

# É

## Études et recherches

■ RAPPORT R-584



**Validation d'un outil d'autodiagnostic  
et d'un modèle de progression de la mesure  
en santé et sécurité du travail**

*Mario Roy  
Jean Cadieux  
Lucie Fortier  
Louise Leclerc*



Solidement implanté au Québec depuis 1980, l'Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST) est un organisme de recherche scientifique reconnu internationalement pour la qualité de ses travaux.

## NOS RECHERCHES

### Mission *travaillent pour vous !*

Contribuer, par la recherche, à la prévention des accidents du travail et des maladies professionnelles ainsi qu'à la réadaptation des travailleurs qui en sont victimes.

Offrir les services de laboratoires et l'expertise nécessaires à l'action du réseau public de prévention en santé et en sécurité du travail.

Assurer la diffusion des connaissances, jouer un rôle de référence scientifique et d'expert.

Doté d'un conseil d'administration paritaire où siègent en nombre égal des représentants des employeurs et des travailleurs, l'IRSST est financé par la Commission de la santé et de la sécurité du travail.

### Pour en savoir plus

Visitez notre site Web ! Vous y trouverez une information complète et à jour.

De plus, toutes les publications éditées par l'IRSST peuvent être téléchargées gratuitement. [www.irsst.qc.ca](http://www.irsst.qc.ca)

Pour connaître l'actualité de la recherche menée ou financée par l'IRSST, abonnez-vous gratuitement au magazine Prévention au travail, publié conjointement par l'Institut et la CSST. Abonnement : 1-877-221-7046

### Dépôt légal

Bibliothèque et Archives nationales  
2008

ISBN : 978-2-89631-311-2 (version imprimée)

ISBN : 978-2-89631-312-9 (PDF)

ISSN : 0820-8395

IRSST - Direction des communications  
505, boul. De Maisonneuve Ouest  
Montréal (Québec)  
H3A 3C2  
Téléphone : 514 288-1551  
Télécopieur : 514 288-7636  
[publications@irsst.qc.ca](mailto:publications@irsst.qc.ca)  
[www.irsst.qc.ca](http://www.irsst.qc.ca)  
Institut de recherche Robert-Sauvé  
en santé et en sécurité du travail,  
octobre 2008

# Études et recherches

RAPPORT R-584

## Validation d'un outil d'autodiagnostic et d'un modèle de progression de la mesure en santé et sécurité du travail

### Avis de non-responsabilité

L'IRSST ne donne aucune garantie relative à l'exactitude, la fiabilité ou le caractère exhaustif de l'information contenue dans ce document. En aucun cas l'IRSST ne saurait être tenu responsable pour tout dommage corporel, moral ou matériel résultant de l'utilisation de cette information.

Notez que les contenus des documents sont protégés par les législations canadiennes applicables en matière de propriété intellectuelle.

*Mario Roy, Jean Cadieux, Lucie Fortier et Louise Leclerc  
Chaire d'étude en organisation du travail,  
Faculté d'administration, Université de Sherbrooke*

Cliquez recherche  
[www.irsst.qc.ca](http://www.irsst.qc.ca)



Cette publication est disponible  
en version PDF  
sur le site Web de l'IRSST.

**CONFORMÉMENT AUX POLITIQUES DE L'IRSS**

Les résultats des travaux de recherche publiés dans ce document  
ont fait l'objet d'une évaluation par des pairs.

## SOMMAIRE

Ce rapport de recherche vise essentiellement à rendre compte de la démarche qui a été réalisée pour la validation d'un outil d'autodiagnostic de l'état de la SST destiné aux établissements de production et de fabrication dans le contexte québécois. La démarche a nécessité (1) la révision d'une version préliminaire du questionnaire, (2) la réalisation d'un test auprès de trois entreprises du secteur de l'imprimerie, (3) des ajustements pour la production d'un questionnaire de deuxième génération, (4) l'utilisation d'une procédure Delphi auprès de cinq experts académiques du domaine, (5) la production d'un questionnaire de troisième génération, (6) et la validation finale du questionnaire auprès d'un échantillon provenant de deux autres établissements du secteur de l'imprimerie.

Cette recherche comporte diverses contributions. D'abord, la validité du questionnaire de troisième génération a été démontrée. Sur le plan scientifique, cet outil est à notre connaissance le premier instrument validé qui permet de rendre compte dans un modèle hiérarchique intégré des observations des employés concernant les pratiques de SST en vigueur au sein de leur établissement. Deuxièmement, la nature formative des échelles de mesure utilisées constitue une innovation particulièrement bien adaptée au contexte de la SST. L'élaboration de ces échelles a nécessité le recours à des procédures de validation statistiques particulières (analyse des tétrades évanescents). Troisièmement, cette étude a confirmé l'existence d'au moins trois groupements d'employés dont les profils de réponses au questionnaire peuvent être classés : les peu attentifs qui ne reconnaissent pas ou très peu les pratiques et les gestes posés pour favoriser la SST, les très attentifs qui reconnaissent toutes les initiatives visant à favoriser la SST et finalement, les baromètres qui adoptent une position beaucoup plus nuancée entre les deux groupes extrêmes. Leur appréciation se rapproche vraisemblablement plus de la situation réelle. L'étude de ces groupements fournit des indicateurs qui ont le potentiel de renseigner sur la progression de la culture SST au sein des établissements, et ce, même en l'absence d'accidents.

Sur le plan de la pratique, cette recherche a permis de développer un outil valide, simple et convivial qui permet de diagnostiquer l'état de la SST au sein des établissements. De plus, la production de tableaux de bord, d'un guide d'intervention et d'une formation au diagnostic destinés aux associations sectorielles paritaires (ASP) et autres intervenants dans le domaine de la SST permettra une utilisation large et appropriée de l'outil qui, nous l'espérons, favorisera l'adoption d'une philosophie d'apprentissage plutôt que de performance en SST.

## REMERCIEMENTS

Nous tenons à remercier tout d'abord madame Marie Ménard, directrice générale ASP Imprimerie, pour son implication dans le recrutement des entreprises de son secteur qui ont participé à la phase finale de la recherche ainsi que pour le financement de départ du projet. Nous remercions également les organisations et les répondants qui ont accepté de participer à l'étude et qui ont contribué à la validation de notre l'outil d'autodiagnostic en santé et en sécurité du travail. Leurs commentaires et suggestions nous ont été très utiles et profitables.

Des remerciements additionnels vont aux membres du comité aviseur du projet. Leur implication et leur expertise ont su influencer l'orientation de cette activité de recherche et la rendre à notre avis encore plus pertinente pour les milieux de travail. Le comité était composé des personnes suivantes : monsieur Arnold Dugas, directeur général ASFETM, Section fabrication d'équipement de transport et de machines; monsieur Waguih Geadah, Ingénieur/coordonnateur ASFETM, Secteur fabrication d'équipements de transport et de machines; madame Isabelle Lessard, directrice Service de formation-information, Conseil patronal en SST; monsieur Michel Pérusse, Conseiller principal, Division SST, AON; monsieur Alain Plourde, directeur général APSST, Secteur Fabrication de produits en métal et électriques; madame Sophie Bergeron, Conseillère principale – SST, Hydro-Québec; monsieur Laurent Gratton, Service valorisation et relations avec les partenaires, IRSST. Nous tenons à remercier aussi madame Ginette Denicourt et les évaluateurs anonymes qui, grâce à leurs suggestions, nous ont permis d'améliorer la version finale de ce rapport.

## TABLE DES MATIÈRES

SOMMAIRE.....	i
REMERCIEMENTS.....	ii
INTRODUCTION.....	1
1. PROBLÉMATIQUE ET ÉTAT DES CONNAISSANCES.....	1
<i>Objectifs poursuivis</i> .....	6
2. DÉMARCHE DE RECHERCHE.....	7
2.1 Version préliminaire du questionnaire de 2 <sup>e</sup> génération.....	7
2.2 Étude de la validité de l'instrument.....	8
2.3 Résultats des tests du questionnaire de 2 <sup>e</sup> génération.....	11
2.4 Méthode Delphi – Élaboration du questionnaire de 3 <sup>e</sup> génération.....	12
<i>Étape 1 : Choix des participants</i> .....	13
<i>Étape 2 : Révision des échelles de mesure</i> .....	14
<i>Étape 3 : Recherche de consensus</i> .....	15
2.5 Seconde collecte des données sur le terrain.....	16
2.6 Validation du questionnaire de 3 <sup>e</sup> génération.....	16
2.7 Analyse de segmentation.....	18
3. DISCUSSION.....	20
<i>Sur le plan de la pratique</i> .....	22
<i>Limites et avenues de recherches futures</i> .....	23
CONCLUSION.....	24
APPLICABILITÉ DES RÉSULTATS.....	25
RETOMBÉES ÉVENTUELLES.....	25
ARTICLES SCIENTIFIQUES PRODUITS DANS LE CADRE DE LA SUBVENTION.....	26
RÉFÉRENCES.....	26

## LISTE DES TABLEAUX

Tableau 1 : Classification des types de mesure de la performance en SST en fonction des niveaux de préoccupations des gestionnaires .....	6
Tableau 2 : Items constitutifs du construit latent « conformité du milieu de travail ».....	10
Tableau 3 : La ventilation de l'échantillon .....	11
Tableau 4 : Les résultats du test des tétrades évanescentes .....	12
Tableau 5 : La ventilation de l'échantillon .....	16
Tableau 6 : Les résultats du test des tétrades évanescentes .....	17
Tableau 7: Définitions des groupements .....	18
Tableau 8 : Table ANOVA – Entreprise A .....	19
Tableau 9 : Table ANOVA – Entreprise B.....	19

## LISTES DES FIGURES

Figure 1 : Hiérarchie des niveaux de préoccupations .....	5
Figure 2 : Différence conceptuelle entre des variables latentes réflexives et formatives .....	9
Figure 3 : Regroupements de répondants .....	20



## INTRODUCTION

Plusieurs observateurs du domaine de la santé et sécurité du travail (SST) remarquent que le niveau des lésions semble avoir atteint un plateau dans le monde industrialisé. Les gains réalisés avec les approches traditionnelles de mesure de fréquence et de gravité des lésions de même qu'avec la mise en place de systèmes de gestion de la SST ont atteint leur limite (Hudson, 2007). Dans le cadre de cette recherche, nous avons consacré nos efforts à développer et à valider un instrument de mesure original qui s'intéresse aux observations des employés et des cadres de premier niveau sur ce qui se fait en matière de SST au sein de leur milieu. L'instrument, qui prend la forme d'un questionnaire, s'inscrit dans une philosophie d'apprentissage d'une culture SST plutôt que dans une philosophie de mesure de performance. Il s'agit d'un outil formatif qui oriente l'action plutôt qu'un outil sommatif qui sanctionne les résultats. L'intention avouée du développement de ce type de diagnostic consiste à favoriser l'adoption d'une véritable culture de SST dans les milieux en mettant l'accent sur les normes, les pratiques et les valeurs des individus au travail.

Le rapport présente dans un premier temps, la problématique à laquelle nous nous adressons de même que l'état des connaissances sur le sujet. Par la suite, nous abordons en détail la démarche de recherche qui a conduit à la validation de l'instrument. La section discussion souligne les contributions et les limites de notre étude. Finalement, la dernière section traite de l'applicabilité des résultats et des retombées au sein des milieux.

### 1. PROBLÉMATIQUE ET ÉTAT DES CONNAISSANCES

L'étude que nous présentons dans ce rapport fait suite à une recherche antérieure dans laquelle nous avons développé la première version d'un instrument de mesure de l'état de la santé et de la sécurité du travail destiné aux entreprises manufacturières organisées en équipes semi-autonomes de travail (Roy, Bergeron et Fortier, 2004). Compte tenu de la pertinence reconnue de l'outil du point de vue des praticiens qui l'ont utilisé, nous avons décidé d'entreprendre une démarche structurée visant à le valider sur le plan scientifique. L'instrument en question s'inscrit dans une philosophie de mesure qui favorise l'émergence d'une culture d'apprentissage de la prévention et de l'élimination des risques au sein des milieux grâce à l'amélioration des processus, des systèmes, des normes et des comportements qui induisent les résultats en matière de SST. Cette philosophie se distingue de la philosophie de mesure de la performance fondée sur les résultats en SST qui est encore largement dominante dans les milieux de travail malgré les effets pervers qu'elle comporte.

Plusieurs auteurs ont signalé que la mesure des résultats est insuffisante pour rendre compte de la performance effective des milieux de travail en SST (Booth, 1993; Mitchell, 2000; O'Brien, 2000; Shaw et Blewett, 1995; Simpson et Gardner, 2001). Booth (1993) présente une série de raisons pour lesquelles il qualifie de « pauvres » les indicateurs de résultats. Les principales sont les suivantes : ils mesurent les échecs plutôt que les succès, ils sont exposés aux fluctuations du hasard, ils mesurent les blessures et non le nombre réel d'accidents, ils sont difficiles à comparer, ils procurent une évidence face au fait que quelque chose ne va pas, mais ils représentent un faible prédicteur de la performance future (Stricoff, 2000). De plus, ces indicateurs sont affectés par l'effet de désirabilité sociale puisqu'il est préférable de présenter une image corporative de

bon citoyen préoccupé par la santé et la sécurité du personnel. Le biais à la baisse est particulièrement important lorsqu'un système de récompense (e.g. bonus) encourage la diminution des événements déclarés (Budworth, 1996). Pour Shaw et Blewett (1995), la mesure des résultats peut même être contre-productive parce que la SST ne devient une priorité de gestion seulement lorsque le taux de fréquence toléré par les gestionnaires a atteint sa limite.

À ce moment, ces derniers accordent l'attention au problème et mettent en place différentes actions pour faire diminuer ce taux. Lorsque l'indicateur redescend à un niveau jugé acceptable, les gestionnaires cessent de lui porter attention en croyant à tort que la situation est sous contrôle. Peu de temps après, le taux remonte et le même scénario se répète. De plus, malgré leur apparente simplicité, ces indicateurs sont peu fiables. Une étude récente menée par Shannon et Lowe (2002) a démontré qu'au Canada, il y aurait une sous-déclaration importante des lésions aux régimes d'indemnisation provinciaux. Ce serait particulièrement le cas lorsque les personnes concernées croient que le retour au travail sera rapide ou lorsqu'elles sont suffisamment couvertes par une autre assurance. Dans de tels cas, plusieurs individus ont tendance à ne pas déclarer l'événement dont ils ont été victimes aux organismes de contrôle provinciaux. Sur un échantillon de 2 500 personnes, 40 % des personnes éligibles à recevoir des prestations n'ont pas soumis de réclamation. La mesure de performance fondée strictement sur les résultats s'inscrit dans une philosophie de commandement et de contrôle qui suppose que les personnes en autorité doivent contraindre les employés pour qu'ils se conforment aux directives sécuritaires (Hudson, 2007). L'individu qui subit une lésion est alors typiquement blâmé pour avoir enfreint les règles de SST.

Si la mesure des résultats et les pratiques d'analyse des accidents ont permis de réaliser des gains appréciables au cours des dernières décennies, il nous faut constater que nous avons atteint un plateau et qu'il faut trouver d'autres façons de s'attaquer à la problématique de SST si nous voulons progresser (Cadieux, Roy et Desmarais, 2006). Plus de contraintes et plus de systèmes visant à rapporter les incidents et accidents ne permettront pas d'obtenir des gains significatifs. Selon divers auteurs, il est temps de considérer la mise en place d'une véritable culture de la SST qui suppose de transformer la façon dont on vit au quotidien la SST dans les milieux de travail (Hudson 2007). Le concept de culture SST n'est pas nouveau et plusieurs auteurs se questionnent sur les caractéristiques d'une telle culture (Hale, 2000; Guldenmund, 2000, Reason, 1998) ou sur la façon de la mesurer (Arezes et Miguel, 2003; Lawrie et al, 2006).

Pour nous, la culture est un phénomène émergent dans les systèmes sociaux qui se définit comme étant l'ensemble des normes, des valeurs et des pratiques qui s'établissent entre les individus à travers le temps. Les règles, les procédures et les systèmes sont des sous-produits de la culture. En se basant sur cette définition, il devient possible d'établir un cadre normatif qui représenterait l'idéal de la culture SST à développer entre les individus membres d'une organisation, c'est-à-dire une culture dans laquelle la SST est prise en considération dans toutes les décisions et toutes les actions quel que soit l'individu impliqué ou le poste qu'il occupe dans l'organisation (Zhang et al., 2002). À partir d'une telle représentation, il est possible de diagnostiquer l'écart perçu entre la culture idéale et la réalité vécue par les membres qui reflète la culture SST ambiante.

Le diagnostic de la culture SST s'inscrit dans le courant de mesure qui s'intéresse aux variables prédictives réputées avoir une incidence directe sur les résultats en SST (Booth, 1993). La mesure des prédicteurs comporte de nombreux avantages. Simpson et Gardner (2001) en ont relevé quelques-uns dont voici les principaux : l'utilisation de ces indicateurs permet de vérifier la réalisation des plans d'action, ils procurent un cadre d'amélioration, ils ont une influence plus positive que les indicateurs de résultats et ils permettent l'évaluation de l'efficacité des mécanismes de contrôle des risques. Pour Gomez (1998) et Toellner (2001), l'utilisation d'indicateurs de processus et de pratiques permet d'obtenir de l'information à un moment où il est encore temps d'intervenir et d'éviter les lésions professionnelles.

L'instrument que nous avons développé et pré-testé dans l'étude initiale se distingue des outils d'audit de gestion existants en s'adressant à l'ensemble des personnels. Il s'intéresse essentiellement aux perceptions des répondants concernant ce qui se fait effectivement en matière de SST au sein de l'établissement. Contrairement aux questionnaires de climat (Flin et al., 2000) d'attitude (Cheyne et al., 2002) ou de motivation (Hudson et al., 1998) envers la SST, il s'agit ici de rendre compte des actions qui sont posées quotidiennement à tous les niveaux de l'organisation pour favoriser la SST. Ces actions constituent les véritables révélateurs de la culture ambiante.

Dans le cadre de notre recherche initiale, nous avons développé une classification (voir tableau 1) qui regroupe en un tout cohérent les outils de mesure répertoriés en fonction des niveaux de préoccupations des gestionnaires. La classification proposée est hiérarchique (voir figure 1), en ce sens qu'il existe un ordre de progression dans la mise en œuvre des catégories de mesure et d'intervention au sein des organisations. Une organisation incapable d'effectuer ses suivis de conformité ou plus simplement de réaliser ses enquêtes d'accidents pourrait difficilement mettre en place avec succès un système de mesure des comportements des travailleurs (Ward, 2000). L'atteinte des niveaux supérieurs de la pyramide théorique implique que l'essentiel du travail à réaliser aux niveaux inférieurs est en bonne partie réalisé. L'idée de classer les diverses cultures de SST en catégories hiérarchisées a initialement été proposée par Westrum (1993) qui a développé un cadre de référence à trois niveaux (pathologique, bureaucratique, générateur). Ce cadre a été repris par la suite et augmenté de deux intermédiaires (réactif et proactif) par Parker et al. (2006). Ces derniers auteurs caractérisent la culture SST des organisations à l'aide d'une typologie hiérarchisée qui illustre les intentions des dirigeants que l'on peut résumer à l'aide de phrases clefs :

- Pathologique : Pourquoi se préoccuper de la SST? En autant qu'on ne se fait pas prendre.
- Réactif : La SST est importante : on en fait beaucoup chaque fois qu'il y a un accident.
- Calculateur : Nous avons les systèmes en place pour gérer tous les risques.
- Proactif : Nous essayons d'anticiper les problèmes de SST avant qu'ils surviennent.
- Générateur : La SST, c'est notre façon d'être en affaires.

L'analyse des façons de faire des organisations permet de les classer dans l'une ou l'autre des catégories. Par la suite, celles qui désirent progresser dans la hiérarchie peuvent adopter les pratiques proposées dans la catégorie supérieure jusqu'à en arriver au niveau cinq, soit générateur. Ce cadre de référence a été développé et proposé avec succès à de grandes

entreprises multinationales dans le but de favoriser leur progression en matière de gestion de la SST.

Nous avons utilisé une logique semblable de hiérarchisation pour développer notre instrument de mesure destiné au diagnostic de la culture SST au sein des établissements (Roy et al., 2004). Les deux catégories de mesure du premier niveau (1a et 1b) sont de type réactif. Elles concernent des résultats du passé sur lesquels il n'est plus possible d'agir. Les gestionnaires qui se limitent à ces mesures sont préoccupés essentiellement par le contrôle des pertes engendrées par les événements accidentels et les lésions professionnelles. Il s'agit du niveau de base dans lequel la culture SST est embryonnaire. La performance des résultats constitue la préoccupation première à ce niveau avec les effets pervers que l'on connaît. Il s'agit ici de rapporter les événements et mettre en place les mécanismes pour éviter qu'ils ne se reproduisent. En deçà de ce niveau, la culture SST n'existe tout simplement pas. C'est le cas pour les organisations qui ignorent tout simplement les règles de SST ou qui considèrent la SST comme une contrainte réglementaire contre laquelle elles doivent faire tout ce qu'elles peuvent pour éviter d'être prise en faute (pathologique).

Au deuxième niveau, l'ordre de préoccupation change et se concentre sur l'idée de prévention. La culture SST est en croissance. La mesure fondée sur les résultats cède le pas à la mesure des procédés, des équipements, des postes de travail, bref sur tout ce qui compose l'organisation du travail et son interface avec les individus. C'est à ce niveau qu'entrent en jeu les variables prédictives qui ont une incidence sur les résultats. En vérifiant si le milieu de travail répond aux exigences réglementaires internes et externes (niveau 2a), il est possible d'apprécier la situation en regard des normes établies et de déterminer les écarts à combler. Au niveau 2b, l'ordre des préoccupations de la mesure ne concerne plus la conformité mais plutôt les processus, les structures et les programmes qui permettent à l'organisation de s'améliorer de façon continue en matière de santé et de sécurité. Les outils et techniques de mesure sont conçus de façon à pouvoir les adapter à chacune des situations investiguées. L'attention est centrée sur les moyens mis en place (programme de prévention, formation, information) pour améliorer la SST plutôt que strictement sur les conditions du milieu. À ce niveau, un engagement élevé de la direction et des employés est requis pour accroître la performance dans la prévention des lésions. Le niveau 2a et 2b recoupe approximativement les niveaux « calculateur » et « proactif » de Parker et al. (2006).

Au troisième et dernier niveau, l'attention se déplace vers les indicateurs de mesure de la culture organisationnelle qui sont reconnus comme étant prépondérants dans le maintien à long terme du succès en matière de santé et de sécurité du travail (Hudson, 2007). À ce niveau, l'ensemble des valeurs, des normes, des pratiques, des systèmes organisationnels et des comportements sont pris en compte dans la mise en place d'un milieu organisationnel sécuritaire qui vise l'élimination des accidents et maladies professionnelles à la source. Au niveau 3a, les mesures de performance s'intéressent aux gestes sécuritaires posés par les travailleurs et aux raisons des défaillances. On s'intéresse aussi aux normes de groupes qui favorisent ou nuisent à l'adoption de comportements sécuritaires, de même qu'à la perception de tous et chacun quant à l'engagement de la direction envers la SST et la participation des personnels en la matière. Au niveau 3b, la mesure de performance atteint son niveau le plus complexe en permettant l'appréciation de l'intégration de l'ensemble des systèmes organisationnels dans une perspective

de santé sécurité. Le système de mesure OHSAS-18001 est un outil typique de ce niveau de préoccupation. L'instrument de mesure universel de la performance des systèmes de gestion de la SST (UAI) développé par les chercheurs de l'Université du Michigan (Redinger et Levine, 1999), demeure à ce jour l'outil le plus complet de sa catégorie. Il s'agit ici de s'assurer que l'ensemble des décisions et des actions à tous les niveaux de l'organisation prennent en compte la SST. La SST devient une façon de vivre au travail, soit le niveau générateur proposé par Parker et al. (2006). Il est intéressant de noter que les deux modèles hiérarchiques ont été conçus de façon totalement indépendante et que néanmoins ils se recoupent largement. Ce constat suggère que non seulement l'idée de hiérarchie dans la progression vers une culture SST semble porteuse, mais que de plus, il apparaît plausible de reconnaître des niveaux de progression relativement définis.

Le choix des outils de mesure en santé sécurité et leur utilisation reflètent de façon tangible les valeurs des organisations et de leurs dirigeants. Plus l'ordre des préoccupations est élevé, plus on peut s'attendre à un accroissement de l'investissement en prévention et à une réduction des coûts de réparation. Il y a là un enjeu majeur pour tous les intervenants concernés par la SST dans les milieux de travail.

Cependant, le véritable enjeu de la mesure en SST consiste, selon nous, à favoriser l'émergence d'une culture d'apprentissage de la prévention et de l'élimination des dangers au sein des milieux de travail grâce à l'amélioration des processus et des systèmes qui déterminent de façon proactive les résultats. Une philosophie d'apprentissage plutôt qu'une philosophie de performance axée sur des indicateurs de résultats doit animer la conception des outils de mesure si l'on veut en arriver à l'élimination éventuelle des lésions professionnelles. Il s'agit en fait d'amener tout le monde à être à l'affût des situations à risques et de valoriser le fait de rendre publique toutes les défaillances potentielles. L'objectif consiste à éliminer au plus tôt tout ce qui peut induire des lésions pour le personnel. C'est dans cet esprit que l'outil proposé s'intéresse essentiellement aux perceptions de l'ensemble des répondants concernant ce qui se fait au sein de leur établissement. Il s'adresse aux niveaux deux et trois de notre modèle hiérarchique.

Figure 1 : Hiérarchie des niveaux de préoccupations

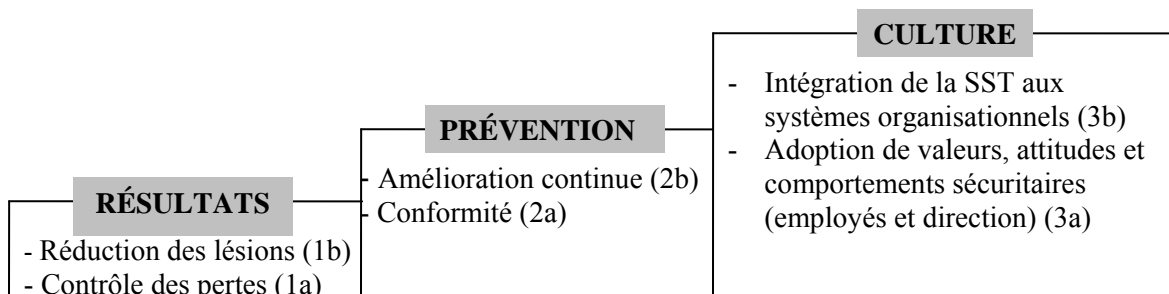


Tableau 1 : Classification des types de mesure de la performance en SST en fonction des niveaux de préoccupations des gestionnaires

	Dimension	Niveaux de préoccupations		Catégories de la mesure
Mesures prédictives (processus et systèmes)	Culture SST	3b	Intégration de la SST aux systèmes organisationnels	-Système de gestion de la SST -Systèmes organisationnels
		3a	Adoption de valeurs, attitudes et comportements sécuritaires (employés et direction)	-Engagement -Normes de groupe -Comportements -Respect des procédures
	Prévention	2b	Amélioration continue	-Organisation de la prévention - Structures -Programmes en SST -Activités (identification, contrôle, support) -Rôles et responsabilités
		2a	Conformité	-Organisation du travail (procédés, procédures sécuritaires de travail, etc.) -Équipements et matériels -Environnement (matières dangereuses, contaminants, aménagement des lieux, etc.)
Mesures réactives (résultats)	Résultats	1b	Réduction des lésions	Résultats raffinés -Incidents -Passés proches -Dommages matériels (bris) -Premiers soins, premiers secours -Assignment temporaire -Coûts indirects
		1a	Contrôle des pertes	Résultats traditionnels a) Données brutes - Nombre d'accidents - Nombre de maladies - Nombre de jours perdus - Coûts directs - Taux de cotisation b) Formules types

### Objectifs poursuivis

Nous reprenons ci-dessous les objectifs tels qu'ils ont été proposés dans le protocole de recherche :

- Procéder à la validation de la deuxième génération d'un instrument d'autodiagnostic qui permettra aux organisations d'apprécier d'une façon pointue leur progression en matière de SST.
- Vérifier la capacité de l'instrument à distinguer les établissements reconnus pour leurs bons résultats de ceux dont les performances sont préoccupantes en matière de SST.
- Démontrer la pertinence du modèle hiérarchique de préoccupation de la mesure en matière de SST.

- Vérifier la capacité de l'instrument à détecter la proportion de répondants dits sensibles et insensibles aux efforts de SST au sein des établissements.
- Procéder à la réalisation de diagnostics organisationnels détaillés à partir de l'interprétation des résultats de l'instrument de mesure.

Le premier objectif de l'étude consiste à procéder à la validation scientifique de la troisième génération de l'instrument d'autodiagnostic de façon à ce que son utilisation puisse être effectuée auprès de l'ensemble des milieux de travail intéressés par la mesure des prédicteurs. En corollaire, nous désirons vérifier la capacité de l'instrument à distinguer les établissements reconnus pour leurs bons résultats de ceux dont la performance est préoccupante en matière de santé et sécurité. De façon secondaire, nous désirons démontrer la pertinence du modèle hiérarchique de préoccupations de la mesure en SST lors de la présentation des diagnostics d'entreprises auprès des comités de SST des établissements qui participent à l'étude. Finalement, nous voulons vérifier la capacité de l'instrument à détecter la proportion de répondants dits très attentifs et peu attentifs aux efforts réalisés en matière de SST au sein des établissements.

## 2. DÉMARCHE DE RECHERCHE

La démarche prévue initialement au protocole de recherche tel qu'approuvé par le comité scientifique de l'IRSSST consistait essentiellement à : - réviser les échelles développées dans le cadre de la recherche initiale à la lumière des résultats préliminaires obtenus; - utiliser une méthodologie de type Delphi pour recueillir l'opinion d'experts sur la validité de construit des indices qui composent chacune des dimensions de l'instrument; - constituer un large échantillon contrasté de répondants provenant du secteur imprimerie pour répondre au questionnaire de deuxième génération; - appliquer la batterie de tests statistiques traditionnellement utilisée pour statuer sur la validité interne et externe de l'instrument et confirmer son potentiel discriminant.

Contrairement à nos attentes, la stratégie traditionnelle de validation statistique de l'outil s'est révélée inadéquate en raison de la nature particulière des échelles que notre instrument comporte. Ce type de situation étant relativement peu connu du milieu de la recherche, nous traiterons de la question de façon détaillée. Nous justifierons la procédure que nous avons adoptée en nous appuyant sur des travaux relativement récents du domaine de la statistique.

### 2.1 Version préliminaire du questionnaire de 2<sup>e</sup> génération

Malgré l'utilisation d'un échantillon relativement petit (70 répondants), l'analyse détaillée des résultats obtenus au questionnaire de première génération nous a permis de constater que nos échelles de mesure étaient d'une nature formative plutôt que réflexive. La différence entre ces deux types de variables est discutée à la prochaine section. La version préliminaire du questionnaire de deuxième génération a aussi pris en compte la recension des écrits dans le domaine et le modèle hiérarchique des préoccupations de mesure en SST que nous avons développé dans l'étude initiale (Roy, Bergeron et Fortier, 2004). De plus, nous avons décidé de modifier le format de l'échelle de mesure de type Likert en cinq points d'ancrage pour une

échelle comportant dix points d'ancrage. Il a été démontré que ce genre de modification apportée au format de réponse n'altère pas la validité interne des instruments (Matell et Jacoby, 1971). Ce changement avait l'avantage de produire une échelle mieux ventilée et adaptée aux techniques d'analyse statistique multivariées que nous entendions utiliser. L'instrument a été soumis pour commentaire à un comité aviseur composé de représentants patronaux et syndicaux bien au fait de la problématique de la mesure en santé et sécurité du travail qui nous ont rassurés sur le caractère pertinent et utile de l'outil du point de vue des praticiens. La version préliminaire a par la suite été adaptée pour être testée auprès de trois entreprises représentatives du secteur de l'imprimerie au Québec. L'un des trois établissements faisait partie du tiers inférieur de son secteur pour ses résultats en SST alors que les deux autres faisaient partie du tiers supérieur. Cette étape de vérification sur le terrain nous est apparue nécessaire pour nous assurer que les modifications effectuées à la version originale n'affectaient pas négativement la validité nomologique de l'instrument. De plus, nous voulions nous assurer que le caractère convivial de l'outil était maintenu du point de vue des répondants.

## 2.2 Étude de la validité de l'instrument

La procédure de validation d'un instrument de mesure dépend de la nature des variables ou des construits étudiés. Comme nos échelles de mesure nécessitent un traitement statistique particulier par rapport à ce que l'on retrouve habituellement dans les études en sciences sociales, nous présentons dans les lignes qui suivent l'argumentation sur laquelle nous nous appuyons pour justifier notre démarche.

Tel que Bollen (1989, p. 222) le précise, l'approche classique de l'analyse de la validité statistique d'un outil de mesure s'effectue via les méthodes d'équations structurelles. Une fois le nombre de construits dont l'existence est à confirmer est fixé, cette approche consiste à étudier les propriétés intrinsèques, ou psychométriques, du modèle de mesure et se déroule essentiellement en cinq étapes.

- 1- La première étape aborde l'étude de l'unidimensionnalité des construits de manière à s'assurer que chaque item soit en lien avec un seul et unique construit latent.
- 2- La seconde étape s'attarde à la cohérence interne des construits par l'entremise de l'appréciation de l'ampleur des coefficients alpha de Cronbach et rho de Joreskog (Roussel et al., 2002).
- 3- La troisième étape s'attaque à la validité convergente où le chercheur tente de quantifier dans quelle mesure les items sont bien en lien avec le construit auquel il se rapporte.
- 4- La quatrième étape s'intéresse à la validité discriminante des construits. Elle a pour but de vérifier que la mesure de chacune des variables latentes est bien spécifique.
- 5- La cinquième et dernière étape s'intéresse à la validité nomologique de l'outil de mesure. Plus précisément, le chercheur s'assure de la cohérence entre ce que mesure l'instrument (les résultats) et la réalité étudiée.

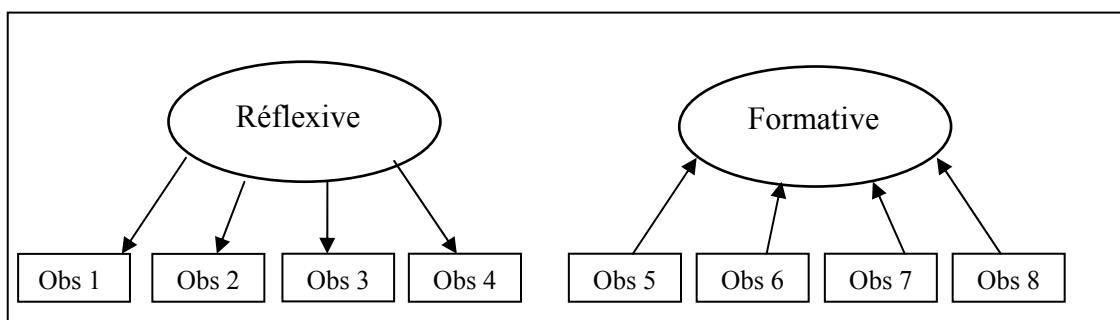


Cependant, Bollen (1989), Chin (1998) et Rossiter (2002) précisent que cette approche classique sous-entend que les variables utilisées dans le modèle de mesure sont de nature réflexive et que cette approche, aussi classique soit-elle, ne s'applique pas lorsque les construits latents sont de nature formative, ce qui est le cas pour l'ensemble des échelles de notre outil de mesure. Rossiter (2002) et de même que Bagozzi (1994, p.333) précisent que, de ces cinq étapes, les quatre premières ne sont pertinentes que dans les cas où les construits latents sont réflexifs.

En s'adressant aux chercheurs, Bollen (1989), Bagozzi (1994), Chin (1998) et Rossiter (2002) estiment qu'un des enjeux les plus importants consiste à faire la distinction entre un construit latent, ou concept, dont l'attribut est réflexif d'un autre dont l'attribut est formatif. La figure 2 présente la différence conceptuelle entre ces deux types de construits. Le sens des flèches est important. Plus précisément, un changement dans une variable latente réflexive aura un impact sur les valeurs des variables observables ( $obs_i$ ,  $i=1,2,3,4$ ), car elles sont le reflet de la variable latente. Dans le cas formatif, ce sont plutôt les changements dans les variables observables ( $obs_i$ ,  $i=5,6,7,8$ ) qui causent (ou forment) des changements au niveau de la variable latente.

Les réponses aux items observables ( $obs_i$ ,  $i=1,2,3,4$ ) liés à un construit intangible ou latent de type réflexif ont tendance à partager une ampleur similaire puisque leurs variations représentent le reflet d'une même entité latente sous-jacente commune. En fait, les variables observées ( $obs_i$ ,  $i=1,2,3,4$ ) auront tendance à être hautement corrélées. Par exemple, l'habileté orthographique (construit latent réflexif) d'un ensemble d'élèves peut être mesurée en exposant ces derniers à divers textes dictés. Les scores obtenus d'une dictée à l'autre devraient partager beaucoup de corrélation puisqu'ils reflètent un même construit latent. En somme, les items servant à la mesure du concept latent réflexif peuvent être interchangeable et le praticien recherche alors une notion d'unidimensionnalité. Basé sur ces constatations, il est compréhensible que les méthodes d'analyse factorielle ainsi que l'alpha de Cronbach, essentiellement basés sur les corrélations, soient des techniques statistiques adaptées aux variables réflexives.

Figure 2 : Différence conceptuelle entre des variables latentes réflexives et formatives



Les construits formatifs sont utilisés lorsque le chercheur désire mesurer un concept intangible, ou latent, qui est composé de plusieurs facettes distinctes. Les items observables ( $obs_i$ ,  $i=5,6,7,8$ ) associés à ce type de construit s'apparentent à une liste de vérification d'activités ou d'items caractérisant le concept latent à mesurer. Et, tel que présenté à la figure 2, c'est la considération et la mesure de l'ensemble de ces items qui forment le construit latent en question. Le construit

traitant de la « conformité du milieu de travail », présenté dans le tableau 2, constitue un exemple de construit latent formatif.

Tableau 2 : Items constitutifs du construit latent « conformité du milieu de travail »

1. Les mécanismes de protection requis sont installés sur les équipements et la machinerie
2. L'entretien préventif est réalisé sur les équipements
3. L'organisation fournit les équipements de protection individuels (EPI) requis pour le travail
4. L'organisation respecte les règlements relativement au bruit, à la qualité de l'air, etc.
5. Les postes de travail sont ajustables aux caractéristiques des employés
6. L'organisation fait respecter les procédures sécuritaires de travail (ex. : cadenassage, espace clos, etc.)

Ce construit dresse la liste des principales variables observables qui, selon les experts, se doivent toutes d'être prises en considération pour qu'un milieu de production soit conforme aux normes en santé et sécurité du travail. Contrairement aux variables réflexives, les réponses aux items liés à un construit latent formatif n'ont pas nécessairement tendance à partager une ampleur similaire, elles peuvent ne pas corrélérer voire même aller dans des sens inverses (Bollen, 1989, p. 223). En effet, dans un même milieu de travail, l'environnement de travail (équipement, air, bruit) peut être conforme aux normes environnementales en vigueur alors que les entretiens préventifs ne sont pas réalisés régulièrement sur les équipements.

Selon Rossiter (2002), à cause de la nature « multidimensionnelle » de ce type de variable latente, les analyses traditionnelles de type factoriel ou encore les techniques usuelles mesurant la cohérence interne des construits (alpha de Cronbach ou le  $\rho$  de Joreskog) ne s'appliquent pas. De plus, Rossiter précise que « *a formed attribute does not follow the domain sampling model. This means that items are not interchangeable, that is, items cannot be added or deleted from the scale (for reliability or any other reason)* ». En somme, Rossiter (2002) et Bagozzi (1994) estiment que pour étudier la validité d'une variable latente formative, seules les validités de contenu et nomologique restent à vérifier.

En réaction à l'article de Howell et al. (2007) qui remettent en question l'utilisation des variables formatives dans les recherches, Bollen (2007) démontre que les problématiques soulevées par ces auteurs sont reliées à des problèmes de spécification des modèles utilisés. À cet effet, afin d'étudier si un modèle formatif ou réflexif est bien adapté à un contexte donné, Bollen et Ting (2000) proposent d'utiliser l'analyse de la « Vanishing Tetrad » que nous traduisons par « tétrade évanescence ». Plus précisément, ce test propose une règle empirique qui permet de vérifier si un modèle réflexif ou formatif est davantage approprié.

Pour un construit donné, cette méthode statistique vise essentiellement à déterminer dans quelle mesure les écarts de covariances entre les groupes d'items, pris quatre à la fois, sont ou non très faibles. En vertu de cette analyse, pour des variables observables X1, X2, X3 et X4, une tétrade  $\tau$  est par définition un produit de quatre covariances et s'exprime ainsi :

$$\tau_{1234} = \sigma_{12}\sigma_{34} - \sigma_{13}\sigma_{24}$$

Algébriquement, pour un même sous-ensemble de variables, il existe une famille de tétrades dont certaines sont redondantes ( $\tau_{1234} = \tau_{1324}$ ). Statistiquement, lorsque les variables X1, X2, X3 et X4 font partie intégrante d'un même construit réflexif, il se trouve que leurs covariances présentent des ampleurs similaires et les tétrades associées à ce modèle auront tendance à être petites ( $\tau_{1234} \rightarrow 0$ ). Puisque des items qui ne sont pas en lien (ne covarient pas ensemble) ont tendance à produire de grandes tétrades, comme c'est le cas pour les variables formatives, le test d'hypothèse proposé par Bollen et Ting (2000) repose sur la faiblesse des ampleurs des tétrades non redondantes, d'où l'appellation de la technique « tétrade évanescence ». Ce type d'analyse statistique s'effectue par l'entremise de la macro SAS développée par Bollen et Ting (1993). Lorsque appliquée à un contexte comme le nôtre, l'hypothèse nulle du test proposé suppose que le modèle spécifié a priori est réflexif. Théoriquement, les items associés, pris par sous-groupes de quatre, produiront de petites tétrades. Statistiquement, cette hypothèse correspond à une tétrade nulle ( $H_0 : \tau = 0$ ). Les auteurs précisent que le rejet de cette dernière correspond à rejeter le modèle réflexif en faveur d'un modèle formatif.

### 2.3 Résultats des tests du questionnaire de 2<sup>e</sup> génération

Le tableau 3 ci-dessous illustre la ventilation de notre échantillon en fonction des entreprises. Le taux moyen de réponse des trois entreprises du secteur de l'imprimerie se situe globalement à 62 % et le nombre total de répondants se situe à 269 employés. Le taux de réponse plus faible dans le cas de la seconde entreprise, soit 42 %, s'explique par le fait que les gestionnaires ont dû limiter la participation des employés afin d'assurer le maintien des activités de production et des conditions de sécurité sur le lieu de travail. Ces taux sont cependant comparables ou supérieurs à ceux que l'on retrouve habituellement dans les études du même type.

Tableau 3 : La ventilation de l'échantillon

	Taux de réponse	Fréquence	% de l'échantillon maître
Entreprise 1	.71 %	.152	56,5
Entreprise 2	.42 %	.47	17,5
Entreprise 3	.73 %	.70	26,0
Total	.62 %	.269	100,0

Compte tenu du stade de développement du questionnaire, la présente démarche statistique se veut exploratoire et l'objectif consiste à donner des pistes d'amélioration du questionnaire aux chercheurs. Les colonnes du tableau 4 présentent les résultats des analyses en tétrade pour chacun des neuf construits initiaux du questionnaire. L'interprétation du tableau 4 s'effectue

ainsi : toutes les dimensions pour lesquelles la valeur-p associée est inférieure à  $\alpha=0.05$ <sup>1</sup> illustrent que le modèle formatif proposé est adéquat. En observant les résultats du tableau 4, il est possible de se rendre compte que, dans le questionnaire complet, les construits relatifs à la responsabilité des employés, aux normes et comportements, à l'amélioration continue et aux activités rejettent l'adéquation d'un modèle formatif et constituent donc des thématiques à revoir en terme de validité de contenu.

Cette étape de l'étude ayant révélée que certaines échelles devaient être revues, l'analyse de la validité nomologique, qui vise à déterminer si l'outil peut être corrélé avec un autre outil qui, tout en étant différent, mesure des aspects comparables de la réalité étudiée, n'a pas été abordée. Malgré tout, les résultats des trois entreprises ont permis aux chercheurs de bien distinguer les entreprises ayant des résultats SST supérieurs à la moyenne du secteur de celles ayant des résultats inférieurs à la moyenne, illustrant un fort potentiel de validité nomologique.

Tableau 4 : Les résultats du test des tétrades évanescentes

Thématiques	Questionnaire 2 <sup>e</sup> génération		
	$\chi^2$	dl	valeur-p
Systèmes	29,998	8	0,000
Engagement de la direction	41,433	20	0,003
Responsabilités des employés	47,707	35	0,074
Normes et comportements	47,121	35	0,083
Amélioration continue	14,741	12	0,256
Activités	0,968	3	0,809
Structures	19,301	9	0,023
Communications	46,434	27	0,011
Conformité du milieu	19,450	8	0,013

\* Pour calculer une tétrade, il faut au moins quatre items dans un construit.

## 2.4 Méthode Delphi – Élaboration du questionnaire de 3<sup>e</sup> génération

Compte tenu des résultats préliminaires encourageants obtenus avec le questionnaire de deuxième génération auprès des répondants du secteur de l'imprimerie, nous avons poursuivi nos efforts de validation des contenus des construits auprès de spécialistes académiques du domaine de la SST. Il s'agissait essentiellement de revoir l'ensemble des thématiques, notamment celles ciblées à l'étape précédente, du questionnaire de deuxième génération et de déterminer s'il était possible de réduire le nombre d'items retenus pour constituer chacune de nos échelles de mesure, sans affecter négativement la validité de contenu des construits du questionnaire. Le jugement

<sup>1</sup> La valeur-p est basée sur la grandeur de la statistique du Chi-deux  $\chi^2$  auquel est associé naturellement un degré de liberté (dl). Plus la valeur du  $\chi^2$  est petite, plus  $H_0 : \tau = 0$  est vraisemblable et plus la valeur-p associée est grande. Pour plus de détails, voir Bollen et Ting (1993).

d'experts était essentiel à cette étape à cause du caractère formatif des indicateurs qui composent nos échelles de mesure. En effet, contrairement aux échelles réflexives dont les indicateurs peuvent être éliminés parce qu'ils sont interchangeables dans le sens où ils sont tous sensés « refléter » le même construit latent, les échelles formatives comportent des indicateurs qui s'additionnent les uns les autres pour « former » le construit latent. Ainsi, chaque retrait ou ajout d'item risque d'affecter la validité de contenu du construit. Nous avons donc orienté la réflexion des évaluateurs sur la détermination de la pertinence des items en les familiarisant avec les définitions des construits formatifs mesurés par le questionnaire.

Nous avons opté pour la méthodologie Delphi qui nous est apparue comme étant bien adaptée à notre situation. La démarche prévue par la méthodologie Delphi consiste essentiellement à recueillir, dans un format structuré, l'opinion anonyme de spécialistes sur un sujet donné pour ensuite permettre à ces mêmes personnes de réagir par écrit aux réponses agrégées de tous les participants et de revoir leur position initiale le cas échéant. L'objectif visé consiste à en arriver à un consensus le plus large possible entre les participants concernant les positions adoptées et ultimement les choix effectués. L'anonymat et la rétroaction permettent aux participants d'apprécier leur niveau de convergence et de débattre des désaccords en se fondant strictement sur la qualité intrinsèque des arguments soulevés. La justification des réponses des participants en réaction au feedback des autres est un avantage particulier de la méthode que l'on ne peut obtenir par des stratégies d'enquêtes ou d'entrevues individuelles.

L'effet de la désirabilité sociale ou du statut des participants est neutralisé par le processus puisque ces derniers ne savent pas qui sont les autres répondants. Cette caractéristique constitue un avantage sur la méthode des groupes de discussions qui peuvent être mobilisés par un participant dominant. Seuls les membres de l'équipe de recherche qui traitent l'information recueillie connaissent les participants. Ainsi, la méthode permet d'obtenir le jugement d'experts sans que ces derniers n'aient à se confronter les uns les autres de façon directe (Dalkey et Helmer, 1963). Cette méthode mise au point dans les années 1950 a largement été utilisée dans le passé pour la réalisation de recherches prospectives et dans des études visant à classer des éléments complexes d'un phénomène par ordre de priorité. Son utilisation est fréquente dans le champ disciplinaire des systèmes d'information de gestion et elle est considérée comme pertinente pour contribuer à la validité de construit dans le design d'études, ce qui correspond bien à notre situation (Okoli et Pawlowski, 2004).

### *Étape 1 : Choix des participants*

Contrairement aux enquêtes par questionnaire, le choix de l'échantillon de participants n'est pas réalisé au hasard puisque l'objectif ne consiste pas à généraliser l'information recueillie à une population donnée. Il s'agit plutôt d'un échantillon délibéré d'informateurs dont le point de vue est jugé opportun compte tenu de l'objectif de l'étude. Dans ce cas-ci, nous devons identifier des personnes compétentes en recherche pour qu'elles exercent leur jugement sur le questionnaire en tant qu'instrument de mesure et compétentes en santé et sécurité du travail puisqu'il s'agissait aussi de porter un jugement sur la pertinence des items retenus pour chacun des construits visés dans le questionnaire. Finalement, comme le questionnaire a été conçu pour être administré à une population d'ouvriers spécialisés et non spécialisés du Québec, les experts devaient être en

mesure d'exercer leur jugement sur le niveau et les particularités du langage utilisé dans la formulation des items rédigés en français.

Une liste de neuf experts correspondants aux critères de sélection a été établie par les membres de l'équipe de recherche. Sur les neuf personnes sollicitées, cinq ont accepté de s'engager dans la démarche compte tenu de leur intérêt pour le sujet à l'étude et de leur disponibilité au moment de la réalisation du projet. Nous avons jugé ce nombre suffisant compte tenu de la spécificité de l'objet d'étude, de l'envergure de la tâche à réaliser pour les participants et du volume d'information à traiter qui serait généré par le processus. Les participants retenus proviennent de deux universités autres que l'université d'attache des auteurs de ce rapport et n'ont aucun lien immédiat avec les auteurs pour la production de communications scientifiques ou l'encadrement d'étudiants gradués. Leur neutralité par rapport au projet fait en sorte qu'ils peuvent être considérés comme étant des consultants recrutés en fonction de leur domaine d'expertise.

Les participants ont accepté que tout le travail réalisé soit complété sur support informatique et que les documents soient transmis à l'équipe de recherche par courriel. Les participants ont aussi accepté de maintenir la confidentialité quant à leur participation et sur le matériel à l'étude dans le cadre du projet. Une rémunération a été prévue pour rétribuer les participants à titre de consultant-expert. Ces derniers devaient s'engager à fournir l'équivalent de sept heures de travail réparties en deux ou trois occasions sur une période de quatre mois.

### *Étape 2 : Révision des échelles de mesure*

Les experts avaient comme tâche de déterminer si les items retenus pour former chacune des dimensions latentes étaient essentiels (à conserver), accessoires (à retirer) ou à reformuler. Ils avaient aussi la possibilité d'ajouter des items qu'ils jugeaient essentiels mais qui n'apparaissaient pas dans la version initiale du questionnaire. Ils pouvaient aussi suggérer de déplacer un item d'un construit à un autre.

Le matériel transmis aux participants pour procéder à la révision des échelles comprenait :

1. une lettre de transmission;
2. un texte d'introduction qui reprenait sommairement la genèse du projet, le cadre théorique à l'origine du questionnaire de même que la particularité des échelles de mesure qui sont de type formatif;
3. des instructions sur la façon de procéder pour appliquer les critères de retrait, de maintien, de reformulation ou d'ajout d'items compte tenu de l'objectif du questionnaire et de son utilisation dans le milieu de travail;
4. une définition de chacun des neuf construits;
5. un formulaire de réponse qui permettait de procéder à l'évaluation des items de même que de l'espace pour reformuler les items et commenter les échelles.

Les experts ont mis en moyenne trois heures pour réaliser leur évaluation dans le délai d'une semaine qui leur était imparti.

### *Étape 3 : Recherche de consensus*

Nous avons noté un degré d'accord élevé entre les experts en ce qui a trait à la pertinence des indicateurs retenus pour former chacune des dimensions qui composent l'instrument de mesure. Plusieurs reformulations d'items ont été proposées pour simplifier, spécifier ou présenter de façon plus concrète et pertinente le contenu visé par chacun des indicateurs. Les répondants ont suggéré de retirer une vingtaine d'items jugés soit accessoires, redondants, inappropriés ou difficiles à répondre par la population visée. Quelques rares ajouts couvrant des aspects jugés essentiels par certains experts ont été intégrés aux endroits proposés. Dans certains cas, ces ajouts ont engendré le retrait d'autres questions qui devenaient alors redondantes.

Toutes les suggestions des experts ont été analysées de façon individuelle par chacun des membres de l'équipe de recherche. Par la suite, des séances de travail en face à face ont permis de prendre des décisions consensuelles sur la formulation proposée pour chacun des items. Ce travail a permis de rédiger un nouveau document de travail qui présente la synthèse des recommandations qui ont été retenues. Ce second document de travail a été conçu de façon à permettre aux participants de signifier leur accord ou leur désaccord face aux modifications proposées par les autres experts et les formulations retenues par l'équipe de recherche.

Le texte original du questionnaire apparaissait dans la première colonne. La seconde présentait les reformulations, les ajouts et les retraits proposés. Les deux colonnes suivantes permettaient aux répondants de signifier leur accord ou leur désaccord face aux modifications. Un espace était prévu au bas de chaque section pour permettre aux répondants de présenter leur argumentation en appui à leur position.

Une nouvelle lettre de transmission indiquait aux répondants les directives pour transmettre leur accord ou désaccord face aux modifications apportées au questionnaire original. De plus, l'échelle de même que la définition conceptuelle des construits ont été à nouveau transmises. Une période de dix jours a été allouée aux experts pour réaliser l'exercice et retourner le document complété.

L'un des participants a dû se retirer à cette étape, parce qu'il n'était pas en mesure de compléter ce travail dans le temps imparti. Les participants ayant une copie électronique de leur réponse initiale étaient en mesure d'apprécier et de déterminer l'écart entre la version proposée, leur position initiale et la première formulation qui leur avait été soumise. Nous avons convenu de poursuivre des itérations auprès de nos participants jusqu'à ce que trois évaluateurs sur quatre soient d'accord avec la formulation proposée sur chacun des items.

Ce niveau a été atteint dès le second tour. Les participants étaient unanimes sur l'ensemble des items sauf 10 indicateurs pour lesquels l'un ou l'autre des participants avait des réserves ou proposait des modifications mineures. À nouveau, l'équipe de recherche a révisé les propositions des experts et pris des décisions consensuelles concernant les recommandations des experts. Lors d'une entrevue de suivi, tous les experts se sont dits satisfaits du déroulement de l'activité et de l'amélioration évidente de la formulation du questionnaire final, soit le questionnaire de 3e génération.

## 2.5 Seconde collecte des données sur le terrain

La collecte des données requises pour procéder à la validation finale de l'outil a été réalisée au sein de deux établissements du secteur de l'imprimerie. Ce secteur a été privilégié à cause du fait que la version préliminaire avait été testée dans trois établissements de ce même secteur et que les regroupements départementaux de même que les postes de travail sont relativement homogènes d'une entreprise à l'autre dans l'ensemble du secteur. Lors de l'étude préliminaire, il avait été possible de sélectionner deux établissements reconnus pour leur performance en SST qui les classait dans le premier tiers de leur secteur pour leurs résultats en SST et un établissement se classait dans le tiers inférieur de son secteur, ce qui suggérait le potentiel de validité nomologique de l'instrument. Deux autres établissements dont les résultats en SST étaient inférieurs à la moyenne du secteur ont refusé de participer à l'étude.

## 2.6 Validation du questionnaire de 3<sup>e</sup> génération

La validation du questionnaire de troisième génération est, à cette étape, davantage une phase de confirmation que d'exploration. La validité de contenu ayant été confirmée lors du Dephi, les analyses en tétrades évanescents ainsi que l'étude de la validité nomologique seront les sujets traités dans cette section.

Le tableau 5 ci-dessous montre que le taux de réponse se situe à 75,17 % dans l'entreprise A et à 64,84 % dans l'entreprise B pour un nombre total de 389 répondants, le taux de participation nous apparaît satisfaisant pour ce type d'étude.

Tableau 5 : La ventilation de l'échantillon

	Taux de réponse	Fréquence	% de l'échantillon maître
Entreprise A	75,17%	220	56%
Entreprise B	64,84%	169	44%
Total		389	100%

Le tableau 6 présente les résultats des analyses en tétrades évanescents pour chacun des neuf construits du questionnaire de 3<sup>e</sup> génération. Ces analyses permettent d'étudier dans quelle mesure l'approche formative est bien adaptée à chacun des construits révisés par l'approche Delphi. L'interprétation des résultats du tableau 6 s'effectue de la même manière que pour le questionnaire de la génération précédente. Toutes les dimensions pour lesquelles la valeur-p associée est inférieure au seuil de signification ( $\alpha=0.05$ ) rejettent le modèle réflexif. Les résultats<sup>2</sup> inclus dans le tableau 6 illustrent que l'approche formative est adéquate à l'ensemble de la problématique étudiée, témoignant par le fait même de la pertinence et de l'efficacité du Delphi.

<sup>2</sup> Les statistiques utilisées ici sont les mêmes que celles du tableau 4.



Tableau 6 : Les résultats du test des tétrades évanescentes

Thématiques	Questionnaire 3 <sup>e</sup> génération		
	$\chi^2$	dl	valeur-p
Systèmes	17.38	5	0.0038
Engagement de la direction	61.73	14	0.0000
Responsabilités des employés	22.47	9	0.0075
Normes et comportements	53.88	14	0.0000
Amélioration continue	15.07	2	0.0005
Activités	30.33	9	0.0004
Structures	45.45	9	0.0000
Communications	30.81	9	0.0003
Conformité du milieu	28.20	9	0.0009

\* Pour calculer une tétrade, il faut au moins quatre items dans un construit.

La validité nomologique d'un outil de mesure consiste à vérifier si les résultats de l'outil sont bel et bien en lien avec la réalité étudiée. Généralement, ce type de validité se discute lorsque d'autres chercheurs utilisent le même outil ou encore lorsque deux outils concurrents sont pris sur le même échantillon et que les résultats sont comparables. Vu le caractère innovateur de l'outil, ces stratégies sont présentement impossibles. En revanche, le développement du questionnaire a fait en sorte que plusieurs entreprises ont dû être étudiées avec des versions antérieures de l'instrument.

Lors du test de questionnaire de deuxième génération, nous avons demandé aux conseillers en SST de sélectionner des établissements considérés parmi le tiers supérieur en matière de performance et d'activités en SST, et des établissements classés dans le tiers inférieur. Les résultats au questionnaire ont permis de retracer le même classement que celui établi par les spécialistes du terrain. De plus, certains établissements avaient procédé à des audits de gestion de la SST dans les semaines ou les mois précédant la passation du questionnaire. De façon typique, les audits vérifient la mise en place d'un ensemble de pratiques, de systèmes et de processus les comparant à un cadre normatif réputé exemplaire en matière de SST. Ici encore, les résultats du questionnaire allaient dans le même sens que les résultats des audits. Ces maintes démonstrations nous permettent d'avancer que l'outil de troisième génération est valide au sens nomologique du terme.

En somme, l'étude de la validité de l'outil de mesure s'est déroulée en trois étapes. Dans un premier temps, la révision des questionnaires de première et de deuxième génération a permis de concevoir un instrument dont la validité de contenu des construits a été rehaussée et reconnue par un panel d'experts dans le cadre d'un Delphi. Ce type de validité signifie que l'outil mesure bien ce qu'il est censé mesurer. Dans un second temps, l'analyse en tétrades évanescentes a permis de démontrer la pertinence de l'approche formative de notre outil de mesure dans le contexte de la SST. Dans un troisième et dernier temps, la validité nomologique vérifiant le lien entre les

résultats et la réalité étudiée est existante. Compte tenu des résultats obtenus, des multiples interactions visant à raffermir l'outil, des précautions méthodologiques adoptées et de la collaboration active des praticiens et des experts académiques tout au long du processus, les chercheurs sont maintenant confiants quant au caractère valide de l'instrument.

## 2.7 Analyse de segmentation

Une fois l'étude de la portée scientifique de l'outil de mesure terminée, une segmentation<sup>3</sup> des employés a ensuite été réalisée. Tel que présenté dans le devis de recherche, les trois groupes pressentis dans les études précédentes étaient les suivants : les peu attentifs ou sévères, les baromètres et les très attentifs ou laxistes. Le tableau 8 présente les définitions préalablement établies dans l'avant-projet pour chacun de ces groupes de travailleurs. Les analyses statistiques de segmentation par profil de répondants ont été menées d'une manière exploratoire et indépendante sur les échantillons maîtres des deux entreprises qui se sont prêtées à notre étude.

Tel qu'anticipé, les analyses ont relevé la présence des trois groupes de répondants dans les deux institutions. De plus, les définitions initialement proposées dans l'avant-projet convenaient tout à fait aux groupements, soulignant ainsi la récurrence du phénomène. Ce résultat intéressant que nous avons déjà noté dans les études exploratoires initiales nous permet de suggérer que l'appréciation de la situation proposée par les observations du groupe des baromètres semble un meilleur reflet de la santé SST par contraste avec les deux autres groupes dont les positions opposées apparaissent trop extrémistes.

Tableau 7: Définitions des groupements

LIBELLÉ	DESCRIPTION
Les peu attentifs ou les sévères	Représente les individus qui jugent les actions menées en SST de manière plutôt sévère. Par exemple, ils perçoivent souvent les activités SST comme étant relativement peu ou pas présentes. Cependant, certains individus peu attentifs aux pratiques peuvent être mêlés à ce groupe.
Les baromètres	Noyau d'individus qui agit à titre de baromètre en ce qui concerne les pratiques SST. Entre le noyau des sévères et celui des laxistes, leur jugement est davantage nuancé concernant la réalité SST dans l'entreprise.
Les très attentifs ou les laxistes	Représente les individus qui perçoivent les actions menées en SST comme étant relativement fréquentes et nombreuses. Cependant, ce groupe contient aussi des employés davantage laxistes qui optent pour la position suivante : « Tout est parfait, ne pas en ajouter ».

Lors d'une analyse de segmentation, on cherche non seulement à nommer et à étudier les groupes, mais on cherche aussi à savoir si chacune des variables incluses dans le processus contribue à la différenciation des groupes. À cet égard, les valeurs-p à 0,000 associées aux valeurs « F de Snedecor » des tableaux 9 et 10 démontrent que chacun des neuf construits pris individuellement est utile pour différencier les groupes, et ce, pour chacune des deux entreprises.

<sup>3</sup> Selon la procédure proposée par Everitt et al. (2001), basée sur la mesure de Ward, une première solution a été obtenue en utilisant la méthode hiérarchique. Les barycentres de la solution optimale à trois groupes ont ensuite été utilisés comme point de départ dans une méthode non hiérarchique. Tel que recommandé, afin de vérifier la stabilité des groupements, cette procédure a été répétée en changeant à chaque reprise la mesure du calcul des distances.

Plus précisément, pour une variable incluse dans une table ANOVA (Analyse de variance), plus la valeur « F » observée est élevée, plus la valeur-p associée est petite et plus les niveaux de perception des groupes diffèrent sur cette variable, montrant par le fait même son pouvoir de différenciation. En rangeant les valeurs « F » en ordre décroissant d'importance, il est possible de noter que la force des différenciations entre les groupes pour chaque thématique varie selon l'entreprise, démontrant la capacité des thématiques proposées à bien capter les différentes réalités SST perçues dans les entreprises.

Finalement, afin de valider la stabilité de la segmentation proposée, des analyses discriminantes ont été réalisées sur les perceptions des employés de chacune des deux entreprises. Cette stratégie statistique consiste à vérifier si, sur la base des réponses initiales des travailleurs, il est possible de retracer le groupe auquel chacun a été identifié par la segmentation. Une statistique importante liée à la performance de ce type d'analyse est le pourcentage d'individus correctement classés. À cet égard, l'algorithme a été en mesure de retracer correctement les groupes dans 96,8 % et 97 % des cas, ce qui suggère la stabilité des groupements d'une entreprise à l'autre.

Tableau 8 : Table ANOVA – Entreprise A

**Sortie 1 : Table ANOVA – Entreprise A**

Établissement		Segment		Erreur		F	Valeur-p
		Carrés moyens	dl	Carrés moyens	dl		
Entreprise A	Activités	215,009	2	1,364	217	157,619	,000
	Responsabilités	216,343	2	1,439	217	150,347	,000
	Structures	231,82	2	1,549	217	149,685	,000
	Engagement de la direction	177,88	2	1,306	217	136,169	,000
	Conformité	154,759	2	1,284	217	120,572	,000
	Normes de groupes	131,285	2	1,342	217	97,804	,000
	Communications	142,583	2	1,477	217	96,547	,000
	Améliorations	131,1	2	1,451	217	90,324	,000

Tableau 9 : Table ANOVA – Entreprise B

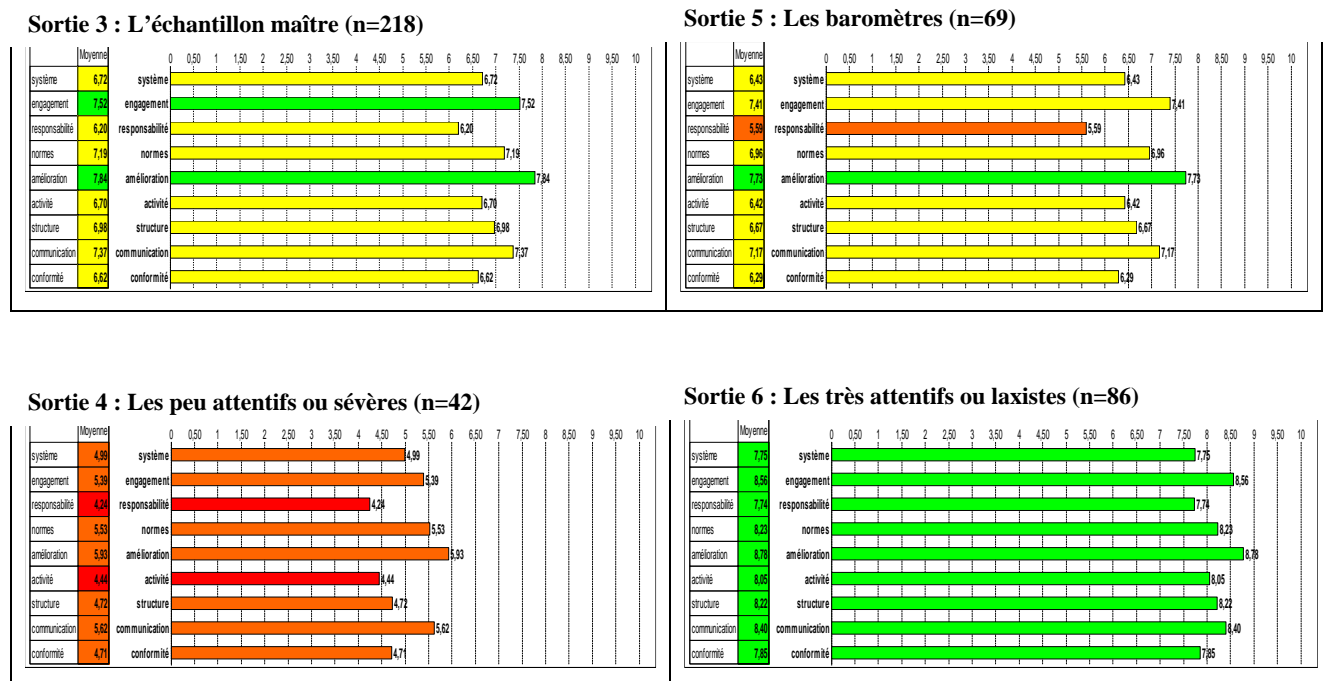
**Sortie 2 : Table ANOVA – Entreprise B**

Établissement		Segment		Erreur		F	Valeur-p
		Carrés moyens	dl	Carrés moyens	dl		
Entreprise B	Engagement de la direction	107,635	2	1,066	166	101,002	,000
	Communications	108,732	2	1,153	166	94,308	,000
	Structures	108,996	2	1,459	166	74,719	,000
	Systèmes de gestion	93,223	2	1,277	166	73	,000
	Activités	101,021	2	1,556	166	64,919	,000
	Améliorations	92,631	2	1,563	166	59,271	,000
	Responsabilités	81,222	2	1,996	166	40,691	,000
	Normes de groupes	64,028	2	1,671	166	38,319	,000
Conformité	48,917	2	1,555	166	31,45	,000	

D'un point de vue pratique, les résultats obtenus lors du diagnostic peuvent être présentés sous la forme de tableaux synthèses ou de tableaux de bord utilisables directement au sein des milieux par les intervenants en SST. À titre d'illustration, la figure 3 présente les tableaux de bord générés pour les neuf thématiques dans chacun des trois groupements de l'établissement B de même que pour l'échantillon complet (maître) de cet établissement. Globalement, les sorties illustrent l'important contraste entre les observations faites d'un groupe à l'autre concernant la même réalité vécue au sein de l'établissement, montrant l'avantage d'étudier plus particulièrement le point de vue du groupe baromètre.

Des tableaux de bord semblables à ceux que nous présentons ci-dessous peuvent être produits pour chacun des départements de l'établissement étudié. Il est ainsi possible de procéder à un diagnostic ciblé dans les unités de proximité de travail des répondants. Un guide d'accompagnement à la démarche du diagnostic SST proposée par cette étude sera mis à la disposition des intervenants en SST.

Figure 3 : Regroupements de répondants



### 3. DISCUSSION

Cette section du rapport traite des contributions de cette recherche sur le plan scientifique et sur le plan de la pratique de la mesure en SST. Sur le plan scientifique, cet outil est à notre connaissance le premier instrument validé qui permet de rendre compte dans un modèle hiérarchique intégré des observations des employés concernant les pratiques de SST en vigueur au sein de leur établissement. Contrairement aux enquêtes d'attitudes ou de climat qui reflètent

les impressions ressenties par les répondants, ce questionnaire permet de procéder à un diagnostic qui illustre l'écart entre la situation actuelle telle qu'observée par les participants et la situation idéale proposée par le cadre normatif qui a servi de base pour la conception de l'instrument. L'adoption d'un tel outil devrait faciliter la réalisation d'études subséquentes qui visent à rendre compte de l'effet de programmes, d'activités et d'interventions destinées à accroître la prévalence d'une culture de SST au sein des établissements.

La démarche utilisée a comporté plusieurs étapes pour en arriver à la production et à la validation du questionnaire final. L'accès à des sources multiples d'informations (comités aviseurs composés de praticiens, panel d'experts académiques, référentiels d'audits de gestion SST, échantillons de répondants provenant des milieux de travail), l'utilisation de méthodes variées de collecte de données (entrevues en profondeur, Delphi, questionnaire itératif, groupes de discussions) et l'utilisation de techniques d'analyse qualitative et quantitative (analyse de contenu, analyse statistique avancée des construits formatifs) nous permettent d'être confiants sur la valeur intrinsèque de l'instrument et sur le caractère convivial de son utilisation dans les milieux. Cet outil vient ajouter une contribution originale au courant de recherche préoccupé par la mesure des variables prédictives réputées avoir une incidence sur les résultats en matière de SST. Il permet notamment de commenter le niveau d'expression de la culture de SST au sein des établissements, qui s'appuie sur les valeurs, les normes et les pratiques des gestionnaires et des employés, de même que sur l'ensemble des systèmes de gestion des établissements.

L'une des avancées de cette étude sur le plan méthodologique a été de concevoir des échelles formatives plutôt que réflexives des construits latents mesurés. Ces échelles apparaissent mieux adaptées pour rendre compte de la réalité de la SST dans les milieux de travail. Par exemple, il apparaît plus juste de prétendre que l'atteinte de la conformité des installations avec la réglementation en vigueur est le résultat additif de l'ensemble des mesures qui ont été déployées à cet effet, plutôt que de prétendre que la conformité se reflète dans les installations. La différence ici est de taille parce que les échelles réflexives partent du principe que l'ensemble des items qui composent les échelles sont équivalents ou interchangeables. Dans le cas de la conformité, ce postulat est faux. En effet, dans certains milieux de travail, la technologie peut faire en sorte qu'il est extrêmement difficile de réduire le bruit ambiant à un niveau conforme alors que la réglementation concernant l'air ambiant peut être facilement rencontrée. L'élaboration des échelles formatives a nécessité l'utilisation de procédures de validation statistiques particulières (analyse des tétrades évanescents). L'utilisation avec succès de cette procédure ouvre la voie à la conception d'autres échelles de type formatif qui nous semblent particulièrement pertinentes dans le domaine de la SST.

Cette étude a aussi permis de confirmer l'existence de groupements de répondants. La récurrence et la stabilité des groupements qui ont émergé suggèrent qu'il sera éventuellement possible d'inférer indirectement la progression de la culture SST dans une entreprise en analysant la composition et le profil de réponse des groupes. On pourra supposer par exemple que la culture SST s'améliore si, entre deux études, le nombre d'individus qui composent le groupe des peu attentifs ou sévères diminue et que les observations du groupe des baromètres montrent des résultats à la hausse. Cette dernière remarque est particulièrement intéressante dans le cas où l'établissement ne déplore aucun accident et qu'il est difficile d'apprécier la progression en SST autrement que par la réalisation d'exercices coûteux d'audits de gestion. Nous sommes

conscients par ailleurs des faiblesses associées à l'utilisation de mesures perceptuelles qui sont particulièrement sensibles aux événements récents qui entourent la passation des questionnaires et au fait que les perceptions sont largement modifiées par les attentes. Ainsi, un groupe très sensibilisé peut devenir moins tolérant face à des événements auxquels il ne portait pas attention auparavant.

La philosophie d'apprentissage qui guide l'utilisation de l'instrument encourage et valorise l'identification des écarts et anomalies en matière de SST. On demande aux répondants de rendre compte de leurs observations de façon à orienter les efforts de leur comité de SST. Il s'agit d'apprendre à améliorer le milieu à partir des indications de toutes les personnes concernées. Cette orientation contraste avec les mesures de performance dont les effets pervers sont bien connus (sous-déclaration, dissimulation, contestation abusive, etc.). L'outil proposé est un instrument formatif qui ne sert pas au contrôle ni à l'appréciation de la performance contrairement à la mesure des résultats et à la réalisation des audits de gestion de la SST. S'il était utilisé avec une perspective évaluative ou de mesure de la performance, il susciterait les mêmes effets négatifs que nous avons signalés plus haut. C'est un outil qui accompagne et qui favorise une autre façon de vivre au travail qui se fonde sur l'apprentissage de la culture de la SST.

### *Sur le plan de la pratique*

De plus en plus d'auteurs reconnaissent l'importance de la culture organisationnelle lorsqu'il est question de SST en soulignant la nécessité d'impliquer les individus (Hudson, 2007). L'utilisation à grande échelle de l'outil au sein d'un établissement constitue à elle seule une intervention de sensibilisation qui permet aux répondants d'apprendre quelles sont les pratiques à adopter à tous les niveaux de l'organisation pour développer une culture de SST. L'instrument permet aussi aux responsables désignés de la SST (coordonnateur à la prévention, comité de SST, représentant à la prévention, etc.) d'apprendre dans quelle mesure leurs interventions sont perçues par les destinataires au sein des milieux de travail.

Le besoin d'avoir accès à des outils à la fois rigoureux et relativement simples d'utilisation est largement exprimé par les praticiens. Le large appui que ce projet de recherche a reçu de la part des associations sectorielles paritaires de SST et des associations syndicales et patronales intéressées par la question nous a conduits à concevoir des tableaux de bord, un guide d'utilisation et une session de formation à l'intention des intervenants en SST.

Depuis quelques années déjà, les entreprises et les observateurs avertis considèrent qu'un plateau a été atteint et qu'il est nécessaire de regarder la question de la SST sous un regard différent, si l'on veut encore améliorer les résultats. La mesure des lésions et des incidents, de même que la mise en place de systèmes de gestion de la SST ne permettent plus d'améliorer la situation de façon significative. Il faut maintenant faire en sorte qu'une culture de la SST s'instaure au sein des milieux si l'on veut aller plus loin. Nous croyons que l'outil de diagnostic pourra contribuer à cet avènement. Le succès d'une telle démarche demande cependant le respect d'un certain nombre de conditions que nous abordons ci-dessous.

La pratique nous a démontré qu'il est impératif que les membres de la direction de l'établissement s'impliquent personnellement et qu'au moins une personne se voit attribuer le mandat de réaliser les recommandations de l'étude au cours de l'année suivant la passation du questionnaire. Les entreprises qui ont utilisé les services d'un expert externe ou de leur association sectorielle ont pu bénéficier d'une expérience encore plus riche dans l'avancement de leur culture SST. Bien que les expérimentations aient été réalisées en nombre limité, nos entrevues de suivi réalisées après quelques mois du dépôt des diagnostics suggèrent qu'un accompagnement externe a avantage à être réalisé, que ce soit par les associations sectorielles ou par d'autres intervenants spécialisés.

Trois générations du questionnaire ont dû être développées pour en arriver à la version validée actuelle. Trois secteurs industriels se sont prêtés à notre étude, soient la foresterie, la fabrication de produits de métal et l'imprimerie. Chaque itération nous a permis d'améliorer la validité et de raffiner notre stratégie d'intervention auprès des milieux étudiés. Le questionnaire a été développé pour adopter une forme suffisamment générique pour être applicable à l'ensemble du secteur manufacturier. Nous croyons qu'il serait possible de concevoir des versions destinées à divers secteurs dont les activités comportent des risques importants pour la SST des travailleurs (ex. : construction, exploitation minière, transport, etc.).

#### *Limites et avenues de recherches futures*

Comme tout instrument de mesure, l'outil que nous proposons comporte un certain nombre de limites. D'abord, mentionnons que les versions de 2<sup>e</sup> et de 3<sup>e</sup> génération ont été validées uniquement dans le secteur de l'imprimerie, milieu pour lequel la chaire a obtenu une commandite et des sites de recherche. En contrepartie cependant, les experts réunis lors du Delphi ont analysé et traité les construits constitutifs de l'outil de mesure de façon à ce qu'ils s'adressent à toutes les entreprises du domaine de la production. Comme l'outil de première génération avait été élaboré à partir de deux autres secteurs d'activités (fabrication de produits en métal et électriques et secteur forestier), nous sommes confiants quant à l'applicabilité de l'outil dans la plupart des organisations de production. Aussi, l'outil s'adresse uniquement aux entreprises qui aspirent volontairement à mettre en place des milieux de travail intrinsèquement sécuritaires. Les entreprises qui ne sont pas prêtes à atteindre et à dépasser le niveau de base de la conformité de leurs installations aux normes en vigueur ne tireront aucun bénéfice de l'utilisation du questionnaire. En fait, la passation d'un tel questionnaire suscite des attentes de changements qui doivent être réalisés concrètement, sans quoi la situation risque d'être jugée plus insatisfaisante par la suite.

Une autre limite concerne la façon dont l'instrument sera déployé dans les milieux. Il existe un risque que les milieux dénaturent l'utilisation de l'outil pour en faire un instrument de contrôle et de mesure de la performance même si cela va à l'encontre de la philosophie d'apprentissage qui l'anime. Les directives données aux répondants lors de la passation du questionnaire leur indiquent que les données vont servir à orienter les efforts de leur comité de SST au cours de l'année. Il ne s'agit pas d'évaluer la performance mais bien d'identifier ce qui se fait et ce qui ne se fait pas au sein du milieu, de façon à initier les changements souhaités. Selon cette approche, les changements devraient avoir une incidence à plus ou moins long terme sur les résultats finaux en SST.

Si l'instrument est utilisé dans une perspective de sanction ou de mesure de performance, il ne pourra plus remplir sa fonction. En effet, la désirabilité sociale associée à l'idée de performance fera en sorte que les réponses seraient vraisemblablement biaisées comme c'est le cas actuellement pour la mesure des résultats.

La hiérarchisation des thématiques n'a pu être étudiée dans le cadre de ce projet. Des études ultérieures utilisant des échantillons appropriés seront nécessaires pour traiter de cette question et pour déterminer si certaines variables jouent le rôle d'antécédents sur le plan structurel. Ce pourrait être le cas par exemple de la dimension qui traite de la communication. De plus, une technique statistique spécialisée en segmentation a été appliquée à notre échantillon de répondants et des groupements ont été découverts. Il a été possible de classer les répondants selon leur niveau de sensibilité aux efforts réalisés en vue d'améliorer la SST. À titre d'illustration, le groupement qualifié de « peu attentif » ne parvient pas à percevoir des interventions reconnues par les autres répondants, le groupement qualifié de « très attentif » perçoit des améliorations là où les autres n'ont rien remarqué et le groupement « baromètre » adopte une position modérée entre ces deux extrêmes. Au-delà des possibilités d'élaboration d'interventions encore plus pertinentes et ciblées auprès de ces groupes, les résultats de cette segmentation des travailleurs offrent notamment une possibilité intéressante d'étudier la progression de la culture SST, et ce, même dans le cas où l'entreprise affiche zéro accident. En effet, la réduction de la proportion de « peu attentif » dans le temps suggère que la population des répondants est plus consciente des efforts déployés en SST, ce qui dénote un intérêt plus grand pour la SST. Actuellement, aucun outil ne permet cette avancée. Des recherches futures seront nécessaires pour explorer cette piste prometteuse.

Finalement, l'introduction des construits formatifs représente à notre avis, une avancée scientifique intéressante dans la mesure de l'état de la SST au sein des milieux de travail. Nous espérons que ces travaux amèneront d'autres chercheurs à produire des outils de même nature qui nous permettront de procéder à des analyses comparatives de validité nomologique et de raffiner nos instruments dans le futur.

## CONCLUSION

Les défis posés par ce projet de recherche étaient multiples. Il s'agissait d'abord de prendre un outil de diagnostic qui avait déjà fait ses preuves sur le plan de l'utilité et de la pertinence pratique, et d'en améliorer le contenu pour lui conférer les attributs de validité scientifique nécessaires à son utilisation dans des recherches futures.

Un second défi était associé à la forme même de l'instrument dont la nature formative des construits a forcé les chercheurs à sortir des sentiers battus pour être en mesure de procéder à des analyses statistiques peu usuelles qui n'auraient pu être envisagées sans la contribution d'un statisticien aguerri.

Le dernier défi visait essentiellement à démontrer qu'il est possible d'aborder la mesure en SST avec une perspective d'apprentissage à l'aide d'un outil formatif plutôt qu'une perspective de



performance qui se situe au niveau inférieur de la hiérarchie des préoccupations dans la mesure de la SST. Cette approche considère que les préoccupations des gestionnaires et des chercheurs doivent progresser vers la création et la mise en place d'une véritable culture de la SST au sein des milieux de travail plutôt que de se limiter aux variations dans les taux de fréquence et de gravité des lésions pour orienter leurs actions en matière de SST.

La poursuite de ces objectifs a nécessité le déploiement d'une stratégie de recherche complexe comprenant diverses méthodes de collecte de données et diverses sources d'information. Nous considérons que les efforts consentis ont porté les fruits attendus et que les modalités de transfert envisagées auprès des ASP et autres intervenants de la SST permettront une diffusion large de l'outil de diagnostic et de la philosophie de mesure qu'il incarne.

## APPLICABILITÉ DES RÉSULTATS

Toutes les étapes de la démarche utilisée, depuis la révision du questionnaire initial, en passant par la méthodologie Delphi jusqu'à la validation du questionnaire de 3<sup>e</sup> génération à l'aide d'un échantillon typique des destinataires à qui s'adresse le questionnaire, ont été réalisées de façon à nous assurer que l'instrument issu de cette recherche serait applicable dans les milieux de travail. Les associations sectorielles paritaires sont d'ailleurs en attente de recevoir la formation requise pour utiliser adéquatement l'instrument et se familiariser avec le guide d'utilisation qui leur servira de support par l'entremise des tableaux de bord. Les résultats du questionnaire seront accessibles en ligne sur le site web sécurisé à cet effet pour chaque organisation qui procédera au diagnostic. De plus, un format électronique du questionnaire permettra la saisie des données à partir des sites intranet des établissements. Le lecteur intéressé à prendre connaissance de notre argumentation sur le caractère éminemment applicable des résultats de cette recherche peut se référer à la section discussion du présent rapport.

## RETOMBÉES ÉVENTUELLES

Bien que l'on ne puisse jamais présumer de l'impact effectif des résultats d'un projet de recherche vers les milieux auxquels il se destine, nous pouvons déjà signaler l'intérêt suscité par l'outil et la philosophie de mesure qui l'inspire. Les réactions des lecteurs à un article récent publié sur notre perspective de mesure dans la revue *Travail et santé* a amené l'éditeur à nous demander de produire un second article qui traitera spécifiquement de l'outil d'autodiagnostic et de son utilisation. Deux ASP ont demandé l'autorisation de reprendre les contenus de l'article pour le diffuser dans leur bulletin d'information.

Une session de formation sera offerte à l'hiver ou au printemps 2008 pour permettre aux intervenants de se familiariser avec la démarche de diagnostic développée dans le cadre du projet. Tous ces intervenants deviendront des multiplicateurs de l'approche auprès des milieux.

Sur le plan scientifique, le *Journal of Safety Research* a déjà publié un article que nous avons soumis qui présentait l'avancement de nos travaux de mesure. La revue *Piste* avait aussi fait de même pour la communauté scientifique francophone. Ces manifestations d'intérêt nous portent à croire que notre instrument de diagnostic a de bonnes chances d'obtenir une diffusion intéressante auprès des milieux.

## ARTICLES SCIENTIFIQUES PRODUITS DANS LE CADRE DE LA SUBVENTION

Cadieux, Jean, Roy, Mario et Desmarais, Lise (2006). A Preliminary Validation of a New Measure of Occupational Health and Safety. *Journal of Safety Research*, 37(4).

Roy, M., Cadieux, J., Desmarais, L. (2005). Amélioration de la performance en SST: les résultats vs les prédicteurs. *PISTES*, 7(2). <http://www.pistes.uqam.ca/v7n2/articles/v7n2a10.htm>

## RÉFÉRENCES

Arezes P. M, Miguel S.A. (2003). The role of safety culture in safety performance measurement. *Measuring Business Excellence*, 7(4), 20-28

Bagozzi R. (1994). *Principle of marketing research*. Cambridge, Ma, Blackwell.

Bollen K.A. (2007). Interpretational confounding is due to misspecification, not to type of indicator: Comment on Howell, Breivik and Wilcox (2007). *Psychological Methods*, 12(2), 219-228.

Bollen K. A., Ting K.-F. (2000). A tetrad test for causal indicators. *Psychological methods*, (5)1, 3-22.

Bollen K. A., Ting K.-F. (1998). Bootstrapping a test statistic for vanishing Tetrad. *Sociological methods & research*. Sage publication, 27(1), 77-102.

Bollen K. A., Ting K.-F. (1993). Confirmatory tetrad analysis. In P.V. Marsden (Ed.). *Social methodology*. Cambridge, MA: Blackwell, 23, 147-176.

Bollen K. A. (1989). *Structural equations with latent variables*. New-York, John Wiley & Sons.

Booth, R. T. (1993). Monitoring Health and Safety Performance – an Overview. *Journal of Health and Safety*, 9, 5-16.

Budworth, N. (1996). Indicators of safety performance in safety management. *The Safety & Health Practitioner*, 14(11), 23-29.

Cadieux, J., Roy, M. et Desmarais, L. (2006). A Preliminary Validation of a New Measure of Occupational Health and Safety. *Journal of Safety Research*, 37(4).

Cheyne, A., Oliver, A., Tomas, J.M., Cox, S. (2002). The architecture of employee attitudes to safety in the manufacturing sector. *Personnel Review*, 31(5–6), 649–670.

Chin W. (1998). Issues and Opinion on structural Equation Modeling. *MIS Quarterly*, 22(1).

- Dalkey, N., Helmer, O. (1963). An Experimental Application of the Delphi Method to the Use of Experts. *Management Science*, 9(3), 458-467.
- Everitt B., Landau S. Leese M. (2001) Cluster Analysis (4th ed), Oxford, Oxford University Press.
- Flin R., Mearns K., O' Connor P., Bryden R. (2000). Measuring the safety climate: identifying the common features. *Safety Science*, 34, 177–192.
- Gomez, M.R. (1998). Exposure Surveillance tools needed in agency GPRA plans. *American Industrial Hygiene association Journal*, 59(6), 371-374.
- Guldenmund, F.W. (2000). The nature of safety culture: a review of theory and research. *Safety Science*, 34, 215–257.
- Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C. (1998). *Multivariate data analysis with reading*. 5e ed. Upper Saddle River, NJ, Prentice Hall.
- Hale, A.R. (2000). Editorial: culture's confusions. *Safety Science*, 34, 1–14.
- Howell R.D., Breivik E., Wilcox J. (2007). Reconsidering formative measurement. *Psychological Methods*, 12(2), 205-218.
- Matell, M.S., Jacobi, J. (1971). Is there an optimal number of alternative for Likert scale Items? *Educational and Psychological Measurement*, 31, 657-674.
- Mitchell, R. (2000). Measuring OHS Performance and Developing Performance Indicators. *Journal of Occupational Health & Safety*, 16(4), 319-323.
- Hudson P. (2007). Implementing a safety culture in a major multi-national. *Safety Science*, 45, 697–722.
- Hudson, P.T.W., Verschuur, W.L.G., Parker, D., Lawton, R. (1998). *Intrinsic motivation for HSE: The Hearts and Minds Program*. Report for Shell International.
- Lawrie, M.J., Parker, D., Hudson, P.T.W. (2006). Investigating employee perceptions of a framework of safety culture maturity. *Safety Science*, 44, 259–276.
- O'Brien, D.P. (2000). *Business measurements for safety performance*. Lewis publishers, Washington, 118 p.
- Okoli, C., Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information & Management*, 42(1), 15.

- Parker, D., Lawrie, M.J., Hudson, P.T.W. (2006). A framework for understanding the development of organisational safety culture. *Safety Science*, 44, 551–562.
- Reason, J.T. (1998). *Achieving a safe culture: theory and practice*. *Work and Stress* 12, 293–306.
- Redinger C.F., Levine S.P. (1999). *Occupational Health and safety Management System Performance Measurement*. A Publication of the American Industrial Hygiene association.
- Rohatgi V.K. (1976). *An introduction to probability theory and mathematical statistics*. NY, John Wiley & Sons.
- Rossiter John R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19(4), 305-417.
- Roussel P., Durrieu F., Campoy É., El Akrem A. (2002). Méthodes d'équations structurelles: recherche and applications en gestion. *Économica*, Paris.
- Roy, M., Bergeron, S. et Fortier L. (2004). *Développement d'instruments de mesure de performance en santé et sécurité du travail à l'intention des entreprises manufacturières organisées en équipes semi-autonomes de travail*. Rapport de recherche R-357 IRSST.
- Shannon, H.S., Lowe, G.S. (2002). How many injured workers do not file claims for workers' compensation benefits? *American journal of industrial medicine*, 42, 467-473.
- Shaw, A., Blewett, V. (1995). Measuring Performance in OHS: Using Positive Performance Indicators. *Journal of Occupational Health and Safety*, 11(4), 353-358.
- Simpson, I., Gardner, D. (2001). Using OHS Positive Performance Indicators to Monitor Corporate OHS Strategies. *Journal of Occupational Health and Safety*, 17(2), 125-134.
- Stricoff, S.R. (2000). Safety performance measurement: Identifying prospective indicators with high validity. *Professional Safety*, 45(1), 36-40.
- Toellner, J. (2001). Improving Safety & Health Performance: Identifying & Measuring Leading Indicators. *Professional Safety*, 46(9), 42-47.
- Ward, S. (2000). One size doesn't fit all: Customizing helps merge behavioural and traditional approaches. *Professional Safety*, 45(3), 33.
- Westrum, R. (1993). Cultures with requisite imagination. In: Wise, J., Stager, P., Hopkin, J. (Eds.). *VeriWcation and Validation in Complex Man–Machine Systems*. Springer, New York.
- Zhang H., Wiegmann D.A., von Thaden T.L., Sharma G., Mitchell A. A. (2002). Safety culture: a concept in chaos? *Proceedings of the 46<sup>th</sup> annual Meeting of the human factor and ergonomics society*, Human factor and Ergonomics Society, Santa Monica.