

CARACTERISTIQUES DE L'INFORMATION, SURCHARGE D'INFORMATIONS ET QUALITE DE LA DECISION

Philémon Rakotonjanahary

*Service de l'enseignement des sciences comptables, HEC Montréal, 3000, chemin de la Côte-Ste-Catherine, Montréal (Québec) Canada H3T 2A7, tél. (514) 340 6517, E-Mail :
Philemon.Rakoto@hec.ca*

Résumé

Dans un contexte de surcharge d'informations, la psychologie cognitive nous apprend que le fait d'augmenter la quantité d'information disponible nuit à la qualité de la décision. Mais il est plus subtil de faire la distinction entre données pertinentes, redondantes et non pertinentes et d'examiner leurs effets respectifs sur la qualité de la décision. L'auteur pose l'hypothèse selon laquelle seules les données redondantes nuisent à la qualité de la décision. L'auteur s'est livré à une expérience dans le cadre de laquelle les sujets simulaient une décision de prêt commercial. La quantité de données disponibles a été manipulée de façon qu'elle soit faible, moyenne ou élevée, ce qui a donné lieu à trois versions de la demande de crédit. Les prêteurs prévoient la santé financière de six entreprises emprunteuses. L'expérience a permis de constater que l'augmentation de la quantité de données redondantes nuit à la qualité de la décision. Les résultats obtenus suggèrent qu'une stratégie d'information possible consiste à cibler les utilisateurs et à leur fournir seulement les données pertinentes à leur tâche. Les résultats démontrent aussi l'utilité des aides à la décision dans un contexte de surcharge d'informations.

Mots clés : *surcharge d'informations, données pertinentes et redondantes, expérience en laboratoire, prédiction de défaillance financière.*

Abstract

In a context of information overload, the cognitive psychology shows that the fact of increasing the quantity of available information harms in decision quality. But it is more subtle to make the distinction between relevant, redundant and not relevant data and to examine their respective effects on decision quality. The author predicts that only redundant data damages decision quality. The author conducted an experiment study within the framework of which the subjects simulated a commercial loan decision. The quantity of available data was manipulated so that there is a weak, an average or a high level, which gave place to three versions of the application for credit. The lenders predicted the financial health of six companies borrowers. The results of the study show that the increase of the quantity of redundant data harms in decision quality. The obtained results suggest that a strategy of information consists in targeting the users and in supplying them only the relevant data in their task. The results demonstrate as well the utility of decision aids in a context of information overload.

Keywords : information overload, relevant and redundant data, laboratory experiment, prediction of financial failure.

Le phénomène de surcharge d'informations est d'actualité dans ce monde où la disponibilité de nouvelles technologies de l'information rend la production d'information de plus en plus rapide et de moins en moins coûteuse. La conséquence observable est que les gestionnaires sont actuellement «submergés» de données et d'informations. Le domaine financier et comptable n'échappe pas à ce phénomène. Depuis la fin des années 1960, les entreprises sont tenues de divulguer de plus en plus d'informations financières. Cette obligation est justifiée selon le Financial Accounting Standards Board (FASB) par l'hypothèse que la présentation d'une quantité plus importante de données facilite la prise de décision (FASB, 1976). Cette tendance persistera dans l'avenir selon l'analyse faite par

un groupe d'étude de l'Institut Canadien des Comptables Agréés chargé d'examiner les objectifs du rapport annuel (ICCA, 1992).

Cette surabondance des divulgations requises par les normalisateurs et les organismes de réglementation comptable a soulevé des questions quant à la capacité des utilisateurs à assimiler cette importante quantité de données. Le phénomène de surcharge d'informations a été invoqué ; en effet, il a été démontré en psychologie que l'abondance de données peut nuire à la prise de décision (voir entre autres Schroder, Driver et Streufert, 1967)¹. On commence à s'inquiéter de la surcharge d'informations financières tant du côté des chercheurs que des professionnels comptables aux États-Unis (Groves, 1994 et Cook et Sutton, 1995), au Royaume-Uni (auteur anonyme dans *Management Accounting*, 1995) et au Canada [monographie commandée par le Comité des normes comptables de l'ICCA sur l'objectif du rapport annuel aux actionnaires pour répondre notamment aux inquiétudes concernant la surcharge d'informations (ICCA, 1992, note 48, p. 246)].

Ainsi, Groves (1994) s'alarmait de l'augmentation constante des divulgations requises (mesurées par le nombre moyen de pages du rapport annuel aux actionnaires) depuis 1972. Selon ses calculs, ce nombre de pages a augmenté à un taux annuel composé de 3,1 %. Si cette tendance se maintient, le nombre de pages du rapport annuel en l'an 2012, selon les projections de Groves, serait de 234 % plus élevé que celui de 1972 (83 % plus élevé que celui de 1992). Bien que le rapport annuel aux actionnaires contienne à la fois des données narratives et des données quantitatives, ces dernières ont aussi augmenté.

L'objectif de la présente recherche est d'analyser si la surabondance des données financières quantitatives fournies aux utilisateurs par l'intermédiaire des rapports financiers influence la qualité de la décision prise par ces utilisateurs. Sa principale contribution est la précision des caractéristiques informationnelles des données disponibles qui sont la pertinence et la redondance. Des hypothèses sont développées et testées sur l'influence de chacune de ces caractéristiques sur la qualité de la décision. Le présent document est structuré de la façon suivante : dans la section qui suit sont décrites les études qui ont observé le phénomène de surcharge d'informations dans une tâche financière. Nous avons formulé quatre hypothèses sur d'une part la relation entre les données disponibles et les données pertinentes, redondantes et non pertinentes, et d'autre part la relation entre chacune des variables énumérées précédemment et la qualité de la décision. La section subséquente contient une description du plan de recherche. Elle est suivie d'une section portant sur les résultats. Enfin, la dernière section est consacrée aux conclusions et aux limites de la présente recherche.

1 Le phénomène de surcharge d'informations dans un contexte de décision financière

Schroder et alii (1967) ont développé un modèle qui relie la quantité d'information disponible au traitement de l'information par les humains. Ce modèle, appelé modèle de la courbe en U inversé,

¹ Snowball (1979) fait une revue détaillée des recherches faites en psychologie sur la surcharge d'informations.

prédit que lorsque l'information disponible augmente, le preneur de décision augmente d'abord son niveau de traitement de l'information, mais seulement jusqu'à un certain point optimal. Si la quantité d'information disponible continue d'augmenter et dépasse ce point optimal, le niveau de traitement de l'information cesse d'augmenter et commence à diminuer : le preneur de décision fait alors face à ce qu'on appelle l'effet de surcharge d'informations. Ce modèle a été appliqué dans le domaine financier et comptable afin de prédire l'influence de la quantité d'information disponible sur la qualité de la décision financière.

1.1 Les études antérieures

Shields (1980 et 1983) et Chewning et Harell (1990) étaient à notre connaissance les seules études qui ont essayé d'appliquer fidèlement dans le domaine de la décision financière le modèle de Schroder et alii (1967). Les résultats obtenus montrent une relation en forme de courbe en U inversé entre la quantité de données disponibles et le niveau de traitement de l'information. Ce dernier est défini comme étant la quantité de données effectivement utilisées dans la prise de décision. Chewning et Harell (1990), dans une deuxième étape de leur étude, ont ensuite constaté que lorsqu'on isole les sujets qui sont les plus « surchargés » de ceux qui ont été moins affectés par la quantité de données disponibles, la qualité de la décision des premiers était significativement moins bonne que celle des seconds.

Les autres études ont essayé de tester l'influence directe de la surcharge d'informations sur la qualité de la décision. Les résultats obtenus ne sont pas toujours convergents. Ainsi, Snowball (1980) et Simnett (1996) n'ont constaté aucun changement significatif de la qualité de la décision à différents niveaux de charge d'information. Par contre, Casey (1980), Shields (1983), Iselin (1988 et 1993) et Gadenne et Iselin (2000) ont constaté une influence négative de la quantité d'information disponible sur la qualité de la décision. Cependant, à l'exception des résultats obtenus par Shields (1983), la relation entre la quantité de données disponibles et la qualité de la décision n'est pas en forme de courbe en U inversé tel que prédit initialement par Schroder et alii (1967) : la qualité de la décision reste stable si on passe d'une quantité faible à moyenne de données disponibles et diminue si cette dernière continue à augmenter à un niveau élevé. Notons aussi que selon certaines études, seules certains types de données influencent de façon négative la qualité de la décision : les données de dimensions diverses (par opposition aux données répétitives) selon Iselin (1988) et les données non pertinentes (par opposition aux données pertinentes) selon Iselin (1993) et Gadenne et Iselin (2000).

Un méta-analyse des résultats de 31 études traitant de prédiction de faillite fait par Hwang et Lin (1999) vient ajouter de la confusion à celles déjà constatées ci-dessus. Selon cette dernière étude, la qualité de la décision se détériore lorsqu'on augmente la quantité de données de dimensions diverses et la quantité de données répétitives.

1.2 Les résultats des études antérieures analysés selon un autre point de vue

Un examen plus approfondi des définitions opérationnelles de chacun des types de données utilisées dans les études antérieures peut jeter une autre lumière aux résultats obtenus. Selon Iselin (1988), les données disponibles proviennent des états financiers comparatifs de plusieurs

années. Les données de dimensions diverses sont composées de tous les postes des états financiers pour une année donnée comptés chacun une fois. Les données répétitives quant à elles sont les postes correspondants aux premiers pour les autres années plus les postes qui reviennent plus d'une fois (par exemple, le bénéfice net qui apparaît à la fois à l'état des résultats et à l'état des bénéfices non répartis est compté une fois comme donnée de diverses dimensions et une fois comme donnée répétitive).

Selon Iselin (1993), les données disponibles sont composées de ratios financiers, du bilan et de l'état des résultats, le tout présenté de façon comparative sur deux ans. Les types de données sont mesurés selon deux méthodes. La mesure objective consiste à identifier les données pertinentes par une analyse discriminante : ce sont celles qui sont considérées comme étant les variables significatives du modèle d'analyse discriminante multiple. Le reste des données entre dans la catégorie des données non pertinentes. La deuxième mesure, la mesure subjective, consiste à demander à l'aide d'un questionnaire aux sujets participants à l'étude leur perception sur les données pertinentes ou non pertinentes.

Simnett (1996) et Gadenne et Iselin (2000) ont utilisé la même méthodologie qu'Iselin (1993) avec un ensemble de données disponibles différentes. Dans Simnett (1996), les données disponibles sont constituées uniquement de ratios financiers. Dans Gadenne et Iselin (2000), on a fourni aux sujets des ratios, le bilan, l'état des résultats, des extraits de l'état des flux de trésorerie et des données sur l'environnement externe (information sur les divers concurrents, le marché financier, l'économie et le secteur).

Enfin, selon le méta-analyse fait par Hwang et Lin (1999), les données de dimensions diverses sont composées de tous les ratios disponibles et les données répétitives le reste des données. Maintenant, ajoutons une autre dimension dans la classification des données disponibles selon la suggestion de Belkaoui (1984). Ce dernier a examiné l'effet de l'ajout de données redondantes (données n'apportant aucune nouvelle information) à un ensemble de données pertinentes disponibles. Les résultats obtenus, contrairement aux prédictions, sont que l'ajout de données redondantes augmente la qualité de la décision. Comme le plan de recherche utilisé par Belkaoui ne lui permettait pas d'observer l'influence des données redondantes à quantité élevée, il suggère d'utiliser sa méthodologie dans une étude traitant de l'influence de la quantité de données redondantes sur la qualité de la décision, ce qui à notre connaissance, n'a pas été fait jusqu'à maintenant.

La classification des données faite par Iselin (1993), Simnett (1996) et Gadenne et Iselin (2000) est intéressante mais incomplète. Selon ces chercheurs, les données pertinentes sont celles qui sont appropriées, qui jouent un rôle distinctif dans le processus de décision. Par contre, les données non pertinentes sont constituées de toutes les données disponibles auxquelles on a enlevé les données pertinentes. Or, dans cette deuxième catégorie, nous pouvons faire la distinction entre les données qui ne sont pas du tout appropriées à la prise de décision (les données non pertinentes) et les données qui sont appropriées mais qui sont superflues : ce sont des données qui expriment sous une forme différente les mêmes traits que les données classifiées comme étant pertinentes. Par définition, ce sont des données redondantes. Les données redondantes sont utiles dans plusieurs domaines. En télécommunication, on les utilise pour dupliquer l'information pertinente à transmettre afin de garantir leur sécurité en cas d'incident. En vérification, la collecte d'information probante nécessite la collecte de données ou de preuves redondantes dont le

recoupement avec les données pertinentes donne une assurance suffisante au vérificateur afin de conclure à la présence ou à l'absence d'inexactitudes dans les états financiers. Par contre, dans le domaine du traitement humain de l'information, la capacité cognitive limitée fait en sorte que l'ajout de données perçues comme étant redondantes nuit au traitement de l'information et de façon ultime à la qualité de la décision qui en résulte. Les données de dimensions diverses qui ont eu une influence négative sur la qualité de la décision dans Iselin (1988) contiennent à la fois des données pertinentes et redondantes. De même, les données classifiées comme étant non pertinentes et qui ont aussi eu une influence négative sur la qualité de la décision dans Iselin (1993) et Gadenne et Iselin (2000) contiennent aussi des données redondantes. Enfin, les deux catégories d'informations développées par Hwang et Lin (1999), à savoir l'information de dimensions diverses et l'information répétitive, toutes les deux ayant une influence négative sur la qualité de la décision, contiennent des données redondantes.

Ainsi, à la lumière de l'analyse faite ci-dessus, il semble approprié de développer une hypothèse qui prédit que seule l'augmentation de la quantité de données redondantes contribue à détériorer la qualité de la décision.

2 Les hypothèses

La figure suivante résume les hypothèses développées dans le présent document.

Variable indépendante → Variable intermédiaire → Variable dépendante

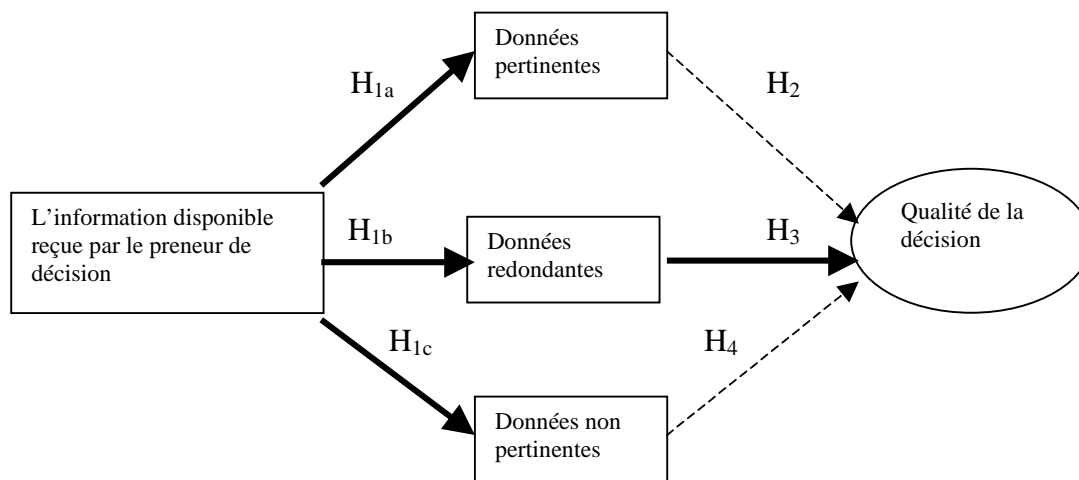


Figure 1 : Relation entre la quantité d'information disponible et la qualité de la décision

Les données disponibles aux décideurs sont classifiées selon les trois caractéristiques suivantes. Tout d'abord, les données pertinentes sont celles qui sont appropriées ou qui jouent un rôle

distinctif dans le processus de décision. Ensuite, nous avons les données non pertinentes qui sont les données qui ne sont pas du tout appropriées à la prise de décision. Enfin, entre ces deux catégories existe la catégorie des données redondantes qui sont des données appropriées mais qui sont superflues : ce sont des données qui expriment sous une forme différente les mêmes traits que les données classifiées comme étant pertinentes. Ces caractéristiques des données disponibles peuvent être mesurées de façon objective à l'aide de techniques statistiques et de façon subjective selon la perception des preneurs de décision. Comme les deux mesures convergent (Iselin, 1993 et Gadenne et Iselin, 2000), nous allons utiliser la mesure subjective afin de développer les hypothèses empiriques. Ainsi, nous prédisons que :

H₁ : Plus la quantité de données dont dispose le décideur pour sa prise de décision est grande, ...

H_{1a} : ...plus il perçoit de données pertinentes.

H_{1b} : ...plus il perçoit de données redondantes.

H_{1c} : ...plus il perçoit de données non pertinentes.

Les données pertinentes ou perçues par les décideurs comme étant pertinentes ne nuisent pas à la qualité de la décision, quelle que soit la quantité disponible. Les études qui ont isolé l'effet des données pertinentes n'ont pas observé de détérioration de la qualité de la décision lorsqu'on a augmenté la quantité de ces données (Iselin, 1993, Simnett, 1996 et Gadenne et Iselin, 2000). Ainsi, nous prédisons :

H₂ : Lorsqu'on augmente la quantité de données perçues comme étant pertinentes par le décideur, la qualité de sa décision ne change pas.

Les données redondantes ont des effets positifs et négatifs sur le processus décisionnel selon Belkaoui (1984) et Gilliland et Schmitt (1993). Les effets positifs sont : (1) la limitation de la recherche d'informations sans qu'il y ait une diminution substantielle de l'exactitude prédictive de la décision, (2) une attention sélective très élevée, (3) une simplification de la représentation mentale du problème à résoudre, prévenant ainsi la surcharge d'informations, (4) une facilité de substitution des pièces d'information entre elles, et (5) une diminution de la non-fiabilité des pièces d'information en ayant des mesures multiples des même variables. Les effets négatifs des données redondantes sont : (1) la difficulté de choisir entre des pièces d'information qui sont corrélées, (2) un gaspillage au niveau de l'effort d'attention consacré aux pièces d'information qui ne sont pas marginalement prédictives, (3) la présence de pièces d'information trop nombreuses résultant en une surconfiance en la décision prise, et (4) la difficulté d'apprentissage par la rétroaction de résultat lorsque les pièces d'information sont corrélées entre elles.

L'effet positif 5 ci-dessus est confirmé par les résultats empiriques d'études en vérification traitant de la révision séquentielle de la croyance (Ashton et Ashton, 1988). Les effets négatifs quant à

eux sont confirmés empiriquement par Lindell et Stewart (1974) et Schmitt et Dudycha (1977) dans des études utilisant des tâches autres que des décisions financières. Nous avons vu dans la section précédente que l'analyse des résultats des études antérieures à la lumière des caractéristiques des données disponibles semble suggérer que la quantité de données redondantes peut nuire à la qualité de la décision si on l'augmente au-delà d'un certain point jugé optimal à partir duquel la capacité cognitive limitée du décideur est dépassée. Concrètement, les données redondantes ne peuvent pas être analysées indépendamment des données pertinentes. En effet, une donnée n'est redondante que relativement à une autre donnée pertinente. Nous prédisons la relation suivante (voir aussi l'illustration de la figure 2 ci-dessous) :

H₃ : Au-delà d'un certain seuil, la quantité de données -- perçues comme étant *pertinentes* et *redondantes* par le décideur -- réduit la qualité de sa décision.

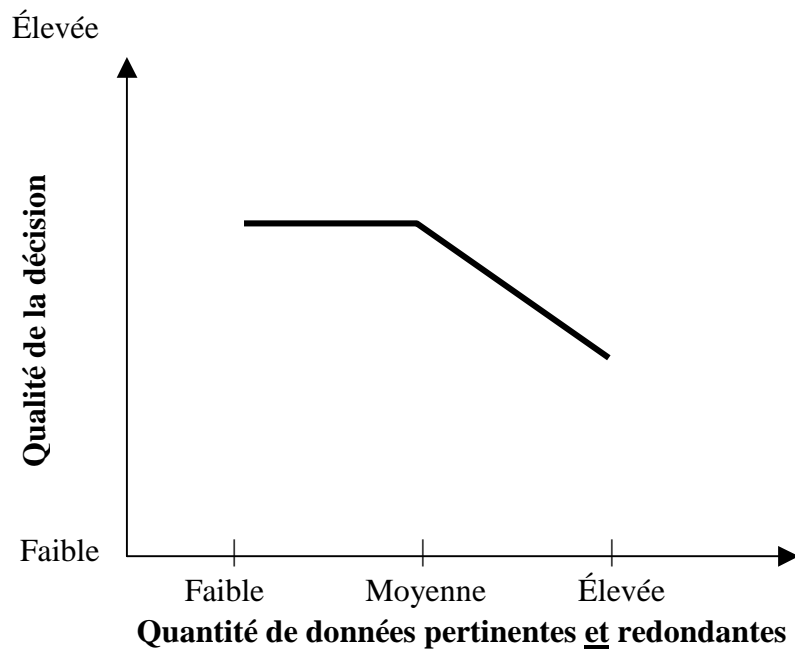


Figure 2 : Relation prédite entre la quantité de données pertinentes et redondantes et la qualité de la décision

Enfin, les données perçues comme étant non pertinentes ne devraient pas influencer la qualité de la décision. Ces données semblent être mises de côté par le décideur dans son processus de décision et n'affectent pas par conséquent sa capacité cognitive. Ainsi, nous prédisons que :

H₄ : Lorsqu'on augmente la quantité de données perçues comme étant *non pertinentes* par le décideur, la qualité de sa décision ne change pas.

3 Plan de recherche

Afin de tester les hypothèses développées dans la section précédente, l'auteur s'est livré à une expérience dans le cadre de laquelle 95 étudiants d'un programme de maîtrise professionnelle en administration qui simulaient une décision de prêt commercial ont lu six demandes de crédit qui contenaient, outre divers renseignements de base, des données financières réelles sur deux ans de six entreprises canadiennes de fabrication. La quantité de données disponibles a été manipulée de façon qu'elle soit faible, moyenne ou élevée, ce qui a donné lieu à trois versions des demandes de crédit. Les prêteurs prévoient la santé financière, en terme de défaillance financière ou non, de chaque entreprise emprunteuse. Ils inscrivaient ensuite sur une échelle à 100 points comment ils percevaient la quantité de données pertinentes, non pertinentes et redondantes contenues dans les données financières disponibles. Ils répondaient enfin à des questions d'ordre général, démographique et technique qui ont été utilisées dans la présente étude comme variables de contrôle. Nous avons utilisé un plan d'expérience formel ou *true experimental design*. Pour cela, les unités-tests, les participants ou sujets de l'étude, sont affectées de façon aléatoire aux traitements. L'expérience s'est déroulée dans des salles de classe et a duré en tout une heure. Les sous sections suivantes décrivent les démarches faites afin de développer le questionnaire utilisé.

3.1 Les données financières disponibles

Les données disponibles étaient constituées selon les trois niveaux de rapports financiers suivants :

1. Niveau faible : les douze ratios financiers divulgués annuellement par Dun & Bradstreet Canada. Ces ratios sont les plus susceptibles d'être connus des utilisateurs de l'information financière. En plus des ratios des entreprises demandant les prêts, ceux de la moyenne du secteur d'activité de chaque entreprise ont aussi été fournis.
2. Niveau moyen : les ratios ci-dessus, plus le bilan et l'état des résultats.
3. Niveau élevé : les ratios, le bilan et l'état des résultats ci-dessus, plus l'état des flux de trésorerie.

Les données contenues dans les rapports financiers sont fournies sous forme comparative sur deux ans. La variation en pourcentage entre les deux années a aussi été préalablement calculée.

3.2 La mesure objective des données pertinentes, redondantes et non pertinentes

La quantité de données disponibles constitue la variable indépendante de notre plan de recherche. Les différents niveaux de quantité de données disponibles constituent les traitements expérimentaux. Afin de pouvoir manipuler la quantité de données disponibles, ces dernières

doivent être dénombrables. Les rapports financiers que nous présentons aux participants ne contiennent que des données numériques afin notamment de pouvoir les dénombrer facilement² pour pouvoir satisfaire à ce critère. Le dénombrement des données contenues dans les états financiers est fait en comptant le nombre de lignes de données numériques uniques dans ces états financiers. Afin de vérifier l'efficacité de la manipulation faite sur les niveaux de quantité d'information contenue dans les trois rapports financiers, nous avons mesuré de façon objective les quantités respectives de données pertinentes, redondantes et non pertinentes.

Les données pertinentes sont obtenues par une analyse discriminante pas à pas (*stepwise discriminant analysis*) : ce sont les variables significatives résultant de ce modèle discriminant. Les données redondantes sont des données fortement corrélées avec les données pertinentes et qui sont aussi significatives lorsqu'on les substitue aux données pertinentes correspondantes dans le modèle discriminant. Les données non pertinentes sont constituées des données restantes (celles qui ne sont ni pertinentes ni redondantes). Le tableau 1 ci-dessous montre les quantités de données objectives telles que mesurées dans la présente étude.

Rapport financier ^a	Quantité de données objectives				
	Totales (1)	Pertinentes (2)	Redondantes (3)	Pertinentes + redondantes (4) = (2) + (3)	Non pertinentes (5) = (1) - (4)
R	48	1	1	2	46
RBE	198	6	3	9	189
RBEF	270	17	4	21	249

^a R : ratios financiers; RBE : ratios financiers + bilan + état des résultats; RBEF : ratios financiers + bilan + état des résultats + état des flux de trésorerie.

Tableau 1 : Quantité objectives de données contenues dans les rapports financiers

² Nous entendons par «données non numériques» les parties du rapport autres que le corps des états financiers : les notes aux états financiers, les graphiques et les commentaires de la direction par exemple. Les libellés des postes des états financiers (encaisse, impôts reportés ou bénéfice net par exemple), qui ne sont pas à proprement parler des données numériques, sont présentés.

3.3 Sélection des six entreprises incluses dans le cas de demande de prêt

Les six entreprises, deux en défaillance financière et quatre en bonne santé, ont été sélectionnées à l'aide d'un échantillonnage stratifié proportionnel à partir d'une population initiale de trente entreprises (10 potentiellement en défaillance financière et 20 potentiellement en bonne santé financière)³. Cette dernière est constituée d'entreprises qui répondent aux critères suivants :

- entreprises listées à la Bourse de Montréal de 1986 à 1993;
- la moitié, soit quinze entreprises satisfont aux critères suivants : (1) un premier critère discriminant de leur santé financière basée sur le rendement boursier, selon la méthodologie utilisée par Cormier, Magnan et Morard (1994) : celles dont le rendement boursier annuel ajusté pour le rendement du marché à un moment donné au cours de la période allant de 1986 à 1990 est inférieur à -50 % (les entreprises potentiellement défailtantes) et celles dont le rendement ajusté est supérieur ou égal à 0 (les entreprises potentiellement en bonne santé financière). La liste des entreprises répondant à ces deux critères nous a été fournie gracieusement par les professeurs Denis Cormier et Michel Magnan. Parmi ces dernières entreprises, nous n'avons retenu, par souci d'avoir une population assez homogène, que les entreprises de fabrication (code SIC 2000 à 3999); (2) pour les entreprises présentant une lourde perte boursière, identification de celles qui ont réellement présenté des signes de défaillance selon les rapports périodiques de la Bourse de Montréal aux dates entourant cette perte boursière : suspension du cours ou radiation pour cause de difficulté financière, retrait de la cote suite à une restructuration, une réorganisation ou une faillite⁴; (3) pour les entreprises présentant un rendement boursier positif, identification de celles dont les titres étaient toujours transigés de façon active pendant les deux années ultérieures; (4) pour les entreprises répondant aux deux critères précédents, identification de celles dont les rapports annuels des trois années précédant le rendement boursier discriminant sont disponibles et qui étaient en affaire au moins cinq ans avant cette dernière date.
- l'autre moitié des entreprises de fabrication composant la population ciblée est constituée d'un échantillon aléatoire de dix entreprises en bonne santé financière et de cinq entreprises en difficulté financières identifiées selon les étapes 2 à 4 ci-dessus. Cet ajout de quinze entreprises supplémentaires a été rendu nécessaire afin de rendre les plus adéquates possibles les inférences statistiques sur les mesures objectives des propriétés informationnelles des données.

Les données de 24 entreprises (l'échantillon d'estimation : 16 en bonne santé et 8 en défaillance financière) ont été utilisées afin d'identifier et de mesurer la quantité de données objectives. Les

³ Dans un échantillonnage stratifié proportionnel, la taille de l'échantillon dans chaque strate (entreprises en défaillance financière et entreprises en bonne santé) est proportionnelle à la taille de la strate dans la population. Un échantillonnage aléatoire simple a été effectué dans chaque strate.

⁴ Ces rapports périodiques sont la publication hebdomadaire Bourse de Montréal : Nouvelles quotidiennes officielles et la publication annuelle Bourse de Montréal : Statistiques.

données des 6 entreprises restantes (4 en bonne santé et 2 en défaillance financière) ont servi à valider les inférences statistiques et ont été incorporées dans le cas de demande de prêt.

3.4 La qualité de la décision

Dans la présente étude, la qualité de la décision est définie par l'exactitude de la décision et par la comparaison entre la décision de l'humain et la décision d'un modèle statistique qui est ici une analyse discriminante. L'exactitude est mesurée par le nombre de prédictions correctes sur les six prédictions faites par chaque répondant. La mesure Modèle vs Humain est la différence entre le nombre de prédictions correctes d'un modèle statistique et le nombre de prédictions correctes faites par l'individu. Ainsi, une valeur négative, nulle ou positive de cette mesure signifient respectivement que l'individu a mieux fait, a fait aussi bien ou a fait moins bien que le modèle statistique⁵.

4 Résultats et analyse

4.1 Statistiques descriptives de l'échantillon

Le tableau 2 ci-dessous résume les caractéristiques descriptives des 95 étudiants en maîtrise en administration qui ont participé de façon volontaire à l'étude.

Variable (<i>test de mesure d'association entre la variable et les groupes expérimentaux : ANOVA ou Khi-carré</i>)	Par groupe expérimental (selon la quantité de données disponibles)			
	Élevé	Moyen	Faible	Total
Sexe (pourcentage des hommes) ($X^2 = 2,54$; $p = 0,28$)	68,80	51,60	51,60	57,40
Âge médian ($X^2 = 7,31$; $p = 0,70$)	31,10	28,90	30,70	30,50
Nombre moyen d'années d'expérience de travail totales ($F = 0,94$; $p = 0,40$)	6,19	5,42	7,09	6,23
Pourcentage de répondants ayant de l'expérience dans une fonction faisant principalement appel à l'interprétation d'information financière ($X^2 = 1,61$; $p = 0,45$)	21,20	35,50	29,0	28,40
Taille de l'échantillon	33	31	31	95

Tableau 2 : Caractéristiques descriptives de l'échantillon

⁵ Pour l'ensemble des 95 répondants de l'étude, l'exactitude moyen est de 4,67 et la mesure Modèle vs Humain est en moyenne de 0,67. Comme cette dernière mesure est significativement supérieure à zéro ($t = 5,89$, $p = 0,00$), dans

4.2 Hypothèse 1 : relation entre la quantité de données disponibles et les quantités de données pertinentes, redondantes et non pertinentes

Une analyse de variance, suivie d'un test de comparaison multiple (test de Duncan) ont été utilisées afin de tester l'hypothèse 1. Les quantités de données pertinentes, redondantes et non pertinentes sont mesurées par la perception des décideurs sur une échelle à 100 points. Les tableaux 3 et 4 ci-dessous résument les résultats de l'hypothèse 1a. L'analyse montre que l'effet de la quantité de données disponibles sur la quantité perçue de données pertinentes est significatif ($p = 0,03$). Les décideurs perçoivent qu'il y a une différence significative entre les quantités de données pertinentes contenues dans le rapport financier de niveau faible et le rapport financier de niveau élevé ainsi qu'entre les quantités de données pertinentes contenues dans le rapport financier de niveau moyen et le rapport financier de niveau élevé. Ces résultats supportent l'hypothèse 1a.

ANOVA — Quantité de données pertinentes par groupe expérimental				
Source de la variation	Degré de liberté	MS	F	p
Erreur résiduelle	92	152,5		
GROUPE	2	533,4	3,5	0,03

Tableau 3 : Relation entre la quantité de données disponibles et la quantité de données pertinentes

Test de comparaison multiple de la quantité moyenne de données pertinentes (PERT) par groupe expérimental : les seules différences significatives (Test de Duncan, $\alpha = 0,05$; différence significative > 6,49)				
GROUPE	PERT	Élevé	Moyen	Faible
Élevé	73,49	N.A	Élevé > Moyen (diff. = 6,71)	Élevé > Faible (diff. = 7,33)
Moyen	66,78	Élevé > Moyen (diff. = 6,71)	N.A	N.S (diff. = 0,62)
Faible	66,16	Élevé > Faible (diff. = 7,33)	N.S (diff. = 0,62)	N.A
N.A : non applicable N.S : non significative				

Tableau 4 : Relation entre la quantité de données disponibles et la quantité de données pertinentes – test de comparaison multiple

la présente étude et conformément aux résultats d'autres études, l'humain fait en moyenne moins bien qu'un modèle statistique.

Les tableaux 5 et 6 ci-dessous résument les résultats de l'hypothèse 1b. Selon ces tableaux, l'effet de la quantité de données disponibles sur la quantité perçue de données redondantes est significatif ($p = 0,02$). Selon la perception des décideurs, il y a une différence significative seulement entre les quantités de données redondantes contenues dans le rapport financier de niveau faible et le rapport financier de niveau élevé. Ces résultats supportent l'hypothèse 1b.

ANOVA — Quantité de données redondantes par groupe expérimental				
Source de la variation	Degré de liberté	MS	F	p
Erreur résiduelle	92	227,4		
GROUPE	2	877,5	3,9	0,02

Tableau 5 : Relation entre la quantité de données disponibles et la quantité de données redondantes

Test de comparaison multiple de la quantité moyenne de données redondantes (RED) par groupe expérimental : les seules différences significatives (Test de Duncan, $\alpha = 0,05$, différence significative > 7,92)				
GROUPE	RED	Élevé	Moyen	Faible
Élevé	36,19	N.A	N.S (diff. = 5,72)	Élevé > Faible (diff. = 11,45)
Moyen	30,47	N.S (diff. = 5,72)	N.A	N.S (diff. = 4,73)
Faible	25,74	Élevé > Faible (diff. = 11,45)	N.S (diff. = 4,73)	N.A
N.A : non applicable N.S : non significative				

Tableau 6 : Relation entre la quantité de données disponibles et la quantité de données redondantes – test de comparaison multiple

Les tableaux 7 et 8 ci-dessous résument les résultats de l'hypothèse 1c. L'analyse montre que l'effet de la quantité de données disponibles sur la quantité perçue de données non pertinentes est marginalement significatif ($p = 0,06$). Alors que le tableau 1 montre que la quantité objective de données non pertinentes varie de 46 à 249 selon les niveaux, la perception des décideurs est différente. Il semble que les décideurs accordent peu d'importance à la quantité réelle de ces données non pertinentes étant donné que cette dernière n'affecte pas leur capacité cognitive. Les décideurs perçoivent quand même qu'il y a une différence significative entre les quantités de données non pertinentes contenues dans le rapport financier de niveau faible et le rapport financier de niveau élevé. Ces résultats supportent partiellement l'hypothèse 1c et met en lumière le peu d'importance accordé par les décideurs aux données perçues comme étant non pertinentes.

ANOVA — Quantité de données non pertinentes par groupe expérimental				
Source de la variation	Degré de liberté	MS	F	p
Erreur résiduelle	92	235,2		
GROUPE	2	678,7	2,9	0,06

Tableau 7 : Relation entre la quantité de données disponibles et la quantité de données non pertinentes

Test de comparaison multiple de la quantité moyenne de données non pertinentes (NPERT) par groupe expérimental : Test de Duncan ($\alpha = 0,05$; différence significative > 8,06)				
GROUPE	NPERT	Élevé	Moyen	Faible
Élevé	30,9	N.A	N.S (diff. = 3,31)	Élevé > Faible (diff. = 9,12)
Moyen	27,6	N.S (diff. = 3,31)	N.A	N.S (diff. = 5,81)
Faible	21,7	Élevé > Faible (diff. = 9,12)	N.S (diff. = 5,81)	N.A
N.A : non applicable N.S : non significative				

Tableau 8 : Relation entre la quantité de données disponibles et la quantité de données non pertinentes – test de comparaison multiple

4.3 Hypothèses 2 à 4 : effet des caractéristiques des données sur la qualité de la décision

Afin de tester les hypothèses sur les effets des caractéristiques des données disponibles sur la qualité de la décision, des régressions multiples sont utilisées. Leur forme est :

$$\text{QUALITÉ} = \alpha_0 + \alpha_1 \text{QUANTITÉ} + \alpha_2 \text{GP1} + \alpha_3 \text{GP2} + \alpha_4 \text{QUANTITÉ} * \text{GP1} + \alpha_5 \text{QUANTITÉ} * \text{GP2} + \alpha_c \text{COVAR} + \alpha_6 \text{PCTD} + \varepsilon \quad (\text{Équation 1})$$

Dans laquelle :

QUALITÉ : qualité de la décision (exactitude de la décision ou mesure Modèle vs Humain)

QUANTITÉ : quantité de données à analyser (pertinentes, pertinentes plus redondantes et non pertinentes)

GP1 : variable qualitative dichotomique (1 si groupe Élevé et 0 si autres)

GP2 : variable qualitative dichotomique (1 si groupe Faible et 0 si autres)

COVAR : variables indépendantes utilisées comme covariables (par exemple, lorsque c'est l'effet de la quantité de données pertinentes qui est analysé, les quantités de données redondantes et de données non pertinentes sont utilisées comme covariables);

PCTD : perception du répondant du pourcentage de défaillance financière dans la population des entreprises canadiennes de fabrication de laquelle l'échantillon de six entreprises étudiées a été tiré (utilisée ici comme covariable).

Après transformation, pour le groupe Élevé, l'équation 1 devient :

$$\text{QUALITÉ} = (\alpha_0 + \alpha_2) + (\alpha_1 + \alpha_4)\text{QUANTITÉ} + \alpha_c\text{COVAR} + \alpha_6\text{PCTD} + \varepsilon$$

(Équation 2)

Pour le groupe Moyen, l'équation devient :

$$\text{QUALITÉ} = \alpha_0 + \alpha_1\text{QUANTITÉ} + \alpha_c\text{COVAR} + \alpha_6\text{PCTD} + \varepsilon$$

(Équation 3)

Et enfin, pour le groupe Faible, l'équation devient :

$$\text{QUALITÉ} = (\alpha_0 + \alpha_3) + (\alpha_1 + \alpha_5)\text{QUANTITÉ} + \alpha_c\text{COVAR} + \alpha_6\text{PCTD} + \varepsilon$$

(Équation 4)

Les coefficients des régressions des équations 2, 3 et 4 s'interprètent comme suit :

- $(\alpha_0 + \alpha_2)$, α_0 et $(\alpha_0 + \alpha_3)$ sont les ordonnées à l'origine des droites correspondantes respectivement aux groupes Élevé, Moyen et Faible;
- $(\alpha_1 + \alpha_4)$, α_1 et $(\alpha_1 + \alpha_5)$ sont les pentes des droites correspondantes respectivement aux groupes Élevé, Moyen et Faible.

Ainsi, les tests des hypothèses relatives à l'effet des quantités de données sur la qualité de la décision se feront par l'analyse des pentes des droites de régression correspondantes aux groupes Élevé, Moyen et Faible. On identifiera pour chaque hypothèse le signe et l'amplitude prévus de chacun des coefficients $(\alpha_1 + \alpha_4)$, α_1 et $(\alpha_1 + \alpha_5)$.

Le modèle de régression utilisé dans la présente étude contrôle l'effet de variables non analysées dans une hypothèse donnée (les covariables) et de la perception des répondants sur le pourcentage de défaillance financière (PCTD) dans la population des entreprises canadiennes dans laquelle les six entreprises du cas ont été tirées. Les corrélations entre les variables indépendantes ne sont pas très élevées ce qui satisfait au critère d'utilisation de covariables⁶.

⁶ La corrélation la plus élevée est entre les variables « données pertinentes » et « données pertinentes et redondantes » (corrélation significative de 0,52). Cependant, ces deux variables ne sont jamais utilisées ensemble dans un même modèle de régression.

4.3.1 Hypothèse 2 : effet de la quantité de données pertinentes sur la qualité de la décision

Selon les prédictions de cette hypothèse, les coefficients ($\alpha_1 + \alpha_4$), α_1 et ($\alpha_1 + \alpha_5$) ne devraient pas être significativement différents de zéro. Selon les résultats obtenus, le modèle de régression utilisant l'exactitude comme variable dépendante n'est même pas significatif (R^2 ajusté = -0,02, $F = 0,84$, $p = 0,58$). Le modèle de régression utilisant la mesure Modèle vs Humain est significatif (R^2 ajusté = 0,26, $F = 4,52$, $p = 0,00$). Pour ce dernier modèle, les statistiques concernant les coefficients des pentes des droites de régression sont les suivantes :

- ($\alpha_1 + \alpha_4$) : coeff. = -0,02, $F = 1,05$, $p = 0,35$;
- α_1 : coeff. = 0,01, $t = 0,32$, $p = 0,75$;
- ($\alpha_1 + \alpha_5$) : coeff. = 0,00, $F = 0,04$, $p = 0,85$.

Ainsi, ces résultats supportent l'hypothèse 2 : la quantité de données perçues comme étant pertinentes n'influence pas la qualité de la décision.

4.3.2 Hypothèse 3 : effet de la quantité de données pertinentes et redondantes sur la qualité de la décision

Selon les prédictions de cette hypothèse, les coefficients α_1 et ($\alpha_1 + \alpha_5$) ne devraient pas être significativement différents de zéro. Par contre, le coefficient ($\alpha_1 + \alpha_4$) devrait être négatif pour la régression utilisant l'exactitude comme variable dépendante et positif pour la régression utilisant la mesure Modèle vs Humain comme variable dépendante. Selon les résultats obtenus, le modèle de régression utilisant l'exactitude comme variable dépendante n'est pas significatif (R^2 ajusté = 0,06, $F = 1,71$, $p = 0,11$). Le modèle de régression utilisant la mesure Modèle vs Humain est significatif (R^2 ajusté = 0,32, $F = 6,18$, $p = 0,00$). Pour ce dernier modèle, les statistiques concernant les coefficients des pentes des droites de régression sont les suivantes :

- ($\alpha_1 + \alpha_4$) : coeff. = 0,02, $F = 4,18$, $p = 0,04$;
- α_1 : coeff. = -0,00, $t = -0,17$, $p = 0,86$;
- ($\alpha_1 + \alpha_5$) : coeff. = -0,02, $F = 1,98$, $p = 0,16$.

Le coefficient ($\alpha_1 + \alpha_4$) positif, correspondant à la pente de la droite de régression du groupe à quantité de données disponibles Élevée, signifie que dans ce groupe, en moyenne l'humain fait moins bien que le modèle statistique.

Ainsi, ces résultats supportent l'hypothèse 3 : la qualité de la décision commence à se détériorer lorsque la quantité de données pertinentes et redondantes passe d'une quantité moyenne à une quantité élevée.

4.3.3 Hypothèse 4 : effet de la quantité de données non pertinentes sur la qualité de la décision

Selon les prédictions de cette hypothèse, les coefficients ($\alpha_1 + \alpha_4$), α_1 et ($\alpha_1 + \alpha_5$) ne devraient pas être significativement différents de zéro. Selon les résultats obtenus, le modèle de régression utilisant l'exactitude comme variable dépendante n'est même pas significatif (R^2 ajusté = 0,04, $F = 1,98$, $p = 0,18$). Le modèle de régression utilisant la mesure Modèle vs Humain est significatif (R^2 ajusté = 0,30, $F = 5,87$, $p = 0,00$). Pour ce dernier modèle, les statistiques concernant les coefficients des pentes des droites de régression sont les suivantes :

- ($\alpha_1 + \alpha_4$) : coeff. = -0,02, $F = 1,78$, $p = 0,19$;
- α_1 : coeff. = 0,01, $t = 0,93$, $p = 0,35$;
- ($\alpha_1 + \alpha_5$) : coeff. = -0,02, $F = 2,08$, $p = 0,15$.

Ainsi, ces résultats supportent l'hypothèse 4 : la quantité de données perçues comme étant non pertinentes n'influence pas la qualité de la décision.

Dans tous les cas, le modèle de régression utilisant l'exactitude comme variable dépendante n'était pas significatif. Il semble, selon les résultats de la présente étude, que l'utilisation de données financières et numériques seules n'est pas suffisante pour prédire de façon cohérente et consistante la défaillance financière d'une entreprise.

5 Conclusion

Le but de la présente étude consistait à démontrer que les données disponibles au décideur se décomposent en données pertinentes, redondantes et non pertinentes et que lorsqu'on dépasse un certain point jugé optimal, l'ajout d'une quantité de données redondantes nuit à la qualité de la décision. Ces principales constatations sont utiles aux préparateurs de l'information comptable et financière. Dans le but d'optimiser leur processus de divulgation d'information, il semble que la meilleure stratégie consiste à cibler les utilisateurs, à identifier les données pertinentes à une tâche donnée et à fournir seulement ces données. De plus, l'utilisation d'une aide à la décision est justifiée dans un contexte où on ne peut pas réduire la surcharge d'informations car c'est dans ce contexte que les modèles statistiques font mieux que les humains.

Cependant, les résultats obtenus sont assujettis à plusieurs restrictions. Tout d'abord, afin de pouvoir dénombrer la quantité de données disponibles, l'auteur s'est limité aux données numériques. Les résultats ne sont pas ainsi généralisables aux données sous forme narrative comme les notes aux états financiers ou les commentaires de la direction incluses dans le rapport annuel aux actionnaires. Ces informations méritent d'être analysées dans un contexte de surcharge d'informations.

Une expérience de type « laboratoire » permet d'isoler l'influence des variables pertinentes à l'étude. Cependant, elle est éloignée du contexte de la prise de décision dans la vraie vie professionnelle. Les résultats de la présente étude ne pourraient ainsi être généralisés avant que des études de type « enquête sur le terrain » viennent les confirmer.

L'étude dont nous avons livré ici les résultats pourrait être poursuivie dans de nombreuses voies. Il serait d'abord intéressant d'étendre l'étude à l'utilisation de données autres que numériques : données sous forme narratives, sous forme graphique ou sous un autre support médiatique. Il serait aussi intéressant d'inclure des données autres que financières dans une étude traitant de défaillance financière afin de se rapprocher plus de la réalité : inclure par exemple des données sur la concurrence ou sur le secteur d'activité. Enfin, dans un but pratique, il serait aussi intéressant d'aller au-delà de la constatation de l'effet négatif de la surcharge d'informations et de tester de façon empirique des moyens de la supprimer ou du moins de l'atténuer.

Références bibliographiques

Ashton, A.H. et R.H. Ashton (1988), «Sequential Belief Revision in Auditing», *The Accounting Review*, octobre, pp. 623-641.

Auteur anonyme (1995), «A Bridge too Far?», *Management Accounting*, Janvier, p. 37.

Belkaoui, A. (1984) «The Effects of Diagnostic and Redundant Information on Loan Officers' Predictions», *Accounting and Business Research*, Été, pp. 259-256.

Casey, J.C. JR. (1980) «Variation in Accounting Information Load: The Effect on Loan Officers' Predictions of Bankruptcy», *The Accounting Review*, vol. 55, No 1, janvier, pp. 36-49.

Chewning, E.G. JR. et A. Harrell. (1990) «The Effect of Information Load on Decision Makers' Cue Utilization Levels and Decision Quality in a Financial Distress Decision Task», *Accounting Organizations and Society*, vol. 15, No 6, pp. 527-542.

Cook, M. et M.H. Sutton (1995), «Summary Annual Reporting: A Cure for Information Overload», *Financial Executive*, janvier/fevrier, pp. 12-15.

Cormier, D., M. Magnan et B. Morard (1994), «L'utilisation de techniques statistiques dans l'évaluation de la pérennité de l'entreprise», *Document de travail # 94-01 : Département des sciences comptables*, Université du Québec à Montréal, janvier.

Dun & Bradstreet Canada (1986 à 1993), *Key business ratios / Principaux coefficients des entreprises*, Dun & Bradstreet Canada, Toronto.

Financial Accounting Standards Board (FASB, 1976), *Discussion Memorandum: Tentative Conclusions on Objectives of Financial Statements of Business Enterprises*, Financial Accounting Standards Board, Stamford, CT.

Gadanne, D. et E.R. Iselin (2000), « Properties of Accounting and Finance Information and Their Effects on the Performance of Bankers and Models in Predicting Company Failure », *Journal of Business Finance & Accounting*, janvier, pp. 155-193.

Gilliland, S.W., et N. Schmitt (1993), « Information Redundancy and Decision Behavior: A Process Tracing Investigation », *Organizational Behavior and Human Decision Processes*, mars, pp. 157-180.

Groves, R.J. (1994), « Financial Disclosure: When More Is Not Better », *Financial Executive*, mai/juin, pp. 11-14.

Hwang, M.I. et J.W. Lin (1999), « Information Dimension, Information Overload and Decision Quality », *Journal of Information Science*, Vol. 25, No 3, pp. 213-218.

Institut Canadien des Comptables Agréés (ICCA, 1992), *L'information à inclure dans le rapport annuel aux actionnaires*, Institut canadien des comptables agréés, Toronto.

Iselin, E.R. (1988), « The Effects of Information Load and Information Diversity on Decision Quality in a Structured Decision Task », *Accounting, Organizations and Society*, vol. 13, No 2, pp. 147-164.

Iselin, E.R. (1993), « The Effects of the Information and Data Properties of Financial Ratios and Statements on Managerial Decision Quality », *Journal of Business Finance and Accounting*, janvier, pp. 249-266.

Lindell, M.K., et T.R. Stewart (1974), « The Effects of Redundancy in Multiple-cue Probability Learning », *American Journal of Psychology*, Vol. 87, No. 3, pp. 393-398.

Schmitt, N., et A.L. Dudycha (1975), « A Reevaluation of the Effect of Cue Redundancy in Multiple-cue Probability Learning », *Journal of Experimental Psychology: Human Learning and Memory*, Vol. 104, No. 3, pp. 307-315.

Schroder, H.M., M.J. Driver et S. Streufert (1967), *Human Information Processing*, New York, Holt, Rinehart et Winston.

Shields, M.D. (1983) « The Effects of Information Supply and Demand on Judgment Accuracy: Evidence from Corporate Managers », *The Accounting Review*, vol. 58, No 2, avril, pp. 284-303.

Shields, M.D. (1980), « Some Effects of Information Load on Search Patterns Used to Analyze Performance Reports », *Accounting, Organizations and Society*, vol. 5, No 4, pp. 429-442.

Simnett, R. (1996), « The Effect of Information Selection and Task Complexity on Predictive Accuracy, Confidence and Calibration of Auditors », *Accounting, Organizations and Society*, octobre/novembre, pp. 699-719.

Snowball, D. (1979), « Information Load and Accounting Reports: Too Much, or Just Right? », *Cost and Management*, mai-juin, pp.22-28.

Snowball, D. (1980) «Some Effects of Accounting Expertise and Information Load: An Empirical Study», *Accounting, Organizations and Society*, vol. 5, No 3, pp. 323-338.