluating level three seems to be verbal communication and reflection.
- Of the eight organizations that shared their thoughts on evaluating level four, the general trend was that it is hard, and that it does not always fit into their roles since they do not have the data needed.

5

Discussion

The results answers the research questions with small amounts of interpretation needed, but for clarity the discussion will return to the research questions together with the interpretation of the authors.

- How do organizations evaluate their training?

Evaluation can be defined in many ways, for the sake of this thesis the authors chose to follow one of the most discussed and implemented models of training evaluation: Kirkpatrick's '4 levels' model (Kirkpatrick, 1959). The organizations that are represented in this thesis are customers of Knowly, and their habits have been observed in two separate ways. First, by gathering data from the Knowly platform, and then by asking respondents of the organizations for their experiences.

Both the data from Knowly and the survey concludes that evaluating courses based on the participants satisfaction is the most common, which confirms prior research (Kirkpatrick & Kirkpatrick, 2006; Blanchard et al., 2000). The data gathered from Knowly suggests that it is very uncommon to evaluate based on learning and behaviors. The data from the survey suggests that it is more common, but both suggests that the frequency of evaluating learning in relation to evaluating on behavior is roughly the same. It is worth noting that the data from Knowly suggest that about half of the courses do not evaluate at all, and that the survey suggest that about 70% of courses evaluate on learning or behaviors.

Since the fourth level, 'Results', is not really possible to evaluate or observe on the Knowly platform, the only source available for the habits of evaluating on such factors is the survey. The survey suggests that the least common way to evaluate training is based on level four, organizational results.

- What factors affect how organizations evaluate their training?

The definition of factors and how they affect the evaluation of training for the purpose of this report is less straight forward. Comparing different categories of courses with each other shows that there are some significant differences in the habits of evaluation. Note that this is based on the data from Knowly only.

When it comes to factors based on the organization, the ones gathered were size and field. The size of the company seems to have some effect on evaluating based on learning and behavior, but there is no linear trend present. The technology

organizations seem to evaluate less than other organizations, but since that result does not translate to level two and three, the trend is most likely only seen on level one.

Courses had the categorical factor 'subject' and binary design choices. The courses with the subject 'Onboarding' saw a significantly higher frequency of evaluation based on learning and behaviors than the subject 'Leadership and Coaching'. Leadership and Coaching did not have any significant difference to any other subject. The binary design choices were 'Self reflection', gathering of 'Expectations', using formative tests, and stipulating transfer goals. We could see that all these design choices had a significant impact on the habits of evaluation. This trend, like with technology organizations, did not translate to levels two and three, and therefore is most likely to affect organizations habits of evaluating on level one.

5.1 Interpretation of the results

Even though the research questions are possible to answer from the results gathered, there are sources of error as well as points of interest to discuss regarding the credibility of these answers.

5.1.1 Organization and course sample

As previously stated, the sample of organizations that was used for this report only included customers of Knowly. This affects the possibility of generalizing these results, since the customers of Knowly are all affected by their relationship with Knowly. However, since the sample is only customers of Knowly a new possibility of interpreting the results as habits of the customers of software solutions for enhancing learning arises, which is of interest. This interpretation also has to be made with caution, since different software solutions can have different effects on the sample.

Worthy of note is that about half of the organizations and courses came from 'Training' and 'Business service' organizations, which implies that they might not create the courses for themselves but for their customers. This might have impact the results, since these organizations could be more pressured with respect to the price of the courses, i.e. the budget they can put into the course as well as the time they spend on each case. They might also not have the same insight into their customers organization and therefore not have the possibility to evaluate the effects of the training. This notion is also suggested by the survey respondents, as the trend on level 4 evaluation was that it is complicated.

5.1.2 The Knowly platform

Even though it is possible to gather evaluation data in Knowly, there is nothing that stops users from doing it elsewhere. This is the reason for the survey and the reason why the results found in the data gathered from Knowly has to be interpreted as evaluation made in Knowly and not general habits of evaluation. The Knowly

platform is based on forms and quizzes for gathering evaluation data, which means that other forms of evaluation would not generally be found on the platform. Since the comments from the survey suggests that evaluation made on level two and three is seldom done in written form, it is probably the case it is not possible to observe on the platform.

Different organizations can have different approaches to using the Knowly platform. Technology organizations show a lower frequency of evaluating on the platform, but these could have adopted evaluation methods outside Knowly to a larger extent, based on factors that was not present in this study. It is also possible to observe that in general, organizations that use more design elements in their courses will have a higher frequency of evaluation. This could be interpreted as a higher level of awareness in the design choices of their courses, but it could also mean that the Knowly platform is more prominent in how they deliver courses. Other organizations that do not show the same course design and frequency of evaluation could have the same course design and evaluation outside of the Knowly platform.

5.1.3 Survey respondents

The strength in gathering data directly from the platform is that every course is analyzed in the same way, and that the bias in the data set can be somewhat controlled. The bias that is present in surveys is harder to control since every respondent has their external and internal influencing factors. The survey conducted for this report requires the respondents to understand the Kirkpatrick model in the same way, and leaves the respondent to interpret the meaning of every level. An effort to explain each level in the survey was made, but it is possible that the levels was interpreted differently by different respondents.

Since the Kirkpatrick model is widely recognized and many are familiar with its concepts (Kirkpatrick & Kirkpatrick, 2006), it is possible that some respondents are also familiar with its criticism and assumptions. This could in turn lead to respondents being more prone to answer that they evaluate on level two, three and four than level one. However, the survey results does in general follow the results of previous studies (Kirkpatrick & Kirkpatrick, 2006; Blanchard et al., 2000).

5.1.4 Comparing the data from Knowly with the survey

Referencing the interpretation of the data from the Knowly platform we can safely say that there should be differences in the results from the platform and the survey. The differences are mostly explained by the evaluation that is done outside the platform. Looking at the trends, the data from Knowly suggests that evaluation on level three is the least common while the survey suggests that level two is less common than level three. This could suggest that courses are evaluated on level three by verbal communication and not written forms, which is also supported by the comments from the survey.

5.1.5 Choice of methods

The results of the study were influenced by the choice to conduct a quantitative study instead of a qualitative one. The quantitative approach this study adapted was based on the four levels of training evaluation provided by Kirkpatrick, but in reality we do not know if the organizations actually follow the same framework. A qualitative approach could be designed to study the actual evaluation practices, by using for example grounded theory, which could help to uncover the practices that are not covered by the Kirkpatrick model. With this approach the results could be more true to reality.

The choice to approach the research questions with a quantitative method was mainly motivated by the relative lack of quantitative research in the field of training evaluation, as well as the unique opportunity to gather large volumes of data presented by the Knowly platform, which could produce significant results. A more experimental qualitative approach would require more time to gather each data point (Esaiasson et al., 2007) which would yield a smaller data set considering the time constraints of the study. This would in turn be less suitable for statistical analysis and could affect the possibility to generalize the results.

5.2 Implications of the results

After interpreting the results the next question to discuss is what the results could mean for those that are affected by them. For the sake of this report, the stakeholders that are taken in to account are the following: learning professionals, i.e. those that in organizations work with learning and course design; Knowly and other suppliers of tools that seek to enhance learning effect, and scholars that seek to research this topic further.

Independent of the stakeholder it is necessary to look at the meta level interpretation of the evaluation habits. What does it mean that a course gathers data to evaluate on a certain level? The reasons for evaluating at a certain level is not sought in this study, and neither are the reasons for not evaluating. Prior research suggests there are many reasons to evaluate on every level, some of which are accounted for in the theory section. Evaluation researchers suggests that the reasons that learning professionals do not evaluate on levels three and four are of lack of knowledge and lack of time (Kennedy et al., 2013; Blanchard et al., 2000). Level one is the only level that scholars argue against evaluating, since it does not correlate with learning (Holton III, 1996).

5.2.1 Implications for learning professionals

The results suggest that the most popular data to gather for evaluation of training is customer satisfaction. This is despite the fact that it arguably does not correlate with learning. It is possible that the goal of the training is to create as much customer satisfaction as possible, since researchers suggest that customer sat-

isfaction correlates to business success rate (McCull-Kennedy & Schneider, 2000). However, if learning professionals only evaluate their success based on customer satisfaction, they can not know if their course will increase the transfer of training, which some evaluation theorists argue is more important (Weinbauer-Heidel & Ibeschitz-Manderbach, 2018; Baldwin & Ford, 1988). If learning professionals seek to optimize their effect on their course participants behavior, research suggests they have to gather data on learning, behaviors or both (Gessler, 2009; Alliger & Janak, 1989).

The results suggest that level four is the least evaluated, by a large margin. This is even though some evaluation theorists suggest that the goals of training should always be mapped to the goals of the stakeholders (for example, see Phillips, 1996). Knowly is a tool that is most commonly used to administrate material for courses, sometimes called "formal" training. Learning in organizations is not only formal training, research says that about 90% of learning in organizations happen outside of formal training, in the flow of work (Jennings, 2013). Of course all learning should map to goals of the ultimate stakeholders, however it could be argued that learning in the flow of work is better suited to be evaluated directly towards organizational results than formal training, since the results sought are in many cases true to standard operations. Training programs that are not a part of standard operations will have an indirect connection to organizational results and some scholars suggest that trying to map training programs to organizational results could have a negative effect on the possibility of measuring the effect of the training (McEvoy & Buller, 1990).

With that said, creating training goals that correspond to level two or three could be more effective than only using level four. Transfer experts propose that transfer goals, i.e. goals that evaluate level three, should be created with the organizational goals in mind (Weinbauer-Heidel & Ibeschitz-Manderbach, 2018). This approach seems fitting with respect to the results gathered by this study, since few seem to evaluate their courses on level four, even though it is important. Evaluating level two through assessments is also something that the results suggests might be underestimated in terms of the value it brings to the course participants. Independent of the actual measured learning that could be evaluated, the act of assessing learning should have a positive effect on the learning itself (Bjork, n.d.; Scanlan, 2012).

5.2.2 Implications for Knowly

The difference between the frequency of evaluation observed on the Knowly platform and the responses from the survey suggests that more evaluation of courses on the Knowly platform are taking place outside the platform than on it. This means that when the Knowly platform is used, the majority of the effects that it has on the training might not be visible on the platform itself. The most common level to evaluate on the Knowly platform is level one, which might suggest that the results that courses have in respect to level one will also be how they measure their success with the platform. Since Knowly aims to deliver a platform that helps learning

professionals to enhance their learning and transfer results, and level two and three are less common to evaluate on the platform than level one, this could mean that there is a conflict of interest between the users and the developers of the Knowly platform.

It was argued in the introduction that Knowly and other tools like it could help their users evaluate more frequently. This was in contrast to what previous studies have shown, since these studies were based on traditional classroom courses and not digitally administrated courses. However, the results suggest that it is still less common to evaluate on level two and three than level one, and that the platform does not help in changing this. There are possible explanations to this, first of all the "evaluation" functionality on the Knowly platform is a new feature, and therefore not adapted by everyone. However, this does not explain that the relationship between level one, two and three in the data from Knowly compared to the survey shows that users are less likely to use Knowly for evaluating on level two and three than to use other methods outside of Knowly. This, again, shows that either Knowly does not provide a attractive way of evaluating on level two or three, or that users do not want to evaluate their courses on the platform for some other reason.

The reason why users do not evaluate their courses on levels two and three on the Knowly platform is likely a combination of the reasons listed above. Comments from the respondents of the survey suggest that verbal evaluation of level three is more common than sending out forms to participants, which means that it is not possible to use Knowly for evaluating level three in their preferred way. However, over 60% of respondents answer that they do evaluate level three, so a majority is interested in the effects on behavior. To be able to develop the Knowly platform, Knowly should also be interested in evaluating the usage of their platform in regards to level three. With that said, it is not known if Knowly or their users can evaluate the usage of the Knowly platform when they evaluate the effects on level three outside of the platform. To be able to evaluate how Knowly helps its users to increase the transfer of training as well as evaluating how Knowly helps to enhance the learning effect, they will need more data which is not present in Knowly today.

5.2.3 Implications for further research

The results of this study confirms prior studies of the same type despite the fact that this study was conducted on digital courses, when most prior studies did not. The question is why tools that can help with evaluation do not seem to affect the habits of learning professionals. There is some evidence that the learning professionals prefer to evaluate behavior via verbal communication, which could be a possible explanation to why digital platforms do not affect their habits. However, it is not clear which method is used the most, and why.

Evaluation theorists, when analyzing and criticizing the Kirkpatrick model, seem to focus on the implications of measuring different types of effect, and the main criticism is often that level one, customer satisfaction, is not a useful effect to measure

(for example, see Holton III, 1996). If learning professionals only measure customer satisfaction, which this study yet again confirms is the most common, they will not be able to make informed decisions on how to develop courses with respect to enhanced learning, transfer, and possibly not organizational outcomes either (Alliger & Janak, 1989; Blanchard et al., 2000; Holton III, 1996). However, it is clear that evaluating customer satisfaction is something that learning professionals are interested in, but it is not clear why. To be able to further suggest effective ways of evaluating in organizations, the different goals of learning professionals are of interest. One specific point of interest for further research could be how learning professionals work with customer success, and how it affects them.

Convincing evidence of whether or not there are factors that affect the habits of evaluation is yet to be found, since the results of the statistical analysis mostly provides insight into how different organizations use the Knowly platform. However, the results suggest that the effect of course subjects could provide an interesting framework for further research into evaluation habits. Course subjects should logically affect the goals of the courses, however, it is not clear how different course subjects correlate to the different levels of evaluation.

References

- Adkins, S. S. (2020a). *The 2019 global learning technology investment patterns: Another record shattering year*. Metaari.
- Adkins, S. S. (2020b). *The worldwide 2020-2025 advanced learning technology market: Edtech in a period of profound innovation*. Metaari.
- Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel psychology*, *42*(2), 331–342.
- Alvarez, K., Salas, E., & Garofano, C. M. (2004). An integrated model of training evaluation and effectiveness. *Human resource development Review*, *3*(4), 385–416.
- Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel psychology*, *41*(1), 63–105.
- Bjork, R. (n.d.). *Applying cognitive psychology to enhance educational practice*. Retrieved 2021-12-17, from <https://bjorklab.psych.ucla.edu/research/>
- Blanchard, P. N., Thacker, J. W., & Way, S. A. (2000). Training evaluation: perspectives and evidence from canada. *International Journal of Training and Development*, *4*(4), 295–304.
- Carnevale, A. P., & Schulz, E. R. (1990). Return on investment: Accounting for training. *Training and Development journal*, *44*(7).
- Esaiasson, P., Gilljam, M., Oscarsson, H., & Wängnerud, L. (2007). Metodpraktikan. *Konsten att studera samhälle, individ och marknad*, *3*(1), 12–19.
- Eseryel, D. (2002). Approaches to evaluation of training: Theory & practice. *Journal of Educational Technology & Society*, *5*(2), 93–98.
- Falconer, L. (2006, mar). Organizational learning, tacit information, and e-learning: a review. *The Learning Organization*, *13*(2), 140–151. doi: 10.1108/09696470610645476
- Gessler, M. (2009). The correlation of participant satisfaction, learning success and learning transfer: an empirical investigation of correlation assumptions in kirkpatrick's four-level model. *International Journal of Management in Education*, *3*(3-4), 346–358.
- Hinton, C., & Fischer, K. W. (2010). Learning from the developmental and biological perspective. *The nature of learning: Using research to inspire practice*, 113–34.
- Holton III, E. F. (1996). The flawed four-level evaluation model. *Human resource development quarterly*, *7*(1), 5–21.
- Jennings, C. (2013). *The 70: 20: 10 framework explained*.
- Kaufman, R., & Keller, J. M. (1994). Levels of evaluation: Beyond kirkpatrick. *Human Resource Development Quarterly*, *5*(4), 371–80.

- Kennedy, P. E., Chyung, S. Y., Winiecki, D. J., & Brinkerhoff, R. O. (2013, nov). Training professionals' usage and understanding of kirkpatrick's level 3 and level 4 evaluations. *International Journal of Training and Development*, 18(1), 1–21. doi: 10.1111/ijtd.12023
- Kirkpatrick, D. (1959). Techniques for evaluating training programs,” a four-part series beginning in the november 1955 issue of the training director’s journal. *Training Directors Journal*.
- Kirkpatrick, D., & Kirkpatrick, J. (2006). *Evaluating training programs: The four levels*. Berrett-Koehler Publishers.
- Kizlik, B. (2012). Measurement, assessment, and evaluation in education. Retrieved October, 10, 2015.
- Lund, S., Madgavkar, A., Manyika, J., Smit, S., Ellingrud, K., Meaney, M., & Robinson, O. (2021). The future of work after covid-19. *McKinsey global institute*, 18.
- McColl-Kennedy, J., & Schneider, U. (2000). Measuring customer satisfaction: why, what and how. *Total quality management*, 11(7), 883–896.
- McEvoy, G. M., & Buller, P. F. (1990). Five uneasy pieces in the training evaluation puzzle. *Training & Development Journal*, 44(8), 39–43.
- Organisation for Economic Co-operation and Development. (n.d.). *Enterprises by business size*. Retrieved 2022-03-21, from <https://data.oecd.org/entrepreneur/enterprises-by-business-size.htm>
- Phillips, J. J. (1996). How much is the training worth? *Training and Development*, 50(4), 20-24.
- Scanlan, C. L. (2012). *Assessment, evaluation, testing and grading*. University of Medicine and Dentistry New Jersey.
- Statistikmyndigheten. (n.d.). *Standard för svensk näringsgrensindelning (sni)*. Retrieved 2022-02-10, from <https://www.scb.se/dokumentation/klassifikationer-och-standarder/standard-for-svensk-naringsgrensindelning-sni/>
- Thalheimer, W. (2018). The learning-transfer evaluation model: Sending messages to enable learning effectiveness.
- Weinbauer-Heidel, I., & Ibeschitz-Manderbach, M. (2018). *What makes training really work: 12 levers of transfer effectiveness*. tredition.
- Whitten II, W. B., & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior*, 16(4), 465–478.
- Wild, R. H., Griggs, K. A., & Downing, T. (2002, oct). A framework for e-learning as a tool for knowledge management. *Industrial Management & Data Systems*, 102(7), 371–380. doi: 10.1108/02635570210439463

A

Appendix 1 - Survey

Undersökning utvärdering av utbildning

Välkommen!

Tack för att du tar din tid att besvara denna undersökning. Målet med denna digitala enkät är att komplettera data som samlats in från utbildningar i Knowly, med information om hur utbildningarna kan utvärderas utanför Knowly.

Studien bygger på Kirkpatrick's utvärderingsmodell (<https://kirkpatrickpartners.com/the-kirkpatrick-model/>) och frågorna ställs i förhållande till modellens fyra nivåer.

Nivåerna förklaras kort i anslutning till varje fråga. Även om du är bekant med Kirkpatrick's utvärderingsmodell kan det vara viktigt att läsa vår förklaring, då vår tolkning av de olika nivåerna kanske inte stämmer överens med din.

Kom ihåg att det inte finns något rätt eller fel, och att vi inte lägger någon värdering i svaren. Vi är endast ute efter att kartlägga hur verkligheten ser ut idag.

Vilken organisation representerar du? (OBS: Ej obligatorisk)

Kort svarstext

Nivå ett: "Reaction"

Förklaring:

Utvärdering baserad på utbildningsdeltagarens upplevelser. Frågor till deltagaren presenteras oftast i en enkät och fokuserar på olika aspekter som innehåll, upplägg och generellt intryck.

I samband med kurser och utbildningar där ni använder Knowly, utvärderar ni i förhållande till Kirkpatrick's nivå ett, "Reaction"?

- Ja

- No
- Jag vet inte

Här är ett fält för ytterligare kommentarer. Berätta gärna på vilket sätt ni utvärderar utifrån upplevelser, eller dela med dig av dina tankar.

Långt textsvar

Nivå två: "Learning"

Förklaring:

Utvärdering baserat på utbildningsdeltagarens kunskaper. Detta kan utvärderas på olika sätt, mest typiskt genom ett kunskapstest i form av ett prov. Ett annat typiskt tillvägagångssätt är att ha en skriftlig eller muntlig presentation.

I samband med kurser och utbildningar där ni använder Knowly, utvärderar ni i förhållande till Kirkpatrick's nivå två, "Learning"?

- Ja
- No
- Jag vet inte

Berätta gärna hur ni utvärderar utifrån kunskaper, eller dela med dig av dina tankar. Utvärderas kanske lärande främst i samband med certifiering eller då diplom ska utfärdas?

Långt textsvar

Nivå tre: "Behavior" / Transfer

Förklaring:

Utvärdering baserat på utbildningsdeltagarens handlingar. Transfer kan utvärderas delvis subjektivt genom att fråga utbildningsdeltagare, chefer eller kollegor om deras upplevelser av utbildningens effekt på deltagarens handlingar. Det kan också utvärderas objektivt genom att observera utbildningsdeltagaren efter utbildningen. En viktig del i att kunna utvärdera transfer är att stipulera mål som sedan kan uppfyllas av kursdeltagaren.

I samband med kurser och utbildningar där ni använder Knowly, utvärderar ni i förhållande till Kirkpatrick's nivå 3, "Behavior"?

- Ja
- No
- Jag vet inte

Berätta gärna hur ni utvärderar utbildningar utifrån handlingar, eller dela med dig av dina tankar.

Långt textsvar

Nivå fyra: "Results"

Förklaring:

Utvärdering baserat på utbildningsdeltagarens organisations framgångsfaktorer. För att utvärdera detta krävs det att isolera organisationsöverspännande faktorer som kan påverkas av att genomföra utbildningsinsatsen. Dessa ska sedan följas upp på innan och efter för att mäta vilken effekt utbildningen har.

I samband med kurser och utbildningar där ni använder Knowly, utvärderar ni utbildningen i förhållande till Kirkpatrick's nivå fyra, "Results"?

- Ja
- No
- Jag vet inte

Berätta gärna hur ni utvärderar utbildningar utifrån organisationens framgångsfaktorer, eller dela med dig av dina tankar.

Långt textsvar

DEPARTMENT OF COMMUNICATION AND LEARNING IN SCIENCE
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se



CHALMERS
UNIVERSITY OF TECHNOLOGY