Transportation Research Part F 73 (2020) 325–347

Contents lists available at ScienceDirect

Transportation Research Part F

journal homepage: www.elsevier.com/locate/trf

Safety culture, safety management and accident risk in trucking companies

Tor-Olav Nævestad *, Jenny Blom, Ross O. Phillips

Institute of Transport Economics, Gaustadalléen 21, NO-0349 Oslo, Norway

ARTICLE INFO

Article history:
Received 22 December 2019
Received in revised form 19 June 2020
Accepted 2 July 2020
Available online 27 July 2020

ABSTRACT

The Safety ladder for goods transport describes an approach with an increasing prevalence of safety structural measures at four Safety ladder levels in trucking companies. This paper validates the Safety ladder approach in empirical research by comparing safety structure, safety culture and accident risk for trucking companies. The study has four aims: 1) To map the safety structure at the different levels of the Safety ladder, 2) Examine whether safety culture is improved with increased structural measures at each Safety ladder level, 3) Examine whether the accident risk decreases at each Safety ladder level, and 4) Discuss practical implications. The study is based on survey data (N = 533) from 17 companies at different levels of the Safety ladder, a Reference sample, and qualitative interviews (N = 30) with management and employee representatives in the companies. Based on the interviews, we describe the safety structural measures at each level of the Safety ladder. Survey results indicate increasing safety culture scores at each level of the Safety ladder, while the accident risk decreases. The safety culture level was 12 points higher at Level 4, than in the Reference sample, and the accident risk at Level 4 was half the risk of Level 2. We conclude by suggesting the concrete management practices related to each level. Thus, the current study identifies and describes management practices in trucking companies that are associated with high safety culture scores and low accident risk.

1. Introduction

1.1. Background

Traffic accidents make up one of the most important causes of death and injuries worldwide, as 1.35 million people die each year on the world's roads, and between 20 and 50 million people sustain non-fatal injures (WHO, 2018). Accidents with heavy goods vehicles (HGVs) represent an important societal challenge. These are generally serious accidents with high proportions of killed and severely injured people, because of the weight of the heavy vehicles. In the EU, about 4000 people were killed in road accidents involving HGVs in 2016, making up about 16% of all road fatalities (ERSO, 2018). In the US, 4761 people were killed in accidents involving large trucks in 2017, and 72% of these were other road users (NHTSA, 2019). In Norway, 20% of the people who were severely injured or killed in traffic in the period 2007–2016 could be attributed to accidents involving HGVs (Nævestad, Phillips, & Hovi, 2018a). Most of the people who were severely injured or killed were other road users.

* Corresponding author.
E-mail address: ton@toi.no (T.-O. Nævestad).

https://doi.org/10.1016/j.trf.2020.07.001
1369-8478/

Robust studies of interventions focusing on safety culture in companies employing drivers at work indicate that such measures may reduce the incidence of traffic accidents with up to 60% (e.g. Gregersen, Brehmer, & Morén, 1996). The relationship between safety culture/climate and safety outcomes is well documented across industries and countries (Zohar, 2010). Studies also indicate a relationship between safety structure and safety outcomes (Naveh & Marcus, 2007; Thomas, 2012, Mooren et al., 2014a). Safety structure refers to the formal aspects of safety management ("how things should be done") as described in procedures, routines and organisational charts, etc. These aspects are often referred to as safety management systems (SMS). Safety culture refers to the informal aspects of safety management ("how things are actually done") (Antonsen, 2009). Safety culture scholars generally seem to agree that safety culture refers to shared ways of thinking and acting that are relevant for safety (Nævestad, 2010). Safety climate refers to quantitative measurements of safety culture, capturing superficial "snapshots" of the culture at a given moment (Flin, Mearns, O'Connor, & Bryden, 2000). In this study, we use the concepts interchangeably.

Despite the potential of preventing accidents involving HGVs, it seems that neither trucking companies, nor regulatory authorities focus sufficiently well on the significance of safety structure and safety culture for transport safety. Panel experts interviewed by Nævestad, Phillips, and Elvebakk (2015) asserted that trucking companies employing drivers at work focus little on safety culture/structure in general. Similar tendencies have been found in research from other countries (e.g. Mooren et al., 2014a). This illustrates the considerable and largely unexploited traffic safety potential of improving the safety of HGV drivers in specific and drivers at work in general (Nævestad et al., 2015). This low focus is especially evident when we compare with the situation in other transport sectors. In aviation, rail and the maritime sector, SMS facilitating positive safety culture are legally required (Lappalainen et al., 2012; Hudson, 2003; Amtrak 2015). In contrast to this, SMS applicable for companies in the road sectors are voluntary (e.g. EN: ISO:39001).

One potential reason for the relatively unexploited potential of safety culture and structure in the road sector, may be research gaps related to the practical utility of these concepts. First, in spite of the high focus on safety structure (SMS) as a way of implementing a positive safety culture in other transport sectors (e.g. aviation, rail), and thereby reducing the accident risk, there is little research explicitly studying the relationship between safety structure and safety culture. Thus, these key relationships are relatively poorly validated in robust studies. One reason for this is that it may be difficult to precisely define and discern between the different aspects of safety structure (SMS) and safety culture (cf. Nævestad et al., 2018b). Some key aspects (e.g. management commitment to safety), may be referred to as both structure and culture/climate (Thomas, 2012, Flin et al., 2000), depending on the perspective. Second, there is little to none research on the relationship between different aspects or elements of SMS and safety culture, although the explicit purpose of SMS implementation in e.g. rail (ERA, 2013), aviation (IATA, 2019) and the maritime sector (Lappalainen et al., 2012) is to facilitate certain aspects of safety culture. Third, there is little to no research describing the analytical/theoretical mechanisms through which SMS implementation may influence safety culture.

In the present paper, we seek to address these knowledge gaps in the heavy goods vehicle transport sector, by studying the outcomes for safety culture and accident risk of a specific approach to safety management which is labeled the Safety ladder for safety management in goods transport.

1.2. Aims

The study has four aims: 1) Map the safety structure at the different levels of the Safety ladder, 2) Examine whether safety culture is improved with increased structural measures at each Safety ladder level, 3) Examine whether the accident risk decreases at each Safety ladder level, and 4) Discuss practical implications.

The Safety ladder for goods transport describes an approach with an increasing prevalence of, and focus on structural safety measures in trucking companies. Thus, based on previous research, we hypothesize that increasing levels of safety measures at each level is associated with decreasing accident risk (cf. Section 2.2.1) and increasing levels of safety culture (cf. Section 2.2.2). This paper validates the Safety ladder approach in empirical research by comparing safety structure, safety culture and accident risk in trucking companies at different levels of the Safety ladder. The purpose of this comparison is to examine whether increasing levels of safety structure (at each Safety ladder level) involves higher safety culture scores, lower accident risk and higher perceived levels of safety management. Moreover, we also collect information about, and describe qualitatively the measures at each level.

1.3. The safety ladder for goods transport

Nævestad et al. (2017) suggest that the size of the companies may be an important factor explaining the relatively low focus on safety structure/culture in trucking companies. In Norway, over 85% of trucking companies employ five or fewer persons (Steen Jensen et al., 2014). Accordingly, similar research from EU-countries with available data indicates that over 80% of the trucking companies have less than 10 employees, while very few companies have more than 50 employees (usually about 1%) (European Commission, 2009). It is not unreasonable to assume that the small trucking companies have fewer resources (time, economy, competence on traffic safety management) compared with larger companies. Nævestad et al. (2017) therefore suggested an approach termed the Safety ladder for goods transport, which consists of a series of measures at different levels/steps. This is suggested on the basis of a systematic literature study of organizational safety measures, an analysis of studies of accidents with drivers at work, and industry characteristics (over 85% of companies have fewer than

five employees). Nævestad et al. (2017) concluded that four main measures aimed at organizational safety management have the greatest transport safety potential and are most realistic for regular trucking companies. These four measures can be arranged on a ladder, starting at the lowest level, before proceeding to the next step, see Fig. 1.

The idea behind the Safety ladder is that companies start at the bottom of the ladder if they have no measures aimed at work-related risk factors in the company. Based on previous research, we assume that the lowest levels are easiest to do something about and that they have the greatest effect. The first step in the ladder, "Managers' commitment to safety", is the most basic step in the Safety ladder, because research shows that this is usually a prerequisite for the company's safety work to be successful. The second step in the safety ladder is "Follow-up of driver speed, driving style and seat belt use". This is aimed at the main risk factors associated with drivers identified in the analysis of fatalities involving drivers in work. This involves e.g. policies for speed, driving style and seat belt use, follow up of driving style with fleet management system, feedback to drivers etc. The third step in the Safety ladder is "Focus on work-related factors influence on transport safety". Given little focus on organizational safety management in trucking companies, it is important that managers and employees in these companies develop an awareness of the importance of work-related factors in transport safety. This applies, for example, to planning, scheduling and organization of transport, with the consequences for drivers' experience of stress, time pressure, fatigue, etc., the extent to which drivers have direct contact with customers or recipients of goods, wage systems, negotiation and contact with transport buyers etc. The fourth step in the Safety ladder is to implement a "Safety Management System", such as ISO:39001, or other similar alternatives. This involves e.g. systematic risk analyses and development of procedures and training systems based on these analyses, systems for reporting safety issues, incidents etc. (Nævestad et al., 2017).

2. Theoretical approach and previous research

2.1. Key concepts

2.1.1. Organisational safety culture

Safety culture can generally be referred to as safety relevant aspects of culture in organisations (Hale, 2000). Discussing different specifications of safety culture, Nævestad (2010) concludes that most of these have in common that they refer to safety culture as shared ways of thinking and acting that are relevant for safety (Nævestad, 2010). Safety culture comprises shared frames of reference which provide ways of seeing (and thereby also ways of not seeing) hazards, which motivate and legitimize certain work practices, thereby proving norms for the preferred ways of doing things, influencing identities, regulating emotions etc. (Antonsen, 2009; Nævestad, 2010). While this definition denotes the deeper layers of culture, whose identification requires time-consuming qualitative methods, the concept of safety climate refers to the more shallow aspects, or manifestations of safety culture (Guldenmund, 2007). This is generally studied by means of safety climate surveys, measuring employees' perceptions of the focus on safety in the organization. Meta studies and reviews find that the most impor-

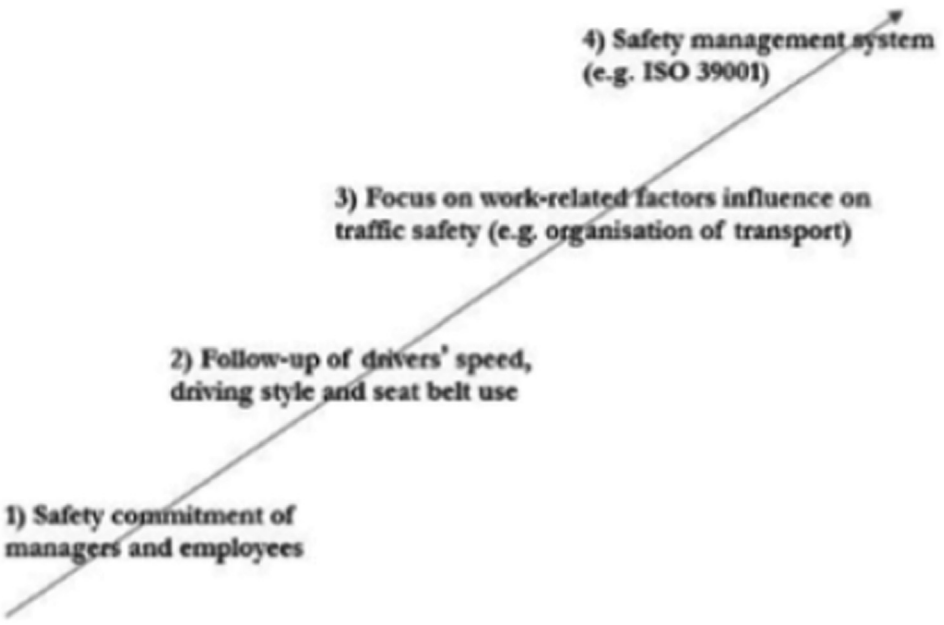

Fig. 1. The Safety ladder for safety management in goods transport.

tant aspect of safety climate is senior managements' commitment to safety, followed by safety system, risk perception, work pressure and competence (cf. Flin et al., 2000).

2.1.2. Organisational safety structure

We may define safety structure as safety-relevant aspects of organizational structure. Organizational structure refers to the way tasks in an organization are divided, how the work flows, how this flow is coordinated, and the forces and mechanisms that allow this coordination to occur (McShane & Travaglione, 2007). Structure in organizations has two basic tasks. The first is "division of labor" into various tasks, which leads to specialization. The second is "coordination of labor" in a way that enables workers to achieve the organization's goals. This tension between differentiation and integration represents a fundamental organizational challenge: the greater the degree of division of labor and specialization, the greater the effort required to coordinate and coordinate each effort. According to McShane and Travaglione (2007), coordination can be attained through: 1) Informal communication, 2) formal hierarchy, involving direct control and 3) standardization, with formal instructions (standardization / specification of tasks), goals (standardization of results) or training (standardization of knowledge).

We may also refer to organisational safety structure as SMS, which use these coordination mechanisms. First, "division of labour" is crucial in SMS, which describe key positions with specific safety responsibilities. Second, "coordination of labour" is also crucial in SMS, especially standardization. SMS include formal routines and measures enabling the organisation to work systematically with safety, e.g. by identifying risk through formal assessments, developing and implementing countermeasures (e.g. procedures, training), defining roles and responsibilities, regularly checking status, monitoring safety performance and adjusting measures, or implementing new measures if necessary (cf. Thomas, 2012). In this paper, we use the terms safety structure and SMS interchangeably, although the former generally refers to aspects of organisational structure with relevance to safety. This could be factors related to organization of transport and scheduling; factors that not necessarily are included in the SMS, but which nevertheless are relevant to safety (cf. Crum & Morrow, 2002; Mayhew & Quinlan, 2006). (These factors relate to the third level in the Safety ladder.)

2.2. Previous research

2.2.1. The relationship between safety structure and safety culture

The implementation of SMS may be viewed as an important strategy to develop safety culture (Nævestad, Hesjevoll, Ranestad, & Antonsen, 2019). This is indicated by studies from other transport sectors, in which SMS facilitating safety culture are legally required, e.g. aviation (IATA 2019), the maritime sector (Lappalainen et al., 2012) and rail (Amtrak 2015). Accordingly, studies from the road sector indicate that SMS implementation and/or structural safety measures may lead to positive safety culture. There are, however, not many studies of this in road, as SMS is not legally required. Naveh and Katz-Navon (2015), examine an intervention involving IS0-39001 certification, studying effects on safety climate and drivers' traffic safety violations. The study finds improved road safety climate in the intervention units, decreased safety climate in the control units, and 75% reduction in traffic safety violations in intervention units, compared with an increase in the control units. Wallington, Murray, Darby, Raeside, and Ison (2014) studies a comprehensive program (2001–2012) in British Telecom involving 95.000 employees driving in their work. The studied intervention describes a process structured by the occupational Haddon matrix, with risk analyses and mitigation ,covering factors like management culture and leadership, journey management, people, vehicle, and society/community. Murray, White, and Ison (2012) describe a comprehensive safety culture/structure program (2005–2009) implemented in an Australian company (Roche), with employees driving in their work. The program involved a mix of cultural and structural elements, but improving safety culture was central. Key initiatives were driver risk assessment, monitoring and improvement; policy development and communication; process and outcomes evaluation; continuous review and refinement of policies, processes, programs etc. Murray, Ison, Sing, and Nijjar, (2009) provide a case study of a comprehensive, 5-year occupational road safety program involving employees driving in their work in a large UK company (Wolseley). The program involves several structural and cultural elements (e.g. competition, handbooks, focus groups, health training, safety climate). Newnam and Oxley (2016) study a "Safety Management for the Occupational Driver" (SMOD) program in an Australian organization. The program focuses on the manager's role in work-related road safety, facilitating four supervisory skills through management training. Supervisors were trained in safety management of drivers, by teaching them to identify situations that pose increased risk to drivers, and effectively managing these situations. The intervention reports of improved safety climate and improvements in the four studied supervisory skills.

These five studies indicate first that it may be hard in practice to discern between pure safety structure and safety culture measures; interventions often include both, and structural measures may be implemented to improve safety culture. Thus, it seems that we currently know little about effective mechanisms. Second, only some of the studies actually measure whether safety culture is improved (e.g. Naveh and Katz-Navon, 2015, Newnam & Oxley, 2016); in the other studies it is just implicitly assumed. Third, few of the studies include control groups, like Naveh and Katz-Navon (2015), or compare the safety improvements with e.g. national trends. Thus, this is a topic which requires more research.

Contrary to the mentioned studies indicating that the introduction of SMS may facilitate the development of a positive safety culture, several accident investigations and studies (cf. Antonsen, 2009) indicate a discrepancy between formal and informal aspects of safety, and thus that the strategy of developing safety culture through SMS implementation may fail. There may be several possible reasons to such discrepancies, located at different analytical levels, e.g. the individual level,

organisational level, framework conditions. First, looking at the individual level, especially young males may intentionally or unintentionally violate procedures (violations), or they may fail to adhere because of inattention or memory failure (lapses) (Reason, 2000). Second, focusing on the organisational level, previous research indicates, that if the SMS is poorly implemented, there will be a gap between formal and informal aspects of safety; i.e. poor safety culture. This could be due to lacking continuity of personnel or implementation activities, conflicts, low trust between managers and employees etc. (Amtrak, 2015). Third, the procedures and routines prescribed by the SMS, could be violated if they are not perceived as meaningful by those to which they are directed (Haukelid, 2008), e.g. if the SMS is poorly adapted to employees' working lives and the company's reality (cf. Almklov, Rosness, & Størkersen, 2014). Fourth, sub-cultures may challenge procedures in specific, and the SMS in general and provide alternative interpretations of them. This could for instance be the case if work groups perceive that increasing proceduralisation threatens their professional competence and autonomy (cf. Haukelid, 2008; Antonsen, 2009). Fifth, the company may not have worked well enough to make the employees familiar with the SMS, so that the employees do not know procedures and routines sufficiently, e.g. due to inadequate training and lack of information (Dahl, 2013). Sixth, challenging framework conditions and production pressure mean that in practice the procedures often must be violated in order to maintain production (cf. Størkersen, Bye, & Røyrvik, 2011). Given relatively low SMS implementation in road, all the above-mentioned studies concerning discrepancies between safety structure and safety culture are from other sectors than road.

2.2.2. The relationship between safety structure, safety culture and accident risk

In a systematic review of the effectiveness of SMS in the transport sectors, Thomas (2012) concludes that there have only been a small number of quality empirical evaluations of SMSs, examining the impact on safety. Although several studies found that organisations with a certified SMS had significantly lower accident rates, there was a lack of agreement about which components of a SMS individually contributed the most to safety performance. The safety outcomes of SMS are studied in a few studies from road transport. In a study of heavy vehicles operators' insurance claims, Mooren et al. (2014b) found that low claimers seemed to focus more strongly on proactive risk assessment. Returning to the above, mentioned studies, we see that Murray et al. (2009) report of an almost 2-fold reduction of third-party collisions per vehicle. Murray et al. (2012) report of a 56% reduction in insurance claims costs and 55% reduction in collision costs. Wallington et al. (2014) found that the overall claim rate (per 1000 vehicles) per year decreased notably and gradually after program initiation. Naveh and Katz-Navon (2015) identified an average 75% reduction in traffic safety violations in intervention units. However, except for the latter, few to none of these studies include both experiment and control groups, and before-after measurements. Moreover, as several measures are implemented at the same time in these case studies, it is difficult to conclude about the outcomes and significance of specific structural measures.

In a literature review examining the relationship between safety culture and safety outcomes, Bjørnskau and Nævestad (2013) conclude that there is a relatively good empirical evidence for the relationship between safety culture and various measures of safety outcomes in correlational studies and studies with retrospective designs in transport, especially when safety is measured through self-reported safety behavior (Wills, Biggs, & Watson, 2005, Davey, Freeman, & Wishart, 2006). However, there is weaker empirical support for the relationship between safety culture and accidents/ injuries. This is to a greater extent due to a of lack high-quality studies than a lack of support in high quality studies. One important exception is provided by Gregersen et al. (1996). This is a study of a safety culture intervention, with pre- and post-measurements, test and control groups, which indicates up to 60% decrease in crash risk (property damage accidents). Christian, Bradley, Wallace, and Burke (2009) also indicate a relationship between safety climate and accident/injuries in a general meta study.

3. Methods

3.1. Survey

3.1.1. Recruitment and sample of companies

We conducted surveys and interviews in 17 trucking companies. The companies in the study were selected with the assistance of our contact person in the Norwegian Hauliers' Association (NLF: "Norges Lastebileier-Forbund"). Our survey consists of four different groups: 1) Reference group, 2) Companies at Level 2 in the Safety ladder, 3) Companies at Level 3 in the Safety ladder and 4) Companies at Level 4 in the Safety ladder.

The data collection period lasted for a period of about six months. To get the highest response rate possible, we used both individual gift card draws (NOK 3000) and collective company rewards for companies with high response rates (NOK 2000). To include drivers who either were unable or unwilling to answer the survey through the web-based scheme, we also offered drivers to answer the survey through paper-based schemes and per telephone interviews. A total of 34 respondents chose the former and 44 respondents chose the latter alternative.

In order to be able to make as controlled a comparison as possible, we also aimed for the different groups of companies to be as equal as possible in terms of: 1) The companies must have a certain share of long-haul transport, and not just distribution transport, 2) We also registered the share of drivers involved in dangerous goods, as this transport is known to have a substantially lower accident risk (Elvik, Høye, Vaa, & Sørensen, 2009), 3) We set a criterion that companies must not have too many foreign drivers, as this may influence our comparisons of safety culture. Previous research shows that foreign employ-

ees (Eastern and Central European) in Northern European companies give higher scores to their managers than Northern European employees (Guldenmund, Cleal, & Mearns, 2013). One likely reason for this is that Eastern and Central European employees have a cultural "deference to authorities", in addition to being in a special situation when working in companies outside their home country (Guldenmund et al., 2013).

3.1.2. Classification of the companies

We have placed the 17 trucking companies in the study at different levels in the Safety ladder, based on what they focus on in their safety work. We used indicators with five criteria for each level on the Safety ladder, based on Nævestad et al. (2017). We had an introductory meeting with two key people from NLF, where we provided descriptions of the characteristics/criteria of these levels (cf. Nævestad et al., 2017). The key people from NLF had good knowledge of the member companies, and based on our description, they selected a total of 77 member companies (in two rounds) and placed them at different levels at the Safety ladder. We said that we wanted to recruit 20 companies, but we chose to contact a larger sample, assuming that only some companies would participate. After we had finished the interviews, the classifications of the companies at different Safety ladder levels were quality assured in the way that two of the researchers classified the companies independently and discussed any disagreements. All re-classifications were discussed with the two mentioned NLF representatives.

In Table 1 we describe the criteria for the classifications of the companies at Level 2–4 on the Safety ladder.

3.1.3. The reference sample

Our network of NLF contacts, union representatives and authority representatives considered the 17 recruited companies from Level 2–4 to focus more on safety than average goods companies. We were unfortunately unable to recruit companies from Level 1 in the Safety ladder, and companies without organizational safety measures ("Level 0"). Although we asked about 20 companies in this category to participate, we were unable to recruit companies at this level. This indicates a correlation between companies' willingness to participate in studies about safety culture and the focus on safety culture in companies. We therefore needed to develop a strategy to also recruit respondents from more "average" trucking companies. We discussed this challenge with our network in the industry. Our contact persons in one of the large trade unions for Norwegian goods drivers suggested that the drivers from companies without a collective agreement would be closer to the average for Norwegian trucking companies than the already recruited respondents/companies. By recruiting these, we could balance our sample better, and have better opportunities to test our hypotheses. We refer to respondents from companies without collective agreements as the "Reference sample". Respondents in the Reference sample were drawn from the union's member list. When drawing respondents to the Reference sample, we had first to remove the respondents from the companies we already had in the original NLF sample. Second, we also had to take out companies that only transports dangerous goods, as these are not average. Third, we also had to consider the share of foreign drivers.

3.1.4. Survey themes

The survey contains seven different themes.

(1) **Background questions.** The survey includes ten background questions, e.g.: Gender, age, experience, seniority, 1000 km driven by heavy vehicle over the last two years, employment status, whether respondents own the vehicle themselves, type of transport, nationality.
(2) **Organisational safety culture.** The GAIN index for safety culture, consisting of 24 questions, are from the Operator's Safety Handbook (GAIN, 2001). The GAIN index is largely influenced by the key elements of Reason (1997) definition of good safety culture. It contains 24 questions, with answer scales ranging from 1 (totally disagree) to 5 (totally agree), on the following five themes:

(A) **Management's attitude to and focus on safety.** The theme is comprised of 8 questions (Cronbach's Alpha: 0.897), e.g. "Management is aware of the most important safety challenges in the company", "Management detects drivers who do not drive safely" "Management often discusses safety issues with the drivers", "Management stops dangerous work assignments and activities".
(B) **Employee attitude to and focus on safety.** The theme is comprised of 3 questions (Cronbach's Alpha: 0.620), e.g. "Drivers encourage each other to drive in a safe way".
(C) **Reporting culture and reactions to incident reporting.** The theme is comprised of 5 questions (Cronbach's Alpha: 0.833), e.g. "After an incident or accident has happened, precautions are taken to prevent it from happening again, "All safety problems and deficiencies that are reported are corrected in a short time"
(D) **Training / training in safety thinking.** The theme is comprised of 4 questions (Cronbach's Alpha: 0.844), e.g. "Drivers in my company receive adequate training to drive safely".
(E) **General safety issues in the organization in question.** The theme is comprised of 4 questions (4 Cronbach's Alpha: 0.677), e.g. "Vehicle safety checks are carried out regularly", "The safety of this company is generally well taken care of".

The GAIN index is constructed by summing the scores on the 24 questions, which show the respondents' in the companies mean scores on the GAIN questions. The minimum score is 24 (1×24) and the maximum score is 120 (5×24). The Cron-

Table 1
Criteria for the classifications of the companies' level at the Safety ladder for safety management in goods transport. Based on Nævestad et al. (2017).

Level 2	1	Policies for speed, driving style and seat belt, which are known to the drivers.
	2	Speed limiter under the factory setting of 89/90 km/h on some/all vehicles.
	3	Fleet management system and continuously monitoring of the drivers' speed and driving style.
	4	Drivers receive regular feedback (daily, weekly, monthly) on speed and driving style from the fleet management system.
	5	Follow up of drivers' seat belt use, and sanctioning of failure to use.
Level 3	1	Only the company's own transport managers have contact with the drivers on a daily basis before delivery, and not customers/receivers.
	2	The wage system is organized to minimize driver stress and fatigue.
	3	Drivers are encouraged to postpone, or abort assignments if they feel that conditions are unsafe, and they do so (e.g. in case of bad weather).
	4	When planning new assignments, consequences for drivers' fatigue and stress are mapped and assessed.
	5	The company systematically negotiates with transport buyers about deadlines, terms, etc. to ensure that safety considerations are properly taken into account.
Level 4	1	The company has a functioning reporting system, which is used by both employees and managers.
	2	The management team regularly reviews reported incidents to learn from them, the learning results in concrete measures, and employees are informed about them
	3	The company regularly conducts formal risk analyzes of all its assignments.
	4	The company has a good set of procedures that are known and experienced as meaningful by the employees.
	5	The company has a good training program, with predefined, theoretical and practical activities, a plan for knowledge goals and designated activities to fulfil the goals and assess goal fulfilment.

bach's Alpha value of the index in our sample is 0.954. We piloted and changed the wording of some of the 24 questions in the GAIN index in a previous study (Nævestad & Bjørnskau, 2014), to make it more relevant and meaningful for drivers in goods transport. We also build on the adaptations made by Bjørnskau and Longva (2009).

(3) **Perceived safety management.** We have included nine safety management questions, which we have used to create a perceived safety management index (Cronbach's Alpha: 0.867). The index has a minimum score of 9 points and a maximum score of 45 points. The index consists of three questions for each level on the Safety ladder (except Level 1). The answer scale ranges from 1 (totally disagree) to 5 (totally agree).[1] Table 6 provides a shortened version of all the questions. Example of Level 2 questions: "Management emphasizes that drivers should not drive faster than speed limits and conditions allow", Example of Level 3 questions: "In my company it is common for drivers to postpone assignments if they feel tired or unfit". Examples of Level 4 questions: "In my company, risk analyzes are performed of potentially dangerous work assignments and activities".

(4) **Work-related conditions and framework conditions.** The survey also includes questions about drivers' type of salaries, how many hours they work on a typical workday, perceived customer pressure and stress e.g. "In my job, I experience that customers are pressing/stressing drivers". Questions also included a "Customer focus on safety" index, comprised of two items (Cronbach's Alpha: 0.774): "Safety is more important than deadlines for our customers", and "Safety is more important than price to our customers".

(5) **Road safety behavior**. The survey contains four questions about drivers' behavior in traffic: "I sometimes disregard the speed limit on a motorway", "I sometimes disregard the speed limit on a residential road", "I sometimes accept some risk because "the situation requires it" (e.g. due to time pressure, bad weather) and "I sometimes break the traffic rules to get ahead faster". The questions were combined into a Road safety behaviour index (Cronbach's Alpha: 0.752).

(6) **Safety outcomes.** The survey also contains questions about accident involvement, type of accident, injuries during loading and unloading, falling asleep behind the wheel etc.

3.1.5. Quantitative analysis

Regression analysis. We have performed two regression analyzes. First, to assess the conditions explaining variation in the organizational safety culture level. We have used linear regression, since the dependent variable is continuous. Second, we conducted a regression analysis to assess the variables explaining whether respondents have been involved in accidents over the past two years while driving a heavy vehicle. We use logistic regression analysis, since the dependent variable is dichotomous (i.e. accident: 1) no, 2) yes). The regression analyzes show the effects of the independent variables that we include, controlled for the other variables in the analysis.

[1] Statements 3, 7, 8 and 9 originally had «do not know» as an answer alternative. This value was removed when we included the statements in the "Perceived safety management index".

3.2. Qualitative interviews

We conducted qualitative interviews with one management representative and one employee representative in all but two companies, which were unwilling, or unable to participate in interviews. The interviews were conducted over the phone and lasted anywhere from 40 to 90 min.

The main purpose of the interviews was to get detailed information about how the companies work with organizational safety management, and to validate their level in the Safety ladder. The interview guide is therefore structured according to the different levels in the Safety ladder. In the interviews, we placed great emphasis on getting specific examples of practices in the companies, and how often certain practices occur.

We started by asking about the company, the number of employees, what is being transported, etc. Then we asked about Level 2 safety management, e.g. whether the companies have policies for speed, driving style and seat belt, which are known to the drivers, whether they have speed limiter under the factory setting of 89/90 km/h on some/all vehicles. We also asked whether they use fleet management systems to continuously monitor the drivers' speed and driving style, and whether drivers receive regular feedback (daily, weekly, monthly) on speed and driving style from the fleet management system. We also asked whether and how drivers are sanctioned for unsafe behaviours behind the wheel.

We then asked questions about Level 3 in the Safety ladder. We started by asking what the company representatives considered to be the most important thing that their company does to facilitate safe driving. Then we asked whether only the company's own transport managers have contact with the drivers on a daily basis before delivery, and not customers/receivers. We also asked whether and how the wage system is organized to minimize driver stress and fatigue. Then we asked whether (and if so, how) drivers are encouraged to postpone, or abort assignments if they feel that conditions are unsafe (e.g. in case of bad weather). We also asked how often this happens, specific examples, and the circumstances leading up to them. Interviewees were also asked whether and how consequences for drivers' fatigue and stress are mapped and assessed when new assignments are planned, and whether the company systematically negotiates with transport buyers about deadlines, terms, etc. to ensure that safety considerations are properly taken into account.

We then asked questions about Level 4 in the Safety ladder; e.g. whether the company has a functioning reporting system, which is used by both employees and managers; how many reports the company gets per employee per year, and the types of reports. We also asked how the company learns from the reports, e.g. whether the management team regularly reviews reported incidents to learn from them, whether the learning results in concrete measures, and whether and how employees are informed about them. We also asked whether the company regularly conducts formal risk analyzes, how often it is done, how it is done, and which operations this applies to. The interviewees were then asked about the procedures that the drivers use in their work, the format of the procedures (e.g. in paper or digitally), what they describe, and whether they are known, and experienced as meaningful by the employees. We also asked what kind of training drivers get when they are hired and possibly later, theoretical and practical elements in the training, predefined knowledge goals etc.

We also asked about safety culture. Among other things, we asked the leaders of the company: a) What they put into the concept of safety culture, b) whether they use the concept in daily work life, c) whether they work to create a good safety culture in their company, and d) how they do it. Finally, we asked the interviewers if their company transports dangerous goods, certifications, e.g. whether they are included in NLF's "Quality and Environment on the Road" (QER), which may function as an instep to ISO:39001, or NLF's Health Safety and Environment (HSE), or ISO 9001, 14001, 39001.

4. Results

4.1. Sample characteristics

Table 2 shows the distributions of respondents in companies at the different levels. The response rate is based on the number of drivers the survey was sent to in the company. We see that there are relatively few respondents in the four companies at Level 2 in the Safety ladder, while most companies and respondents are at Level 4. There are four companies at Level 2 and also four at Level 3, while the sample includes nine companies at Level 4 in the safety ladder.

The overall response rate for Level 2–4 is 38%, despite the fact that the data collection period lasted over a period of about six months. When we also include the Reference sample, the overall response rate in the study is 28%.

A total of 4% (23) of the 533 respondents are women. Table 3 shows the age groups in the sample. We see that almost half of the respondents are over 46 years (47%), and that there is no clear pattern when looking at age differences between the groups. The differences are significant at the 10% level (P = 0.058).

Table 4 shows the respondents' experience, distributed on the different groups. There seems to be a higher proportion of drivers with the shortest experience (0–5 years) at Level 2, compared to the other groups, and a lower proportion with the longest experience (>16 years). It is difficult to see any clear pattern when we compare the groups. The differences are significant at the 5% level (P = 0.040).

Table 5 shows the respondents' distribution on the different types of transport.

As expected, we see that Level 4 has the largest proportion of drivers transporting dangerous goods. It is somewhat surprising that 11% of the drivers in the Reference sample transport dangerous goods, but this due to the fact that many companies have some drivers primarily transporting dangerous goods, although this is not the primary business of the company.

Table 2
The distributions of respondents and response rates in companies at the different levels in the Safety ladder. The response rate is based on the number of drivers the survey was sent to in the companies.

	Total	Response rate
Reference sample	80	12%
Level 2 (Company A-D)	39	32%
Level 3 (Company E-H)	126	40%
Level 4 (Company I-Q)	288	38%
Total	**533**	**28%**

Table 3
Respondents' age groups in the different levels of the Safety ladder.

Age	Reference	Level 2	Level 3	Level 4	Total	Total
<26 years	5%	13%	19%	8%	11%	56
26–35	21%	23%	17%	22%	21%	110
36–45	26%	31%	16%	22%	22%	117
46–55	29%	21%	31%	31%	30%	160
56 +	19%	13%	18%	16%	17%	90
Total	**80**	**39**	**126**	**288**	**533**	**533**

Table 4
Respondents' experience in the different groups.

Experience	Reference	Level 2	Level 3	Level 4	Total	Total
0–5 years	14%	38%	27%	16%	20%	107
6–10 years	23%	13%	11%	16%	16%	84
11–15 years	13%	13%	13%	15%	14%	73
16–20 years	21%	15%	16%	16%	17%	90
> 20 years	30%	21%	34%	36%	34%	179
Total	**80**	**39**	**126**	**288**	**533**	**533**

Table 5
Respondents' involvement in different types of transport at each level.

Age	Reference	Level 2	Level 3	Level 4	Total	Total
Long distance	25%	18%	36%	12%	20%	106
Distribution	35%	42%	35%	32%	34%	180
Long/distribution	29%	39%	25%	21%	25%	131
Dangerous goods	11%	0%	4%	35%	22%	116
Total	**80**	**39**	**126**	**288**	**533**	**533**

Table 6
Shares in each group agreeing (somewhat/totally) on the different statements.

	Reference	Level 2	Level 3	Level 4
1) Management focus on seat belt use	54%	61%	75%	84%
2) Management focus on speed	60%	77%	84%	91%
3) Policies for speed and driving style	52%	66%	86%	89%
4) Focus on drivers' private life for safety	23%	28%	51%	45%
5) Manager is genuinely concerned about wellbeing	44%	56%	62%	53%
6) Drivers postpone assignments if unfit	25%	29%	42%	33%
7) Well-functioning reporting system	72%	71%	82%	88%
8) Risk analyses are performed	50%	67%	56%	71%
9) We have job descriptions	50%	41%	54%	66%
Perceived safety management index	31.2	32.9	35.7	36.7
Total number of respondents	80	39	126	288

4.2. Safety structure at the different levels

The first aim of the study is to map the safety structure at the different levels of the Safety ladder. This is presented in Sections 4.2.1–4.2.4.

4.2.1. Follow up of speed, driving style and seat belt use

(1) *Policy for speed, driving style and seat belt use.* The companies at Level 2 have policies for speed, driving style and seat belts, that are known to the drivers. The drivers are minimally informed about these policies when they are hired, and often they must also sign that they are familiar with the policies, and that they are committed to following them when hired:

When hiring, we also inform the drivers of our policy and they must sign that they agree with the regulations. It is a working rule that they sign when they begin. Those who possibly started before we got it, have had to sign it afterwards. It clearly states what we accept and do not accept. (Management Representative, Company A).

The companies at levels 3 and 4 also have such policies, which the drivers are made aware of when hired, in addition to, for example, also training in how to take risk assessments and assessments about whether they are tired and fit to drive (e.g. in Company E).

(2) *Speed limiters.* The companies' use of, and views on, speed limiters limiting the speed beyond the statutory limit of 90 km/h varied. The Level 2 companies generally do not have a speed limiter lower than the required factory settings of 90 km/h. There were slightly different views on how appropriate speed limiter are among the different companies, including among the companies at Level 4. About half of the participating companies at Level 4 have speed limiters below 90 km/h, often at 80, partly due to requirements from oil companies that they transport goods for.

(3) *Fleet management system*. The majority of Level 2 Companies have fleet management systems which record several aspects of driving style (braking, speed, acceleration, etc), and they use them to continuously monitor the drivers' speed and driving style. They often have Volvo or Scania's systems, but there are also several other fleet management systems in use among the companies. Fleet management systems calculate scores for driving styles according to defined criteria, and drivers are informed of these at specific intervals (daily, weekly, monthly).

It is important to mention that the systems are used both to follow up driving with primary implications for safety (abrupt braking, G-forces in curves) and driving with primary implications for the environment (idle driving, diesel consumption). In addition, several types of driving styles often have implications for both (e.g. acceleration, speed). This means that having such systems can be motivated both by finances and by safety, and the importance of these factors as motivation varied in the companies.

(4) *Feedback to the drivers about fleet management scores.* Whether and what kind of feedback given to drivers from the fleet management system varies among the companies at Level 2. The drivers in the studied companies mainly received feedback in two ways. First, they can get information (their scores on different aspects and their general score) about their driving style daily, weekly or monthly. Second, they can receive feedback through conversations with management and their employee representative, e.g. if they have negative scores, or if violations are detected.

Several of the other companies, at levels 3 and 4, have systems that give drivers frequent feedback, e.g. after each trip. The management representative for Company G said, for example. that the system is connected to a mobile "application", that drivers can watch it all the time, and that they can also see their score compared to others in the company. For example, they can see what number they have, or how they are ranked relative to their colleagues. Some of the companies at Level 3 also linked financial bonuses to the drivers' scores. Drivers who have poor scores also often received special follow-up, e.g. through conversations and/or special training.

The companies that actively use fleet management systems tend to have more or less official ratings of all the drivers, and many of the interviewees said that employees often "compete" to get good scores and to have a high position on the rankings. In several cases, the drivers also received bonuses for good scores.

(5) *Seat belt use*. The companies generally have no systematic way of actively monitoring drivers' seat belt use, except that the vehicles are generally equipped with seat belts that give an alarm if not in use, and that it is sanctioned if it is discovered accidentally. At least one company has, however, actively controlled this themselves, by making observations during one day at a loading point.

4.2.2. Focus on work-related factors with implications for traffic safety

(1) *Contact with customers.* Several of the interviewees said that only their own transport managers have direct contact with the customers, and that the drivers are never in direct contact with the customers. This criteria was included at level 3 in the Safety ladder, based on an assumption that drivers' direct customer contact may induce stress (Nævestad et al.,

2017). Interviews results indicate however that this is a relatively complex question. The type of contact with the customers often varies according to the type of transport, and there are many different ways of handling this internally in the companies, depending of what is being transported and depending on the customer.

First, the customer or the client is not necessarily the same as the recipient. Second, several of the companies (e.g. at Level 3, groupage shipments) emphasized the importance of giving drivers increased responsibility for the planning and organization of the transport assignments, including contact with customers, because the drivers have the best knowledge of the optimal organization of the transport route, as well as involving them increases their well-being. Third, the importance of customer contact for stress also depends on the customers' attitude and the type of customers. Fourth, the degree of customer contact also depends on the type of transport. In some industries, the driver operates relatively independently throughout the week, and has close contact with the customers.

(2) *Wage system.* We also asked the interviewees if they have set up the wage system to minimize the drivers' stress and fatigue. This means that they do not have any elements of commission driving, e.g. bonus for miles driven and number of completed assignments. None of the companies said that they have performance-based pay, even though some of them had included this in the payroll system in the past. All of the companies in our sample at levels 2, 3 and 4 have hourly pay and overtime pay for work over eight hours and trip price for long haul with payment for both transport and rest away from home. In addition, at least two of the companies (E and H) have salary bonuses for good scores in the fleet management systems.

(3) *Postpone assignments.* Several of the interviewees mentioned examples that customers and receivers of goods may "push" to get goods delivered. This especially applies to drivers who transport groupage shipments.These drivers emphasized that this transport often may involve time pressure.

We asked the interviewees whether drivers postpone assignments, if or when they feel it is not safe to complete them. However, the answers to this question can be interpreted in several ways, and it does involve some uncertainty. Some said that if conditions were unsafe, they would not have agreed to them in the first place. Others said it is "the driver who makes the decision there and the", about possibly postponing or canceling an assignment, and that "no one can contest or undo his decision". We could look at how often it happens, but this could also indicate bad planning. The crucial issue seems to be related to the culture, acceptance and possibilities for postponing or aborting assignments. The employee representative from Company P (Level 4) said for instance they have included a time margin in the schedule which means that drivers should be able to stop and sleep for an hour or two during the night shift:

(. . .) Fatigue is the greatest danger of the night shift. Our policy is that if you are tired you should stop and sleep. There is room for the schedule as well. (. . ..) You should be able to sleep for an hour or two during a shift. (Employee Representative, Company P).

(4) *Planning of assignments.* We asked the interviewees several questions about whether and how they take into account the drivers' potential stress and fatigue when planning assignments. When asked what is the most important thing they do to facilitate drivers to drive safely, Company H's employee representative replied:

Minimizing stress. It is important that we have arranged it so that it is natural to manage it within a good timeframe, and without violating the driver's hour rules and without violating the speed limit to fulfil the deadline. That's the most important thing. They [the recipients] don't call us to when we're done. We may also get sick, and then we must call and say that today I cannot drive. And it's always okay. Then the load lies until we can drive again. The company calls the customers and explains that the goods are a little delayed due to illness. It is an honest case. I have experienced that myself and it works. (Employee Representative, Company H).

Several of the companies emphasized that they involve the employees in the mapping of potential challenges that a new assignment may cause in terms of fatigue and stress. When asked how his company seeks to minimize drivers' stress, fatigue, time pressure, etc. when planning transportation assignments, the Company F (Level 3) management representative emphasized engaging the drivers and taking into account their views, their private needs and tasks.

(5) *Negotiate with transport buyers.* We also asked the interviewees whether they "negotiate" with the transport buyers about deadlines, terms of transport, facilitation, etc. for reasons of safety. The purpose of this question was to get an assessment of whether they feel they can do it, and whether they have a culture for doing so, if necessary. All the interviewees emphasized that they have clear limits on what kind of assignments they take and do not take, and that they do not accept terms that they believe conflict with safety. It seems however, that type of client and the duration of contracts and assignments were of great importance to the companies' experience of whether they can negotiate with clients about deadlines, terms, etc. and whether they can influence the terms of the market in which they operate. Several interviewees mentioned that there are certain expectations that the goods should arrive at the agreed time. It is perhaps the drivers transporting groupage shipments who have the most direct experiences with this:

It's a hectic profession. There is time pressure and there is stress, especially on long routes, so it is not possible to conceal that. (. . .) It's price sensitive and there is focus on price and delivery time. Special delivery time. If it should be delivered

tomorrow at 10, I will be there before ten. We need time-efficient plans that add up. A filled car provides a better and more efficient everyday life. (Employee Representative, Company A).

Others mentioned that it was more difficult to negotiate with the large customers, because they get "too big", and sometimes "do not have the same ownership of safety", or that it is a challenge to be "service minded". The issue of negotiation is complex, as it is influenced by the type of transport buyers, type of transport, and as it may be unclear what is negotiated and when it could be negotiated.

4.2.3. Safety management systems

(1) *Well functioning reporting system.* Companies at levels 2 and 3 generally do not have functioning systems for reporting safety-related issues. Although several of them said that the technical infrastructure (e.g. e-mail, web page) and reporting routines exist in the company, since they are all part of the QER system, they said that they receive very few reports a year (e.g. about 10 in total). In these companies, several of the reported conditions are non-conformities related to delivery conditions (quality) and not necessarily safety.

The companies at Level 4 are characterized by their functioning reporting systems, involving e.g. a few hundred reports a year. The management representative in Company Q said that he has done a lot to develop a "living" reporting system:

There were few [reports] in 2016 when I started. I made it alive and started commenting with text messages immediately when someone wrote a report. We have a KPI of three reports per employee per year: 500, and now we can fulfil it. This is subject to continuous improvement based on suggestions from those who drive. It requires simplification. . ..and that's how they constantly work on suggestions. . ..It is important that they see something happening. (Management Representative Company Q).

In Company P, a key employee sits at one of the loading points two to three days a week. All the drivers come to see him when they are there and talk about driving style, handling of non-conformities, loading and unloading routines etc. They talk to him for about 10–15 min while they are there, and he may talk to at least all the drivers during a three-week period.

(2) *Management review of reported incidents.* Companies at levels 2 and 3 have few reported incidents per year, and therefore have no great need for regular reviews of reported incidents, or methods to learn from them. However, some mentioned that this will change soon.

Interviewees from the companies at Level 4 have a greater number of reported incidents than the others, and they also have routines for analyzing and learning from the reports, and informing the employees about the result (i.e. analyzes and measures). This feedback seems to be a prerequisite for getting a relatively large number of reports. Several of the interviewees mentioned that information to employees about the results of reports is important for employees to see that the reporting system works; that the reports are taken seriously by management; that they go through them and take steps to improve the organization.

(3) *Risk analyses.* Interviewees from the companies at levels 2 and 3 said that they do not carry out formalized risk analyzes of normal assignments, because they are not required by transport buyers or authorities.

Level 4 companies carry out risk analysis more often, and they do so even if it is not required by a third party. For example, the management representative from Company I said. that:

We review [a transport route or assignment] and consider; how long does it take: time to load and unload, driving, time of year, winter? etc. Then you get feedback. We do this together with the employee representatives, everything from daily to weekly . . . in new assignments or in pressing situations . . . then risk analysis must be conducted to see if we may end up in something [e.g. dangerous situations]. (Management Representative Company I).

The companies at Level 4 usually carry out risk analyzes of driving routes, loading, delivery and other work activities, at least annually or semi-annually. The Employee Representative in Company M emphasized that this thinking "permeates all day-to-day operations. . ...with all kinds of incidents that may occur." The management representative from Company J mentioned that they carry out risk analyzes together with the drivers, department by department, as the routes are so different, and that they adapt procedures and training based on risk analyzes.

(4) *Procedures.* Interviewees from the companies at Level 2 responded that they see some challenges in getting the drivers to know and follow the procedures. The company representative from Company D said that the drivers "are not very fond of paperwork" ("they cannot do paperwork") and that they must therefore assist them with different types of documentation:

We see a "disconnect" from the drivers and the system. It must be implemented to a greater extent in their everyday lives so that it is the driver and not always us who follow up and do the job for them. We have a new HSE and quality system for the drivers, but they do not have access to this yet. Some have a good understanding of the systems and participate in debates, while others are not interested. Old habits die hard. They are used to being a driver, and digitalization and new demands are not easy. (Management Representative, Company D).

Companies at levels 3 and 4 generally reported somewhat fewer challenges related to this type of discrepancy between the procedures and the drivers, although some of them also mentioned this. Companies at levels 3 and 4 often have comprehensive and well-structured driver manuals.

The management representative from Company Q (Level 4) also stated that in his company there is some discrepancy between the drivers and the system. Several of the drivers perceive the requirements for documentation as hassle:

They drive a truck because they like it; they will not sit and write. In a busy day, I see that it can be perceived as hassle. We also tried to read the procedures into audio files, due to drivers with reading and writing difficulties. This received positive feedback. If that is not good enough, I have spent 4–5 h explaining it to people who do not understand. (Management Representative Company Q).

(5) *Training.* Level 2 companies generally do not have comprehensive training programs structured in a set of predefined sequences. If there are people who have worked for several years as a driver who should have training, for example, they will receive practical training in one day or two, and they do not undergo theoretical training. In companies at Level 3, the training often lasts a little longer, and it also contains more theory and information about the company.

What sets the training at Level 4 apart from the lower levels in particular is the systematic approach: that the companies have a set of predefined knowledge and skills requirements. These are e.g.: a) a plan for what their drivers should be able to do, and b) a plan for how to make sure that the drivers have (or gain) this knowledge. Another difference is the focus on the theoretical knowledge and the attitudes. At Level 2, the training is less systematic when it comes to points a) and b). This is partly due to the fact that these companies have fewer formalized requirements for what their drivers should be able to do, and that the training of experienced drivers may be mainly about learning the driving routes, in addition to the management expressing what kind of behavior and attitudes are expected. In addition, the training is more individualized and coincidental in the sense that its content and quality depends on who you get it from, e.g. which driver you learn the route from.

The training programs in the companies at Level 4 also last longer than the training programs of the companies at Level 3. This especially applies to those only transporting dangerous goods. Company Q, for instance, has a training course which lasts for half a year with three meetings during the trial period. In the most comprehensive training programs, it is also possible for the drivers to "fail", which means that they will not be hired, because the companies believe that they do not have the personal qualities required to work in the company. This applies for example to Company P.

4.2.4. Perceived safety management at the different levels

The survey includes nine statements that measure perceived safety management: three questions for each level on the Safety ladder (Level 2–4). Statements 1–3 measure Level 2, 4–6 Level 3 and 7–9 measure Level 4. Answer alternative ranged from 1 (totally disagree) to 5 (totally agree). Table 6 shows the shares in each group agreeing (somewhat/totally) on the different statements.The table shows a clear increase in the proportions agreeing with the statements for each level in the Safety ladder. The exception is the statements regarding Level 3, in which the companies at Level 3 have higher proportions agreeing than Level 4.

We also made a perceived safety management index based on the sum-scores of the nine statements. The Cronbach's Alpha value for the index is 0.867, indicating that the statements to a high degree measure the same construct. The minimum score is 9 points and the maximum is 45 points. Table 6 indicate that the mean scores for the index increases for each Safety ladder level, although the difference between Level 3 and 4 is small. A Tukey post-hoc test, in which we examine whether the differences between the different means are significant, shows that the difference between the mean score of the Reference Group and Level 2 is not significant, probably due to the low number of respondents at Level 2. The differences between the mean of the Reference Group and Level 3 and Level 4 are significant at the 5% level. The difference between the means of levels 3 and 4 is small and not significant. The difference between the means of levels 2 and 4 is significant at the 10% level ($p = 0.054$).

4.3. Safety culture at the different levels

The second aim of the study is to examine whether safety culture is improved with increased structural measures at each Safety ladder level. The majority of the interviewed management representatives said that they do not use the safety culture concept on a daily basis. When asked what they put in the concept of safety culture, many of the interviewees mentioned e.g. common attitudes and common ways of behaving. They emphasized that having a good safety culture means that focusing on safety should "lie in the backbone" of the drivers. Several of the interviewees said that the work to create a good safety culture in the company "starts from the time a driver is hired", indicating the importance of training as a tool for developing safety culture. They said that creating a good safety culture in the company is about communicating the norms that apply in the company when it comes to safety and the behavior they expect from the drivers. Several of the management representatives, especially at levels 3 and 4 emphasized that they build safety culture by involving employees, e.g. in continuous improvement processes focusing on analysis of safety data, development of measures and follow-up of their effects.

We measure safety culture using the GAIN index, which consists of 24 questions that make up five themes (see Section 3.1.4 for a presentation of the themes, examples of questions and the Cronbach's Alpha values for the five theme indexes). Table 7 presents the mean scores on the five theme indexes and on the organizational safety culture index.

Fig. 2 shows the mean scores of the GAIN-index for each Safety ladder level, and the mean scores for the perceived safety management index.

Fig. 2 shows increasing safety culture scores for each level in the Safety ladder, and that the difference between the Reference group at the lowest level and Level 4 is 12 points. The perceived safety management index shows the same trend, with a five-point difference between the lowest and the highest level. A Tukey post-hoc test, where we examine whether the differences between the different mean safety culture scores are significant, shows that the difference between the mean scores of the Reference Group and Level 2 is not significant (P = .327), probably due to the low number of respondents at Level 2. The difference between the mean scores of the Reference Group and Level 3 and Level 4 are significant at the 1% level (each with P = .001). The difference between the mean scores of levels 3 and 4 is small and not significant (P = .881). The difference between the means of levels 2 and 4 is significant at the 10% level (P = .067). Although not all differences are significant, we see a clear pattern of increasing safety culture scores for each level.

4.3.1. Multivariate analysis: Which factors influence safety culture scores?

The safety culture scores at the different levels may be influenced by several factors. To control for this, we conduct linear regression analyses. Table 8 shows the results of eight regression models with the respondents' GAIN safety culture index as dependent variable.

The analyzes primarily show that perceived safety management is the variable that contributes most to explaining variation in safety culture among the respondents. We see that the adjusted R^2 value, which indicates the proportion of variation in the dependent variable explained by the independent variables is close to zero in the first two models, but that it rises to 0.749 in Model 3 when we include perceived safety management. This means that Model 3, i.e. perceived safety management, explains 75% of the variation in the dependent variable (safety culture).

Second, we see that the type of transport (ADR) and customer focus on safety contribute significantly in the analyzes. The latter variable consists of two questions that are combined: "Safety is more important than deadlines for our customers" and "Safety is more important than price for our customers" Pearson correlation analyses show that both are moderately correlated with safety culture (Pearson's R 0.585 ** and 0.576 **, respectively). ADR contributes negatively, and this is difficult to explain. This is by and large due to the fact that the safety culture score in Company L was somewhat lower than expected. This company only transports dangerous goods (ADR), and has a total of 80 respondents in the sample, and 68% of the ADR drivers in our study.[2]

The contribution of perceived safety management is reduced somewhat in Model 5, when we include customer's focus on safety, indicating a relationship between perceived safety management and customer focus on safety (a bivariate analysis shows moderate correlation: Pearson's R: 0.659 **). This could indicate that transport buyers may influence the level of safety management in trucking companies. (Negative customer influence does not contribute significantly, cf. Model 6).

In model 7, we include the variable "Level 4 in the Safety ladder" to examine whether, and the extent to which other aspects of the companies at Level 4 than those we have measured by means of the perceived safety management index, contribute to explaining the variation in safety culture. Results indicate only a small contribution which is significant at the 10% level, and which ceases to contribute significantly, when we include Company I in Model 8. The weak contribution of Level 4 indicates that our variable Perceived safety management measures the most important safety relevant aspects or differences between the companies located at the different levels in the Safety ladder.

In model 8, we also include Company I, which had the highest safety culture score in the study. This variable measures whether respondents work in Company I (Value = 2), or not (Value = 1). The purpose of including this variable was to assess whether there are other aspects with Company I than what we measure by means of the "Perceived safety management index" (that we control for in the analyses) that might explain the high safety culture score of Company I. The variable "Company I" contributes significantly at the 1% level, but it has the smallest contribution in the model. Thus, we may perhaps conclude that our variable "Perceived safety management" explains most of the safety culture score in this company, but that there still are some important factors related to Company I contributing to its high safety culture score, that we have been unable to measure.

Finally, neither nationality nor age contributes significantly in the analyses, although respondents from Eastern Europe (95.6) had a higher safety culture score than the Norwegian average (93.5), in line with the mentioned deference for authority mechanism (Guldenmund et al., 2013).

[2] As this result could indicate multicollinearity, which occurs when independent variables in a regression analysis are (strongly) correlated, we also checked for multicollinearity. This might be a problem in the analyses, as coefficients may become very sensitive to small changes in the model, reducing precision of the estimate coefficients, which may weaken the statistical power of the regression (Frost, 2020). Multicollinearity can be checked by examining the variance inflation factors (VIF), which identifies the correlation between independent variables, and the strength of that correlation. A VIF value of 1 indicates no correlation, VIF values between 1 and 5 suggest moderate correlation, but it is not severe enough to warrant corrective measures, while VIFs>5 represent critical levels of multicollinearity where the coefficients are poorly estimated, and the p-values are questionable (Frost, 2020). The highest VIF value in Table 8, Model 8, were Perceived safety structure (1.844), Customer focus on safety (1.781), Safety ladder level 4 (1.315) and Type of transport (ADR=2) (1.313). These, levels are far lower than the critical levels of multicollinearity, and thus it does not seem that the independent variables in Table 8 are strongly affected by multicollinearity.

Table 7
Mean scores on the five organizational safety culture themes and the GAIN index for organizational safety culture.

	Reference	Level 2	Level 3	Level 4
1) Management's attitude and focus on safety	27	29	32	31
2) Employees' attitude and focus on safety	12	12	12	12
3) Reporting culture and reactions to incident reporting	18	19	21	21
4) Training in safety thinking	13	15	15	15
5) General safety issues in the organization	14	15	16	17
Scores on the GAIN index for organisational safety culture	84	89	95	96
Total number of respondents	80	39	126	288

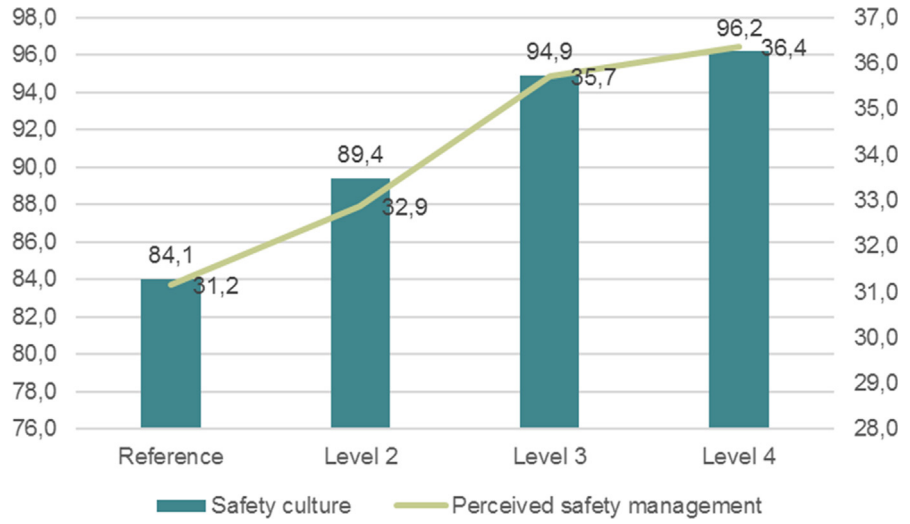

Fig. 2. Mean scores of the GAIN index for safety culture in the four groups, and average scores on the index for safety management The GAIN index is comprised of 24 questions with five answer alternatives (Min: 24, Max: 120). The index for safety management is comprised of three questions for each level (2, 3 and 4) of the Safety ladder (Min: 9, Max: 45). Reference (N = 80), Level 2 (N = 39), Level 3 (N = 126) and Level 4 (N = 288).

Table 8
Linear regression. Dependent variable: GAIN safety culture index. Standardized beta coefficients.

Variable	Mod. 1	Mod. 2	Mod. 3	Mod. 4	Mod.5	Mod. 6	Mod. 7	Mod. 8
Age (46–55 = 2)	0.048	0.047	−0.009	−0.004	−0.005	−0.008	−0.006	0.003
Nationality (East−EU = 2)		−0.014	0.021	0.014	0.010	0.009	0.010	0.013
Perceived safety management			0.893***	0.915***	0.809***	0.802***	0.796***	0.791***
Type of transport (ADR = 2)				−0.139***	−0.131***	−0.140***	−0.155***	−0.136***
Customer focus on safety					0.162***	0.156***	0.157***	0.162***
Customer pressure/stress						−0.038	−0.033	−0.033
Safety ladder Level 4 (=2)							0.045*	0.023
Company I (=2)								0.073***
Adjusted R^2	−0.001	−0.003	0.749	0.813	0.827	0.828	0.829	0.834

* p < 0.1 ** p < 0.05 *** p < 0.01.

4.4. Accident risk at the different levels

The third aim of the study is to examine whether the accident risk decreases at each Safety ladder level. The survey contains questions about accident involvement with heavy vehicles over the past two years, and type of accident (property damage, personal injury, fatal accident).[3] A share of 25% in the Reference group had been involved in a property damage accident. At Level 2, 28% had been involved in a property damage accident and 3% in a personal injury accident. The corresponding shares at Level 3, were 21% and 2%, while it was 14% and 2% at level 4. Thus, Level 4 had the lowest proportion involved in accidents in recent years. When comparing the groups' accident involvement, it is very important to also consider the kilometers driven in

[3] We also asked respondents if they have been injured at work during the last two years while loading or unloading (i.e. not in traffic). The level answering no was 66% in the reference group, 58% at level 2, 70% at level 3 and 76% at level 4. Again, we see that level 4 had the lowest levels of injured respondents at, although differences were not statistically significant.

the same period. We do this in Table 9. Additionally, the different accident types have been combined into the value "accident involvement", combining (minimum) property damage, (minimum) personal injury and fatal accident.

Table 9 indicates that the level of accident risk at Level 4 is almost half of the level of accident risk at Level 2. However, the differences between accident risks (e.g. between levels 2 and 4) are not statistically significant, because of very small numbers of accidents in the different groups, e.g. only 12 accidents at Level 2. We still emphasize the differences in risk between the groups. First, as the numbers are small. Second, because we see a clear trend, or a pattern with an expected reduction in risk from Level 2 to Level 4. Third, because Level 4 accident risk is about half of Level 2 accident risk. Surprisingly, Table 9 indicates that Level 2 has a higher risk than the reference group. This may be partly because there is less long-distance transport at Level 2 (and a high number of kilometers traveled in the Reference Group), but this is uncertain. The reference group also has the highest mean for kilometers driven per driver.

Fig. 3 illustrates the relationship between safety culture and (traffic) accident risk in the studied groups.

Although we see that the accident risk is higher at Level 2 than it is in the reference group, Fig. 3 indicates a relationship between safety structure, safety culture and accident risk. In Table 10, we examine these relationships further in a multivariate analysis with accident involvement as the dependent variable.

4.4.1. Multivariate analysis: Which factors influence drivers' accident involvement?

We conduct a logistic regression analysis of the variables predicting variation in accident involvement over the past two years. We use logistic regression, because we have recoded the accident variable into a dichotomous dependent variable (accidents: no/yes). Table 10 shows the results of nine regression models with respondents' accident involvement over the past two years as a dependent variable. Odds ratios are presented, and they indicate whether the odds of an accident is reduced (negative odds ratios) or increased (positive odds ratios), when the independent variables increase with one value. Thus, the significant odds ratios in Model 9 represent the effects of a specific variable, adjusted for eight other factors.

As expected, Table 10 shows that the respondents' kilometers traveled over the past two years affect their accident involvement. The contribution of this variable is small, because of the coding of this variable. One unit on the variable is 1000 kms driven. The coefficients presented are odds ratios (exp(b)), and they indicate that when the Kms driven variable increases with one value (1000 kms), the odds of an accident increase with a factor of 1.005. The confidence interval of this coefficient ranges from 1.001 to 1.008.

Second, we see that age contributes significantly to accident involvement. The coefficient is lower than one, indicating that drivers over 26 years old have lower accident risk than drivers younger than 26 years old. When the age variable increases with one value (to over 26 years) the odds of an accident decrease. The confidence interval of this coefficient is wide, ranging from 0.132 to 0.969.

Third, as expected, we see that type of transport also contributes significantly to respondents' accident involvement: non-ADR drivers are (2.4 times) more likely to be involved in accidents (up to Model 7). This variable ceases, not unexpectedly, to contribute significantly when the variable "Safety ladder Level 4" is included in the analysis in Model 8. The reason is that the ADR companies are at Level 4, although there are some individual ADR drivers in companies at the lower levels

Fourth, we see that higher safety culture scores contributes to lower odds of accident involvement, controlled for the other variables in the model (e.g. type of transport, safety management and kilometers driven). The reduction in the odds of an accident is small, as the coefficient shows reduced odds per unit on the safety culture index, which ranges from 24 to 120 points. Safety culture ceases, however, to contribute significantly when safety behaviour is included, in Model 9.

Fifth, we see that drivers' safety behavior contributes significantly at the 5% level to explain accident involvement in Model 9. Safety behaviour is an index comprised of four behaviour variables: speeding in residential areas, speeding on motorways, whether drivers sometimes violate traffic rules to get quicker to their destination and whether they accept "a little" risk if the situation demands it. It seems that these behaviours are related to accident involvement: A one-point increase on the safety behaviour index (min:4, max:20) increases the odds of an accident with a factor of 1.1. The confidence interval of this coefficient ranges from 1.005 to 1.210. The result that safety culture ceases to contribute significantly when safety behavior is included in Model 9 indicates a relationship between safety culture and safety behaviour. It seems that the influence of safety culture on accident involvement is mediated through safety behaviours; safety culture influences safety behaviours which influences accident involvement.

The Nagelkerke R value in model 9 is 0.139, which indicates that the model explains about 14% of the accident involvement of respondents in the sample.

5. Discussion

5.1. Safety structure at the different levels of the safety ladder

The first aim of the study was to map safety structure at the different levels of the Safety ladder. The results of the interviews show that what characterizes the companies at Level 2 in the Safety ladder are that: 1) They have policies for speed, driving style (and seat belts), 2) Fleet management system recording drivers' speed, driving style and diesel consumption, and 3) On the basis of this, drivers receive regular (daily, weekly, monthly) feedback on their driving style. Reviewing the

Table 9
Accidents, kilometres driven and accident risk per million kilometres driven with a heavy vehicle over the last two years at the various Safety ladder levels.

	Reference	Level 2	Level 3	Level 4	Total
Share of respondents involved in accidents	25	31	23	16	20
Total number of respondents	80	39	126	288	533
Risk (accidents per million km)	2,06	2,80	2,09	1,58	1,86
Avg.: 1000 km per driver in the last 2 years	122	110	110	103	109
Avg.: Accidents per driver	0,25	0,31	0,23	0,16	0,20
Kms driven (millions)	9,725	4,287	13,890	29,775	58,077

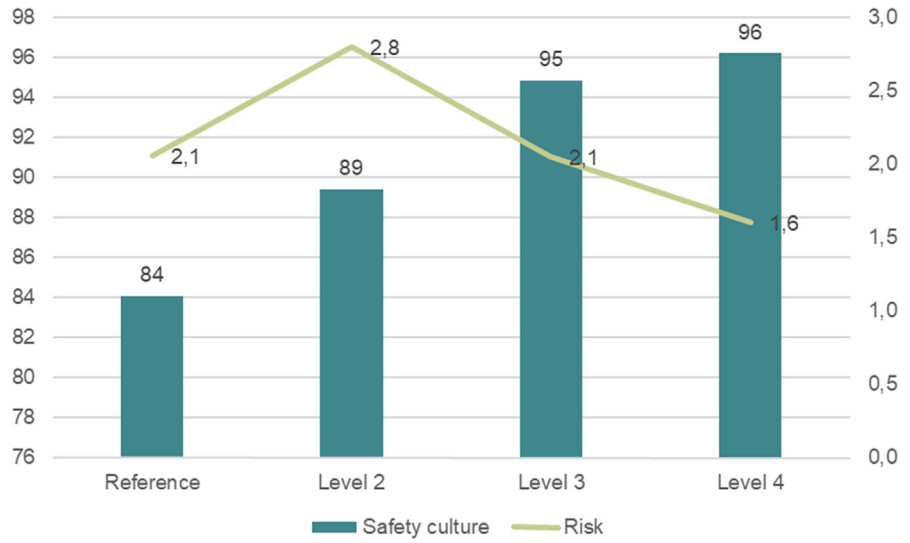

Fig. 3. Average scores of the GAIN index for safety culture in the four groups, and accident risk based on numbers from the survey. The GAIN index is comprised of 24 questions with five answer alternatives (Min: 24, Max: 120). Reference (N = 80), Level 2 (N = 39), Level 3 (N = 126) and Level 4 (N = 288).

Table 10
Logistic regression. Dependent variable: Accident involvement last 2 years (No = 0, Yes = 1). Odds ratios Exp(B).

Variable	Mod. 1	Mod. 2	Mod. 3	Mod. 4	Mod. 5	Mod. 6	Mod. 7	Mod. 8	Mod. 9
Kms driven	**1.005*****	**1.006*****	**1.006*****	**1.006*****	**1.005*****	**1.005*****	**1.005*****	**1.005*****	**1.005*****
Age (<26 = 0,>26 = 1)		**0.333*****	**0.388****	**0.384****	**0.361****	**0.363****	**0.360****	**0.344****	**0.390****
Type of transport (ADR = 0, Not ADR = 1)			**2.121****	**2.006***	**2.451****	**2.415****	**2.419****	1.987	1.565
Perceived safety management				0.988	1.050	1.050	1.050	1.048	1.047
Safety culture					**0.962***	**0.962****	**0.963***	**0.966***	0.972
Customer pressure/stress						1.019	1.017	0.996	0.976
Customer focus on safety							0.991	0.983	0.996
Safety ladder Level 4 (=0, other = 1)								1.502	1.524
Safety behaviour									**1.103****
Nagelkerke R	0.046	0.074	0.093	0.097	0.114	0.114	0.114	0.121	0.139

* p < 0.1 ** p < 0.05 *** p < 0.01.

Safety ladder criteria (cf. Table 1) in light of the interview results, it seems that criteria 2 (speed limiter) should be discarded, as the interview results indicate that views and practices related to this issue were mixed among companies at all levels.

Interview results indicate that the companies at Level 3 in the Safety ladder are characterized by the following aspects (in addition to the aspects at Level 2): 1) The wage system is organized to minimize driver stress and fatigue. 2) Potential consequences for fatigue and stress are assessed when new assignments are planned. 3) Drivers are encouraged to postpone, or abort assignments if they feel that conditions are unsafe, and they do so. Interview results do not support criteria 1: Drivers only have contact with their own transport managers and not customers/receivers before delivery. The prevailing impression is that the type of contact with the customers often varies according to the type of transport. Interview results also indicate that it is difficult to discern the importance of criteria 5, related to Level 3. The type of client and the duration of contracts and assignments are of great importance for the companies' experience of whether they can negotiate with clients about deadlines, terms, etc.

Interview results indicate that the companies at Level 4 in the Safety ladder are characterized by the following aspects (in addition to the aspects at Level 3): 1) Functioning reporting systems, which often involve a few hundred reports a year (e.g.

2–3 per employee per year). 2) Reported incidents and safety issues are regularly reviewed to learn from them, and employees are informed about the results. 3) The companies regularly conduct formalized risk assessments of all activities. Level 4 companies conduct risk analyzes more often, and they do so even if not required by a third party. 4) They have a well-developed set of procedures that are known and perceived as meaningful. 5) They have systematic training programs with predefined theoretical and practical activities, plan for knowledge goals, activities to fulfil the goals and assess goal development.

5.2. The relationship between safety structure and safety culture

The second aim of the study was to examine whether safety culture is improved with increased structural measures at each Safety ladder level. Results indicate increasing safety culture scores for each level in the Safety ladder, and that the difference between the Reference group at the lowest level and Level 4 is 12 points. Comparing the scope of structural safety measures for each Safety ladder level, we have seen that, while Level 2 is characterized by five structural safety measures, Level 4 is characterized by (at least) 15 safety structural measures. The result that the companies at Level 4 in the Safety ladder score higher than the companies at step 2 both on the index for perceived safety management and safety culture is not surprising in this perspective. The companies at Level 4 generally have three times more structural safety measures than those at Level 2, and they also work more systematically with them. This is the most relevant explanation to the relationships that we have seen between safety culture and safety structure.

This main results in line with previous research, indicating that increasing structural safety measures (and SMS implementation) is related to improvements in safety culture scores (e.g. Naveh and Katz-Navon, 2015; Newnam & Oxley, 2016). As we have seen, these two studies are the only studies actually measuring improvements in safety culture among the few studies focusing on safety structure, culture and safety outcomes. Although Murray et al. (2009, 2012) and Wallington et al. (2014) also study safety structure/culture interventions, they do not measure potential improvements in safety culture like the present study does. Thus, the present study contributes to the research in this field by documenting improvements in safety culture with increasing safety structural measures, at the different Safety ladder levels.

The previous studies indicate that it may be hard in practice to discern between aspects of safety structure and safety culture, and the outcomes of specific aspects of SMS (cf. Nævestad et al., 2018b). By relating the specific safety management measures at each level (gradual implementation) to increasing safety culture scores, the present study contributes to some extent to reducing this research gap. It must, however, be noted that it is difficult to know whether structure influences culture, or whether culture influences structure (i.e. "what comes first?"). It is conceivable that companies with a good safety culture are motivated to introduce several measures aimed at safety management. This contests the hypothesis that SMS facilitates culture, suggesting that it may also be the other way around.

5.2.1. Companies with both positive safety structure and positive safety culture

There were three companies at Level 4 in the study that have both relatively comprehensive systems and good safety culture scores: Company J, I and P (>100 points on the GAIN index for safety culture). How can we explain the positive safety culture scores of these companies? Based on the interviews, we may conclude that two of the things that these three companies have in common are that they appear to have: 1) a relatively strong involvement of the employees in the safety work, and 2) that they seem to use a systematic "continuous improvement" approach. These three companies appear to have a strong focus on developing risk awareness through joint discussions about hazards in the work between managers and employees. The management representative for Company I underlined the importance of involving all parties in the process of safety culture development, e.g. through continuous improvement, underlining that he is very conscious of working with small groups. Likewise, a key employee in Company P regularly sits at one of the loading points to talk to all the drivers about all kinds of safety issues. A continuous improvement approach means that the company continuously monitors the hazards it faces through reporting incidents and safety issues, the use of key statistics, etc., that counter-measures are developed and monitored, based on these data, and that employees are involved in all of the phases in this cycle. In addition, this approach involves good communication between managers and employees on safety issues. These key features (employee involvement, continuous improvement, communication) seem to be in accordance with the previous research on this issue, which all describe systematic processes of continuous improvement over several years, involving both employees and managers (Murray et al., 2009, 2012; Wallington et al., 2014; Naveh and Katz-Navon, 2015; Newnam & Oxley, 2016).

5.2.2. Companies with safety management systems and low safety culture scores

Company Q, O and K at Level 4, all have less than 90 points on the GAIN index, in spite of well-developed safety structures, in accordance with Level 4 in the Safety ladder. Results from these companies indicate that it is not always the case that high safe culture scores follow from several structural measures. This result is also in accordance with previous studies, which point to several potential explanations and symptoms, e.g. low SMS ownership due to implementation difficulties, conflicts, low trust (Amtrak, 2015), employee resistance (Haukelid, 2008), poorly adaptions of SMS to work life realities (Almklov et al., 2014), "unnecessary" formalization threatening professional competence (Antonsen, 2009) etc. Based on the interviews, it was difficult to evaluate which of these explanations that apply to company Q, O and K. It should also be noted that response rates in these companies were low (perhaps also indicating lower safety engagement). The relatively low safety culture scores in these Level 4 companies, and the mentioned studies indicating discrepancies between formal and informal aspects

of safety, seem to indicate that it is of little use to have sophisticated formal SMS, if they are not sufficiently implemented and embraced by the employees, perhaps illustrating that it is more important to have a good safety culture than a good safety system.

5.3. The relationship between safety culture, safety structure and safety outcomes

The third aim of the study is to examine whether the accident risk decreases at each Safety ladder level (with improved safety culture and safety structure). Results indicate that the accident risk in average decreases for the companies at each level of the Safety ladder (although we see that the accident risk is higher at Level 2 than it is in the Reference group). This indicates a relationship between safety structure, safety culture and accident risk. Previous research also indicates that increasing safety structural measures are related to reductions in accidents (e.g. Murray et al., 2009, 2012; Wallington et al., 2014; 2014b). The result that the accident risk decreases with increasing safety culture scores is also in accordance with previous research (cf. Christian et al., 2009). Few of the previous studies of safety culture/structure from the road sector do, however, include accident risk as an outcome measure.

Moreover, few to none of the previous studies provide systematic analyses of the relationship between gradual implementation of safety structure (Safety ladder levels), safety culture and accident risk, like we do in the present study. Moreover, as several structural measures often are gradually implemented in the existing case studies, it is difficult to conclude about the outcomes and significance of specific structural measures (Murray et al., 2009, 2012, Wallington et al., 2012). In the present study, we examine the outcomes of different types of measures (at different levels) with a Reference sample, which has few measures, and which presumably represents an industry average. The main challenge with the present study is, however, that it is cross sectional, and that we are unable to conclude about causality, as we have not conducted before and after measurements at the different Safety ladder levels. Thus, we are unable to conclude about what comes first: safety culture or safety structure. However, based on our two regression analyses, which focus on the relationship between structure, culture and accidents, we can conclude that it seems that safety structure is an important predictor of (or a least strongly correlated with) safety culture, and that safety culture is related to accident involvement, but that this relationship is mediated by road safety behaviours. This is in accordance with our general conclusion in section 5.2.2 above, that safety culture seems more important for safety outcomes than safety structure.

5.4. Questions for future research

5.4.1. How can we theoretically explain the relationship between structure and culture?

Despite the focus on SMS implementation or safety structural measures as a strategy for facilitating safety culture in air, rail and maritime transport, theoretical explanations of the relationship between structure and culture seem to be lacking in the literature. We propose two theoretical mechanisms to understand this relationship. The first focuses on sensemaking (Weick, 1995) and cognitive dissonance. SMS formally prescribe certain shared ways of acting. Based on (Weick & Sutcliffe, 2007) and Reason (1997), we may hypothesize that organisations "may act into what they become". Describing the sensemaking mechanism, Weick (1995) contend that when members in organisations start to act in new ways (e.g. as prescribed by the SMS) they may look for explanations to this, to reduce cognitive dissonance. If they adopt ways of thinking supporting the SMS, focusing on "continuous improvement", and the importance of a systematic approach to safety, they may not only start to act differently, but also think differently. In this way, shared ways of thinking and acting in relation to safety (i.e. safety culture) may be implemented through SMS.

The second theoretical explanation, focusing on how safety structural measures may be used to develop safety culture is based on Edgar Schein (2004) «primary embedding mechanisms» that managers can use to shape culture. Schein contends that managers signal basic assumptions, norms and values (i.e. organisational culture) through e.g. what they pay attention to, measure and control on a regular basis, how they allocate resources, role model, allocate rewards and status. The companies' safety measures can be analyzed considering this, e.g. the use of fleet management system, with continuous recording of driving style, ranking of drivers and feedback to drivers. It seems that using the fleet management system actively to focus on speed and driving style may also represent an important way to develop a safety culture, as this is a very concrete way of monitoring behaviour, communicating norms, rewarding, sanctioning etc.

5.4.2. Small differences between the safety culture scores of Level 3 and 4

We find small and non-significant differences between the safety culture scores at levels 3 and 4. This could potentially be due to the above-mentioned companies at Level 4 with low safety culture scores. It could also be related to the size of the companies at Level 4 and their degree of formalization. The companies at Level 4 are generally larger than those at Level 3. With increased company size, there is an increased need to use formalized methods (SMS) to coordinate and control employees (cf. McShane & Travaglione, 2007). One possible challenge with formal management through systems (and size) is that management becomes more distant for employees. For smaller companies, with informal personal relationships and short lines of communication, formal SMS may be perceived as less relevant. If the manager can manage and coordinate through direct daily contact with a handful of drivers, they may not see the need to create comprehensive formal procedures to standardize driver behaviour. Employees' assessments of management's commitment to safety is the most important aspect of safety culture/climate (cf. Flin et al., 2000). Against this backdrop, we may hypothesize that in smaller companies, we may

find a high safety culture level despite less sophisticated SMS (e.g. at Level 3). This is, however, mere speculation. Moreover, we should also remember that our results indicate smaller companies with low formalization and relatively low safety culture scores, and larger companies with relatively high formalization and high safety culture scores.

5.4.3. The importance of framework conditions

Our regression analyses of factors influencing safety culture indicate that type of transport (ADR) and customer focus on safety contribute significantly. We also see important differences related to customer pressure and stress at the different Safety ladder levels. Examining the shares agreeing with the statement that customers put pressure on and stress drivers, we see that 46% agreed in the Reference sample, 45% agreed at Level 2, 29% agreed at Level 3 and 22% agreed at Level 4. These findings are in line with previous research, which find that type of transport (Elvik et al., 2009) and framework conditions (Bjørnskau & Longva, 2009) influence safety culture and accident risk. Our regression analyses indicate relationships between framework conditions, safety culture, safety management and accident involvement in the studied companies. This is illustrated in Fig. 4.

5.5. Practical implications: Good practices for safety management

Our study indicates that the management practices at each level of the Safety ladder for goods transport are related to increasing levels of safety culture and decreasing levels of accident risk

(We have excluded the management practices that were related to significant ambiguities):

(1) Policy for speed, driving style and seat belt use.
(2) Fleet management system recording drivers' speed, driving style and fuel consumption.
(3) Regular (daily, weekly, monthly) feedback about the driving style to drivers
(4) Wage system developed to minimize drivers' stress and fatigue.
(5) Drivers are encouraged to stop and postpone assignments, if they perceive that it is unsafe to complete them.
(6) When assignments are planned, companies conduct an assessment of the strain (e.g. stress, fatigue) on the driver that the new assignment will involve.
(7) Functioning reporting system, which is used by employees.
(8) Reported incidents are regularly analyzed (by the company) to learn.
(9) The company regularly conducts formalized risk assessments (based on predefined methods) for all their assignments.
(10) Well-developed set of procedures that are known and perceived as meaningful by the employees.
(11) Well-developed training program (theoretical, practical) of a certain duration, with predefined requirements to driver competence and knowledge, and activities to fulfil the requirements.

It is important to stress that there is considerable uncertainty concerning the specific importance of each of these management practices for safety and safety culture, and that they must be examined further in future research.

5.6. Methodological challenges and weaknesses

1) We cannot be certain whether the identified mechanisms actually explain the identified differences between the different steps of the Safety ladder. The differences between the levels may be due to third variables that we have been unable to measure, e.g. related to framework conditions, technology and vehicles. In this study, we only focus on some aspects of safety management, and these are aspects related to organisational safety management: safety culture and SMS.

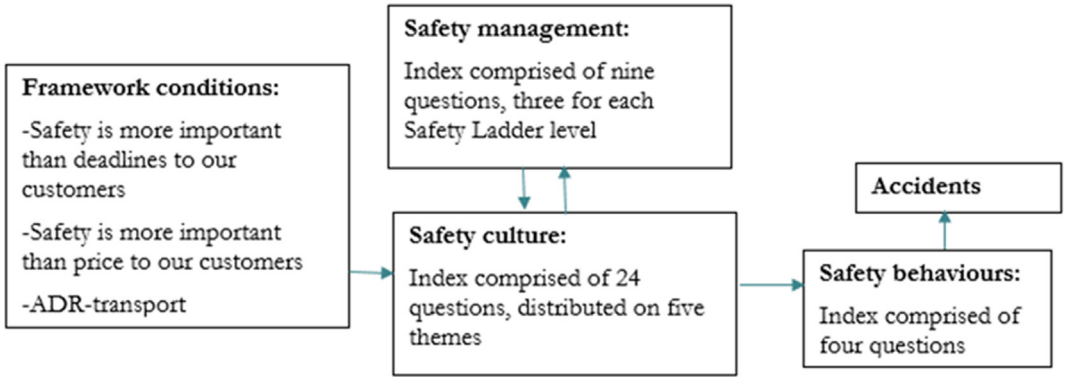

Fig. 4. Relationships between framework conditions, safety culture, safety management and accident involvement in the studied companies, as indicated in the regression analyses.

2) We cannot know what affects what. We suggest that SMS may provide a way of implementing safety culture in companies, but we cannot conclude about causality in the present study. Safety culture and safety structure mutually affect each other, and it is therefore difficult to separate analytically between them. It is for instance conceivable that companies with a good safety culture are motivated to implement more safety management measures. On the contrary, we may also hypothesize that companies with good safety structure are unable to develop good safety culture, despite several measures.

6. Conclusion

Accidents with HGVs represents an important societal challenge. Efforts are therefore needed to reduce the accident risk of HGV transport. Despite the potential of preventing HGV accidents through organizational safety management, it seems, however, that neither trucking companies, nor regulatory authorities focus sufficiently well on this. To compensate for known implementation challenges in the sector, Nævestad et al. (2017) suggested the Safety ladder for safety management in trucking companies. The Safety ladder for goods transport describes an approach with an increasing prevalence of, and focus on structural safety measures at four management levels, or Safety ladder levels in trucking companies (cf. Nævestad et al., 2017). The present paper has validated the Safety ladder approach in empirical research by comparing safety structure, safety culture and accident risk for trucking companies at different levels in the Safety ladder. Based on our interview results, we have described increasing safety structural measures for the companies on each level of the Safety ladder. Survey results indicate that safety culture scores increase at each level of the Safety ladder, while the accident risk decreases. The safety culture level was 12 points higher at Level 4, than in the Reference sample, and the accident risk at Level 4 was half of that of Level 2. We conclude by suggesting the concrete management practices related to each level as good practices, as these seem to be related to the increases in safety culture and decreases in accident risk.

CRediT authorship contribution statement

Tor-Olav Nævestad: Conceptualization, Funding acquisition, Methodology, Project administration, Writing - original draft, Writing - review & editing. **Jenny Blom:** Methodology, Writing - review & editing. **Ross O. Phillips:** Methodology, Writing - review & editing.

Acknowledgements

The present study has been funded by the Norwegian Public Roads Administration (NPRA), and our contact person has been Hans-Petter Hoseth. The companies have been recruited with good help from the Norwegian Hauliers' Association (NLF), represented by Jens Olaf Rud. We have also received help from a drivers' union to recruit the reference sample. We are very thankful for the funding from NPRA and for all the help and interesting discussions. The paper is based on a project, which also is presented in a comprehensive Norwegian report (Nævestad et al., 2018c).

Appendix A

See Table A1.

Table A1
Mean scores for safety culture in the companies, measured by means of the GAIN-index, number of respondents, standard deviation and response rates in each company. Response rates are calculated based on the number of drivers who received the surveys in the companies, and this number may in some cases deviate from the total number of drivers in the companies. (In some companies, only some branches received questionnaires.)

Level	Company	Mean score	S.D.	N	Response rate
	Reference	84	22.36	80	12%
Level 2	Company A	81	9.19	6	33%
	Company B	95	17.62	15	38%
	Company C	87	24.28	12	30%
	Company D	88	16.38	6	24%
Level 3	Company E	94	15.17	28	20%
	Company F	97	11.73	26	87%
	Company G	95	17.54	33	30%
	Company H	94	15.22	39	98%
Level 4	Company I	101	10.54	92	71%
	Company J	102	7.62	28	80%
	Company K	89	15.69	12	24%
	Company L	92	16.47	80	40%
	Company M	93	13.26	22	63%
	Company N	94	15.08	21	32%
	Company O	89	16.66	12	12%
	Company P	102	7.47	13	14%
	Company Q	88	11.80	8	18%
	Total	94	16.65	533	28%

References

Almklov, P. G., Rosness, R., & Størkersen, K. (2014). When safety science meets the practitioners: does safety science contribute to marginalization of practical knowledge?. *Saf. Sci., 67*, 25–36.

Amtrak (2015) Safety and security: opportunities exist to improve the Safe-2-Safer program, (Audit Report OIG-A-2015-007, February 19, 2015).

Antonsen, S. (2009). *Safety culture: Theory, Method and improvement* (tion1st Edition). London: CRC Press.

Bjørnskau, T., & Nævestad, T.-O. (2013). *Safety culture and safety performance in transport – A literature review, TØI Working-paper-50267*. Oslo: Transportøkonomisk institutt.

Bjørnskau, T., & Longva, F. (2009). *Safety Culture in Transport. (Norwegian language). TØI rapport 1012/2009*. Oslo: Transportøkonomisk institutt.

Christian, M. S., Bradley, J. C., Wallace, J. C., & Burke, M. J. (2009). Workplace safety: A meta-analysis of the role of person and situation factors. *Journal of Applied Psychology, 94*, 1103–1127.

Crum, Michael R., & Morrow, Paula C. (2002). The influence of carrier scheduling practices on truck driver fatigue. *Transportation Journal*, 20–41.

Dahl, Ø. (2013). Safety compliance in a highly regulated environment: A case study of workers' knowledge of rules and procedures within the petroleum industry. *Safety Science, 60*, 185–195.

Davey, J., Freeman, J. & Wishart, D. (2006). A study predicting crashes among a sample of fleet drivers, In Proceedings Road Safety Research, Policing and Education Conference, Gold Coast, Queensland.

Elvik, R., Høye, A., Vaa, T., & Sørensen, M. (2009). *The Handbook of Road Safety Measures* (2nd edn). Bingley: Emerald Insight.

ERA (2013). EUROPEAN RAILWAY AGENCY Safety Unit Application guide for the design and implementation of a Railway Safety Management System DEVELOPING AND IMPROVING SAFETY CULTURE IN THE ORGANISATION.

ERSO(2018) https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/statistics/dacota/bfs20xx_hgvs.pdf.

European Commission. (2009). *Road freight transport vademecum*. European Commission, Directorate General Energy and Transport, Directorate E – Inland Transport, Unit E.1 – Land Transport Policy.

Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Measuring safety climate: Identifying the common features. *Safety Science, 34*(1–3), 177–192.

Frost, J. (2020) https://statisticsbyjim.com/regression/multicollinearity-in-regression-analysis/.

GAIN (Global Aviation Network) (2001). Operator's Flight Safety Handbook.

Gregersen, N. P., Brehmer, B., & Morén, B. (1996). Road safety improvement in large companies. An experimental comparison of different measures. *Accident Analysis & Prevention, 28*(3), 297–306.

Guldenmund, F. W. (2007). The use of questionnaires in safety culture research – An evaluation. *Safety Science, 45*, 723–743.

Guldenmund, F., Cleal, B., & Mearns, K. (2013). An exploratory study of migrant workers and safety in three European countries. *Saf. Sci., 52*, 92–99.

Hale, A. R. (2000). Culture's confusions. *Safety Science, 34*(1–3), 1–14. https://doi.org/10.1016/S0925-7535(00)00003-5.

Haukelid, K. (2008). Theories of (safety) culture revisited—An anthropological approach. *Safety Science, 46*(3), 413–426.

Hudson, P. (2003). Applying the lessons of high risk industries to health care. *Quality and Safety in Health Care, 12*, i7–i12.

IATA (2019) Creating a positive safety culture. Best practices to align with Annex 19's new recommendations, https://go.updates.iata.org/safety-culture?_ga=2.26439165.1941341082.1571071913-370565390.1570852392.

Lappalainen, F. J., Kuronen, J., & Tapaninen, U. (2012). Evaluation of the ISM Code in the Finnish shipping companies. *Journal of Maritime Research, 9*(1), 23–32.

Mayhew, C., & Quinlan, M. (2006). Economic pressure, multi-tiered subcontracting and occupational health and safety in Australian long-haul trucking. *Employee Relations, 28*(3), 212–229.

McShane, S., & Travaglione, A. (2007). *Organisational behaviour on the pacific rim* (2nd ed). New South Wales, Australia: McGraw-Hill.

Mooren, L., Grzebieta, R., Williamson, A., Olivier, J., & Friswell, R. (2014a). Safety management for heavy vehicle transport: A review of the literature. *Safety Science, 62*, 79–89.

Mooren, L., Williamson, A., Friswell, R., Olivier, J., Grzebieta, R., & Magableh, F. (2014b). What are the differences in management characteristics of heavy vehicle operators with high insurance claims versus low insurance claims? *Safety Science, 70*, 327–338.

Murray, W. S. Ison, P. Gallemore & H Sing. Nijjar (2009) Effective Occupational Road Safety Programs. A case study of Wolseley, Transportation Research Record, NO: 2096, pp. 55-64.

Murray, W., White, J., & Ison, S. (2012). Work-related road safety: A case study of Roche Australia. *Safety Science, 50*(1), 129–137.

Naveh, E., & Marcus, A. (2007). Financial performance, ISO 9000 standard and safe driving practices effects on accident rate in the U.S. motor carrier industry. *Accident Analysis and Prevention, 39*(4), 731–742.

Naveh, E., & Katz-Navon, T. (2015). A Longitudinal Study of an Intervention to Improve Road Safety Climate: Climate as an Organizational Boundary Spanner. *Journal of Applied Psychology, 100*(1), 216–226.

Newnam, S., & Oxley, J. (2016). A program in safety management for the occupational driver: Conceptual development and implementation case study. *Safety Science, 84*, 238–244.

NHTSA (2019) https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812663.

Nævestad, T.O. (2010). Culture, crises and campaigns: examining the role of safety culture in the management of hazards in a high risk industry. PhD, University of Oslo: Oslo.

Nævestad, T.O. & Bjørnskau, T. (2014). Kartlegging av sikkerhetskultur i tre godstransportbedrifter.Oslo: TØI rapport 1300/2014.

Nævestad, T. O., Phillips, R. O., & Elvebakk, B. (2015). Traffic accidents triggered by drivers at work–A survey and analysis of contributing factors. *Transportation Research Part F: Traffic Psychology and Behaviour, 34*, 94–107.

Nævestad, T. O., Phillips, R. O., & Elvebakk, B. (2017). The safety ladder: Developing an evidence-based safety management strategy for small road transport companies. *Transport Reviews*.

T.-O. Nævestad, R. O. Phillips, I. B. Hovi, G .N. Jordbakke og R Elvik (2018a) Miniscenario: Sikkerhetsstigen. Innføre tiltak for sikkerhetsstyring i godstransportbedrifter. TØI rapport 1620/2018, Oslo: Transportøkonomisk institutt

Nævestad, T.-O, R. Phillips, I. Storesund Hesjevoll (2018b), How can we improve safety culture in transport organizations? A review of interventions, effects and influencing factors in Transportation Research Part F: Psychology and Behaviour. Volume 54, P. 28–46.

Nævestad, T.O., J. Blom., R.O. Phillips (2018c) Sikkerhetskultur, sikkerhetsledelse og risiko i godstransportbedrifter på veg; TØI rapport 1659/2018, Oslo: Transportøkonomisk institutt.

Nævestad, T.-O., Hesjevoll, I. S., Ranestad, K., & Antonsen, S. (2019). Strategies regulatory authorities can use to influence safety culture in organizations: Lessons based on experiences from three sectors. *Safety Science, 118*(2019), 409–423.

Reason, J. (1997). *Managing the risks of organizational accidents*. Aldershot: Ashgate Publishing Ltd.

Reason, J. (2000). Human error: models and management. *British Medical Journal, 320*, 768–770. https://doi.org/10.1136/bmj.320.7237.768.

Schein, E. (2004). Organizational Culture and Leadership (Third Edition ed.). San Francisco: Jossey-Bass.

Jensen, Steen, Bråten, R. M., Jordfald, B., Dotterud Leiren, M., Nævestad, T.-O., Skollerud, K. H., et al (2014). Arbeidsforhold I gods og turbil. *Fafo rapport, 2014*, 58.

Størkersen, K.V.; Bye, R.J.; Røyrvik, J.O.D. Sikkerhet i fraktefarten. Analyse av drifts- og arbeidsmessige forhold på fraktefartøy. In NTNU Samfunnsforskning AS; Studio Apertura: Trondheim, Norway, 2011

Thomas, M. J.W. (2012). A systematic review of the effectiveness of safety management systems. No. AR-2011-148. Australian Transport Safety Bureau.

Wallington, D., Murray, W., Darby, P., Raeside, R., & Ison, S. (2014). Work-related road safety: Case study of British Telecommunications (BT). *Transport policy, 32*, 194–202.

Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.

Weick, K.E. & K.M. Sutcliffe (2007): Managing the unexpected. Resilient performance in an age of uncertainty. (2nd ed.) San Fransisco: Jossey Bass.
Wills, A. R., Biggs, H. C., & Watson, B. (2005). Analysis of a safety climate measure for occupational vehicle drivers and implications for safer workplaces. *Australian Journal of Rehabilitation Counselling, 11*(1), 8–21.
Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions. *Accident Analysis & Prevention, 42*(5), 1517–1522.