Accident Analysis and Prevention 68 (2014) 57–74

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Determining the causal relationships among balanced scorecard perspectives on school safety performance: Case of Saudi Arabia

Turki Alolah[a,b], Rodney A. Stewart[a,*], Kriengsak Panuwatwanich[a], Sherif Mohamed[a]

[a] Griffith School of Engineering, Griffith University, Gold Coast Campus, Queensland 4222, Australia
[b] Saudi Ministry of Education, Saudi Arabia

ARTICLE INFO

Article history:
Received 11 April 2013
Received in revised form 30 January 2014
Accepted 3 February 2014
Available online 11 February 2014

Keywords:
Safety
Schools
Safety performance evaluation
Saudi Arabia
Balanced scorecard
Systems theory
Partial least squares
Exploratory factor analysis
Structural equation model

ABSTRACT

In the public schools of many developing countries, numerous accidents and incidents occur because of poor safety regulations and management systems. To improve the educational environment in Saudi Arabia, the Ministry of Education seeks novel approaches to measure school safety performance in order to decrease incidents and accidents. The main objective of this research was to develop a systematic approach for measuring Saudi school safety performance using the balanced scorecard framework philosophy. The evolved third generation balanced scorecard framework is considered to be a suitable and robust framework that captures the system-wide leading and lagging indicators of business performance. The balanced scorecard architecture is ideal for adaptation to complex areas such as safety management where a holistic system evaluation is more effective than traditional compartmentalised approaches. In developing the safety performance balanced scorecard for Saudi schools, the conceptual framework was first developed and peer-reviewed by eighteen Saudi education experts. Next, 200 participants, including teachers, school executives, and Ministry of Education officers, were recruited to rate both the importance and the performance of 79 measurement items used in the framework. Exploratory factor analysis, followed by the confirmatory partial least squares method, was then conducted in order to operationalise the safety performance balanced scorecard, which encapsulates the following five salient perspectives: safety management and leadership; safety learning and training; safety policy, procedures and processes; workforce safety culture; and safety performance. Partial least squares based structural equation modelling was then conducted to reveal five significant relationships between perspectives, namely, safety management and leadership had a significant effect on safety learning and training and safety policy, procedures and processes, both safety learning and training and safety policy, procedures and processes had significant effects on workforce safety culture, and workforce safety culture had a significant effect on safety performance.

Introduction

Safety in Saudi school environments

The Saudi Ministry of Education recognises that the health and safety of its teachers, school executives, students, parents, visitors and the broader school community are of paramount importance. In fact, the ministry is committed to providing safety conditions in each school and seeks to minimise school accidents by using successful and accurate safety regulations. A number of regulations have been introduced in recent years in a bold attempt to rapidly improve safety in Saudi schools. The ministry, for instance, implemented regulations that enforce heightened expectations of student safety to ultimately reduce the rates of accidents and incidents. These regulations include compulsory first-aid courses for teachers, laboratory safety guidelines and student supervision, to name a few (Bendak, 2006).

Despite the implementation of stricter regulations, safety has not improved to the degree expected due to a lack of enforcement practices and ongoing safety performance evaluations. The application of these regulations by people who do not have safety knowledge leads to disasters such as the 2002 fire in a girls' school that resulted in 15 deaths (Prokop, 2003). Moreover, in the past decade (2002–2012), there were 19 student and teacher fatalities

* Corresponding author. Tel.: +61 7 5552 8778.
E-mail addresses: t.alolah@griffith.edu.au (T. Alolah), r.stewart@griffith.edu.au (R.A. Stewart), k.panuwatwanich@griffith.edu.au (K. Panuwatwanich), s.mohamed@griffith.edu.au (S. Mohamed).

http://dx.doi.org/10.1016/j.aap.2014.02.002
0001-4575/

and more than 150 student injuries, of which 12 were very serious (Saudi Ministry of Education, 2013).

Poor communication is the main reason behind the lack of ongoing enforcement of safety regulations and performance evaluation. In Saudi Arabia, male and female roles are viewed as complementary rather than equal (Esposjto, 1982; Keddie, 1979; Saleh, 1972). El Guindi (1981) noted that "*one basic feature of Arab socio-cultural organisation is the division of society into two separate and complementary worlds, the men's and the women's*". In Saudi Arabia, this division is relatively rigid and strict; sexes do not mix (El Guindi, 1981). This underlying feature of Saudi culture means that communication between the male dominated Ministry of Education and the female school system is often limited and a step removed from the school operational environment. Other issues include a general lack of safety awareness from teachers and school executives and regional schools often not being included in centrally administered safety initiatives. Concerns about school safety are growing in the wider community; however, there are still limited studies addressing the issue of school safety, particularly studies targeted at Saudi Arabia. Although Western countries have conducted numerous studies on school safety (Nansel et al., 2001; Pellegrini and Long, 2002; Pepler et al., 2006), the majority of this research focuses on developed Western countries, whose results are often not transferable to the Middle East region's entirely different cultural context.

Saudi Arabia's 33,000 existing schools (Saudi Ministry of Education, 2013), as well as future schools that will be built, must implement a more robust safety management system. Indeed, to gain a clearer understanding of school safety in the Saudi context, it is important that further studies be conducted in this specific context. The following additional aspects should also be considered and taken into account when developing such a system: management (Choudhry et al., 2007; Mearns et al., 2003); safety training (Cooper and Phillips, 2004; Ng et al., 2005); safety processes (Clarke, 2010; Teo and Ling, 2006); safety culture (Grabowski et al., 2007; Mayze and Bradley, 2008; Wadsworth and Smith, 2009); and finally, the most important aspect, safety performance (Grabowski et al., 2007). These aspects cover the key elements of the school safety system and must be concurrently addressed in order to achieve a desired overall level of safety performance.

Evaluating school safety systems holistically

Senge (2006) advocate that to achieve any shared goal of an organisation, the fifth discipline of 'systems thinking' is paramount to ensure that all stakeholders learn from each other. Rasmussen (1997) argues for a system-oriented approach to safety risk management and accident prevention that is based on abstraction rather than structural decomposition. His research helped move the narrowly focused safety control system agenda to one that was broader and covered the entire socio-technical system. This transition from creating safety management control systems to strategic safety management systems that considers a destination statement (i.e. safety goal) for the organisation is imperative to create long-term sustainable improvements in safety outcomes.

A number of systems thinking tools have been developed in recent years in order to conceptualise, operationalise, model, diagnose or evaluate natural, scientific, engineered, social or conceptual systems. Systems thinking theory has been applied to safety problems across a range of disciplines including road safety (Larsson et al., 2010), hazardous waste fire causation (Goh et al., 2010), aerospace safety (Leveson, 2007), to name a few. All of these studies highlight that all aspects of the multi-faceted socio-technical system must be considered in order to interpret safety outcomes. Safety systems theory is often applied after a series of significant incident or accidents occur in order to determine the key determinants of accident causation within that system. Such accident causation mapping exercises can also be utilised to build holistic safety system evaluation frameworks that encapsulate all pieces of the safety puzzle and lead to sustainable reductions in incidents and accidents. Such safety systems can be underpinned by systems theory but also need to be operationally definable and user friendly for laypersons to interpret and implement on a day-to-day basis.

The Balanced Scorecard (BSC) developed by Kaplan and Norton (1992) is an internationally recognised and commonly utilised tool for evaluating performance across range of disciplines and areas (e.g. Mohamed, 2003; Stewart, 2008). The first generation BSC was control orientated with only a simplistic attempt to encapsulate all aspects of a system through including both lead and lag indicators of a particular performance outcome. This original BSC framework has four performance perspectives (i.e. internal business, financial, customer and, innovation and learning) and attempted to integrate all the interests of key stakeholders on a scorecard. The term '*balanced*' in the name reflects the balance provided between short- and long-term objectives, between quantitative and qualitative performance measures, and between different performance perspectives. The diverse interests and measures are categorised in the above-mentioned four performance perspectives of the scorecard.

However, over time the BSC has evolved into one that can map entire systems and also embed complex systems dynamic modelling techniques (e.g. Capelo and Dias, 2009). Lawrie and Cobbold (2004) argue that the third-generation balanced scorecard has transitioned from a strategic control tool to one that enables organisations to strategically map causal relationships between particular factors to some key goals related to a 'destination statement'. Kunc (2008) also demonstrates how the balanced scorecard provides a robust foundation for developing causal system models as they include all facets related to a particular desired outcome. In this study, the authors have attempted to conceptualise and empirically construct a safety scorecard for Saudi schools that includes the causal relationship between enabling and outcome perspectives of safety performance.

Developing a new safety performance evaluation framework for Saudi schools

The Saudi Ministry of Education has introduced safety rules to regulate aspects related to school safety and associated risk factors. The aim of these regulations is to ensure student safety and to decrease the rate of accidents and incidents. However, these systems have not generated the desired level of safety in schools because the regulations neither address the system-wide leading and lagging measures of safety performance nor provide methods for evaluating safety issues in Saudi schools. Applying the concept of a BSC could address this gap by effectively providing a measure of school safety. According to Kaplan and Norton (1996), the BSC provides employees (i.e. school personnel) with the necessary knowledge and skills acquired through learning and growth. The BSC approach also can innovate and build appropriate strategic capabilities and efficiencies. Some scorecard elements would also identify cause-and-effect relationships, that is, the impact of school safety processes and improvements on overall safety performance, goals, measurements, and perspectives. Strategic aspects would include the development of areas that define the scope of a BSC system, the strategic grid or logical framework for organising the various objectives, and the strategic framework, which is the combination of all the strategic grid's objectives. Incorporating such strategic aspects would provide one complete framework to manage the overall strategy and its constituent elements. The BSC system would also include the desired targets and templates, i.e., the visual tools to guide people in constructing and achieving the desired long-term objectives. The fact that performance is always an integral part of the BSC would help ensure that appropriate methods were identified to measure safety aspects, and importantly, to improve future performance. In addition, the causal relationships and interactions between the enablers and the goals of the developed safety performance BSC framework can be examined, thus providing a greater understanding of their interdependence and facilitating safety performance improvement in each school. As noted earlier, the aim of this research is to devise a safety BSC framework for measuring safety performance in Saudi schools and to develop a causal relationship among the developed safety measurement perspectives.

Research objectives

The overarching goal of this research investigation was to develop a robust safety performance BSC framework that could be later applied to benchmark, monitor and effectively manage safety in the growing Saudi public school system. Specific objectives completed to achieve that stated goal included:

- Develop a conceptual safety performance BSC that includes diverse perspectives that will encapsulate the system-wide leading and lagging indicators of safety performance in Saudi schools;
- Develop an operationalised safety performance BSC with statistically validated perspectives, factors and measurement items; and
- Determine the statistically significant causal relationships among safety performance BSC perspectives.

Conceptual safety performance BSC for Saudi schools

Numerous aspects of safety performance evaluation were considered and critiqued in the development of a safety performance BSC framework for Saudi schools. These aspects include school structure, processes, and staff hierarchy. An extensive review of the completed safety evaluation literature yielded five perspectives that encapsulate the system-wide leading and lagging indicators of safety performance. These perspectives are: safety management and leadership; safety learning and training; safety policy, procedures, and processes; workforce safety culture; and safety performance. Each perspective must be understood, and the factors and items related to each perspective should be sufficiently defined. The developed perspectives included a number of factors, and each factor included a number of measurement items. The key factors of each BSC perspective are discussed and described in the following sections and a summary of derived framework perspectives, factors and items is detailed in Table 1.

Safety management and leadership

According to Mearns et al. (2003), management relates to the actual practices, management roles, and functions associated with safe workplace practices. Furthermore, Choudhry et al. (2007) note that management plays a key role in promoting a positive safety culture. For example, when employees are aware of their responsibilities regarding incident and injury prevention, they will exhibit more interest in maintaining a safe and healthy work site. However, Håvold and Nesset (2009) indicate that management is concerned with the extent to which employees perceive that the organisation provides an effective information exchange regarding safety matters. Mearns et al. (2003) note that management also relates to the level of employees' trust in their supervisor, the competence of the supervisor to support safety practices, and the willingness of the supervisor to accept responsibility for mistakes. Teo et al. (2005) indicate that management should introduce incentives to improve safety practices.

Safety learning and training

Management's attitude and actions are devoted to providing the necessary job-related training and promoting the importance of safety training to all school users (Cooper and Phillips, 2004). To implement training and education, the development of safety training programs and the allocation of resources should be additional concerns (Ng et al., 2005). The enhancement of safety awareness among school users should include promotional strategies such as mission statements, published materials, and media (Choudhry et al., 2007). To incorporate the lessons learned from accidents and near accidents, safety improvement strategies might include the use of incident/accident reports and feedback (Håvold and Nesset, 2009). Overall, this perspective is concerned with school users' perception, knowledge, and competency regarding safety practices (Håvold and Nesset, 2009).

Safety policy, procedures, and processes

To effectively evaluate safety performance and work practices and to improve the effectiveness of safety management systems, Teo and Ling (2006) suggest that management should comply with organisational safety policies, procedures and processes. These researchers also indicate that safety audits and reviews are a structured process of collecting independent information on the efficiency, effectiveness, and reliability of the total safety management system, and of drawing up plans for correction and preventative action. This perspective is also concerned with the level of information documented in an audit/accident investigation report, users' satisfaction with follow-up actions taken by management, and supervisors' interest and ability to take necessary action (Clarke, 2010; Grabowski et al., 2007; Huang et al., 2004; Kines et al., 2010; Lawrie et al., 2006). Safety policies and procedures are considered to be one of the most influential elements driving safety performance, because organisational policies regarding safety have a significant influence in cultivating a positive and healthy safety culture (Clarke, 2010; Kines et al., 2010; Ng et al., 2005; Teo et al., 2005). Additionally, safety processes can take responsibility for safety equipment, tools, and other accessories (Glendon and Litherland, 2001; Kines et al., 2010; Mcdonald et al., 2000; Sawacha et al., 1999; Teo and Ling, 2006; Teo et al., 2005). Overall, this perspective is concerned with regular maintenance and the reinforcement of positive achievement (Kines et al., 2010; Teo and Ling, 2006).

Workforce safety culture

Mearns et al. (2003) defined safety culture as "that assembly of characteristics and attitudes in organisations and individuals, which establishes that, as an overriding priority, safety issues receive the attention warranted by their significance". The safety culture is particularly important in Saudi Arabia because it forms the context within which individual safety attitudes develop and persist and safety behaviours are promoted. Additionally, this perspective is concerned with the openness and the effectiveness of the organisation's reporting system and users' propensity to report accidents (Grabowski et al., 2007; Mayze and Bradley, 2008; Wadsworth and Smith, 2009). Additionally, this perspective focuses on users' perception of the safety of the work environment, including feedback, responsibility, empowerment, and reporting (Clarke, 2010; Mayze and Bradley, 2008; Wadsworth and Smith, 2009). The work situation and the effect of pressure on individuals' behaviours, attitudes, and safety practices are also examined (Glendon and Litherland, 2001; Håvold and Nesset, 2009). Finally, this perspective is concerned with the causes of accidents/incidents

Table 1
Synthesis of literature supporting safety performance BSC perspectives.

Perspectives	Factors	Items	Authors
Safety management and leadership	Management commitment to safety	Management actions safety issues Management promotes a safety culture Management provides adequate resources to safety Management participation in risk assessments, consultative committee meetings, and inspections Management encourages employees to voice concerns and safety improvement proposals	Choudhry et al. (2007); Mearns et al. (2003); Rundmo and Hale (2003); Wadsworth and Smith (2009)
	Safety communication	Management and supervisors have an open door policy Safety information is brought to employees' attention by their supervisors Safety information is provided (e.g. media, mission statements, accident statistics, etc.) Safety information (i.e. procedures) is visibly present in the workplace	Choudhry et al. (2007); Håvold and Nesset (2009); Kines et al. (2010)
	Employee involvement in safety	All levels of employees are empowered to be involved in safety management Employees are involved in setting safety objectives, decision making, and improvement plans	Choudhry et al. (2007); Mearns et al. (2003)
	Perceived supervisor competence	Supervisor is more attentive to safety issues than the average employee Supervisor is trusted and can relate safety-related information to employees Supervisor has the adequate skills and authority to tackle safety issues	Choudhry et al. (2007); Mearns et al. (2003); Wadsworth and Smith (2009)
	Safety practice incentives	Monetary incentives (e.g. bonuses) for employees for good safety practices Recognition incentives (e.g. safe employee of the month) for employees for good safety practices Punitive measures for continued poor safety practices (e.g. fines, demotions, etc.)	Teo et al. (2005)
Safety learning and training	Safety training and seminars	Development of safety training (e.g. short talks, group meetings, and workplace safety responsibilities) Provision of safety training to all employees Resource allocation for safety training Adequate and up-to-date safety training	Choudhry et al. (2007); Cooper and Phillips (2004); Håvold and Nesset (2009); Ng et al. (2005)
	Safety promotional strategies	Enhance safety awareness through clear mission statements (e.g. slogans and logos) Provision of published materials (e.g. books, statistics, and newsletters) Provision of media promotion (i.e. posters, displays, audio-visual media, e-mail, and Internet)	Choudhry et al. (2007); Sawacha et al. (1999)
	Safety learning openness	Employees give tips to each other on how to work safely Accident/incident reports are used to improve safety. Employees learn lessons from near misses and incident reports Feedback is used to improve safety	Håvold and Nesset (2009)
	Safety knowledge and competence	Employees are familiar with the company's safety policy Employees understand the purpose of the Safety Management System Employees know when to report near accidents	Håvold and Nesset (2009); Huang et al., 2013
Safety policy, procedures and processes	Safety audits and reviews	Conducting safety inspections and supervision Employment of safety office and supervisor Familiar with the company's safety policies and procedures Audit program is conducted regularly	Håvold and Nesset (2009); Lawrie et al. (2006); Ng et al. (2005); Teo and Ling (2006)
	Safety accountability and feedback	Publication of safety issues to staff members and parents The results of accident investigation are fed back to the supervisory level Employees are satisfied with the feedback given on accidents/incidents Employees are satisfied with follow-up measures taken after accidents/incidents	Clarke (2010); Grabowski et al. (2007); Huang et al. (2004); Kines et al. (2010); Lawrie et al. (2006); Mcdonald et al. (2000)
	Safety policies and procedures	Development of emergency plans and procedures Implementation of safety audits to the safety management system Supervisor monitors progress towards safety improvement goals based on the feedback and weekly meetings Safety policies/procedures can be followed without conflicting with work practices	Clarke (2010); Glendon and Litherland (2001); Kines et al. (2010); Ng et al. (2005); Teo and Ling (2006)
	Safety operations and governance	Equipment, tools, and other accessories are maintained regularly Conducting training on the use of safety equipment Safety officer's attitude has great influence on others' safety attitudes	Glendon and Litherland (2001); Kines et al. (2010); Ng et al. (2005); Sawacha et al. (1999); Teo and Ling (2006)
	Built environment maintenance	Development of a safety checklist to actively maintain facilities before accidents occur Identification of any safety area problems and respond in a timely manner	Kines et al. (2010); Teo and Ling (2006)

Workforce safety culture	Propensity to report accidents and incidents	Employee perceptions of the effectiveness of the reporting system Employee willingness to report a co-worker's failure Employee perceptions of the company's ability to correct mistakes	Grabowski et al. (2007); Mayze and Bradley (2008); Mearns and Håvold (2003); Wadsworth and Smith (2009)
	Individual responsibility to safety	The influence of personality on safety practices Value placed on personal safety responsibility High-quality work environment leads to better personal safety responsibility Employee involvement in informing management of safety issues Employee ability to consider safety as a top priority	Clarke (2010); Grabowski et al. (2007); Mayze and Bradley (2008); Wadsworth and Smith (2009)
	Perceptions of work situation and pressure	There are enough employees to carry out the required work Employees have enough time to carry out their tasks Realistic times are scheduled for completing assigned tasks Work procedures are presented clearly	Glendon and Litherland (2001); Håvold and Nesset (2009); Nansel et al. (2001)
	Fatalism	Accidents are unavoidable The use of machines and technical equipment makes accidents unavoidable Accidents seem inevitable despite the school's efforts to avoid them	Håvold and Nesset (2009)
Safety performance	General safety behaviours of staff	The influence of safety attitudes and behaviours Behaviour checklist towards safety practice The willingness to comply with procedures and policies, and to participate in safety practices Usage of personal protection equipment	Clarke (2010); Glendon and Litherland (2001); Mayze and Bradley (2008)
	Accidents and Incident rates	Measurement of accidents and incident rates Recordable accidents/incident frequency Total injury frequency	Glendon and Litherland (2001); Grabowski et al. (2007); Mayze and Bradley (2008)
	Emergency response	Emergency response preparedness Employees can identify the cause of an incident in a timely manner Employees competency handling emergency situations	Cooper and Phillips (2004); Glendon and Litherland (2001); Håvold and Nesset (2009); Teo and Ling (2006)
	Safety reporting, response and corrective actions	There are clear and well-documented procedures for developing remedial actions based on the incidents' causes An effective documentation management system ensures the availability of procedures Auditors' reports can provide valuable feedback and a basis for corrective and preventive actions	Clarke (2010); Cooper and Phillips (2004); Glendon and Litherland (2001); Mayze and Bradley (2008)

and managerial and individual efforts regarding safety preservation (Håvold and Nesset, 2009).

Safety performance

Grabowski et al. (2007) define safety performance as a measured element using organisation users' perceptions of safety in their work environment. Several researchers have considered organisation users' perceptions of workplace safety as a safety performance variable, along with other variables such as historical data on accidents and other incidents, the development of a safety system, individual behaviours and attitudes towards safety practices, the monitoring of safety compliance, the establishment of safety committees, the communication of safety policies, and organisation users' participation (Glendon and Litherland, 2001; Mayze and Bradley, 2008). The safety performance perspective is concerned with individual safety behaviours, the willingness to participate, and the implementation of a checklist before performing safety practices (Clarke, 2010; Glendon and Litherland, 2001), as well as the frequency of accidents and incidents in the workplace (Grabowski et al., 2007). This perspective is viewed as the most important aspect in Saudi schools because this outlook is concerned with emergency plans, response times to an accident/incident, and emergency training (Cooper and Phillips, 2004; Teo and Ling, 2006), which are often lacking in the Saudi educational environment. The safety performance perspective is concerned with safety feedback and reporting in the workplace and the assessment of corresponding corrective and preventive actions (Clarke, 2010; Cooper and Phillips, 2004; Glendon and Litherland, 2001; Mayze and Bradley, 2008; Teo et al., 2005).

Conceptualised causal relationships among safety performance BSC perspectives

The enablers of safety performance outcomes

The BSC framework developed by Kaplan and Norton (1992) provides the theoretical foundation that underpins the leading indicators of the safety performance BSC perspectives. Apart from the field of commerce in which the framework originated, the BSC has been adapted to the following range of performance evaluation areas: information technology (Stewart, 2008); healthcare (Inamdar et al., 2002); and tourism (Phillips and Louvieris, 2005), to name a few. The BSC perspectives intentionally focus on both leading and lagging indicators to provide a broader indication of both current and likely future performance trends, in order to go beyond traditional safety performance evaluation approaches that are too narrowly focused on lagging indicators (e.g. no. of accidents). This feature of the BSC framework was critical for this particular study as there needs to a much greater understanding on the relationships among the developed safety BSC perspectives in order to unpack the mechanisms for improving safety performance in Saudi schools over the longer term. Thus, the following sections will present arguments for the hypothesised role of leading safety BSC perspectives and their associated factors in influencing the mediating and lagging perspectives of safety performance.

The role of safety management and leadership

Concerning safety learning and training, every school requires practical and knowledgeable staff. The calibre of the staff is the backbone of any school. Safety learning and training will occur when personnel are motivated, encouraged to be innovative and have been empowered by management (Yang et al., 2009). According to Brown and Kenney (2006), management and leadership influence the quality of learning and training in any organisation. For example, management levels influence learning as a consequence of their effect on the collection and analysis of information. Additionally, leaders and managers provide training courses and seminars, ensuring that these learning opportunities are understandable and applicable. Managers' can ensure that safety training courses are appropriately selected and removed from the training schedule if they are not achieving targeted outcomes (Brown and Kenney, 2006). Moreover, management can accelerate information processing and decision-making, thereby diffusing safety knowledge more rapidly (Argyris and Schön, 1996). Beyond the effects of leadership in learning and training, management also is characterised by a unity of command, a win-lose competitive dynamic, a specialisation of tasks, and a focus on the rationality of ideas that excludes feelings, all elements that improve the learning and training culture (Argyris, 1999). This argument led to the following hypothesis:

H1. Safety management and leadership influences safety learning and training activities.

Safety management and leadership also have a powerful oversight on the development of safety policy, procedures, and processes. Additionally, Argyris (1999) indicates that good management is commonly characterised by a skill set that can have a very positive influence on the safety policy, procedures, and processes of an education system. According to Hayman (2007), the influence of management and leadership on policy, procedures and processes to improve safety systems and to raise school safety awareness also leads to staff taking on board higher levels of safety accountability. The above arguments led to the second hypothesis:

H2. Safety management and leadership influences safety policy, procedures, and processes.

The role of safety learning and training

As shown in numerous studies, the concept of learning and training has attracted considerable attention from leading management and organisational behaviour thinkers such as Argyris (1999), Senge (2006), Marquardt (2000), Garratt (1999), and Pedler and Aspinall (2000). Although quantitative data is limited, anecdotal evidence indicates that learning-oriented leaders and managers promote a learning organisation culture that often leads to improved performance and over the medium to long-term sustainable competitive advantage (Brown and Kenney, 2006). In any organisation, there are varied perceptions regarding the ideal learning and training courses and seminars to gain sufficient learning traction on a particular topic. Learning is associated with the intensity and quality of training provided in a safe environment, followed by an appropriate practical experience on the basis of theory taught in the class. The learning process is not the same for everyone because the perception, exposure and mental states of the individuals affect the progression of learning. Furthermore, training involves both skills development and risk assessment associated with the task as well as an understanding of appropriate safety measures to prevent any calamity. During the process, employees are encouraged to be self-motivated to direct their actions towards accomplishing goals and keeping the constraints of risk and safety in consideration (International Atomic Energy Agency, 2012). Effective safety learning and training programs can enhance several aspects of the workforce safety culture, including heightened confidence to report safety issues, higher personal safety responsibility and a less fatalistic view of life, to name a few. This argument provides the foundations for the third hypothesis:

H3. Safety learning and training influences the workforce safety culture.

The role of safety policy, procedures, and processes

The installation of appropriate safety policy, procedures and processes in an organisation can often be an effective means to address inadequate safety performance. Well-designed safety procedures are imperative since the safety risks must be concurrently assessed while associated safety procedures are devised and the required safety measures are implemented. Designing safety action processes must also consider staff preparedness to deal with any kind of mishap (Dyson, 1999). For example, school safety policy must be able to deal with a broad range of violent behaviours, such as provocation, aggression, violence, and discrimination (Timmerman, 2003). According to Netshitahame and Van Vollenhoven (2006), to ensure that behaviour expectations and procedures are clearly communicated, consistently enforced and fairly applied, each school should draw up an explicit school safety policy and provide enforcement procedures that are in line with school policy. Kauffman (1997) and Scott (2001) indicate that the link between poor educational achievement and the types of behaviour that threaten school safety should be considered in school safety policy. A review article by Zohar (2010) illustrates a safety pyramid that shows safety policies and departmental priorities being key enablers to enhanced safety performance in organisations. These aforementioned aspects are of paramount significance because they directly portray the safety in Saudi schools. For example, the influence of policies, procedures, and processes on workforce safety culture has been noted through an increase in safety awareness, a focus on preventing incidents and accidents, and an increase in the safety culture for school users. Additionally, this influence also improves the safety culture of students, teachers, and school executives and increases their safety knowledge to prepare for and address emergency situations. This discussion provides the persuasion for the fourth hypothesis:

H4. Safety policy, procedures and processes influence the workforce safety culture.

The role of workforce safety culture

Workforce safety culture refers to common behaviour patterns within the group of people constituting any organisation. People who demonstrate similar positive beliefs and behaviour patterns toward safety within the school generally reflect a strong workforce safety culture. A good safety culture is generally reflected by workers and managers placing a high degree of importance on safety for all work practices and there is greater propensity to report incidents and accidents. When a critical proportion of employees start to demonstrate good safety practices, a strong workforce safety culture tends to permeate across the entire organisation. The strength of the workforce safety culture is an important aspect in determining safety performance, as is the extent to which the pattern of behaviour aligns with school safety requirements. This situation requires a safety culture that promotes people who are willing to be proactive at trying new approaches for preventing school incidents and accidents. Moreover, a strong culture tends to bind together the people of the school, contributing to smoother functioning and built-in rigidities.

Workforce safety culture is a perspective that can block, welcome, or reshape performance reforms in ways that are consistent with organisational norms (Jennings and Haist, 2004). This view has been supported by both qualitative and quantitative research. Yang and Hsieh (2007) determined that a strong workforce culture supports performance reforms. Julnes and Holzer (2002) and Moynihan and Landuyt (2009) indicate that a mission-oriented culture is positively associated with the success of performance reforms. For instance, a case study by Broadnax and Conway (2001) covering the topic of performance management in the Social Security Administration portrays a leader actively reshaping the organisational workforce culture using newsletters, e-mail, and other communication methods. These researchers found that a regular meeting with field officers and querying them with regard to performance indicators was considered to be the most important approach. In other words, "You don't change culture through memos" (Broadnax and Conway, 2001). Schneider (2004) suggests that the professional norms inherent in a workforce culture influence performance. Zohar and colleagues have completed a number of studies which link safety culture and/or climate factors with safety performance outcomes (e.g. Zohar, 2010; Zohar et al., 2014).

Such findings indicate that fostering safety performance requires a supportive workforce safety culture. There are a number of theoretical reasons to believe that a developmental workforce safety culture is associated with safety performance. In a developmental culture, safety performance is integrated into management decisions as formative rather than summative feedback. This approach encourages school users to learn and improve their school safety performance. As a result, school users are more honest about safety weaknesses and more open to discussions on safety problems and alternative processes (Meyer, 1991; Meyer et al., 1965; Moynihan, 2005), which will improve school safety performance. The above arguments led to the final hypothesis:

H5. The workforce safety culture influences safety performance.

Summary of causal relations hypotheses

In summary, the following five hypotheses resulted from this study (Fig. 1):

- H1: Safety management and leadership influences safety learning and training activities;
- H2: Safety management and leadership influences safety policy, procedures and processes;
- H3: Safety learning and training influences the workforce safety culture;
- H4: Safety policy, procedures, and processes influence the workforce safety culture; and
- H5: The workforce safety culture influences safety performance.

Research method

Research sample

The participants in this study included Saudi teachers, school executives, and Ministry of Education officers. A province-wide survey from the Ministry of Education was conducted among the participants. The sample selection for any empirical study is an important issue, and sufficient variability should be ensured in the sample selection so that statistical analyses will produce justified predictable values. The reliability of the results and theory formation depends heavily on sample selection criteria. Moreover, sample selection should be based on a theoretical paradigm rather than on statistical specifications (Yin, 2003).

In this study, the respondents are Saudi teachers, school executives, and Ministry of Education officers who participated in the survey. The study was conducted in five cities in Saudi Arabia that represent the five main Saudi Arabian regions. For each region, the largest and most populous cities were chosen for the survey as follows: Riyadh, the capital and the largest city in Saudi Arabia; Jeddah, the second largest city in Saudi Arabia and the capital city of the western region; Dammam, the most populous city in the eastern region; Hail, the most populous city in the northern region; and Abha, one of the most populous cities in the southern region.

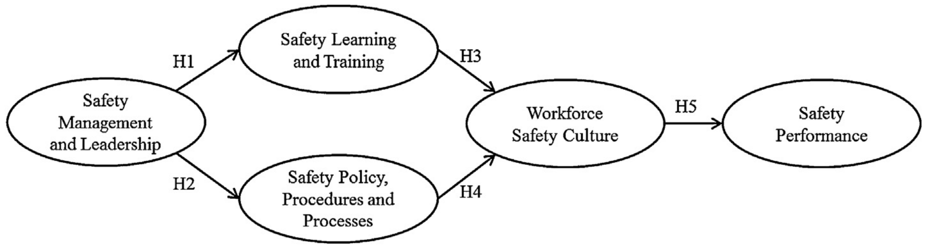

Fig. 1. Safety performance BSC framework hypotheses.

Questionnaire development and survey

Creswell (2003) noted that careful questionnaire design is an important requirement for an effective and justified quantitative study. The survey questions are categorised to ensure that responses are based on past experiences, accurately reflect the participants' perceptions, and increase measurement reliability. Therefore, several aspects were considered when developing the questionnaire. First, the questionnaire was separated into the following two main sections: (1) gathered demographic information about the respondents and their experiences in the Saudi education environment, (2) elicited respondent opinions on the importance of each perspective in the proposed research framework. The respondents' opinions were based on their safety performance experience. Second, five-point scales were used to measure the operationally defined elements of the perspectives within the proposed research framework. Items for each perspective were generated and tested. Five-point Likert scales were applied to measure the developed safety performance BSC items in the Saudi schools using two columns. Column A was concerned with the importance of each item, whereas Column B dealt with the performance of each item, according to the experience of participants. The perspective consists of a set of statements that express the importance and the performance of each item as either favourable or unfavourable attitudes toward the object of the interest.

Statistical techniques

First, descriptive statistical analyses such as frequencies, percentages, and means were conducted regarding the sample profile (Sekaran, 2003). In addition to a visual inspection of frequencies for all perspective items, the data analysis included the measurement of central tendency and dispersion. Following this procedure, exploratory factor analysis (EFA) was conducted to examine the factor structures of the safety BSC perspectives, to reduce redundant items (Neuman, 2006), and to improve the reliability of each perspective. According to Hair et al. (2010), the exploratory analysis procedure is a powerful tool that can address a wide range of theoretical questions. This procedure also defines potential relationships in general form, which then allows the multivariate techniques to estimate relationships (Hair et al., 2010). In other words, by allowing the method and the data to define the nature of the relationships, exploratory analyses can then specify the relationships (Hair et al., 2010).

To test the developed hypotheses, appropriate multivariate statistical techniques were performed, including EFA, PLS confirmatory analysis, and PLS based structural equation modelling (SEM). This study employed an EFA to identify the underlying structure of data for each perspective. Based on the nature of the proposed relationships, the small sample size, and the research objective, the PLS analysis technique was considered an appropriate tool to evaluate the revised theoretical framework. PLS, which was developed by Wold (1980), is a second-generation multivariate analysis technique (Barclay et al., 1995) used to analyse statistical frameworks that involve a set of perspectives and multiple items (Chin, 1998b; Fornell and Bookstein, 1982). In addition, the SEM analysis technique has been applied. According to Chin (1998b) and Hulland (1999), using SEM enables an assessment of the psychometric properties of the measures and facilitates simultaneous tests of the measurement and structural framework (Barclay et al., 1995). SEM is compatible with interval-style data and can assess a framework with a relatively small sample size (Barclay et al., 1995; Chin, 1998a,b; Gefen et al., 2000). SEM is also an appropriate technique when the major concern is predicting the dependent variables (Chin, 1998a,b; Fornell and Bookstein, 1982). In summary, EFA, PLS, and SEM were chosen as statistical techniques in this study. The assessment criteria of each technique is presented and detailed in Table 2.

Data analysis and results

Descriptive statistics

Two-hundred participants from the Saudi educational environment, including teachers (55%), school executives (21%), and Ministry of Education officers (24%), participated in this study to express their perceptions of the proposed items regarding the balanced scorecard. Of the total, 70% were male and 30% were female. Approximately 85% of the participants had been working in an educational environment for at least 3 years, 78% had six to over 20 years of experience, and 22% had less than five years of experience. The majority of participants had from six to over 20 years of experience, whereas the remainder had less than six years including 38 with less than one year of experience. Regarding the participants' regional locations, 69% were in the central Saudi region, 11% in the east, 7% in the west and south, and 6% in the north.

The descriptive statistics provided preliminary findings that evaluated and interpreted the means values of all items in columns A and B. In addition, the data were free from extreme outliers and upheld the assumptions of normality and linearity. To check whether the sample used in this study sufficiently represented the participants, the assessment of the standard deviation and the standard error of the mean value indicated that a mean value could be used as a representative score for each item. Although participants were different (i.e., teachers, school executives, and Ministry of Education officers), the analysis of variance (ANOVA) results confirmed that the data set could be treated as a single sample. Therefore, the data set in the study satisfied the condition required for the three analysis techniques (EFA, PLS, and SEM). EFA was performed using the column A data (the importance of each item) to explore and confirm the structure of the factors for each individual perspective. PLS and SEM were performed on the column B data (the performance of each item) to confirm the EFA results regarding the relationship between perspectives and to test the proposed hypotheses. Correlation coefficients of each perspective and the overall framework were carried out to determine whether all the perspectives represent the whole BSC framework. The results reveal that the correlation coefficients of each perspective (with

Table 2
Assessment criteria used in EFA, PLS and PLS-based SEM.

Validity type	Criterion	Description	Authors
Exploratory factor analysis (EFA)	Bartlett's test	Bartlett's (1954) test of sphericity can be used to determine the factorability of data ($p < 0.001$).	Field (2009); Hair et al. (2010); Tabachnick and Fidell (2007)
	KMO	Kaiser–Mayer–Olkin measure of sample adequacy (KMO) can be used to determine the factorability of data, presenting the minimum acceptable sample (0.60).	Hair et al. (2010); Tabachnick and Fidell (2007)
	Anti-image correlation matrix	Can be used to test the factorability of data. Each matrix should be inspected for correlations above (0.30).	Hair et al. (2010); Pellegrini and Long (2002); Tabachnick and Fidell (2007)
	Cut-off factor loading	Each loading should be above (0.50).	Hair et al. (2010)
	Cronbach's Alpha	The scales for the reliability coefficients (0.70).	Pellegrini and Long (2002)
	Cumulative variance	The total variance extracted by successive items of its perspective (60%).	Hair et al. (2010)
Partial least squares (PLS)	Item loadings	Measures how much of the items are explained by the corresponding factor. Values should be significant at (0.50).	Chin (1998b)
	Cross-loadings	Cross-loadings are obtained by correlating the component scores of each item with all other factors. If the loading of each item is higher for its designated factor than for any of the other factors, it can be inferred that the frameworks of BSC perspectives differ sufficiently from one another.	Chin (1998b)
	Fornell–Larcker criterion	The average explained variance (AVE) of the correlate of each factor must be higher than the correlation between the other factors.	Pepler et al. (2006)
	Cronbach's alpha (CA)	Alpha values ranges from 0 (completely unreliable) to 1 (perfectly reliable). Proposed threshold value for confirmative (explorative) research: CA > (0.80).	Cronbach (1951); Nunnally and Bernstein (1994)
	Composite reliability (CR)	CR values between 0 (completely unreliable) and 1 (perfectly reliable). As an alternative to Cronbach's Alpha, it allows items to not be equally weighted. Proposed threshold value for confirmative (explorative) research: CA > (0.80).	Ando et al. (2005); Nunnally and Bernstein (1994)
	Average variance extracted (AVE)	Attempts to measure the amount of variance that BSC perspective captures from its items relative to the amount due to measurement error. Proposed threshold value: AVE > (0.50).	Pepler et al. (2006)
Structural equation modelling (SEM)	Coefficient of determination (R^2)	Attempts to measure the explained variance of a perspective relative to its total variance. Values of approximately 0.670 are considered *substantial*, values around 0.333 *moderate*, and values around 0.190 *weak*.	Chin (1998b); Ringle (2004)
	Path coefficients	Path coefficients between the perspectives should be analysed in terms of their *algebraic sign*, *magnitude*, and *significance*.	Huber et al. (2007)
	Effect size (f^2)	Measures if perspective has a substantial impact on its factors. Values of 0.02, 0.15, 0.35 indicate the predictor factor's *low*, *medium*, or *large* effect in the structural framework.	Chin (1998b); Cohen (1988); Ringle (2004)
	Predictive relevance (Q^2)	The Q^2 statistic is a measure of the predictive relevance of a block of manifest perspectives. A tested model has more predictive relevance the higher Q^2 is, and modifications to a model may be evaluated by comparing the Q^2 values. The proposed threshold value is $Q^2 > 0$. The predictive relevance's relative impact can be assessed by means of the measure Q^2.	Fornell and Cha (1994); Geisser (1975); Stone (1974)
	Goodness-of-Fit (GoF)	Defined as the geometric mean of the average communality and average R^2 (for endogenous perspectives) GoF ($0 < \text{GoF} < 1$), GoF = 0.1 is small, GoF = 0.25 is medium, and GoF = 0.36 is large.	Fornell and Larcker (1981); Tenenhaus et al. (2005); Wetzels et al. (2009)

Table 3
Correlation coefficients of each perspective and the overall framework.

Perspective	Number of Items	Column A		Column B	
		Pearson correlation coefficients of each perspective with perspective-total for the framework	Cronbach's Alpha (α)	Pearson correlation coefficients of each perspective with perspective-total for the framework	Cronbach's alpha (α)
Safety management and leadership	17	0.636**	0.863	0.627**	0.859
Safety learning and training	16	0.753**	0.885	0.651**	0.866
Safety policy, procedures and processes	19	0.760**	0.929	0.536**	0.913
Workforce safety culture	16	0.755**	0.898	0.722**	0.894
Safety performance	11	0.579**	0.894	0.553**	0.821

** Correlation is significant at the 0.01 level (2-tailed).
Total Cronbach's alpha: 0.967.

Table 4
EFA results for all perspectives.

Perspective	Cumulative variance	Cronbach's alpha	KMO	Bartlett's Test	Total correlations between items	Items deleted
SML (5F, 16I)	70.889%	0.854	0.736	$p < 0.001$	0.319 to 0.686	A3.1
SLT (3F, 14I)	64.995%	0.872	0.812	$p < 0.001$	0.301 to 0.785	B2.3 and B4.1
SPPP (3F, 15I)	69.416%	0.917	0.870	$p < 0.001$	0.384 to 0.804	C1.4, C2.1, C3.3, and C3.4
WSC (3F, 13I)	66.037%	0.886	0.823	$p < 0.001$	0.324 to 0.706	D1.4, D2.3, and D4.3
SP (3F, 8I)	78.906%	0.854	0.835	$p < 0.001$	0.445 to 0.817	E1.3, E2.3, and E4.1

items in each perspective aggregated) with those of all perspectives combined as a single framework were statistically significant at a level of 0.01 (Table 3). These results indicate a high degree of internal consistency among the perspectives and confirm the internal correlation among all perspectives in the framework, thus confirming the internal validity of the perspectives used in this study. Additionally, the reliability was significant through Cronbach's alpha (0.967) (Pellegrini and Long, 2002).

Scale reliability analysis

As discussed earlier, column A data (the importance of each item) was using to measure the five proposed perspectives in the conceptual framework (SML, SLT, SPPP, WSC, and SP), five independent scales were used in the survey questionnaire. The scale reliability analysis was performed to assess the internal consistency of these perspectives. Table 3 presents the Cronbach's alphas for the measurement scales of the perspectives in column A. The values of the alpha coefficient of the five scales range from 0.863 to 0.929, which is much higher than the low value limit of 0.7, indicating very strong consistency among the scales. The high alpha value assumes that the scores of all items with one perspective have the same range and meaning (Cronbach, 1951).

These results support the use of exploratory factor analysis to determine whether any items should be removed and to accurately capture the meaning of the framework perspectives (Gerbing and Anderson, 1988; Hair et al., 2010; Koufteros, 1999). Because the main objective of this technique is to group the items into a set of latent dimensions rather than to condense individual respondents into distinct groups (factors), an *R factor analysis* was employed. This type of factor analysis is used to examine the relationships among variables to identify groups of variables that form latent dimensions or factors (Hair et al., 2010). Each perspective was processed to determine whether the items in each perspective could be represented by their factors. Additionally, to ensure that the items in each factor had practical significance, cut-off loading of 0.50 was used (Hair et al., 2010).

Confirming model architecture: exploratory factor analysis

Based on the previous results, EFA was performed separately for each perspective using a principle axis extraction method with Varimax rotation. The results of the EFA for each perspective are presented in Table 4. For the SML perspective, five factors were identified and explained 70.889% of the total variance. For the SLT perspective, the three identified factors explained 64.995% of the total variance. The SPPP perspective's three factors explained 69.416% of the total variance, whereas the WSC perspective's three factors explained 66.037% of the total variance. For the SP perspective, the three factors explained 78.906% of the total variance. Thirteen items were removed because the factor loading for these items was less than 0.50. These items included the following: (A3.1), (B2.3), (B4.1), (C1.4), (C2.1), (C3.3), (C3.4), (D1.4), (D2.3), (D4.3), (E1.3), (E2.3), and (E4.1). The other items had factor loadings that were greater than 0.50.

The Cronbach's alpha for each perspective was greater than 0.70, indicating a good reliability of scale. To confirm the significance of the factor loading results, an evaluation of the correlation matrix was conducted using the Kaiser–Nayer–Olkin (KMO) and Bartlett's test. For each perspective, the results show that the KMO was greater than 0.70, which is acceptable, and the Bartlett's test was $p < 0.001$, a significant probability level. Finally, when assessing the factorability of the data, all remaining items fell within the acceptable range for a factor analysis of 0.3; although there were high correlations between items and other factors, the data still must have more than 0.50 factors loading.

Validating model architecture: partial least squares method

As previously described, column B was concerned with the performance of each item, according to the experience of the

Table 5
Cronbach's alphas for the perspective performance measurement scales.

Perspective	Number of items (refined)	Cronbach's alpha (α) Column B
Safety management and leadership (SML)	16	0.845
Safety learning and training (SLT)	14	0.854
Safety policy, procedures, and processes (SPPP)	15	0.889
Workforce safety culture (WSC)	13	0.844
Safety performance (SP)	8	0.807

Table 6
Cross-loading framework at the item level.

Item	SML1	SML2	SML3	SML4	SML5	SLT1	SLT2	SLT3	SPPP1	SPPP2	SPPP3	WSC1	WSC2	WSC3	SP1	SP2	SP3
SML11	0.74	0.64	0.20	0.25	0.13	0.54	0.34	0.22	0.07	0.13	−0.03	0.16	0.16	0.09	0.12	0.13	0.30
SML12	0.79	0.73	0.21	0.20	0.20	0.37	0.12	0.14	0.08	0.16	0.06	0.10	0.11	0.10	0.17	0.07	0.16
SML13	0.71	0.62	0.19	0.25	0.22	0.41	0.15	0.17	0.00	0.15	0.01	0.11	0.09	0.09	0.12	0.09	0.19
SML14	0.74	0.53	0.59	0.30	0.31	0.45	0.39	0.25	0.17	0.39	0.06	0.25	0.35	0.30	0.10	0.09	0.01
SML15	0.49	0.29	0.37	0.37	0.22	0.32	0.39	0.15	0.24	0.36	0.16	0.33	0.42	0.24	0.10	0.21	0.17
SML21	0.65	0.78	0.14	0.15	0.00	0.48	0.29	0.18	0.20	0.05	−0.04	0.12	0.02	0.13	0.04	0.10	0.22
SML22	0.59	0.78	0.23	0.11	0.06	0.37	0.25	0.06	−0.01	0.12	−0.11	0.17	0.10	0.00	0.06	0.09	0.19
SML23	0.64	0.77	0.31	0.19	0.28	0.49	0.25	0.16	0.08	0.15	0.08	0.16	0.13	0.15	0.14	0.21	0.25
SNL31	0.38	0.28	0.72	0.32	0.19	0.07	0.04	0.03	0.19	0.25	0.01	0.16	0.20	0.27	0.02	0.04	0.01
SML32	0.51	0.36	0.78	0.35	0.18	0.38	0.28	0.12	0.14	0.12	0.01	0.11	0.20	0.28	0.08	−0.01	−0.05
SML33	−0.05	−0.06	0.51	0.32	0.33	0.11	0.12	0.10	0.44	0.26	0.27	0.20	0.30	0.32	0.25	0.27	0.09
SML34	0.23	0.11	0.72	0.37	0.64	0.30	0.25	0.24	0.28	0.40	0.29	0.23	0.41	0.28	0.13	0.16	0.10
SML41	0.35	0.16	0.39	0.86	0.47	0.41	0.30	0.08	0.21	0.38	0.29	0.19	0.43	0.20	0.13	0.24	0.08
SML42	0.30	0.17	0.42	0.82	0.29	0.45	0.32	0.22	0.34	0.30	0.21	0.19	0.35	0.26	0.22	0.24	0.17
SML51	0.11	−0.07	0.42	0.42	0.84	0.27	0.23	0.13	0.21	0.39	0.33	0.05	0.39	0.22	0.16	0.18	0.06
SML52	0.39	0.29	0.41	0.41	0.93	0.41	0.30	0.13	0.15	0.39	0.20	0.14	0.36	0.23	0.22	0.29	0.20
SLT11	0.52	0.49	0.34	0.40	0.29	0.85	0.59	0.26	0.29	0.25	0.12	0.32	0.33	0.32	0.16	0.32	0.27
SLT12	0.56	0.56	0.36	0.44	0.28	0.81	0.53	0.23	0.29	0.20	0.18	0.33	0.29	0.32	0.14	0.33	0.27
SLT13	0.52	0.51	0.19	0.34	0.30	0.80	0.52	0.21	0.16	0.28	0.18	0.29	0.33	0.18	0.10	0.31	0.26
SLT14	0.52	0.53	0.17	0.43	0.44	0.80	0.40	0.18	0.11	0.20	0.13	0.25	0.28	0.16	0.07	0.23	0.28
SLT15	0.34	0.29	0.26	0.40	0.30	0.70	0.55	0.27	0.26	0.32	0.18	0.41	0.50	0.40	0.25	0.36	0.26
SLT16	0.37	0.33	0.23	0.40	0.29	0.74	0.49	0.16	0.15	0.24	0.14	0.31	0.41	0.22	0.18	0.25	0.33
SLT21	0.24	0.19	0.20	0.27	0.20	0.52	0.76	0.19	0.14	0.26	0.00	0.30	0.34	0.13	0.22	0.31	0.11
SLT22	0.20	0.22	0.22	0.22	0.16	0.43	0.64	0.27	0.19	0.31	0.10	0.22	0.27	0.15	0.05	0.18	0.06
SLT23	0.43	0.39	0.21	0.25	0.28	0.54	0.80	0.25	0.01	0.24	0.02	0.31	0.30	0.18	0.17	0.22	0.17
SLT24	0.27	0.17	0.12	0.32	0.22	0.37	0.66	0.66	−0.02	0.17	−0.05	0.27	0.31	0.17	0.31	0.21	0.16
SLT31	0.25	0.18	0.13	0.14	0.17	0.27	0.44	0.97	0.03	0.25	−0.01	0.20	0.30	0.08	0.24	0.17	0.28
SLT32	0.27	0.16	0.21	0.20	0.11	0.28	0.47	0.97	0.18	0.31	0.02	0.23	0.33	0.15	0.19	0.14	0.15
SPPP11	0.17	0.12	0.25	0.24	0.12	0.15	0.10	0.14	0.90	0.54	0.50	0.41	0.37	0.55	0.23	0.45	0.16
SPPP12	0.13	0.08	0.24	0.16	0.00	0.10	0.17	0.20	0.75	0.50	0.43	0.40	0.27	0.52	0.13	0.32	0.01
SPPP13	0.12	0.10	0.30	0.29	0.13	0.24	0.02	0.06	0.74	0.31	0.40	0.37	0.36	0.55	0.36	0.53	0.37
SPPP14	0.02	0.01	0.20	0.16	0.09	0.13	−0.01	0.10	0.74	0.33	0.43	0.28	0.19	0.44	0.12	0.26	0.10
SPPP15	0.29	0.20	0.40	0.24	0.25	0.27	0.10	0.00	0.54	0.46	0.24	0.26	0.31	0.36	0.08	0.23	0.06
SPPP16	0.11	0.09	0.23	0.21	0.19	0.25	0.16	−0.02	0.81	0.33	0.43	0.38	0.29	0.52	0.18	0.41	0.19
SPPP17	0.00	−0.02	0.14	0.35	0.25	0.23	−0.04	−0.05	0.71	0.29	0.47	0.33	0.31	0.45	0.19	0.42	0.23
SPPP18	0.10	0.09	0.22	0.30	0.15	0.30	0.12	0.20	0.72	0.31	0.50	0.37	0.33	0.40	0.14	0.36	0.19
SPPP21	0.20	0.10	0.22	0.17	0.26	0.19	0.11	0.03	0.46	0.69	0.30	0.15	0.32	0.29	0.11	0.26	0.02
SPPP22	0.25	0.10	0.16	0.28	0.33	0.21	0.26	0.22	0.24	0.74	0.26	0.24	0.38	0.18	0.27	0.38	0.21
SPPP23	0.38	0.18	0.38	0.43	0.42	0.28	0.31	0.32	0.44	0.84	0.33	0.30	0.49	0.31	0.19	0.36	0.12
SPPP24	0.29	0.13	0.32	0.39	0.38	0.27	0.32	0.32	0.37	0.83	0.45	0.31	0.58	0.27	0.28	0.37	0.18
SPPP25	0.14	−0.01	0.21	0.22	0.23	0.22	0.27	0.17	0.41	0.64	0.33	0.21	0.36	0.42	0.14	0.25	0.05
SPPP31	−0.01	−0.03	0.17	0.17	0.12	0.12	0.03	−0.05	0.60	0.33	0.89	0.33	0.28	0.43	0.01	0.25	0.06
SPPP32	0.18	0.01	0.07	0.35	0.39	0.21	−0.01	0.10	0.19	0.39	0.59	0.20	0.39	0.28	0.19	0.29	0.26
WSC11	0.23	0.10	0.27	0.25	0.15	0.32	0.25	0.19	0.46	0.31	0.39	0.82	0.59	0.42	0.25	0.52	0.33
WSC12	0.25	0.20	0.25	0.21	0.15	0.34	0.34	0.16	0.40	0.27	0.30	0.86	0.49	0.41	0.30	0.57	0.37
WSC13	0.26	0.22	0.19	0.15	0.16	0.42	0.47	0.32	0.17	0.31	0.10	0.54	0.48	0.30	0.05	0.33	0.25
WSC14	0.06	0.08	−0.05	0.00	−0.13	0.15	0.10	0.01	0.25	0.01	0.19	0.55	0.16	0.26	0.17	0.35	0.37
WSC15	0.00	0.04	−0.02	0.02	−0.07	0.01	0.05	−0.02	0.25	0.08	0.16	0.54	0.15	0.22	0.37	0.39	0.28
WSC21	0.19	0.11	0.27	0.36	0.19	0.39	0.39	0.28	0.37	0.26	0.26	0.65	0.67	0.44	0.29	0.52	0.35
WSC22	0.27	0.02	0.43	0.42	0.37	0.34	0.25	0.30	0.40	0.48	0.36	0.41	0.79	0.41	0.31	0.39	0.39
WSC23	0.35	0.19	0.21	0.42	0.36	0.40	0.30	0.14	0.21	0.40	0.35	0.54	0.78	0.28	0.33	0.39	0.43
WSC24	0.15	−0.01	0.24	0.25	0.28	0.24	0.32	0.30	0.25	0.53	0.37	0.25	0.73	0.31	0.21	0.31	0.11
WSC25	0.24	0.10	0.27	0.28	0.37	0.31	0.34	0.24	0.30	0.52	0.24	0.41	0.80	0.41	0.30	0.47	0.20
WSC31	0.12	0.11	0.24	0.03	0.08	0.15	0.12	0.01	0.50	0.16	0.29	0.38	0.21	0.75	0.19	0.30	0.16
WSC32	0.16	0.12	0.30	0.26	0.12	0.28	0.15	−0.01	0.59	0.21	0.31	0.47	0.35	0.80	0.26	0.45	0.33
WSC33	0.24	0.05	0.33	0.28	0.35	0.30	0.21	0.27	0.32	0.49	0.44	0.23	0.50	0.64	0.19	0.33	0.16
SP11	0.11	0.01	0.13	0.23	0.22	0.16	0.24	0.24	0.19	0.25	−0.01	0.17	0.34	0.22	0.82	0.48	0.34
SP12	0.20	0.13	0.19	0.16	0.20	0.17	0.29	0.23	0.21	0.24	0.00	0.27	0.37	0.28	0.79	0.39	0.36
SP13	0.10	0.11	0.01	0.08	0.09	0.11	0.08	0.05	0.15	0.11	0.23	0.30	0.18	0.16	0.68	0.50	0.56
SP21	0.28	0.16	0.26	0.41	0.42	0.39	0.32	0.12	0.35	0.44	0.22	0.47	0.60	0.40	0.58	0.79	0.46
SP22	0.14	0.25	−0.02	0.12	0.06	0.35	0.25	0.07	0.41	0.17	0.24	0.57	0.29	0.37	0.32	0.78	0.57
SP23	−0.13	−0.07	−0.01	0.03	0.07	0.03	0.09	0.17	0.34	0.31	0.28	0.34	0.28	0.30	0.39	0.58	0.37
SP31	0.34	0.32	0.10	0.18	0.26	0.47	0.32	0.34	0.12	0.26	0.14	0.36	0.42	0.29	0.37	0.47	0.79
SP32	0.06	0.17	−0.03	0.07	0.03	0.15	0.00	0.05	0.23	0.02	0.14	0.40	0.26	0.22	0.54	0.60	0.86

participants. Moreover, the Cronbach's alphas for the measurement scales of the perspectives in column B, after deleting previous items, are shown in Table 5. The values of the alpha coefficient for all five scales ranged from 0.807 to 0.889, and were much higher than the low value limit of 0.7, which indicated a very good consistency of the scales.

In conclusion, the developed scales comprised reliable and valid items that adequately captured the meaning of the framework perspectives and their related factors. These scales were used in the further analyses (i.e. PLS and SEM) by confirming the finding of the framework perspectives (this section), and by identifying the relationships between the framework perspectives and the framework validation hypotheses testing (Section 6.5). Based on an assessment of Table 2, the PLS algorithm can be run to calculate the framework parameters' estimates. Wold has developed PLS as a general method for the estimation of path frameworks involving

Table 7
Correlation among perspective scores (square root of AVE in diagonal).

Item	SML1	SML2	SML3	SML4	SML5	SLT1	SLT2	SLT3	SPPP1	SPPP2	SPPP3	WSC1	WSC2	WSC3	SP1	SP2	SP3
SML1	0.70																
SML2	0.81	0.77															
SML3	0.46	0.30	0.69														
SML4	0.39	0.20	0.48	0.84													
SML5	0.31	0.16	0.47	0.46	0.89												
SLT1	0.60	0.58	0.33	0.51	0.40	0.79											
SLT2	0.40	0.34	0.26	0.37	0.30	0.66	0.72										
SLT3	0.27	0.17	0.18	0.18	0.14	0.28	0.47	0.97									
SPPP1	0.16	0.11	0.33	0.32	0.19	0.27	0.11	0.11	0.74								
SPPP2	0.34	0.14	0.35	0.40	0.44	0.32	0.34	0.29	0.52	0.75							
SPPP3	0.07	−0.02	0.17	0.30	0.28	0.20	0.02	0.01	0.58	0.45	0.76						
WSC1	0.27	0.20	0.24	0.23	0.12	0.40	0.39	0.22	0.47	0.33	0.36	0.68					
WSC2	0.32	0.11	0.38	0.46	0.42	0.45	0.43	0.33	0.41	0.58	0.42	0.61	0.75				
WSC3	0.24	0.13	0.40	0.27	0.25	0.34	0.22	0.12	0.64	0.40	0.48	0.49	0.49	0.73			
SP1	0.18	0.11	0.14	0.21	0.22	0.19	0.26	0.22	0.24	0.26	0.10	0.32	0.39	0.29	0.76		
SP2	0.17	0.18	0.12	0.28	0.28	0.38	0.33	0.16	0.50	0.43	0.34	0.64	0.56	0.50	0.61	0.72	
SP3	0.23	0.29	0.04	0.15	0.16	0.35	0.17	0.22	0.22	0.15	0.17	0.46	0.40	0.30	0.55	0.65	0.83

latent perspectives indirectly measured by multiple items (Wold, 1966, 1979, 1980, 2006). The assessment of the measurement frameworks is considered a systematic application of the reflective criteria. In this study, cross-loading was higher than 0.05 (Chin, 1998b) except for two items that were less than 0.50 and therefore omitted (Table 6).

The second measure is the Fornell–Larcker criterion (Fornell and Larcker, 1981), which requires that the average explained variance (AVE) of the correlate of each factor must be higher than the correlation between the other factors (Table 7). The next measure is Cronbach's alpha (CA), the traditional criterion for assessing reliability, wherein a high alpha value assumes that the scores of all items with one perspective have the same range and meaning (Cronbach, 1951). The composite reliability (CR) is an alternative measure to Cronbach's alpha (Werts et al., 1974). Because CR overcomes some of CA's deficiencies, Chin (1998b) recommends CR as a measure. In this study, CR values for all perspectives above 0.80 indicate an acceptable range (Nunnally and Bernstein, 1994). The last measure is convergent validity, which involves the degree to which individual items reflecting a perspective converge in comparison to items measuring different perspectives. The AVE proposed by Fornell and Larcker (1981) is considered a commonly applied criterion of convergent validity. Urbach and Ahlemann (2010) suggest that to demonstrate sufficient convergent validity, an AVE value of at least 0.50 indicates that the perspective is on average able to explain more than half of the variance of its items. The AVE value for all perspectives in this study are greater than 0.50 (Table 8).

Structural equation modelling using PLS: hypotheses testing

After the measurement of the safety performance BSC has been successfully validated through EFA and PLS, the structural framework can be analysed using SEM to determine the significance of the paths and the predictive power of the framework. This statistical technique was selected for the treatment of the previously explained EFA and PLS results. The technique includes structural equation modelling, along with a focus on the special form of SEM known as partial least squares.

According to Wetzels et al. (2009), PLS path modelling can be used to estimate the parameters in a reflective, hierarchical framework perspective. The second order in the conceptual framework perspectives also comprise several first-order factors (Wetzels et al., 2009). To illustrate the assessment of a reflective, third-order hierarchical framework, PLS has been used through conceptualisation (Mathwick et al., 2001). Fig. 2 presents the conceptual safety performance BSC orders and also shows the loading of each factor to its perspective. The loading for each factor was higher than 0.05 (Chin, 1998b), which was acceptable.

The different criteria for assessing a PLS framework on the structural level are summarised in Table 2. The first essential criterion for the assessment of the PLS structural equation model is each endogenous perspective's coefficient of determination (R^2). R^2 measures the relationship of safety performance BSC perspectives, which explains the perspective with regard to its total factors. To obtain a minimum level of explanatory power, the values should be sufficiently high for the framework. A value of approximately 0.670 is considered substantial, whereas a value of approximately 0.333 is average and a value of 0.190 and lower is weak (Chin, 1998b) (see Table 10).

The next step in the structural framework assessment is the evaluation of the path coefficients between the framework's perspectives. Because PLS makes no distribution assumptions, to test the effects and statistical significance of the path coefficients, the bootstrapping re-sampling technique was used (Macmillan et al., 2005; Ribbink et al., 2004; White et al., 2003). Five-hundred random samples from the original data set were generated to conduct the bootstrapping process (Macmillan et al., 2005; Wixom and Watson, 2001). The path coefficient's algebraic signs, magnitude, and significance were checked. Contrary to the theoretically assumed relationship, the signs' paths do not support the prepostulated hypotheses (Wetzels et al., 2009). Wetzels et al. (2009) also indicate

Table 8
AVE, CR, and CA values.

Perspective	AVE	Composite reliability (CR)	Cronbach's alpha (CA)
SML	0.51	0.84	0.85
SLT	0.65	0.84	0.88
SPPP	0.67	0.86	0.89
WSC	0.66	0.85	0.85
SP3	0.73	0.89	0.81

Table 9
Critical *z*-values for 1-tail and 2-tail tests.

Sig. level (p)	Symbol	Critical z-values	
		1-tail test	2-tail test
0.001	****	3.090	3.290
0.010	***	2.326	2.576
0.050	**	1.645	1.960
0.100	*	1.282	1.645
Not significant	n.s	–	–

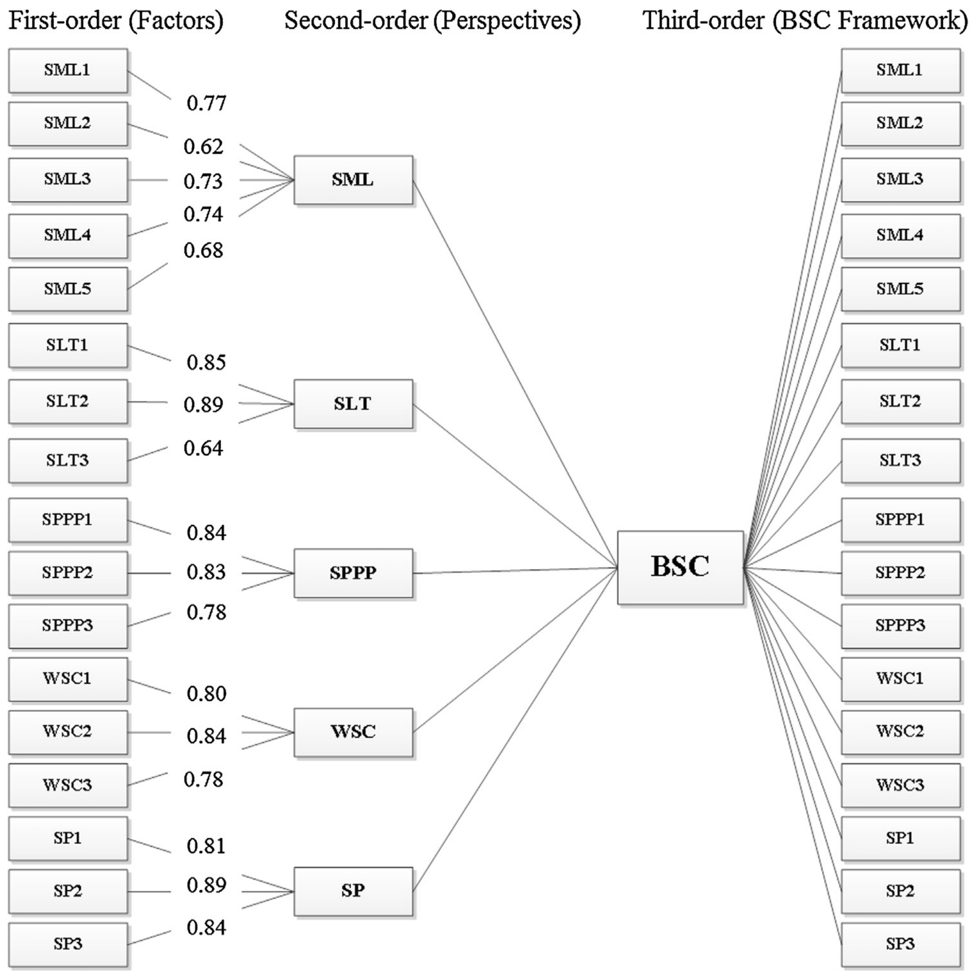

Fig. 2. Conceptual safety performance BSC framework hierarchy.

that the strength of the relationship between perspectives can be measured by the path coefficient's magnitude. Several authors, e.g., Huber et al. (2007), suggest that path coefficients should exceed 0.100 to account for a certain impact within the framework. Additionally, at least at the 0.050 level, path coefficients should be significant. Re-sampling techniques such as bootstrapping (Efron, 1979; Efron and Tibshirani, 1994) have been used in this study to determine the significance. Moreover, according to (Chin, 1998a,b), the bootstrap critical ratios determine the stability of the estimates, and the ranges between −1.96 and +1.96 were acceptable. Alternately, the average variance accounted for (AVA) represented the average of R^2 of the structural framework and indicated the overall predictive power of the framework (Fornell and Bookstein, 1982). In this study, the AVA for the endogenous variables was 0.38, and the R^2 values for the predicted variables were all greater than Falk and Miller (1992) recommended level of 0.10; therefore, it was appropriate to examine the significance of the paths associated with these variables. Falk and Miller (1992) recommend a level of 0.015, and in this study, all of the paths were above the recommended level; additionally, all variables had bootstrap critical ratios above the acceptable level (greater than 1.96, $p < 0.05$).

Based on the observed p values, Table 9 contains the extracts from statistical tables that are used for specifying the significance level. The table also contains the critical p values for rejecting the unacceptable hypotheses at different levels of significance and for 1-tail and 2-tail tests. However, Venaik (1999) suggests that the observed p values are comparable to the critical p values in the table, given that when the sample size is large, the observed p values can be approximated as p values. Furthermore, Sørensen et al. (2004) recommend the use of 2-tailed to obtain a clearly directional hypothesis.

An additional important step, the goodness-of-fit (GoF) test, has been applied to the SEM assessment. Tenenhaus et al. (2005) define this test as the geometric mean of the average communality and

Table 10
PLS results for hypotheses testing.

Dependent (predicted) variables	Independent (predictor) variables	Hypotheses	Path coefficient (β)	Critical ratio (CR)	Sig. level (p)	Tails	R^2	AVE	GoF
SLT	SML	H1	0.618	6.472	****	2-tail	0.38	0.65	0.50
SPPP	SML	H2	0.443	3.677	****	2-tail	0.20	0.67	0.37
WSC	SLT	H3	0.319	3.709	****	2-tail	0.53	0.66	0.59
	SPPP	H4	0.566	7.773	****	2-tail			
SP	WSC	H5	0.636	8.526	****	2-tail	0.40	0.73	0.54
AVE							0.38		

Table 11
Effect size for the structural relationships.

Factor	Perspective	R^2		f^2	Effect size
		Included	Excluded		
SML 1	SML	0.99943	0.96226	>0.35	Large
SML 2		0.99943	0.98381		
SML 3		0.99943	0.96198		
SML 4		0.99943	0.97773		
SML 5		0.99943	0.97809		
SLT 1	SLT	0.99997	0.80922	>0.35	Large
SLT 2		0.99997	0.95791		
SLT 3		0.99997	0.98140		
SPPP1	SPPP	0.99974	0.81779	>0.35	Large
SPPP2		0.99974	0.92887		
SPPP3		0.99974	0.99108		
WSC1	WSC	0.99990	0.93732	>0.35	Large
WSC2		0.99990	0.87337		
WSC3		0.99990	0.96735		
SP1	SP	0.99923	0.93041	>0.35	Large
SP2		0.99923	0.92880		
SP3		0.99923	0.95705		

average R^2 for endogenous perspectives. These researchers suggest the global fit for PLS path modelling of more than 0 and less than 1 ($0 < \text{GoF} < 1$). Because communality equals AVE in the PLS path modelling approach, Wetzels et al. (2009) propose a cut-off value of 0.5 for communality, as suggested by Fornell and Larcker (1981). As Cohen (1988) propose in line with the effect sizes for R^2 (small: 0.02, medium: 0.13, and large: 0.26), the following GoF criteria for small, medium, and large effect sizes of R^2 are obtained by Wetzels et al. (2009) by substituting the minimum average AVE of 0.50 and the effect sizes for R^2 in the following equation: $\text{GoF} = \sqrt{\text{AVE} \times R^2}$. GoF = 0.1 is considered as small, GoF = 0.25 as medium, and GoF = 0.36 as large. For validating the PLS model globally, these scales may serve as baseline values (Wetzels et al., 2009). Table 10 details the hypotheses of the study, the path coefficient between the exogenous and endogenous perspectives, the average variance extracted, the average account for R^2, and bootstrap critical ratios, which were acceptable, and the GoF, which was large.

An additional assessment of the structural framework involves the effect size f^2. Cohen's f^2 was computed to assess the effect of each set of predictors while controlling for all other factors in the framework; f^2 of 0.02, 0.15, and 0.35 reflect small, medium, and large effect sizes, respectively (Cohen, 1988). The value of f^2 can be calculated according to Eq. 1.

$$f^2 = \frac{R^2(included) - R^2(excluded)}{1 - R^2(included)} \quad (1)$$

In the above, R^2_{included} and R^2_{excluded} are the R-squares that provide the dependent perspective when the predictor factor is used or omitted in the structural equation, respectively. Table 11 shows that there was a large effect on size between each factor and its perspective.

Finally, an assessment of the structural framework involves the framework's capability to predict. The predominant measure of predictive relevance is Stone–Geisser's Q^2 (Geisser, 1975; Stone, 1974), which postulates that the framework must be able to adequately predict each endogenous latent perspective's items. The Q^2 value is obtained by using a blindfolding procedure, a sample reuse technique that omits every dth data point and uses the resulting estimates to predict the omitted portion (Chin, 2010). Notably, the omission distance d must be chosen so that the number of valid observations divided by d is not an integer. Hair et al. (2011) indicate that the omission distance d values commonly range from 5 to 10.

The blindfolding procedure is only applied to endogenous latent constructs that have a reflective measurement framework specification. Q^2 comes in the following two forms: the cross-validated redundancy and communality. Hair et al. (2011) recommend using the cross-validated redundancy, which unlike the cross-validated communality, uses the PLS-SEM estimates of both the structural framework and the measurement frameworks for data prediction and, thereby, perfectly fits the PLS-SEM approach. Moreover, Chin (2010) suggests that $Q^2 > 0$ implies that the framework has predictive relevance, whereas $Q^2 < 0$ implies a lack of predictive relevance. Table 12 shows that the values of cross-validated redundancy represented by Q^2, which can be obtained from the SmartPLS output, were found to range from 0.298787 to 0.436045. The results indicated that SML, SLT, SPPP, WSC, and SP have predictive relevance to BSC. In sum, the results signified the research framework has a good predictive relevance.

Table 12
Blindfolding results.

Perspective	Omission distance (d) = 5		Omission distance (d) = 10	
	Cross-validated communality Q^2	Cross-validated redundancy Q^2	Cross-validated communality Q^2	Cross-validated redundancy Q^2
SML	0.311259	0.299755	0.311108	0.298787
SLT	0.436045	0.427943	0.436036	0.431652
SPPP	0.404610	0.391293	0.404565	0.395948
WSC	0.363917	0.351628	0.363890	0.355287
SP	0.428356	0.412933	0.428318	0.421140

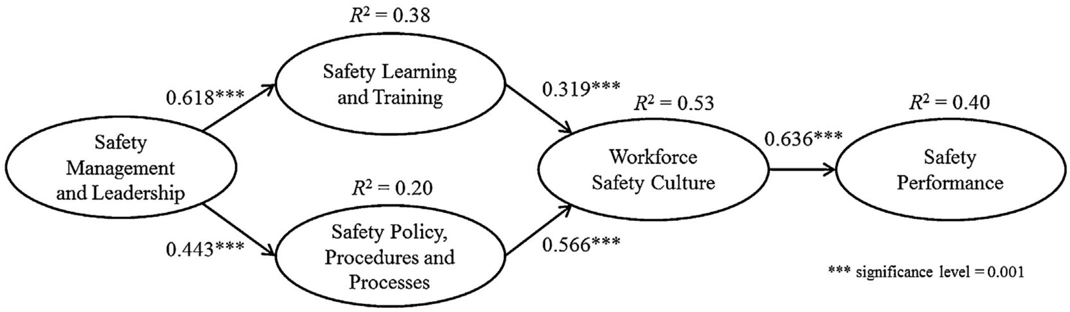

Fig. 3. Causal relationship between school safety BSC perspectives.

Table 13
Hypotheses testing results summary.

No.	Hypotheses	Results
H1	Safety management and leadership has significant effect on SLT	Supported
H2	Safety management and leadership has significant effect on SPPP	Supported
H3	Safety learning and training has significant effect on WSC	Supported
H4	Safety policy, procedures, and processes has significant effect on WSC	Supported
H5	Workforce safety culture has significant effect on safety performance	Supported

Hypotheses testing summary

The SEM analysis results of the proposed framework provide support for the five hypotheses for the inner framework. Table 13 presents the results of the hypotheses testing. The proposed framework, with an illustration of the path coefficients within the inner framework and R squared values for the endogenous perspectives, is presented in Fig. 3.

Conclusions

In the past two decades, several research studies have attempted to examine and/or create a framework for school safety performance processes. However, none of these existing frameworks can be immediately adopted to explain the causal safety system performance process in Saudi schools. There also were numerous applications of the balanced scorecard in measuring safety performance, although none of these applications can be used to examine school safety in Saudi Arabia. In an attempt to fill this gap in knowledge, this study developed a balanced scorecard framework for safety performance in Saudi schools that captures all of the relevant perspectives that influence the effectiveness of the safety performance process. Additionally, the developed framework provides evidence that the knowledge level of Saudi teachers, school executives, and Ministry of Education officers has improved after implementing safety performance BSC initiatives.

This paper details the results of the development and measurement scale analysis through the conceptual safety performance BSC framework, the assessment of scale reliability, and an exploratory factor analysis of the survey data. The study also confirms the structure of the perspectives and their causal relationship using PLS-based structural equation modelling.

The measurement scale analysis was conducted to identify the structure among the set of measurement items for each perspective in the framework and also to reduce the data. The conceptual definition, dimensionality, reliability, and face validity for these scales were established through EFA. In two sequential stages, the PLS framework was then analysed and interpreted. In stage one, the reliability and validity of the measurement framework was presented, and in stage two, an examination of the structural framework was undertaken by interpreting the path coefficients and identifying the adequacy of the research framework. Importantly, the results of the measurement framework indicated that all perspective measures were reliable and valid. Furthermore, the results of the structural framework analysis indicated that all research hypotheses were supported.

Traditional safety management practices in many Schools, particularly in developing and newly industrialised countries such as Saudi Arabia, are often solely focused on incident and accident orientated performance indicators. This approach to safety management has meant that safety performance improvement has been slow. Purely focusing on current safety performance outcomes often meant that the enabling indicators of future safety performance such as safety leadership, training, policies, procedures and culture, were not adequately addressed. This study provides empirical evidence of the causal relationships between the leading and lagging perspectives of safety performance. Specifically, it shows that the Saudi Ministry of Education needs to examine, monitor and evaluate the entire safety system within their burgeoning number of schools. By considering all the perspectives of the developed safety BSC system developed herein, and through an understanding of their causal relationships, they will be able to map out a streamlined path to achieving the safety goals of the Ministry and country. Taking a systems view of safety for public schools in Saudi Arabia will undoubtedly achieve these desired goals in a shorter time.

Study limitations

Despite the large number of studies addressing the concept of safety, little research has focused on the safety performance BSC process in schools. Furthermore, to the best of the researcher's knowledge, such research investigations have not resulted in the development of a statistically verified safety performance BSC framework for Saudi schools. The new framework provides a systematic approach that the Saudi Ministry of Education can implement to minimise school accidents and incidents. Although the current research method was exhaustive, and permitted the generation of the most comprehensive safety performance BSC framework possible, the research had a few limitations, such as, the lack of Saudi Ministry of Education department officers in some cities, or in the different geographical regions in Saudi Arabia, constrained the number and location of appropriate study sites. Hence, the study was limited to large population cities (i.e. Riyadh, Dammam, Jeddah, Hail, and Abha). To enable the wider application, and to gain a better generalisability, the new safety performance BSC framework would benefit from being tested on a larger sample size, as well as a more diverse sample diversity, especially in terms of regional location and cultural context.

Also, the developed structural equations within the safety performance BSC framework were only applied in Saudi Arabia. Hence, they need to be tested within education sectors in other countries. It is expected that the outcomes will differ, depending on the maturity of the host nation and its education sector, and the strength of the relationship in the verified safety performance BSC framework. Finally, the safety performance items were evaluated only through a quantitative assessment scale. Thus, the findings could be enhanced through the application of a qualitative assessment; this analysis could improve the strength of adopting the benchmarking method.

Finally, the present BSC framework and causal links verified in this study are linear (Fig. 3) where in practice some are complex two-way relationships that interact over long periods of time. The research method adopted (i.e. single snapshot measurement using questionnaire survey) limited the findings to those presented in Fig. 3. However, a longitudinal study over many years could enable the formulation of functions explaining all direct relationships and also feedback loops, enabling the creation of a more sophisticated Systems Dynamic model. Robust SD models require extensive data collection over many years, which was outside the scope and budget of this present study. Nonetheless, the present study begins to unpack evidence that safety in schools needs to be thought about more systematically. Relying solely on measuring incidents and accidents will not improve safety performance; instead the focus must be on the upstream enablers of better safety performance such as safety leadership, training, policies and procedures, to name a few.

Implications and future research

This research study has a number of implications for the Saudi Arabian educational sector that is attempting to develop and promote an effective safety performance process in across their rapidly growing portfolio of primary, elementary and secondary schools. The developed method is particularly important for publicly funded schools in which the Saudi Ministry of Education is concerned. It is hoped that advanced safety monitoring and management approaches such as this will be willingly and effectively transferred to teachers, school executives, and Ministry of Education officers. The safety performance BSC framework should be considered as the initial foundation for a more comprehensive approach. The developed framework may also be of particular interest to multilateral funding agencies such as the United Nations Educational, Scientific and Cultural Organisation (UNESCO), which might seek tools to better monitor the performance of the school safety process when the organisation provides services to schools in developing countries.

Study findings and implications provide a solid foundation for future work. Future work includes the application of the developed safety performance BSC framework through numerous case studies in Saudi schools in order to validate the framework perspectives and their causal linkages. Also, causal mapping at the item level adopting the interpretative structural modelling approach is proposed in order to intricately map the Saudi school safety system. Furthermore, a systematic benchmarking approach will be developed based on the perspectives, factors and items of the safety performance BSC framework and will be applied to longitudinally monitor the safety performance of the Saudi schools of different characteristic groups (i.e. region, gender, type, etc.). Safety performance benchmarking and its longitudinal monitoring will assist the Ministry of Education to focus on deficient critical lagging factors that may need to be improved in order to boost leading indicators of safety performance such as incidents and accident rates. Also, an SD model could be formulated with this extensive longitudinal dataset enabling Ministry of Education officers to better prioritise strategies for sustainably enhancing safety performance over the long term.

References

Ando, M., Asakura, T., Simons-Morton, B., 2005. Psychosocial influences on physical, verbal, and indirect bullying among Japanese early adolescents. The Journal of Early Adolescence 25 (3), 268–297.

Argyris, C., 1999. On Organizational Learning. Oxford University Press.

Argyris, C., Schön, D., 1996. Organizational Learning II: Theory, Method, and Practice. Addison-Wesley, Wokingham.

Barclay, D., Higgins, C., Thompson, R.C., 1995. The partial least squares (PLS) approach to causal modeling: personal computer adoption and use as an illustration. Technology Studies 2 (2), 284–324.

Bendak, S., 2006. Evaluation of school safety in Riyadh. International Journal of Injury Control and Safety Promotion 13 (4), 264–266.

Broadnax, W.D., Conway, K.J., 2001. The social security administration and performance management. In: Quicker, Better, Cheaper: Managing Performance in American Government., pp. 143–175.

Brown, M.L., Kenney, M., 2006. Organizational learning: theoretical and methodological considerations. In: Leann Brown, M., Michael Kenney, Zarkin, M.J. (Eds.), Organizational Learning in the Global Context. Ashgate Publishing Limited, pp. 01–17.

Capelo, C., Dias, J.F., 2009. A system dynamics-based simulation experiment for testing mental model and performance effects of using the balanced scorecard. System Dynamics Review 25 (1), 1–34.

Chin, W.W., 1998a. Commentary: issues and opinion on structural equation modeling. MIS Quarterly 22 (1), 07–16.

Chin, W.W., 1998b. The partial least squares approach to structural equation modeling. In: Marcoulides, G.A. (Ed.), Modern Methods for Business Research. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 295–336.

Chin, W.W., 2010. How to write up and report PLS analyses. In: Handbook of Partial Least Squares., pp. 655–690.

Choudhry, R.M., Fang, D., Mohamed, S., 2007. The nature of safety culture: a survey of the state-of-the-art. Safety Science 45 (10), 993–1012.

Clarke, S., 2010. An integrative model of safety climate: linking psychological climate and work attitudes to individual safety outcomes using meta-analysis. Journal of Occupational and Organizational Psychology 83 (3), 553–578.

Cohen, J.B., 1988. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Lawrence Erlbaum Associates, Hillsdale, NJ.

Cooper, M.D., Phillips, R.A., 2004. Exploratory analysis of the safety climate and safety behavior relationship. Journal of Safety Research 35 (5), 497–512.

Creswell, J.W., 2003. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 2nd ed. Sage Publications, Inc, Thousand Oaks, CA.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16 (3), 297–334.

Dyson, M. (1999). How and when to write policies and procedures (2nd ed.). Kingsgrove, NSW: Australian Council for the Rehabilitation of Disabled. http://www.safework.sa.gov.au/contentPages/docs/labrWritingPolicy.pdf

Efron, B., 1979. Bootstrap methods: another look at the jackknife. The Annals of Statistics 7 (1), 01–26.

Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap. Chapman Hall, New York.

El Guindi, F., 1981. Veiling infitah with Muslim ethic: Egypt's contemporary Islamic movement. Social Problems, 465–485.

Esposjto, J.L., 1982. Islamization: religion and politics in Pakistan. The Muslim World 72 (3-4), 197–223.

Falk, R.F., Miller, N.B., 1992. A Primer for Soft Modeling. University of Akron Press, OH, USA.

Field, A., 2009. Discovering Statistics Using SPSS. Sage Publications Ltd, Los Angeles.

Fornell, C., Bookstein, F.L., 1982. Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. Journal of Marketing Research 12 (4), 440–452.

Fornell, C., Cha, J., 1994. Partial least squares. In: Bagozzi, R. (Ed.), Advanced Methods of Marketing Research. Blackwell, Cambridge, MA, pp. 52–78.

Fornell, C., Larcker, D.F., 1981. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 18 (1), 39–50.

Garratt, B., 1999. The learning organisation 15 years on: some personal reflections. The Learning Organization 6 (5), 202–207.

Gefen, D., Straub, D., Boudreau, M., 2000. Structural equation modelling techniques and regression: guidelines for research practice. Communications of AIS 7 (7), 01–78.

Geisser, S., 1975. The predictive sample reuse method with applications. Journal of the American Statistical Association 70, 320–328.

Gerbing, D.W., Anderson, J.C., 1988. An updated paradigm for scale development incorporating unidimensionality and its assessment. Journal of Marketing Research 25 (2), 186–192.

Glendon, A.L., Litherland, D.K., 2001. Safety climate factors, group differences and safety behaviour in road construction. Safety Science 39 (3), 157–188.

Goh, Y.M., Brown, H., Spickett, J., 2010. Applying systems thinking concepts in the analysis of major incidents and safety culture. Safety Science 48 (3), 302–309.

Grabowski, M., Ayyalasomayajula, P., Merrick, J., Harrald, J., Roberts, K., 2007. Leading indicators of safety in virtual organizations. Safety Science 5 (10), 1013–1043.
Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., 2010. Multivariate Data Analysis, 6th ed. Prentice Hall, Upper Saddle River, NJ.
Hair, J.F., Ringle, C.M., Sarstedt, M., 2011. PLS-SEM: Indeed a silver bullet. The Journal of Marketing Theory and Practice 19 (2), 139–152.
Håvold, J.I., Nesset, E., 2009. From safety culture to safety orientation: validation and simplification of a safety orientation scale using a sample of seafarers working for Norwegian ship owners. Safety Science 47, 305–326.
Hayman, A., 2007. National Report on Management Effectiveness Assessment and Capacity Development Plan for Jamaica's System of Protected Areas. The Nature Conservancy, Kingston, Jamaica, 111 pp.
Huang, Y.H., Chen, P.Y., Krauss, A.D., Rogers, D.A., 2004. Quality of the execution of corporate safety policies and employee safety outcomes: assessing the moderating role of supervisor safety support and the mediating role of employee safety control. Journal of Business and Psychology 18 (4), 483–506.
Huang, Y.-H., Zohar, D., Robertson, M.M., Garabet, A., Murphy, L.A., Lee, J., 2013. Development and validation of safety climate scales for mobile remote workers using utility/electrical workers as exemplar. Accident Analysis & Prevention 59, 76–86.
Huber, F., Herrmann, A., Meyer, F., Vogel, J., Vollhardt, K., 2007. Kausalmodellierung mit partial least squares: eine anwendungsorientierte einführung. Gabler Verlag, Wiesbaden.
Hulland, J., 1999. Use of partial least squares (PLS) in strategic management research: a review of four recent studies. Strategic Management Journal 20 (2), 195–204.
Inamdar, N., Kaplan, R.S., Bower, M., 2002. Applying the balanced scorecard in healthcare provider organizations. Journal of Healthcare Management/American College of Healthcare Executives 47 (3), 179.
International Atomic Energy Agency, 2012. Establishing the safety infrastructure. In: Safety Package for Module 3: Leadership and Management for Safety; Safety in the Operating Organization and Preparation for Commissioning, Retrieved from http://www-ns.iaea.org/tech-areas/safety-infrastructure/ssg-16-module-3.asp?s=0&l=94
Jennings, E.T., Haist, M.P., 2004. Putting performance measurement in context. In: Ingraham, P.W., Lynn, L.E. (Eds.), The Art of Governance: Analyzing Management and Administration. , pp. 152–194.
Julnes, P.L., Holzer, M., 2002. Promoting the utilization of performance measures in public organizations: an empirical study of factors affecting adoption and implementation. Public Administration Review 61 (6), 693–708.
Kaplan, R.S., Nortan, D.P., 1996. The Balanced Scorecard: Translating Strategy into Action. Harvard Business Press, United States.
Kaplan, R.S., Norton, D.P., 1992. The balanced scorecard-measures that drive performance. Harvard Business Review 70 (1), 71–79.
Kauffman, J.M., 1997. Characteristics of Emotional and Behavioral Disorders of Children and Youth, 6th ed. Prentice Hall, Upper Saddle River, NJ.
Keddie, N.R., 1979. Problems in the study of middle eastern women. International Journal of Middle East Studies 10 (2), 225–240.
Kines, P., Andersen, L.P.S., Spangenberg, S., Mikkelsen, K.L., Dyreborg, J., Zohar, D., 2010. Improving construction site safety through leader-based verbal safety communication. Journal of Safety Research 41 (5), 399–406.
Koufteros, X.A., 1999. Testing a model of pull production: a paradigm for manufacturing research using structural equation modeling. Journal of Operations Management 17 (4), 467–488.
Kunc, M., 2008. Using systems thinking to enhance strategy maps. Management Decision 46 (5), 761–778.
Larsson, P., Dekker, S.W.A., Tingvall, C., 2010. The need for a systems theory approach to road safety. Safety Science 48 (9), 1167–1174.
Lawrie, G., Cobbold, I., 2004. Third-generation balanced scorecard: evolution of an effective strategic control tool. International Journal of Productivity and Performance Management 53 (7), 611–623.
Lawrie, M., Parker, D., Hudson, P., 2006. Investigating employee perceptions of a framework of safety culture maturity. Safety Science 44 (3), 259–276.
Leveson, N.G., 2007. Technical and managerial factors in the NASA challenger and Columbia losses: looking forward to the future. In: Fleishman, E., Handelsman, J. (Eds.), Controversies in Science and Technology, From Chromosomes to the Cosmos, vol. 2. Mary Ann Liebert Inc..
Macmillan, K., Money, K., Money, A., Downing, S., 2005. Relationship marketing in the not-for-profit sector: an extension and application of the commitment–trust theory. Journal of Business Research 58 (6), 806–818.
Marquardt, M.J., 2000. Action learning and leadership. The Learning Organization 7 (5), 233–241.
Mathwick, C., Malhotra, N., Rigdon, E., 2001. Experiential value: conceptualization, measurement and application in the catalog and Internet shopping environment. Journal of Retailing 77 (1), 39–56.
Mayze, B.R., Bradley, L.M., 2008. Safety culture: a multilevel assessment tool for the construction industry. In: Third International Conference of the Cooperative Research Centre (CRC) for Construction Innovation–Clients Driving Innovation: Benefiting from Innovation, 12–14 March 2008, Gold Coast, Queensland.
Mcdonald, N., Corrigan, S., Daly, C., Cromie, S., 2000. Safety management systems and safety culture in aircraft maintenance organisations. Safety Science 34 (1–3), 151–176.
Mearns, K., Håvold, J.I., 2003. Occupational health and safety and the balanced scorecard. The TQM Magazine 15 (6), 408–423.
Mearns, K., Whitaker, S.M., Flin, R., 2003. Safety climate, safety management practice and safety performance in offshore environments. Safety Science 41 (8), 641–680.
Meyer, H.H., 1991. A solution to the performance appraisal feedback enigma. Academy of Management Executive 5 (1), 68–76.
Meyer, H.H., Kay, E., French, J.R.P., 1965. Split roles in performance appraisal. Harvard Business Review 43, 123–129.
Mohamed, S., 2003. Scorecard approach to benchmarking organisational safety culture in construction. Journal of Construction Engineering & Management 129 (1), 80–88.
Moynihan, D.P., 2005. Goal-based learning and the future of performance management. Public Administration Review 65 (2), 203–216.
Moynihan, D.P., Landuyt, N., 2009. How do public organizations learn? Bridging cultural and structural perspectives. Public Administration Review 69 (6), 1097–1105.
Nansel, T.R., Overpeck, M., Pilla, R.S., Ruan, W.J., Simons-Morton, B., Scheidt, P., 2001. Bullying behaviors among US youth: prevalence and association with psychological adjustment. The Journal of the American Medical Association 285 (16), 2094–2100.
Netshitahame, N.E., Van Vollenhoven, W.J., 2006. School safety in rural schools: are schools as safe as we think they are? South African Journal of Education 22 (4), 313–318.
Neuman, W., 2006. Social Research Methods: Qualitative and Quantitative Approaches, 4th ed. Allyn & Bacon, Boston, MA.
Ng, S.T., Cheng, K.P., Skitmore, R.M., 2005. A framework for evaluating the safety performance of construction contractors. Building and Environment 40 (10), 1347–1355.
Nunnally, J.C., Bernstein, I.H., 1994. Psychometric Theory, 3rd ed. McGraw-Hill, New York.
Pedler, M., Aspinall, K., 2000. A concise guide to the learning organisation. The Learning Organization 7 (1).
Pellegrini, A.D., Long, J.D., 2002. A longitudinal study of bullying, dominance, and victimization during the transition from primary school through secondary school. British Journal of Developmental Psychology 20 (2), 259–280.
Pepler, D.J., Craig, W.M., Connolly, J.A., Yuile, A., Mcmaster, L., Jiang, D., 2006. A developmental perspective on bullying. Aggressive Behavior 32, 376–384.
Phillips, P., Louvieris, P., 2005. Performance measurement systems in tourism, hospitality, and leisure small medium-sized enterprises: a balanced scorecard perspective. Journal of Travel Research 44 (2), 201–211.
Prokop, M., 2003. Saudi Arabia: the politics of education. International Affairs 79 (1), 77–89.
Rasmussen, J., 1997. Risk management in a dynamic society: a modelling problem. Safety Science 27 (2–3), 183–213.
Ribbink, D., Van Riel, A.C.R., Liljander, V., Streukens, S., 2004. Comfort your online customer: quality, trust and loyalty on the Internet. Managing Service Quality 14 (6), 446–456.
Ringle, C.M., 2004. Gütemaße für den partial least squares-ansatz zur bestimmung von kausalmodellen. In: Hansmann, K.W. (Ed.), Industrielles Management. Universität Hamburg, Hamburg.
Rundmo, T., Hale, A.R., 2003. Managers' attitudes towards safety and accident prevention. Safety Science 41 (7), 557–574.
Saleh, S., 1972. Women in Islam: their status in religious and traditional culture. International Journal of Sociology of the Family 2 (1), 35–42.
Saudi Ministry of Education, 2013. Project Department, Retrieved from http://moe-b.com/INDEX.HTM [reviewed in 21.03.13].
Sawacha, E., Naoum, S., Fong, D., 1999. Factors affecting safety performance on construction sites. International Journal of Project Management 17 (5), 309–315.
Schneider, M., 2004. Performance management by culture in the national labor relations board's division of judges and the German labor courts of appeal. Journal of Public Administration Research and Theory 14, 19–32.
Scott, T.M., 2001. A schoolwide example of positive behavioral support. Journal of Positive Behavior Interventions 3 (2), 88–94.
Sekaran, U., 2003. Research Methods for Business: A Skill Building Approach, 4th ed. John Wiley & Sons Inc, New York.
Senge, P., 2006. The Fifth Discipline: The Art & Practice of the Learning Organization. Doubleday & McClure Company, New York.
Sørensen, T., Mastekaasa, A., Kleiner, R., Sandanger, I., Bøe, N., Klepp, O.M., 2004. Local community mobilisation and mental health promotion. International Journal of Mental Health Promotion 6 (4), 05–16.
Stewart, R., 2008. A framework for the life cycle management of information technology projects: project IT. International Journal of Project Management 26 (3), 203–212.
Stone, E.F., 1974. The moderating effect of work-related values on the job scope-job satisfaction relationship. In: PhD Dissertation. University of California, Irvine.
Tabachnick, B.G., Fidell, L.S., 2007. Using Multivariate Statistics, 5th ed. Pearson Education, Inc., Boston.
Tenenhaus, M., Vinzi, V.E., Chatelin, Y.M., Lauro, C., 2005. PLS path modeling. Computational Statistics & Data Analysis 48, 159–205.
Teo, E.a.L., Ling, F.Y.Y., 2006. Developing a model to measure the effectiveness of safety management systems of construction sites. Building and Environment 41 (11), 1584–1592.
Teo, E.a.L., Ling, F.Y.Y., Chong, A.F.W., 2005. Framework for project managers to manage construction safety. International Journal of Project Management 23 (4), 329–341.

Timmerman, G., 2003. Sexual harassment of adolescents perpetrated by teachers and by peers: an exploration of the dynamics of power, culture, and gender in secondary schools. Sex Roles 48 (5/6), 231–244.

Urbach, N., Ahlemann, F., 2010. Structural equation modeling in information systems research using partial least squares. Journal of Information Technology Theory and Application 11 (2), 05–40.

Venaik, S., 1999. A Model of Global Marketing in Multinational Firms: An Empirical Investigation. The University of New South Wales, PhD Dissertation.

Wadsworth, E.J.K., Smith, A.P., 2009. Safety culture, advice and performance. Policy and Practice in Health and Safety 7 (1), 05–32.

Werts, C.E., Linn, R.L., Jöreskog, K.G., 1974. Intraclass reliability estimates: testing structural assumptions. Educational and Psychological Measurement 34, 25–33.

Wetzels, M., Odekerken-Schröder, G., Van Oppen, C., 2009. Using PLS path modelling for assessing hierarchical construct models: guidelines and empirical illustration. MIS Quarterly 33 (1), 177–195.

White, J.C., Varadarajan, P.R., Dacin, P.A., 2003. Market situation interpretation and response: the role of cognitive style, organizational culture, and information use. Journal of Marketing 67 (3), 63–79.

Wixom, B.H., Watson, H.J., 2001. An empirical investigation of the factors affecting data warehousing success. MIS Quarterly 25 (1), 17–41.

Wold, H., 1966. Estimation of principal components and related models by iterative least squares. In: Krishnaiah, P.R. (Ed.), Multivariate Analysis. , pp. 391–420.

Wold, H., 1979. Model Construction and Evaluation When Theoretical Knowledge is Scarce: An Example of the Use of Partial Least Squares. University of Geneva, Geneva.

Wold, H., 1980. Model construction and evaluation when theoretical knowledge is scarce: theory and application of partial least squares. In: Kmenta, J., Ramsey, J.B. (Eds.), Evaluation of Econometric Models. Academic Press, New York, NY, pp. 47–74.

Wold, H., 2006. Partial least squares. In: Kotz, S., Johnson, N.L. (Eds.), Encyclopedia of Statistical Sciences. , pp. 581–591.

Yang, C.C., Wang, Y.S., Chang, S.T., Guo, S.E., Huang, M.F., 2009. A study on the leadership behavior, safety culture, and safety performance of the healthcare industry. World Academy of Science, Engineering and Technology 2 (8), 87–94.

Yang, K., Hsieh, J.Y., 2007. Managerial effectiveness of government performance measurement: testing a middle-range model. Public Administration Review 67 (5), 861–879.

Yin, R.K., 2003. Case Study Research: Design and Methods, 2nd ed. Sage Publications Inc., Thousand Oaks, CA.

Zohar, D., 2010. Thirty years of safety climate research: reflections and future directions. Accident Analysis & Prevention 42 (5), 1517–1522.

Zohar, D., Huang, Y.-H., Lee, J., Robertson, M.M., 2014. A mediation model linking dispatcher leadership and work ownership with safety climate as predictors of truck driver safety performance. Accident Analysis & Prevention 62, 17–25.