

**Direction des bibliothèques**

**AVIS**

Ce document a été numérisé par la Division de la gestion des documents et des archives de l'Université de Montréal.

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

**NOTICE**

This document was digitized by the Records Management & Archives Division of Université de Montréal.

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

**Expérimentation d'un modèle d'évaluation permettant  
de juger du développement d'une compétence  
d'investigation scientifique en laboratoire**

par  
Eric Dionne

Département d'administration et de fondements de l'éducation  
Faculté des sciences de l'éducation

Thèse présentée à la Faculté des Études Supérieures en vue de l'obtention du grade de  
Philosophia Doctor (Ph.D.) en science de l'éducation  
option Mesure et évaluation

7 mai, 2008

© Eric Dionne, 2008



**Identification du jury**

**Université de Montréal  
Faculté des études supérieures**

**Cette thèse intitulée :**

**Expérimentation d'un modèle d'évaluation permettant de juger du développement d'une  
compétence d'investigation scientifique en laboratoire**

**présentée par**

**Eric Dionne**

**a été évaluée par un jury composé des personnes suivantes :**

<b>Blais, Jean-Guy</b>	<b>Président-rapporteur</b>
<b>Laurier, Michel D.</b>	<b>Directeur de recherche</b>
<b>Vazquez-Abad, Jesus</b>	<b>Co-directeur de recherche</b>
<b>Thouin, Marcel</b>	<b>Membre du jury</b>
<b>Simon, Marielle</b>	<b>Examinatrice externe</b>
<b>ARCHAMBAULT, JEAN</b>	<b>Représentant du doyen de la FES</b>

**Thèse acceptée le 17, Juin 2008**

## Résumé

---

Les défis liés à l'évaluation des compétences et des apprentissages complexes sont nombreux et variés. Depuis plusieurs années, on cherche à définir des méthodologies afin de permettre aux enseignants et à toutes les personnes concernées de faire reposer leur jugement sur des bases solides surtout quand les enjeux sont critiques comme dans le cas de l'évaluation certificative. Cette recherche à caractère exploratoire trouve appui sur les travaux de Rey, Carette, Defrance & Kahn (2003, p. 21) et propose un modèle multi niveaux d'évaluation certificative que nous avons expérimenté dans le contexte de l'enseignement de la science et de la technologie au premier cycle du secondaire au Québec. Compte tenu du contexte, nous avons profité de l'occasion pour vérifier l'effet de différentes variables comme le contexte disciplinaire, l'effet enseignant ou la différence entre les garçons et les filles sur la robustesse du modèle.

Afin de recueillir les données, nous avons conçu et administré trois situations d'évaluation qui comportaient chacune trois niveaux pour un total de neuf situations d'évaluation distinctes. Ces épreuves<sup>1</sup> ont été administrées à 560 élèves de première secondaire en mai et juin 2006. Après avoir corrigé les épreuves et colligé les données, nous avons par la suite procédé à leur analyse au moyen de statistiques non paramétriques afin de vérifier l'efficacité du modèle.

Les résultats que nous avons obtenus indiquent qu'un modèle multi niveau permet aux élèves en difficulté de fournir davantage d'indicateurs aux évaluateurs afin de guider ces derniers dans le jugement qu'ils doivent porter sur le niveau de développement de la compétence d'investigation scientifique. Nous avons également remarqué que les résultats étaient semblables peu importe le contexte disciplinaire associé à chacune des situations d'évaluation. Nous n'avons noté aucune différence importante entre les garçons et les filles. De plus, les variables liées à l'enseignant ne semblent pas avoir un effet important sur les résultats des élèves.

---

<sup>1</sup> Dans le cadre de cette recherche, nous considérons équivalentes les expressions « situations d'évaluation » et « épreuves ».

Les analyses nous indiquent que le modèle utilisé permet d'obtenir des résultats intéressants surtout auprès de certains profils d'élèves qui se sont dessinés lors de l'analyse des données. Des suites à cette recherche pourraient toucher à la robustesse du modèle d'évaluation dans d'autres domaines d'enseignement comme par exemple en mathématique. D'autre part, les résultats sont suffisamment encourageants pour procéder à une expérimentation dans un contexte plus confirmatoire.

**Mots clés :**

**Évaluation, compétences, investigation scientifique, science, enseignement secondaire**

## Abstract

---

The challenges in relation with the assessment of the competencies and complex learning are major and diverse. For many years, various methods were proposed in order to allow teachers and other stakeholders to get high quality assessment particularly in the context of high stake assessment. This exploratory research is based on Rey, Carette, Defrance & Kahn's work (2003, p. 21) and proposes a multi level model for certification which we tried out in the context of science and technology teaching in the first cycle of secondary school in Quebec. Given the context, we took advantage of the opportunity to study the effect of various variables (disciplinary context, teaching effect, difference between boys and girls) on the robustness of the model.

In order to collect data, we conceived and administered three (3) three-level assessment situations for a total of nine distinct assessment situations. These tests were administrated during the months of May and June 2006 to 560 students in their first year of secondary school. After marking the tests, we carried out the data analysis with nonparametric statistics in order to check the model effectiveness.

The results we obtained indicate that a multi level model provides the teachers with more information about students with learning problem helping them to judge the level of development of the scientific inquiry competence. We also noticed that the results were similar no matter of the disciplinary context associated with each assessment situation. We did not observe any important differences between boys and girls. Moreover, the variables related to the teacher do not have an important effect on the student's results. The analyses indicate that the model gives interesting results particularly with certain student's profiles which emerged during data analysis. Follow-up studies could focus on robustness of the assessment's model in other fields such as mathematics. In addition, the results are sufficiently encouraging to consider a confirmatory study.

**Key words:****Assessment, competencies, scientific inquiry, science, secondary level**

## Table des matières

<b>IDENTIFICATION DU JURY .....</b>	<b>III</b>
<b>RÉSUMÉ .....</b>	<b>IV</b>
<b>ABSTRACT .....</b>	<b>VI</b>
<b>LISTE DES TABLEAUX.....</b>	<b>XIII</b>
<b>LISTE DES FIGURES .....</b>	<b>XVII</b>
<b>LISTE DES SIGLES ET ABRÉVIATIONS .....</b>	<b>XXI</b>
<b>DÉDICACE .....</b>	<b>XXII</b>
<b>REMERCIEMENTS .....</b>	<b>XXIII</b>
<b>INTRODUCTION.....</b>	<b>1</b>
<b>CHAPITRE 1. PRÉSENTATION ET JUSTIFICATION DE LA RECHERCHE ....</b>	<b>3</b>
<b>1.1 Problématique .....</b>	<b>3</b>
1.1.1 Contexte du renouveau en enseignement des sciences au Québec .....	4
1.1.2 L'évaluation des compétences en science et technologie .....	13
1.1.2.1 La cohérence du triangle curriculum-apprentissage-évaluation .....	13
1.1.2.2 L'évaluation des apprentissages en science : des modèles à développer .....	15
1.1.2.3 L'évaluation certificative dans une approche par compétences en science et technologie.....	18
1.1.3 Proposition d'un modèle d'évaluation certificative .....	20
1.1.3.1 Un modèle d'évaluation certificative par cycle .....	21
1.1.3.2 Cadre dans lequel se situent les situations d'évaluation .....	23
1.1.3.3 Un modèle d'évaluation à trois niveaux .....	27
<b>1.2 Objectifs de la recherche.....</b>	<b>31</b>
<b>1.3 Questions de recherche.....</b>	<b>32</b>
<b>1.4 Pertinence de la recherche .....</b>	<b>36</b>

<b>CHAPITRE 2. RECENSION DES ÉCRITS .....</b>	<b>37</b>
<b>2.1 Nouvelles tendances en évaluation .....</b>	<b>37</b>
<b>2.2 Caractéristiques et définition de l'évaluation basée sur la performance (EBP). 39</b>	
2.2.1 Les avantages .....	44
2.2.2 Les limites .....	45
<b>2.3 L'investigation scientifique en laboratoire de type hands-on .....</b>	<b>47</b>
2.3.1 Activités <i>hands-on</i> : définitions et caractéristiques .....	48
2.3.2 Types de situations <i>hands-on</i> .....	49
<b>2.4 État de la question concernant les variables investiguées .....</b>	<b>52</b>
2.4.1 Liens entre le modèle d'évaluation, la forme de l'épreuve, le contenu disciplinaire et les résultats des élèves en science .....	53
2.4.2 Liens entre le sexe de l'élève, les caractéristiques de l'enseignant et les résultats des élèves en science .....	58
<b>CHAPITRE 3. MÉTHODOLOGIE .....</b>	<b>62</b>
<b>3.1 Les situations d'évaluation expérimentées .....</b>	<b>62</b>
3.1.1 La conception des situations d'évaluation .....	63
3.1.2 Caractéristiques de chacun des niveaux .....	67
3.1.3 Description des situations d'évaluation .....	73
<b>3.2 Validation des situations d'évaluation .....</b>	<b>74</b>
3.2.1 Validation de contenu des situations d'évaluation .....	74
3.2.2 Pré-expérimentation .....	76
3.2.3 Modalités d'expérimentation et échantillonnage .....	78
3.2.3.1 Collecte des données .....	78
3.2.3.2 Échantillonnage .....	79
3.2.3.3 Les consignes données aux enseignants lors de l'expérimentation .....	80
<b>3.3 Variables investiguées .....</b>	<b>82</b>
3.3.1 Les variables principales .....	82
3.3.1.1 Le niveau de maîtrise de la compétence .....	82
3.3.1.2 L'effet du contexte disciplinaire sur la maîtrise de la compétence .....	82
3.3.2 Les variables contextuelles .....	83
3.3.2.1 L'expérience de l'enseignant .....	83
3.3.2.2 La formation initiale de l'enseignant .....	84
3.3.2.3 Le sexe des élèves .....	84
<b>3.4 Traitement et analyse des données .....</b>	<b>84</b>
3.4.1 La correction des situations d'évaluation .....	86
3.4.2 Contrôle de la qualité concernant la fiabilité de la notation .....	88
3.4.3 Logiciels utilisés pour l'analyse et la présentation des données .....	90

<b>CHAPITRE 4. ANALYSE DES DONNÉES .....</b>	<b>91</b>
<b>Introduction.....</b>	<b>91</b>
<b>Description de l'échantillon .....</b>	<b>91</b>
<b>Première partie : Analyse des scores aux différentes situations d'évaluation .....</b>	<b>99</b>
Introduction.....	99
1. Vue d'ensemble des trois situations d'évaluation.....	100
1.1 Analyse descriptive des données.....	100
1.2 Interprétation des données .....	111
2. Analyse détaillée des résultats par composantes et critères de notation .....	113
2.1 Analyse descriptive des données.....	113
2.1.1 Analyse descriptive de la distribution des scores bruts moyens pour la situation d'évaluation 1 (SE1) pour chacun des critères de notation pour les trois niveaux .....	115
2.1.2 Analyse descriptive de la distribution des scores bruts moyens pour la situation d'évaluation 2 (SE2) pour chacun des critères de notation pour les trois niveaux .....	121
2.1.3 Analyse descriptive de la distribution des scores bruts moyens pour la situation d'évaluation 3 (SE3) pour chacun des critères de notation pour les trois niveaux .....	126
2.1.4 Analyse comparative de la distribution des scores bruts moyens concernant les trois situations d'évaluation pour chacun des critères de notation.....	132
2.2 Interprétation des données .....	136
3 Analyse des résultats pour chacun des profils de passation.....	140
3.1 Analyse descriptive des données.....	140
3.2 Interprétation des données .....	167
4. Analyse détaillée des résultats par composantes et critères de notation pour chacun des niveaux : comparaison des élèves en réussite et des élèves en difficulté .....	169
4.1 Analyse descriptive des données.....	169
4.2 Interprétation des données .....	179
<b>Deuxième partie : Robustesse du modèle au regard de certaines variables contextuelles.....</b>	<b>181</b>
1. Les différences entre les garçons et les filles.....	181
1.1 Analyse descriptive des données.....	181
1.2 Interprétation des données .....	183
2. L'effet de l'expérience des enseignants .....	184
2.1 Analyse descriptive des données.....	184
2.2 Interprétation des données .....	188
3. L'effet de la formation initiale des enseignants.....	190
3.1 Analyse descriptive des données.....	190
3.2 Interprétation des données .....	195
<b>Troisième partie : Commentaires des enseignants sur la passation des situations d'évaluation .....</b>	<b>197</b>
1. Gestion de classe et aspects techniques liés à la passation .....	197
1.1 Le suivi auprès des élèves.....	197
1.2 La gestion de classe .....	198

1.3 L'aide apportée par le technicien de laboratoire.....	198
2. Nature des situations d'évaluation.....	199
2.1 Le passage d'un niveau à l'autre.....	199
2.2 L'effet de contamination.....	200
2.3 L'intérêt suscité par chacune des situations d'évaluation.....	200
2.4 L'inadéquation entre les situations d'évaluation et les situations d'apprentissage réalisées durant l'année scolaire.....	201
3. Facteurs liés aux élèves.....	201
3.1 La frustration chez certains élèves.....	202
3.2 La démotivation liée au contexte de passation.....	202
<b>CHAPITRE 5. DISCUSSION DES RÉSULTATS .....</b>	<b>204</b>
1. Résultats concernant la première question de recherche .....	204
1.1 Les informations recueillies.....	204
1.2 Les profils d'élèves.....	206
1.3 L'équivalence des situations de niveau 1 et 3 pour les élèves performants .....	206
1.4 L'efficacité relative du niveau 2.....	208
1.5 Le syndrome de la page blanche.....	209
2. Résultats concernant la deuxième question de recherche.....	210
3. Résultats concernant la troisième question de recherche.....	211
3.1 Première sous-question.....	212
3.2 Deuxième sous-question.....	213
3.3 Troisième sous-question.....	214
Limites de la recherche.....	216
<b>CONCLUSION .....</b>	<b>220</b>
Résumé de la recherche.....	220
Pertinence de la recherche.....	222
Recherches futures.....	223
Pistes de recherche éventuelle nécessitant une nouvelle expérimentation.....	223
Pistes de recherche éventuelle nécessitant le recours à d'autres outils d'analyse.....	224
<b>RÉFÉRENCES.....</b>	<b>226</b>
<b>ANNEXES .....</b>	<b>240</b>
<b>APPENDICE A : Liste des membres du comité d'experts</b>	<b>xxv</b>
<b>APPENDICE B : Outils d'évaluation remis aux membres du comité d'experts</b>	<b>xxvi</b>
<b>APPENDICE C : Lettre de présentation du projet et formulaires de consentement</b>	<b>xli</b>

<b>APPENDICE D : Les situations d'évaluation</b>	<b>xlvi</b>
<b>APPENDICE E : Le DVD contenant les consignes concernant la passation des situations d'évaluation</b>	<b>cliv</b>
<b>APPENDICE F : Le questionnaire administré aux enseignants</b>	<b>clv</b>

## Liste des tableaux

---

Tableau I	Synthèse des situations d'évaluation qui ont été expérimentées. ....	29
Tableau II	Nouvelles tendances en évaluation des apprentissages à l'échelle de la classe .....	38
Tableau III	Description des catégories d'investigation (Watson et al., 1999)....	50
Tableau IV	Principales dimensions caractérisant chacun des niveaux pour une situation d'évaluation donnée. ....	69
Tableau V	Échelle de notation générale des situations d'évaluation .....	87
Tableau VI	Description de chacun des échelons de la grille de notation .....	88
Tableau VII	Répartition des élèves en fonction de l'âge et du sexe. ....	92
Tableau VIII	Répartition des élèves en fonction de la vocation attirée à leur école. ....	93
Tableau IX	Répartition des enseignants en fonction du sexe et de l'expérience.	94
Tableau X	Répartition des enseignants en fonction de leur formation initiale..	95
Tableau XI	Répartition des enseignants en fonction de leur profil de formation continue.....	96
Tableau XII	Répartition des élèves ayant réalisé chacun des profils de passation des situations d'évaluation.....	97
Tableau XIII	Répartition des élèves ayant réalisé chacune des neuf versions des situations d'évaluation. ....	98
Tableau XIV	Statistiques descriptives des scores bruts pour chacune des situations d'évaluation (par niveau). ....	101
Tableau XV	Comparaison deux à deux des moyennes pour toutes les situations d'évaluation et par niveau.....	110
Tableau XVI	Proportion des élèves ayant obtenu un vecteur réponse nul pour chacune des situations d'évaluation (SE) et chacun des niveaux (N). ....	111

Tableau XVII	Correspondance entre les critères de notation et les composantes de la compétence d'investigation scientifique telles que formulées dans le Programme de formation de l'école québécoise, enseignement secondaire (1 <sup>er</sup> cycle), discipline science et technologie. ....	114
Tableau XVIII	Scores moyens pour chacun des critères de notation pour chacun des niveaux de la SE1. ....	117
Tableau XIX	Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE1N1. ....	118
Tableau XX	Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE1N2. ....	119
Tableau XXI	Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE1N3. ....	120
Tableau XXII	Scores moyens des critères de notation pour chacun des niveaux de la SE2. ....	122
Tableau XXIII	Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE2N1. ....	123
Tableau XXIV	Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE2N2. ....	124
Tableau XXV	Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE2N3. ....	125
Tableau XXVI	Scores moyens des critères de notation et de chacune des composantes pour chacun des niveaux de la SE3. ....	128
Tableau XXVII	Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE3N1. ....	129
Tableau XXVIII	Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE3N2. ....	130
Tableau XXIX	Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE3N3. ....	131
Tableau XXX	Scores moyens à chacun des critères de notation pour les trois situations d'évaluation (regroupement par niveau). ....	136
Tableau XXXI	Statistiques descriptives des scores bruts pour chacun des profils de passation pour la situation d'évaluation 1 (SE1). ....	141

Tableau XXXII	Nombre d'élèves en réussite et en échec pour chacun des profils de passation pour chacun des niveaux dans le cadre de la SE1.....	146
Tableau XXXIII	Statistiques descriptives des scores bruts pour chacun des profils de passation pour la situation d'évaluation 2 (SE2). .....	148
Tableau XXXIV	Nombre d'élèves en réussite et en échec pour chacun des profils de passation pour chacun des niveaux dans le cadre de la SE2.....	152
Tableau XXXV	Statistiques descriptives des scores bruts pour chacun des profils de passation pour la situation d'évaluation 3 (SE3). .....	154
Tableau XXXVI	Nombre d'élèves en réussite et en échec pour chacun des profils de passation pour chacun des niveaux dans le cadre de la SE3.....	158
Tableau XXXVII	Statistiques descriptives des scores bruts pour le profil de passation #1 pour chacune des situations d'évaluation. ....	161
Tableau XXXVIII	Statistiques descriptives des scores bruts pour le profil de passation #2 pour chacune des situations d'évaluation. ....	162
Tableau XXXIX	Statistiques descriptives des scores bruts pour le profil de passation #3 pour chacune des situations d'évaluation. ....	165
Tableau XL	Statistiques descriptives des scores bruts pour le profil de passation #4 pour chacune des situations d'évaluation. ....	167
Tableau XLI	Scores moyens pour chacun des critères d'évaluation pour les élèves en réussite et en échec à chacun des niveaux de la SE1. ....	171
Tableau XLII	Scores moyens pour chacun des critères d'évaluation pour les élèves en réussite et en échec à chacun des niveaux de la SE2. ....	175
Tableau XLIII	Scores moyens pour chacun des critères d'évaluation pour les élèves en réussite et en échec à chacun des niveaux de la SE3. ....	177
Tableau XLIV	Résultats du test de Kruskal-Wallis* pour chacun des critères de notation pour les situations d'évaluation 1, 2 et 3 (comparaison des rangs moyens entre les garçons et les filles). ....	183
Tableau XLV	Résultats du test de Kruskal-Wallis* pour les situations d'évaluation 1, 2 et 3 : comparaison des rangs moyens des scores des élèves pour les enseignants ayant moins de 5 ans d'expérience, entre 5 et 10 ans d'expérience et ceux ayant plus de 10 ans d'expérience. ....	185
Tableau XLVI	Résultats du test de Mann-Whitney* pour les situations d'évaluation 1, 2 et 3 : comparaison des rangs moyens des scores des élèves pour	

**les enseignants ayant une formation initiale en enseignement des sciences ou en science par rapport à ceux qui n'en ont pas..... 191**

## Liste des figures

---

Figure 1	Schéma synthèse de la compétence 1 en science et technologie au premier cycle du secondaire. ....	7
Figure 2	Compétences du programme d'éveil scientifique (premier degré) de la Belgique. ....	17
Figure 3	Schéma synthèse du modèle d'évaluation par cycle en science et technologie. ....	22
Figure 4	Types de situations d'évaluation retenues pour évaluer la première compétence du programme de science et technologie (1 <sup>er</sup> cycle)....	25
Figure 5	Séquences possibles d'administration des situations d'évaluation..	30
Figure 6	Démarche pour la conception initiale des situations d'évaluation-bilan. ....	64
Figure 7	Processus de développement d'une situation d'évaluation de performance (Solano-Flores & Shavelson, 1997).....	76
Figure 8	Score moyen par niveau pour chacune des situations d'évaluation (SE). ....	101
Figure 9	Distribution des scores pour la situation d'évaluation 1 (SE1) pour chacun des niveaux (N1, N2 et N3). ....	103
Figure 10	Distribution des scores bruts pour la situation d'évaluation 1 (SE1) pour chacun des niveaux (N1, N2 et N3). ....	104
Figure 11	Distribution des scores pour la situation d'évaluation 2 (SE2) pour chacun des niveaux (N1, N2 et N3). ....	105
Figure 12	Distribution des scores bruts pour la situation d'évaluation 2 (SE2) pour chacun des niveaux (N1, N2 et N3). ....	106
Figure 13	Distribution des scores pour la situation d'évaluation 3 (SE3) pour chacun des niveaux (N1, N2 et N3). ....	107
Figure 14	Distribution des scores bruts pour la situation d'évaluation 3 (SE3) pour chacun des niveaux (N1, N2 et N3). ....	108

Figure 15	Diagramme en moustaches présentant la distribution des scores moyens pour les trois situations d'évaluation pour chacun des niveaux.....	109
Figure 16	Distribution des scores moyens par critère de notation (A) et par composante (B) pour chacun des niveaux de la SE1.....	116
Figure 17	Distribution des scores moyens par critère de notation (A) et par composante (B) pour chacun des niveaux de la SE2.....	121
Figure 18	Distribution des scores moyens par critère de notation (A) et par composante (B) pour chacun des niveaux de la SE3.....	127
Figure 19	Distribution des scores moyens pour chacun des critères de notation de chacune des situations d'évaluation (par niveau).....	133
Figure 20	Distribution des scores pour la situation d'évaluation 1 (SE1) pour les niveaux 1, 2, et 3 (N1, N2 et N3) par profil de passation.....	142
Figure 21	Distribution des scores moyens par profil de passation et par niveau pour SE1.....	143
Figure 22	Distribution par niveau des scores moyens pour les élèves appartenant au profil 3 de la situation d'évaluation 1.....	144
Figure 23	Variation des taux de réussite (et d'échec) pour chacun des niveaux de la situation d'évaluation 1 (SE1).....	146
Figure 24	Distribution des scores pour la situation d'évaluation 2 (SE2) pour les niveaux 1, 2, et 3 (N1, N2 et N3) par profil de passation.....	149
Figure 25	Distribution des scores moyens par profil de passation et par niveau pour SE2.....	150
Figure 26	Distribution par niveau des scores moyens pour les élèves appartenant au profil 3 de la situation d'évaluation 2.....	151
Figure 27	Variation des taux de réussite (et d'échec) pour chacun des niveaux de la situation d'évaluation 2 (SE2).....	153
Figure 28	Distribution des scores pour la situation d'évaluation 3 (SE3) pour les niveaux 1, 2, et 3 (N1, N2 et N3) par profil de passation.....	155
Figure 29	Distribution des scores moyens par profil de passation et par niveau pour SE3.....	156
Figure 30	Distribution par niveau des scores moyens pour les élèves appartenant au profil 3 de la situation d'évaluation 3.....	157

Figure 31	Variation des taux de réussite (et d'échec) pour chacun des niveaux de la situation d'évaluation 3 (SE3).....	159
Figure 32	Distribution des scores moyens pour chacune des situations d'évaluation (profil 1).....	160
Figure 33	Distribution des scores moyens pour chacune des situations d'évaluation (profil 2).....	163
Figure 34	Distribution des scores moyens pour chacune des situations d'évaluation (profil 3).....	164
Figure 35	Distribution des scores moyens pour chacune des situations d'évaluation (profil 4).....	166
Figure 36	Variation du score moyen des élèves en échec et en réussite pour chacun des niveaux de la situation d'évaluation 1.....	172
Figure 37	Écart entre les scores moyens des élèves en réussite et ceux en échec pour chacun des critères de notation et pour chacun des niveaux de la SE1.....	173
Figure 38	Variation du score moyen des élèves en échec et en réussite pour chacun des niveaux de la situation d'évaluation 2.....	174
Figure 39	Écart entre les scores moyens des élèves en réussite et ceux en échec pour chacun des critères de notation et pour chacun des niveaux de la SE2.....	176
Figure 40	Variation du score moyen des élèves en échec et en réussite pour chacun des niveaux de la situation d'évaluation 3.....	178
Figure 41	Écart entre les scores moyens des élèves en réussite et ceux en échec pour chacun des critères de notation et pour chacun des niveaux de la SE2.....	179
Figure 42	Distribution des scores obtenus aux neuf situations d'évaluation pour les garçons et les filles.....	182
Figure 43	Diagrammes en forme de boîte présentant la distribution des scores moyens pour la situation d'évaluation 1 en fonction du critère d'expérience des enseignants.....	186
Figure 44	Diagrammes en forme de boîte présentant la distribution des scores moyens pour la situation d'évaluation 2 en fonction du critère d'expérience des enseignants.....	187

Figure 45	Diagrammes en forme de boîte présentant la distribution des scores moyens pour la situation d'évaluation 3 en fonction du critère d'expérience des enseignants.....	188
Figure 46	Diagrammes en forme de boîte présentant la distribution des scores moyens pour la situation d'évaluation 1 en fonction de la formation initiale des enseignants. ....	192
Figure 47	Diagrammes en forme de boîte présentant la distribution des scores moyens pour la situation d'évaluation 2 en fonction de la formation initiale des enseignants. ....	193
Figure 48	Diagrammes en forme de boîte présentant la distribution des scores moyens pour la situation d'évaluation 3 en fonction de la formation initiale des enseignants. ....	195

## Liste des sigles et abréviations

---

É.-T.	Écart-type
EBP	Évaluation basée sur la performance
MELS	Ministère de l'Éducation, du Loisir et des Sports
MEQ	Ministère de l'Éducation du Québec
Moy	Moyenne
NSP	Ne s'applique pas
PEI	Programme d'éducation internationale
SE	Situation d'évaluation
S&T	Science et Technologie

## Dédicace

---

*Cette thèse est dédiée à ma fille Marianne ainsi qu'aux  
prochain(e)s Dionne qui s'ajouteront éventuellement,  
nous l'espérons, à notre famille ...*

## Remerciements

---

Mes études doctorales ont débuté en janvier 2002. Au cours des cinq dernières années, et bien avant, plusieurs personnes ont été importantes dans ma vie personnelle et professionnelle.

Dans un premier temps, vous me permettrez de remercier d'abord mes proches qui ont subi les contraintes liées à une recherche doctorale. Merci à mon épouse Catherine pour son soutien indéfectible et pour sa patience légendaire. Merci aussi d'avoir réalisé la révision linguistique de ce texte. Une pensée toute spéciale à mes parents, Gérard et Jeannette. Issus d'un milieu modeste, ils ont eu l'intelligence de valoriser l'éducation de leur fils sans trop savoir où cela le mènerait... Merci infiniment pour tout. Merci également à Thérèse et Roger qui m'ont largement aidé afin que je puisse accéder, sans trop me ruiner, aux études avancées. Merci à Joëlle Dewasmes pour ses encouragements incessants et son optimisme... débordant ! Une pensée aussi pour Laurent, Sissy et Yumin. Merci à mes amis pour m'avoir changé les idées dans les moments « d'essoufflement » : France Dubé et son Harold, Marc Daviau et sa conjointe Brigitte, Marc Chartrand et France Charbonneau, Bernard Thibodeau et Linda St-Jean, Phylippe Laurendeau, Marie-Claude Chartrand, Suzanne Gosselin et Eric Muloin. Merci à Gino Ciarlo pour avoir prit soin de ma santé physique sur les courts de tennis... Merci à Sylvie Chartrand pour ses nombreux et généreux conseils fiscaux. Certaines personnes m'ont marqué au niveau intellectuel au cours de mes études. Je pense à Martin Provost, à Sami Haddad et à Réjean Labrie, de véritables modèles pour moi. Merci à Marino Tremblay qui m'a accompagné à une époque particulièrement difficile où j'ai abandonné l'école et qui a fait en sorte que je puisse y retourner et accéder ainsi au dernier palier du système éducatif...

Dans un second temps, j'aimerais remercier les acteurs principaux et de soutien liés à cette recherche. D'abord, merci aux élèves et aux enseignants d'avoir accepté de prendre le risque de participer à cette étude. Merci également aux conseillers pédagogiques qui ont bien voulu m'épauler dans cette aventure. Je pense particulièrement à Daniel Ricard,

Martine Filiatrault, Brigitte Lussier et Erick Sauvé. Un merci tout spécial à Myriam Larue pour son dévouement exceptionnel et pour son dynamisme contagieux ! Sans toutes ces personnes, il ne fait aucun doute que cette recherche n'aurait jamais pu être menée à terme. Mille mercis.

Dans un troisième temps, mes sincères salutations s'adressent à mes différents collègues qui ont contribué à ma réflexion au cours des nombreuses années. En premier lieu, je pense à mes amis Paule Bellehumeur, Julie Cartier, Luc Prud'homme et Marie-Claude Veilleux. Ces excellents pédagogues ont été, et sont encore, une source constante de ressourcement. Merci également à Patrice Potvin de l'UQAM, à Pierre Lachance du RÉCIT-MST et à Denis Fyfe du CDP qui m'ont permis de confronter mes idées et parfois mes lubies ! Merci à Julie Grondin qui a réalisé une lecture et de nombreuses relectures qui ont permis d'améliorer la qualité de la langue de ce texte ainsi que sa clarté. Les erreurs et les aspects ambigus demeurent de mon entière responsabilité.

Enfin, ces remerciements ne seraient pas complets sans une pensée pour mes directeurs qui m'ont accompagné de main de maître au cours de ce processus. Merci à Michel D. Laurier qui a accepté de superviser cette thèse malgré une charge de travail déjà colossale. Ses réflexions brillantes, ses encouragements et son soutien ont été appréciés tout au long de mes travaux. Merci également à Jesus Vazquez-Abad pour son soutien moral et pour ses analyses toujours pertinentes. Ces deux personnes ont fait de moi un bien meilleur chercheur. Merci. Je voudrais également souligner la contribution de Jean-Guy Blais au cours de mes études doctorales. Ses encouragements à la publication scientifique et au dépassement ont été, et sont encore, grandement appréciés. Merci aussi de m'avoir permis, sans jugement et sans rancœur, de reprendre un nouveau départ à un moment où j'en avais besoin. À tous ceux que j'ai pu oublier, mes regrets les plus sincères et... Merci !

## Introduction

---

Plusieurs pays procèdent depuis quelques années à de profondes réformes touchant, entre autres, l'enseignement des disciplines scientifiques et technologiques. Ces réformes visent essentiellement à mieux préparer les jeunes aux nombreux défis qui les attendent au cours du vingt et unième siècle en leur fournissant une culture scientifique et technologique suffisante afin qu'ils puissent comprendre et agir de façon éclairée dans le cadre d'enjeux sociétaux. Mais préparer les jeunes à faire face à ces défis n'est pas une mince tâche. Au Québec, les autorités concernées ont décidé de faire *tabula rasa* en remplaçant l'ensemble des programmes d'études scientifiques et technologiques autant au primaire qu'au secondaire. Les anciens programmes formulés par objectifs ont maintenant fait place à des programmes formulés par compétences. Les approches pédagogiques à tendance behavioriste ont été reléguées au second plan à la faveur des approches socio-constructiviste, constructiviste et cognitiviste. La perspective disciplinaire a été supplantée à la faveur d'une perspective interdisciplinaire. Plusieurs considèrent d'ailleurs que ces changements représentent, ni plus ni moins, une révolution dans le paysage de l'enseignement scientifique et technologique québécois. Pour les enseignants<sup>2</sup> du secondaire, ces changements ont des impacts à plusieurs niveaux. Premièrement, ils doivent s'approprier de nouvelles connaissances pédagogiques, scientifiques et technologiques. En effet, on peut dire – sans exagérer – que les connaissances dans ces trois domaines ont explosé au cours des dernières décennies. D'ailleurs, les réformes s'appuient de plus en plus sur des résultats de recherche émanant du secteur de l'éducation. En ce qui concerne les sciences et la technologie, l'explosion des connaissances est particulièrement manifeste depuis environ un siècle. Deuxièmement, ils doivent intégrer ces nouvelles connaissances à leur enseignement tout en s'adaptant afin de tenir compte des nombreuses contraintes pédagogiques et administratives auxquelles ils sont confrontés dans leur milieu respectif. Troisièmement, ils doivent modifier la façon d'évaluer leurs élèves afin de tenir compte des fondements des nouveaux programmes d'études. Dans le cadre de cette recherche, nous

---

<sup>2</sup> Dans cet ouvrage le masculin est utilisé à titre épique dans le seul but d'alléger le texte.

nous penchons précisément sur ce dernier aspect lié à l'évaluation. En fait, comme nous le verrons dans le premier chapitre, nous essayons de mieux comprendre les différentes modalités d'évaluation qui peuvent affecter l'évaluation d'une compétence scientifique en laboratoire dans le contexte d'évaluation particulièrement rigide que constitue l'évaluation certificative.

Cette thèse est divisée en cinq chapitres. Le premier est consacré à la problématique qui présente le contexte général de la recherche ainsi que les problèmes auxquels nous nous attaquons dans le cadre de cette étude. Le second chapitre est dédié au cadre théorique qui présente le point de vue des différents auteurs qui se sont intéressés à des problématiques semblables à la nôtre. Le troisième chapitre traite de la méthodologie sur laquelle nous nous sommes appuyés afin de collecter, traiter et analyser les données. Le quatrième chapitre est dédié à la présentation des résultats. Enfin, le cinquième chapitre présente les principales interprétations issues du traitement des données.

S'il n'y a pas de solution c'est qu'il n'y a pas de problème.  
Jacques Rouxel

## **Chapitre 1. Présentation et justification de la recherche**

---

### ***1.1 Problématique***

Nous avons choisi de présenter la problématique en trois parties. La première est consacrée à la description du renouveau qui touche l'enseignement des sciences au Québec depuis quelques années. Plus particulièrement, nous présentons les aspects qui militent en faveur du développement de compétences d'investigation scientifique en laboratoire. En seconde partie, nous abordons les facteurs liés à l'évaluation des apprentissages et des compétences dans le contexte de l'enseignement des sciences que nous avons précédemment exposé. Plus précisément, nous traitons des considérations liées à l'évaluation certificative et des défis qui marquent ce contexte d'évaluation. Enfin, dans la troisième partie, nous présentons le modèle que nous avons développé afin de tenir compte des multiples contraintes liées à l'évaluation certificative d'une compétence d'investigation scientifique en laboratoire.

Avant d'aborder le cœur de la problématique, nous pensons qu'une parenthèse s'impose. L'investigation scientifique en laboratoire est une partie importante du développement de la culture scientifique des élèves. Dans la très grande majorité des programmes d'études nationaux, on retrouve des apprentissages à développer qui touchent à l'investigation scientifique. En parcourant les programmes d'études nationaux de différents pays occidentaux (Belgique, Australie, États-Unis, etc.), on constate que l'investigation scientifique en laboratoire occupe une place de choix. Inévitablement, les choix politiques colorent le sens à donner à la compétence d'investigation scientifique en laboratoire. À certains endroits, on met l'accent sur l'aspect technique de l'investigation en demandant aux élèves, par exemple, de maîtriser les techniques de laboratoire. À d'autres endroits, on met plutôt l'accent sur des aspects liés à la démarche afin de développer les habiletés de résolution de problème dans un contexte bien précis à savoir celui du travail en laboratoire. Puisque le Québec procède présentement à une vaste opération de remplacement de ses

programmes d'études en science et technologie, nous avons fait le choix d'essayer de nous arrimer le plus possible à ces nouveaux programmes afin que nos résultats et nos conclusions soient le plus possible signifiants dans le contexte actuel qui prévaut au Québec. Par contre, nous avons aussi fait des choix qui peuvent s'éloigner des prescriptions ministérielles mais que nous jugeons absolument essentiels à faire afin de maintenir la cohérence de notre modèle. Ainsi, nous pensons que le cadre théorique de notre recherche se situe à niveau mitoyen entre un arrimage strict aux nouveaux programmes et une absence totale d'arrimage à la réalité pédagogique qui marque actuellement l'enseignement de la science et de la technologie au Québec.

### **1.1.1 Contexte du renouveau en enseignement des sciences au Québec**

La réforme de l'éducation québécoise, par le biais des prescriptions des nouveaux programmes d'études, exige maintenant des enseignants du secondaire qu'ils développent chez leurs élèves des compétences (Ministère de l'Éducation du Québec, 2001, 2003c). Certaines sont plutôt génériques, les compétences transversales, alors que d'autres sont plus spécifiques aux différentes disciplines. En science et technologie, les élèves du premier cycle ont deux ans pour développer les trois compétences qui suivent (Ministère de l'Éducation du Québec, 2003c, p. 275 à 281) :

- |                      |   |
|----------------------|---|
| Première compétence  | Chercher des réponses ou des solutions à des problèmes d'ordre scientifique ou technologique. |
| Deuxième compétence  | Mettre à profit des connaissances scientifiques et technologiques.                            |
| Troisième compétence | Communiquer à l'aide des langages utilisés en science et technologie.                         |

La première compétence consiste pour les élèves à résoudre des problèmes scientifiques ou technologiques de type *hands-on*. Il s'agit donc pour les élèves de concevoir et de réaliser un dispositif expérimental en laboratoire permettant d'apporter une ou des solution(s) au problème qui leur est présenté. Nous reviendrons subséquemment sur la définition du concept d'activités de type *hands-on*. Il existe plusieurs définitions associées à ce concept

mais dans le cadre de ce texte nous souscrivons à la proposition de Lump et Oliver qui sera expliquée plus loin dans la présente section et qui met l'accent sur trois dimensions : un aspect de découverte, l'autonomie laissée à l'élève et la mise en œuvre d'un dispositif de preuve expérimentale. Ces dimensions nous apparaissent étroitement liées aux prescriptions du MELS. La section 2.3 présente une synthèse des principales définitions et caractéristiques du concept d'activités *hands-on*. La seconde compétence exige des élèves qu'ils expliquent soit des phénomènes scientifiques, soit le fonctionnement d'un objet technique en ayant recours aux concepts pertinents et aux éléments contextuels appropriés (ex. aspects éthiques, aspects historiques, etc.). La troisième compétence, quant à elle, implique que les élèves communiquent adéquatement l'information scientifique ou technologique en ajustant, par exemple, leur discours à l'interlocuteur. En contexte d'apprentissage, nous pouvons facilement supposer que ces compétences ne se développent pas de façon séquentielle. De la même manière, une activité pédagogique peut cibler le développement de plusieurs compétences disciplinaires et de quelques compétences transversales. En contexte d'évaluation, la situation est différente puisqu'il s'agit alors de rendre compte du développement de chacune des compétences. Cette recherche s'intéresse à l'évaluation de la compétence d'investigation scientifique en laboratoire ce qui correspond à la première compétence du programme de premier cycle en science et technologie. Nous prendrons soin d'analyser celle-ci en omettant – pour des raisons d'économie – de faire de même pour les deux autres compétences qui ont pour nous moins d'intérêt à l'intérieur de cette recherche.

Pour les fins de notre analyse, nous présenterons d'abord les aspects importants de la compétence en question en nous référant aux prescriptions du Programme de formation et en relevant les choix que nous avons réalisés afin de palier à certaines omissions ou certains aspects nébuleux du Programme de formation. Par la suite, nous identifierons les aspects problématiques principalement au regard de l'évaluation.

Le ministère de l'Éducation définit opérationnellement la première compétence au moyen de deux types de démarche de résolution de problèmes que cette compétence vise à développer chez les élèves : la démarche d'investigation en science et la démarche de

conception d'un objet technique en technologie. En ce sens, nous pouvons dire que cette compétence a la particularité d'être duelle comme l'illustre la figure 1. Les éléments de démarche se retrouvant dans les rectangles noircis sont communs aux deux démarches (scientifique et technologique). Les éléments apparaissant dans les rectangles grisés sont propres à la science alors que ceux se retrouvant dans des rectangles sur fond blanc sont caractéristiques de la démarche technologique. En analysant la figure 1, on constate que les éléments de démarche qui diffèrent ont trait à la concrétisation de la démarche c'est-à-dire au moment où les élèves doivent véritablement mettre la main à la pâte. En science, il s'agit de réaliser une expérience ou de collecter les données et ensuite d'analyser les résultats alors qu'en technologie il s'agit de fabriquer le prototype et de procéder à la mise à l'essai. Pour des raisons qui seront expliquées plus loin dans cette section, nous avons choisi de ne pas aborder le volet technologique dans le cadre de cette recherche. Ainsi, à partir de maintenant, nous ne nous référerons qu'au volet scientifique lié à la première compétence donc à la partie de gauche (rectangles noirs et gris) de la démarche générique illustrée à la figure 1.

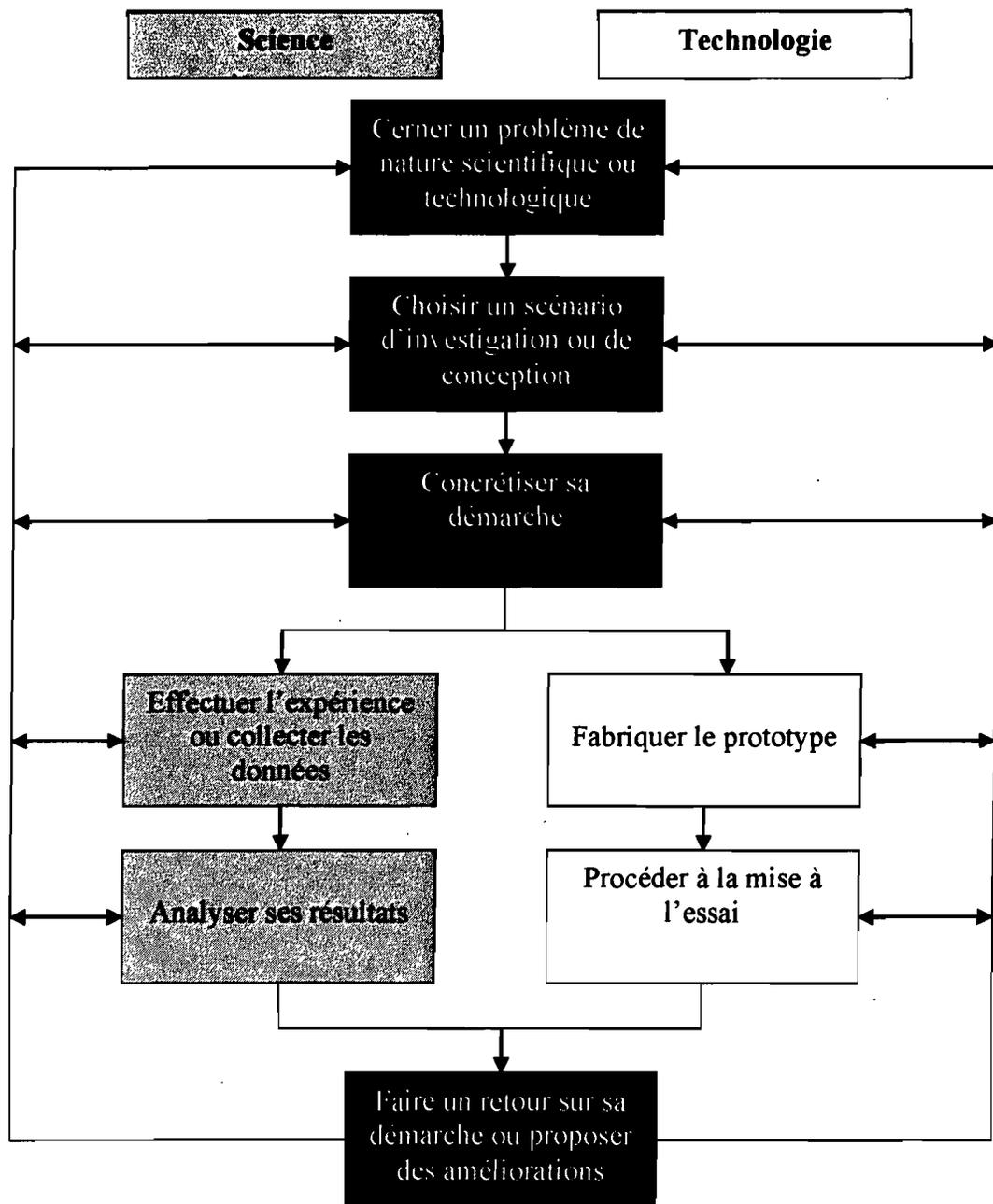


Figure 1  
Schéma synthèse de la compétence 1 en science et technologie au premier cycle du secondaire<sup>3</sup>.

<sup>3</sup> Adapté du Programme de formation de l'école québécoise (Ministère de l'Éducation du Québec, 2003c, p. 276).

Les nombreuses flèches de rétroaction apparaissant dans la figure 1 témoignent du caractère itératif des démarches impliquées dans le développement de cette compétence. L'élève peut, à différentes étapes de sa démarche, faire un retour en arrière afin, par exemple, d'explorer une nouvelle avenue ou réaliser à nouveau une étape antérieure pour se rassurer par rapport à sa réflexion ou à ses résultats. Ces flèches permettent de rendre compte du caractère « résolution de problème » accolé à cette compétence. Cet aspect, qui n'est pas nouveau dans les programmes d'études actuels de science et de technologie, est cependant bien mis en évidence dans le nouveau programme dont il est ici question. L'idée est de mieux représenter la nature véritable d'une démarche scientifique qui s'éloigne nettement, dans la pratique de la recherche scientifique, de la linéarité du modèle classique : observation → hypothèse → manipulation → analyse → conclusion. En ce sens, cette compétence rejoint plusieurs des caractéristiques associées au travail scientifique en laboratoire. Bien que la littérature n'offre pas un consensus clair, nous nous référons à la définition de Lump et Oliver (1991) qui ont identifié trois dimensions caractéristiques d'activités en laboratoire de type *hands-on* et que nous pouvons adapter à l'approche par compétences retenue dans les nouveaux programmes : (1) une dimension liée à la découverte (*inquiry dimension*), (2) une dimension liée à l'autonomie de l'élève (*structure dimension*) et (3) une dimension liée à la mise en œuvre d'un dispositif de preuve expérimentale (*experimental dimension*). Examinons le lien unissant chacune de ces dimensions avec les fondements de la première compétence du programme. La première dimension fait référence à l'aspect « découverte » qui devrait être associé à une investigation scientifique. En effet, une investigation scientifique ne permet généralement pas de déterminer à l'avance et de façon précise les résultats qui seront obtenus au terme du processus de recherche. Au niveau pédagogique, cette dimension a un impact majeur sur le rôle de l'élève mais aussi sur le rôle de l'enseignant. Ici comme ailleurs, on reproche régulièrement aux activités scientifiques se déroulant en laboratoire de ne servir qu'à valider la théorie enseignée par les enseignants (Dionne, 1999; Duschl, 2003). On peut

qualifier ces activités de laboratoires « recettes<sup>4</sup> ». Cette situation n'est cependant pas propre aux écoles secondaires québécoises. Aux États-Unis, on fait des constats fort similaires :

*The usual high-school science "experiment" is unlike the real thing: The question to be investigated is decided by the teacher, not the investigator; what apparatus to use, what data to collect, and how to organize the data are also decided by the teacher (or the lab manual); time is not made available for repetitions or, when things are not working out, for revising the experiment; the results are not presented to other investigators for criticism; and; to top it off, the correct answer is known ahead of time. (American Association for the Advancement of Science, 1993, p. 9)*

La seconde dimension est étroitement liée à la première et concerne deux aspects. D'abord, en s'en tenant exclusivement à la définition de Lump et Oliver, il est question de l'autonomie à laisser à l'élève dans l'exercice de cette investigation scientifique. Ici, on fait face à un dilemme. Pour ne pas dénaturer l'activité d'investigation scientifique, il faut nécessairement laisser suffisamment d'autonomie à l'élève. Par contre, un niveau d'autonomie trop grand peut faire en sorte que certains élèves soient désorientés et ne sachent ni quoi faire et ni comment le faire. Normalement, un élève compétent devrait être en mesure de réaliser une démarche d'investigation en contexte d'autonomie parfaite. Ensuite, cette dimension peut se distinguer quant au niveau d'ouverture inhérent à l'énoncé du problème. À ce sujet, Arsac, Germain et Mante (1988) (cité dans Astolfi et al. (1997, p. 141)) ont identifié trois caractéristiques associées à un problème ouvert :

L'énoncé doit être court. L'énoncé ne doit donner aucun indice quant à la méthode, ni la solution (pas de questions intermédiaires ni de questions du type « montrer que »). En aucun cas, cette solution ne doit se réduire à l'utilisation ou l'application immédiate des derniers résultats présentés au cours. Le problème se trouve dans un domaine conceptuel avec lequel les élèves ont assez peu de familiarités. Ainsi, peuvent-ils prendre facilement « possession » de la situation et s'engager dans des essais, des conjectures, des projets de résolution, des contre-exemples. (Arsac et al., 1988)

---

<sup>4</sup> Un laboratoire de type « recette » est une activité pédagogique où les étapes expérimentales sont fournies et souvent expliquées aux élèves avant qu'ils ne procèdent eux-mêmes aux manipulations. Le nombre d'hypothèses plausibles et vraisemblables est généralement limité. Les résultats expérimentaux sont généralement communs à tous les élèves (problèmes fermés) et une seule conclusion est possible au terme de la démarche.

Cette caractéristique d'ouverture associée aux problèmes est particulièrement présente et mise en évidence dans le nouveau programme. On vise ainsi à donner à l'élève des problèmes véritables pour lesquels une démarche scientifique rigoureuse sera incontournable à réaliser pour accomplir la tâche. Enfin, la troisième dimension concerne la mise en oeuvre de mécanismes permettant à l'élève d'appuyer ses réflexions sur des preuves expérimentales. Ces preuves peuvent reposer par exemple sur des observations qualitatives ou quantitatives. Cette dimension est caractéristique de l'activité scientifique et c'est une des caractéristiques qui la distingue d'ailleurs d'activités non-scientifiques ou pseudo-scientifiques.

Les dimensions retenues par Lump et Oliver (1991) nous apparaissent intéressantes, et ce, à deux niveaux. D'abord, elles ne mettent pas l'accent uniquement sur le caractère « pratique » où l'élève doit mettre la « main à la pâte » comme plusieurs auteurs le suggèrent. Une activité en laboratoire de type *hands-on*<sup>5</sup>, caractéristique de la première compétence, implique aussi le déploiement d'un raisonnement rigoureux mais aussi imaginatif afin de trouver des solutions méthodologiques ingénieuses pour résoudre le problème. Il ne s'agit donc pas uniquement de suivre un protocole, où toutes les étapes sont fournies et qui consiste pour l'élève à réussir chacune de ces étapes sans commettre d'erreurs de manipulation. Ensuite, ces dimensions mettent en lumière le fait que l'élève doit, à son niveau, mettre en oeuvre une ou des démarches associées aux sciences. Il ne s'agit donc pas uniquement et exclusivement de reproduire des démarches apprises en classe mais de mettre en oeuvre des démarches intellectuelles qui s'apparentent aux démarches des scientifiques professionnels.

La compétence, telle que définie par le MELS – et dont il est ici question – souffre d'un certain nombre de difficultés. Dans un premier temps, abordons les difficultés liées à son caractère duel. Ce choix est discutable à au moins deux égards. Premièrement, le type de situation d'apprentissage associé à chacune de ces démarches (scientifique et

---

<sup>5</sup> Nous avons constaté que l'expression « *hands-on* » ne faisait pas consensus dans la littérature. Dans le cadre de cette recherche, nous considérons qu'une activité « *hands-on* » implique la manipulation de matériel de laboratoire en vue de résoudre un problème à caractère scientifique.

technologique) est clairement différent selon que l'on se trouve dans un contexte ou dans l'autre. En empruntant le concept de « familles de situation » à Roegiers (2000), on constate que les situations d'investigations scientifiques ne peuvent s'assimiler aux démarches de conception en technologie. Ensuite, les éléments de performance qu'il est possible d'observer, dans un contexte d'évaluation, sont différents selon que l'on s'inscrit dans un contexte scientifique ou technologique. Par exemple, les aspects liés à l'esthétisme sont généralement peu signifiants en science alors qu'ils ne peuvent être négligés en technologie. Cette dichotomie est d'ailleurs clairement visible dans les attentes de fin de cycle qui sont distinctes pour le volet scientifique et le volet technologique (Ministère de l'Éducation du Québec, 2003c, p. 277). C'est pourquoi nous avons fait le choix de ne pas considérer la dimension technologique dans le cadre de cette recherche. À notre avis, la position du Ministère au regard des compétences en science est nébuleuse à propos du rôle de la communication dans la mise en œuvre, entre autres, de la première compétence. Nous considérons que la compétence liée à la communication (compétence 3) devrait être jugée en concomitance avec, d'une part, la première compétence et, d'autre part, avec la seconde compétence. En effet, juger de la compétence de communication scientifique est sans intérêt si l'élève n'a pas une intention précise à poursuivre. Cette intention est, la plupart du temps, explicite dans le cadre d'une activité qui vise à développer ou évaluer les démarches mises de l'avant lors de la compétence 1 ou de la compétence 2. Par conséquent, nous avons fait le choix de considérer les aspects liés à la compétence de communication scientifique dans le cadre du modèle que nous avons développé.

Le programme de science et technologie intègre des contenus provenant de cinq disciplines scientifiques (chimie, physique, biologie, géologie et astronomie) en plus de la technologie. Les concepts issus de ces disciplines ont été organisés en quatre « univers » : univers matériel, univers vivant, Terre et espace et univers technologique (Ministère de l'Éducation du Québec, 2003c). Le contenu de formation regroupe également les stratégies, techniques et attitudes qui sont autant d'éléments prescriptifs du programme.

Le programme de science et technologie se caractérise, comme nous l'avons vu précédemment, par des démarches à acquérir mais aussi par son contenu de formation.

L'un ne va pas sans l'autre et les deux ont une importance indéniable au regard de l'évaluation des compétences des élèves. De fait, puisque le contenu de formation est prescriptif, on doit s'assurer de la maîtrise des concepts de la part des élèves. Cependant, afin de respecter les orientations du programme, il importe que l'élève fasse montre de la maîtrise des concepts dans des situations qui permettent le déploiement de compétences. Il ne s'agit donc pas de recourir à des situations d'évaluation « décontextualisées » qui agiraient en parallèle à l'évaluation des compétences. Autrement dit, les contenus doivent être évalués en contexte dans des situations d'évaluation qui ressemblent le plus possible aux situations d'apprentissage. Cet attrait pour l'approche interdisciplinaire, bien qu'intéressante d'un point de vue didactique, pose cependant des problèmes supplémentaires à la problématique déjà lourde de l'évaluation des compétences en science et technologie. Examinons quelques-uns de ces problèmes. D'abord, il n'existe aucune proposition émanant du Ministère sur le modèle d'interdisciplinarité à privilégier en science et technologie. Pourtant, les modèles d'intégration pour un curriculum ne manquent pas. Fogarty (1991) propose, par exemple, trois modèles d'intégration intra-disciplinaires qui pourraient être fort pertinents pour structurer le niveau d'interdisciplinarité attendu. D'ailleurs, l'absence d'un modèle d'intégration se fait particulièrement sentir quand vient le moment de planifier l'évaluation certificative, ce qui nous amène à nous poser des questions importantes : les situations d'évaluation doivent-elles être impérativement interdisciplinaires ? Si oui, combien de champs disciplinaires doivent être touchés dans une telle situation d'évaluation ? Le second problème qui se pose concerne la difficulté à bâtir des situations interdisciplinaires pour des élèves du premier cycle (13-14 ans). À nouveau, on constate qu'il n'existe présentement aucun modèle permettant de donner des indications méthodologiques concernant la rédaction de situations d'apprentissage et d'évaluation interdisciplinaires. Ce vide méthodologique se ressent singulièrement quand vient le moment de concevoir des situations pédagogiques qui tentent de créer des liens entre des Univers pour lesquels les liens sont moins naturels, par exemple entre l'Univers technologique et l'Univers Terre et Espace. Pour illustrer cette difficulté, mentionnons qu'au cours d'une simulation d'implantation du programme de science et technologie, Dionne et Potvin (2007) ont constaté que, sur les trois situations d'apprentissage

développées par une équipe d'enseignants du premier cycle du secondaire, aucune ne faisait appel spécifiquement à des concepts de deux Univers ou plus.

### **1.1.2 L'évaluation des compétences en science et technologie**

L'évaluation des compétences, et plus globalement l'évaluation des opérations cognitives complexes, représente un défi important autant pour les chercheurs que pour les enseignants des différents réseaux scolaires. Dans les sections qui suivent, nous présenterons trois dimensions que nous jugeons importantes en ce qui concerne l'évaluation des compétences en science et technologie. Dans un premier temps, nous discuterons de la cohérence du triangle curriculum-apprentissage-évaluation. Dans un second temps, nous discuterons de la nécessité de proposer et d'expérimenter de nouveaux modèles d'évaluation qui tiennent compte des nombreuses contraintes théoriques, méthodologiques et pragmatiques liées à l'évaluation de compétences. Enfin, nous terminerons cette section en présentant les principales contraintes que nous avons identifiées dans ce contexte.

#### **1.1.2.1 La cohérence du triangle curriculum-apprentissage-évaluation**

Le choix de formuler des programmes d'études par compétences a un impact évident et majeur sur l'évaluation des apprentissages des élèves. En effet, comme le souligne le Ministère dans sa Politique d'évaluation des apprentissages (Ministère de l'Éducation du Québec, 2003b), il est de toute première importance que l'évaluation des apprentissages soit rigoureuse, transparente et cohérente. La rigueur fait référence aux standards de qualité et aux mesures prises permettant de s'assurer que l'évaluation soit valide et fidèle. La transparence, quant à elle, souligne l'importance que les différentes composantes de la démarche d'évaluation soient connues à l'avance et bien communiquées aux différents intervenants (élèves, parents, administrateurs scolaires, public en général, etc.). Enfin, la cohérence permet de s'assurer du lien important unissant la démarche et le produit de l'évaluation avec les fondements des nouveaux programmes. Dans le passé, la cohérence

entre les fondements d'un programme et l'évaluation qui lui était associée a parfois fait défaut. L'application du programme d'études de sciences physiques 416-436 (Dionne, 1999) en est d'ailleurs un exemple éloquent. Nous avons alors montré que l'une des causes expliquant la faible application du programme (approche constructiviste, accent sur des activités de résolution de problème, etc.) est la nature des évaluations ministérielles qui s'éloignaient (et qui s'éloignent encore...) nettement des visées éducatives du programme. Par exemple, les épreuves ministérielles, composées majoritairement d'items à correction objective, mesurent essentiellement - si l'on se réfère à la taxonomie de Bloom (1975) - la connaissance et la compréhension. À ce sujet, on constate que les initiateurs de la réforme québécoise (Ministère de l'Éducation du Québec, 1997, p. 95) sont d'ailleurs sensibles à ces dérives :

Ainsi, nous recommandons que les épreuves d'évaluation développées tant par les écoles que par le ministère de l'Éducation, aux fins de contribuer à l'évaluation individuelle des apprentissages des élèves, délaissent la forme dite objective (à choix multiple, par exemple) et qu'elles soient plutôt composées de *questions ouvertes*, c'est-à-dire empruntant aux modèles qui demandent, par exemple, des réponses avec argumentation, des réponses libres, des réponses illustrées par des données, des réponses orales, des activités pratiques, des portfolios particulièrement pour les disciplines artistiques, etc. Ce type d'instrument de mesure correspond davantage aux visées d'intégration des savoirs et de développement de l'esprit de synthèse que préconise la réforme que nous proposons.

Plusieurs auteurs (Anderson, 2002; Auger & Dassa, 1992; Bain, 1999; Baker, 2004; Beckers, 2002; Bélair, 1999; Dionne, 2005; Hadji, 1997; Morissette, 1984; Nitko, 1993; Popham, 2001b; Wiggins, 1993b; Wilson & Scalise, 2003) soulignent l'importance de la cohérence (ou de la congruence) unissant le triangle programme d'études - enseignement et apprentissage - évaluation, ajoutant que cette condition est essentielle pour assurer la validité de construit. Ainsi, les situations d'évaluation que nous proposons dans le cadre de cette thèse ont été construites afin de minimiser les écarts entre les trois pôles de ce triangle.

Ajoutons en terminant que la problématique associée à l'évaluation des apprentissages - dans le cadre de la réforme - est particulièrement vive au secondaire (Dionne, 2004). En ce sens, les propositions émises ci-dessus par le MEQ représentent certainement des

changements radicaux par rapport aux pratiques traditionnelles concernant l'évaluation des apprentissages. La culture des enseignants du secondaire, la complexité de leur tâche et le contexte dans lequel ils œuvrent (ex. le nombre d'élèves par classe, le nombre d'élèves par enseignant, disponibilité parfois restreinte des classes-laboratoires, etc.) sont quelques-uns des facteurs qui expliquent pourquoi les changements en évaluation exigent autant de temps et d'énergie. Notons que ces aspects ne sont pas exclusifs à la réforme québécoise. On arrive à des constats similaires autant en Europe (Beckers, 2002) qu'aux États-Unis (Baker, 2004).

### **1.1.2.2 L'évaluation des apprentissages en science : des modèles à développer**

Dans l'application de la présente réforme, l'évaluation est sans contredit l'un des aspects qui inquiète le plus les enseignants, les conseillers pédagogiques et les administrateurs scolaires (Dionne & Blais, 2004; Dionne & Potvin, 2005). La plupart de ces acteurs souhaitent obtenir de l'aide concernant l'instrumentation et les différentes méthodologies sur lesquelles il est possible de s'appuyer pour inférer le niveau de développement des compétences des élèves. Le Ministère est d'ailleurs sensible à ces inquiétudes et semble conscient des travaux à mener dans ce domaine car il s'est engagé à soutenir des initiatives de recherches permettant de proposer des pistes de solution utiles pour les praticiens :

Les recherches-actions, menées en partenariat avec les universités, seront orientées vers la mise au point de modèles d'évaluation et vers des solutions concrètes aux problèmes que rencontrent les milieux scolaires. » (Ministère de l'Éducation du Québec, 2003a, p. 5)

Ces craintes ne sont pas sans fondement. Dans une approche par compétences, la place du jugement professionnel est, par exemple, nettement plus visible que dans une approche par objectifs. Dans une telle approche, le jugement de l'enseignant est souvent masqué par des illusions de rigueur basées sur des sommations mathématiques liées aux « bonnes réponses ». Comme le souligne Scallon (2001), « dans une pédagogie par objectifs, il s'agissait de dénombrer les tâches réussies pour marquer à divers moments le nombre d'objectifs maîtrisés ». Bref, le jugement était souvent limité à une mesure chiffrée dite

objective basée sur le « nombre de bonnes réponses ». L'évaluation du niveau d'atteinte d'une compétence est possible lorsque l'on met les élèves dans un contexte où ils seront en mesure d'exercer la dite compétence. En ce sens, l'approche évaluative basée sur des tests et des examens ne permet pas de recueillir l'ensemble de l'information nécessaire pour inférer le niveau de développement de la compétence.

Comme le souligne avec justesse Scallon (2004), le champ de l'évaluation des compétences – et plus globalement l'évaluation à travers des tâches complexes – est un domaine qui a été relativement peu étudié par les chercheurs, surtout dans le contexte de l'évaluation à l'échelle de la classe. Une rapide recension des écrits nous amène à constater qu'il existe une littérature relativement abondante concernant, par exemple, les différents outils qu'il est possible d'utiliser pour noter la performance d'un élève à une tâche complexe telles les échelles de notation de type « analytique » ou de type « holistique »<sup>6</sup> (Andrade, 2000; Arter & McTighe, 2001; Brookhart, 1994, 1999; Hunter, Jones, & Randhawa, 1996; Mertler, 2001; Walvoord & Anderson, 1998). D'autres recherches, réalisées pour la plupart aux États-Unis (Brookhart, 2003; Cizek, 2001; Popham, 1999b), s'intéressent à la dualité « évaluation à grande échelle (*large-scale assessment*) et évaluation à l'échelle de la classe (*classroom assessment*) ». Enfin, d'autres études (Brown & Shavelson, 1996; Crahay & Detheux, 2005; Leighton, Rogers, & Maguire, 1999; Rey et al., 2003), certainement plus rares, s'intéressent à la méthodologie de l'évaluation des tâches complexes. Ces dernières études, bien qu'intéressantes, trouvent difficilement écho dans le contexte de la discipline « science et technologie » telle qu'elle est définie ici au Québec et ce, pour deux raisons principales. D'abord, les études dont il est question sont souvent réalisées dans un contexte disciplinaire associé aux langues ou aux mathématiques et sont donc rarement effectuées dans un contexte d'enseignement scientifique ou technologique. Ainsi, on peut se demander dans quelle proportion les résultats présentés s'appliquent également aux disciplines scientifiques. Ensuite, ces études sont basées sur un référentiel théorique qui n'est pas nécessairement le même que celui utilisé au Québec. Par exemple, les recherches menées par Rey *et al.* (2003) de la Belgique s'intéressent à l'évaluation des compétences.

---

<sup>6</sup> Selon Arter et McTighe (2001, p. 18), une échelle de type analytique se caractérise par une fragmentation des principales dimensions associées à la tâche qui sont jugées individuellement. Une échelle holistique permet de porter le jugement sur une impression générale de la performance de l'élève.

Cependant, la définition du concept de « compétence » retenue par les Belges s'éloigne considérablement de celle retenue par les Québécois. Par exemple, l'équivalent belge du programme du premier cycle de science et technologie prescrit le développement de 17 compétences – schématisées à la figure qui suit – regroupées sous l'égide d'un « socle de compétences » (Ministère de la communauté française de Belgique). Notons qu'au Québec il n'y a que trois compétences à développer.

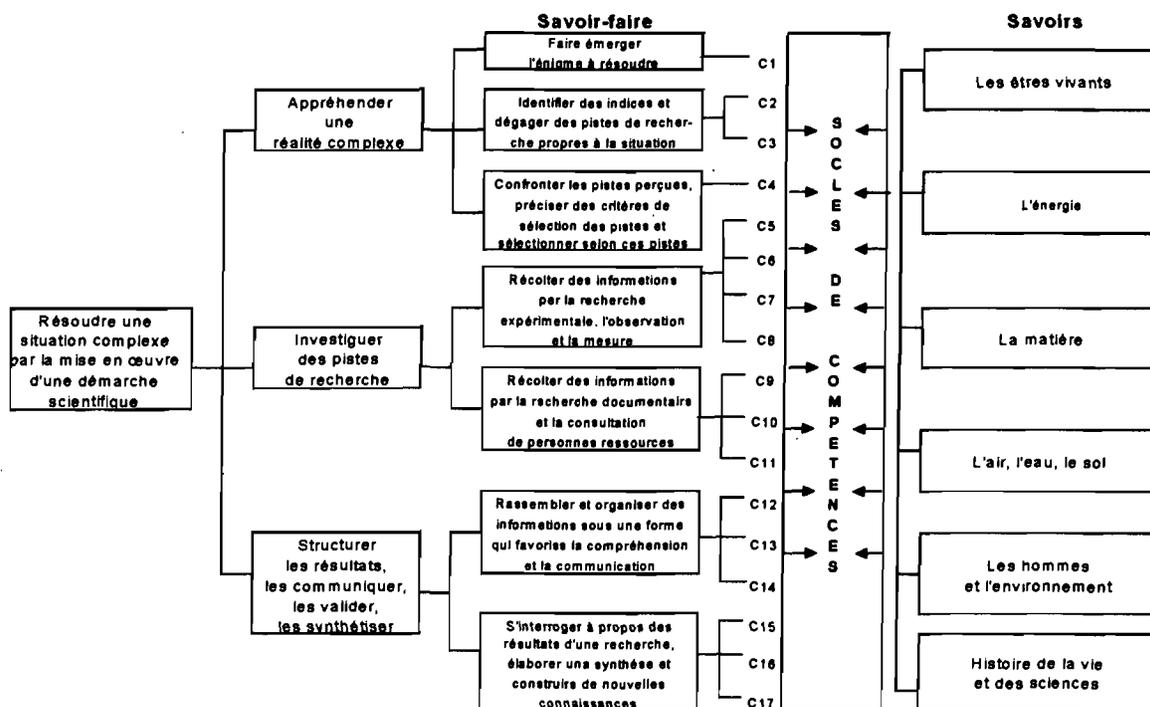


Figure 2  
Compétences du programme d'éveil scientifique (premier degré) de la Belgique.

En ce sens, même si dans les deux cas nous parlons de « compétence », il n'en demeure pas moins que le cadre conceptuel auquel on se réfère est nettement différent, ce qui laisse planer un doute sur le caractère généralisable des résultats obtenus en Belgique.

### 1.1.2.3 L'évaluation certificative dans une approche par compétences en science et technologie

Traditionnellement, on considère que l'évaluation peut servir à remplir trois fonctions principales. Selon Scallon (1988, p. 12), on peut généralement se référer à la typologie de Bloom, Madaus et Hastings (1971) en invoquant trois grandes catégories de fonctions liées à l'évaluation en éducation : l'évaluation diagnostique, l'évaluation formative et l'évaluation sommative. Comme le soulignent Laurier, Tousignant et Morissette (2005), cette classification peut amener un certain nombre de problèmes en ce qui concerne particulièrement l'interprétation de ces différents concepts et leur opérationnalisation. Ces auteurs (Laurier *et al.*, 2005, p. 51) soulignent certaines dérives face à la compréhension, par exemple, du concept d'évaluation formative :

Bloom et ses collaborateurs (1971) ont repris la distinction (entre évaluation formative et évaluation sommative) en la généralisant à l'évaluation des apprentissages et en faisant de cette distinction un élément clé de la pédagogie de la maîtrise. C'est pourquoi l'évaluation formative a toujours été plus ou moins associée à une perspective mécaniste de l'apprentissage. (...) Des pratiques souvent peu conformes à l'esprit de l'évaluation formative (...) expliquent peut-être que, dans la perception de beaucoup d'élèves, l'évaluation formative n'est rien de plus qu'une forme d'entraînement en vue de l'évaluation sommative (Lusignan et Goupil, 1997).

Le ministère de l'Éducation du Québec a, quant à lui, choisi d'éviter de recourir à ces concepts afin de justement contourner ces écueils. Dans sa Politique d'évaluation des apprentissages (Ministère de l'Éducation du Québec, 2003b), il identifie deux fonctions liées à l'évaluation à savoir : l'aide à l'apprentissage et la reconnaissance des compétences. L'aide à l'apprentissage peut se manifester de plusieurs façons. Il peut s'agir de réaliser un diagnostic en vue de mieux planifier les apprentissages. Elle peut aussi se manifester en vue de permettre une régulation des apprentissages qui peut agir autant pour l'élève que pour l'enseignant. Quant à la reconnaissance des compétences, elle vise essentiellement à rendre compte du niveau de développement des compétences au terme, par exemple, d'une longue séquence d'apprentissage ou au terme d'un cycle.

La fonction de « reconnaissance des compétences » pourrait, pour certains, être associée de près ou de loin au concept « d'évaluation sommative ». Scallon (2000, p. 18), reprenant une suggestion attribuée au départ par Weiss (1977), n'est pas d'avis d'associer ces deux concepts. Il propose de substituer à l'expression « évaluation sommative » l'expression « évaluation certificative ». Cette dernière serait, selon lui, plus acceptable conceptuellement parlant puisqu'elle ne renvoie pas à un principe de « sommation ». On retrouve également l'expression « évaluation-bilan » qui semble être en tout point équivalente à l'expression « évaluation certificative ». Dans le cadre de cette recherche, nous emploierons ces deux expressions en les considérant équivalentes. Ainsi, l'évaluation certificative a donc pour principal objectif de rendre compte des apprentissages qui étaient prévus. Ce type d'évaluation est socialement très important pour l'élève. Plusieurs parcours académiques sont accessibles sur la seule voie de la performance de l'élève à ce type d'évaluation. Ce dernier comporte donc des enjeux importants qui obligent les enseignants et les administrateurs scolaires à la plus grande prudence quant à l'élaboration et à l'utilisation des outils d'évaluation qui seront utilisés dans ce contexte.

Dans le cadre de cette recherche, nous nous intéressons résolument au contexte de l'évaluation certificative. D'autres chercheurs, comme l'équipe de Rey (2003), ont opté pour une autre perspective, soit celle associée à l'évaluation diagnostique. De notre côté, il nous est apparu plus intéressant de bâtir un modèle d'évaluation dans un contexte plus rigide, comme celui de l'évaluation certificative. Trois raisons principales expliquent ce choix. Dans un premier temps, nous pensons que les contraintes liées au contexte de l'évaluation certificative sont beaucoup plus importantes que dans tout autre contexte ce qui implique qu'il faille s'interroger sur les modalités pratiques liées à l'administration des situations d'évaluation servant à l'évaluation certificative. À titre d'exemple, les enseignants doivent – dans cette perspective – user de leur jugement professionnel pour déterminer le niveau d'atteinte de chaque élève au regard des trois compétences disciplinaires évaluées. Pour ce faire, ils doivent faire en sorte de minimiser l'effet de contamination entre les élèves. Cet effet peut être sans importance dans une perspective d'apprentissage mais on ne peut le négliger dans une perspective d'évaluation certificative où les enjeux sont importants pour tous les acteurs concernés (élèves, parents, enseignants,

direction scolaire, etc.). Gott et Duggan (2002, p. 197) soulignent l'importance de tenir compte des différentes contraintes que rencontrent les enseignants en s'assurant que, peu importe le modèle d'évaluation employé, ce modèle soit gérable par l'enseignant et qu'il assure une validité apparente (*face validity*) suffisante. Roberts et Gott (2006) soulignent également les contraintes de temps associées à l'évaluation de l'investigation scientifique. Ils rapportent qu'il faut parfois plus de quatre heures pour accomplir une investigation scientifique, ce qui représente une contrainte majeure dans le contexte scolaire.

Dans un deuxième temps, l'importance accordée à l'évaluation certificative au secondaire a amené dans le passé (et encore à l'heure actuelle) de nombreuses dérives. Nous n'avons qu'à penser aux difficultés qu'entraînent les pratiques liées au *teach to the test* qui consiste à court-circuiter le processus d'apprentissage, au moyen de différentes stratégies<sup>7</sup> (ex. enseigner les questions qui feront l'objet de l'évaluation), en vue de mieux réussir aux épreuves certificatives. En l'absence de modèles expérimentés et éprouvés, il y a des risques que de telles dérives se produisent à nouveau dans le futur. Dans un troisième temps, le présent contexte de réforme pédagogique qui souffle sur le Québec nous invite à nous attarder aux difficultés liées à l'évaluation certificative afin de mieux outiller les enseignants à s'appropriier les nouveaux curriculums en science et technologie. Cette préoccupation rejoint celle de Porter (1999, p. 205) qui recommande à ce sujet d'être prudent dans le choix du type d'évaluation : « *Care must be taken during the implementation process to develop or adopt assessment measures which drive the curriculum in a prescribed direction.* ».

### 1.1.3 Proposition d'un modèle d'évaluation certificative

Après avoir exposé le cœur de notre problématique, nous souhaitons maintenant présenter le modèle d'évaluation que nous avons expérimenté dans le cadre de cette recherche. Pour ce faire, nous présenterons d'abord les aspects macroscopiques qui situent notre modèle à

---

<sup>7</sup> Le phénomène du « *teach to the test* » est particulièrement important aux États-Unis en raison de la culture de testing à laquelle sont confrontés les étudiants américains tout le long de leur cursus scolaire. Pour plus de détails sur ce sujet, le lecteur peut se référer aux ouvrages de Popham (2001a) et de Gordon et Reese (1997).

l'échelle d'un cycle d'apprentissage de deux ans. Par la suite, nous exposerons le cadre dans lequel les situations d'évaluation se situent en expliquant les choix que nous avons fait pour déterminer ce cadre. Nous terminerons par la présentation du modèle à trois niveaux que nous proposons.

### **1.1.3.1 Un modèle d'évaluation certificative par cycle**

Avant de procéder à la description détaillée du modèle d'évaluation que nous comptons expérimenter, il convient de situer ce modèle dans un cadre plus large. Pour l'instant, il n'existe aucune véritable proposition, toutes disciplines confondues, concernant les conditions d'évaluation certificative dans le cadre de programmes formulés par cycle. Les tenants de l'approche par compétences soulignent, avec justesse d'ailleurs, qu'il faut beaucoup de temps pour développer des compétences. D'autres ajoutent qu'il est difficile de porter un jugement sur le développement d'une compétence en se référant à une seule situation d'évaluation. De plus, les enseignants ne se sentent généralement pas à l'aise à l'idée de poser un jugement s'ils craignent de ne pas avoir suffisamment d'informations ou s'ils doutent de la qualité de cette information. Fort de ces différents constats, nous pensons qu'un modèle d'évaluation certificative par cycle, tel que schématisé à la figure 3, peut s'avérer une proposition intéressante. À l'intérieur de ce modèle, nous identifions trois « moments » d'évaluation, schématisés par les encadrés grisés aux étapes 1, 4 et 8<sup>8</sup> de la figure 3 qui représentent des moments formels d'évaluation à caractère « certificatif ». Le premier moment survient au début du cycle à l'arrivée des élèves en première secondaire. Cette situation pourrait être à vocation diagnostique afin de déterminer le niveau de compétence de chacun des élèves au début du cycle. Il s'agirait donc de la première cueillette d'information concernant le niveau de compétences de chacun des élèves. Le second moment apparaît au terme de la première année du cycle c'est-à-dire à la quatrième étape. Il s'agit, à ce moment, de faire le point à mi-cycle afin de préparer la dernière année d'étude. Enfin, le dernier moment survient au terme du cycle, soit à la huitième étape, et à

---

<sup>8</sup> Le modèle que nous proposons s'appuie sur un calendrier scolaire découpé en quatre étapes par année. Chacune des étapes comporte généralement environ le même nombre de journées de classe. Ce modèle est de loin le plus répandu dans les commissions scolaires et les écoles secondaires privées québécoises.

ce moment il s'agit de réaliser une évaluation de type bilan afin de prendre des décisions administratives concernant la suite du cursus scolaire de chaque élève.

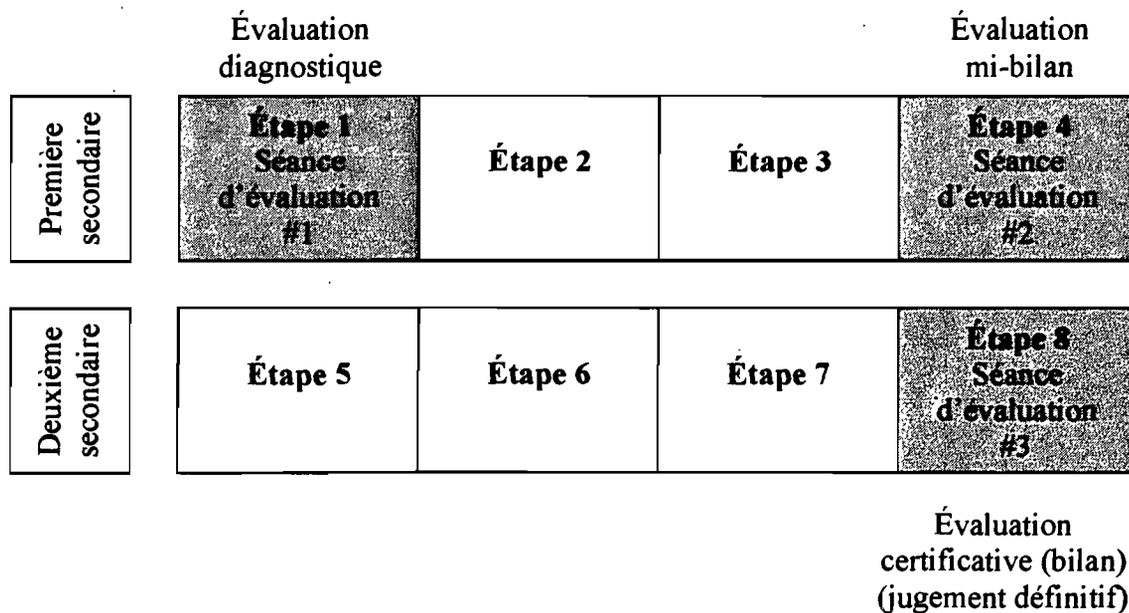


Figure 3  
Schéma synthèse du modèle d'évaluation par cycle en science et technologie.

Dans la perspective d'une évaluation bilan, on peut penser que les informations recueillies à la fin du cycle auront une plus grande valeur que celles recueillies durant le cycle. En effet, les attentes de fin de cycle et les échelles de niveaux de compétence élaborées par le Ministère ont été construites afin de rendre compte de performances attendues au terme d'un cycle de deux ans. Cependant, les informations recueillies au début du cycle et à mi-cycle pourraient être une source intéressante de données secondaires afin de mieux appuyer le jugement des enseignants à la fin du cycle.

Dans le cadre de cette recherche, nous avons expérimenté des situations d'évaluation qui ont été développées dans un contexte d'évaluation certificative de fin de cycle mais elles ont été administrées à mi-cycle. Deux raisons expliquent ce choix. D'abord, pour des raisons éthiques puisqu'il aurait été hasardeux d'expérimenter des situations d'évaluation devant servir à prendre des décisions administratives auprès des élèves. Ensuite, nous

souhaitions expérimenter ce modèle d'évaluation après une année d'appropriation de l'approche par compétences en science et technologie. Nous pensions qu'une recherche comme celle-ci devait être menée le plus rapidement possible afin de mieux cerner la problématique de l'évaluation des compétences dans un contexte d'enseignement scientifique. Bien sûr, ce choix implique un certain nombre de conséquences et de limites inhérentes à cette recherche. Au chapitre 5, nous présentons les principales limites associées à cette recherche et dont plusieurs sont liées étroitement au choix que nous avons dû faire et que nous venons tout juste de présenter.

Enfin, nous pensons que cette recherche pourrait être à nouveau réalisée dans quelques années afin de voir si le modèle est toujours adéquat après quelques années d'appropriation. Une telle recherche menée de façon plus longitudinale pourrait certainement être pertinente afin de voir si le modèle peut tenir la route à différents niveaux d'application d'un programme par compétences.

### **1.1.3.2 Cadre dans lequel se situent les situations d'évaluation**

Pour les fins de notre recherche, nous proposons un cadre définissant les types de situations d'évaluation pouvant être rencontrées en science et technologie. Puisqu'il n'existe que peu de littérature à ce sujet, nous nous sommes inspirés des écrits de Roegiers sur la notion de « famille de situations » afin de catégoriser chacun des types de situations d'évaluation que nous avons définis. Selon Roegiers (2000, p. 130), une compétence se définit, entre autres, « par rapport à une famille de situations ». Autrement dit, une compétence s'exerce et se développe véritablement dans la mesure où l'on propose des situations d'apprentissage qui partagent un certain nombre de caractéristiques communes. Selon lui, certaines compétences peuvent se développer au moyen d'un nombre important de situations alors que d'autres ne s'opérationnalisent qu'au moyen de quelques situations relativement limitées en nombre. Cet auteur définit également deux paramètres permettant de circonscrire une famille de situations :

- ce que l'on donne à l'élève;
- le type de tâche offerte à l'élève.

Le premier paramètre correspond davantage aux contraintes liées à la tâche. Roegiers (2000, p. 133) donne l'exemple de la conception d'un dessin technique où les pièces doivent avoir un maximum de «  $x$  » faces, ces dernières doivent être planes et tous les angles doivent être droits. Ces contraintes peuvent être perçues comme des attributs discriminants afin de marquer la différence ou la ressemblance entre des situations d'apprentissage. Le second paramètre fait référence aux éléments qui agiront sur la démarche que l'élève entreprendra. Roegiers (p. 133) donne l'exemple du degré de complexité de la tâche, du niveau de précision attendu, du volume de la production, etc. Nous avons retenu les propositions de Roegiers pour établir notre cadre puisqu'elles permettent de définir des balises à la fois rigoureuses et souples pour la confection de situations d'évaluation certificative. En effet, un cadre nous est apparu indispensable puisque le Programme de formation ne propose aucun modèle d'intégration de ses différentes composantes (compétences disciplinaires, contenu de formation, etc.). La figure 4 présente le cadre que nous proposons et qui s'inspire des propositions de Roegiers.

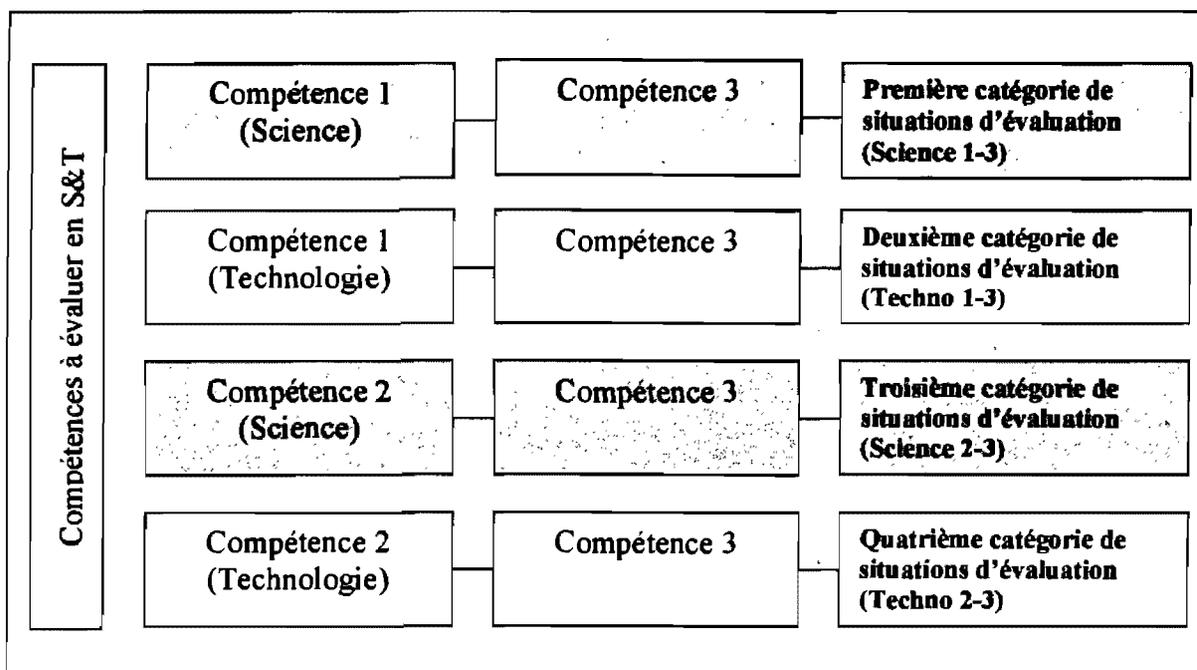


Figure 4

Types de situations d'évaluation retenues pour évaluer la première compétence du programme de science et technologie (1<sup>er</sup> cycle).

Ainsi, quatre catégories de situations d'évaluation ont été définies. La première catégorie de situation (Science 1-3) propose à l'élève de mettre en œuvre la première compétence dans un contexte scientifique tout en faisant appel à la troisième compétence axée sur la communication. La seconde catégorie (Techno 1-3) fait appel aux mêmes compétences mais cette fois-ci dans un contexte technologique. Comme nous le mentionnions auparavant dans ce chapitre, les nuances entre les démarches d'investigation scientifiques et les démarches de conception en technologie nous sont apparues suffisamment importantes pour créer une division à l'intérieur même de la première compétence. Nous sommes conscients que certains auteurs, comme Fourez (1994) par exemple, considèrent qu'il est inutile de créer une telle division entre la science et la technologie. Bien que nous respectons ce point de vue, nous demeurons sceptiques quant à l'équivalence des inférences qu'il est possible de réaliser selon que la situation d'évaluation appartienne davantage à la science ou davantage à la technologie. C'est pourquoi nous avons choisi de scinder la première compétence afin de nous attarder uniquement au contexte scientifique.

La troisième catégorie (Science 2-3) propose des situations où l'élève doit exercer sa compétence scientifique en expliquant des phénomènes naturels tout en dégagant des retombées de la science (Ministère de l'Éducation du Québec, 2003c, p 279). À nouveau, cette compétence peut difficilement s'exercer seule sans l'apport de la communication. C'est pourquoi la compétence 3 apparaît dans cette catégorie de situations. Enfin, la dernière catégorie (Techno 2-3) repose sur la compréhension du fonctionnement d'objets techniques et sur la compréhension des retombées de la technologie. Cette compétence est à nouveau amalgamée avec la compétence de communication. Cette division de la compétence 2 (science/technologie) nous est apparue incontournable puisque les démarches d'analyse en technologie sont, quand on les analyse, relativement éloignées des démarches utilisées en science. Bref, le contexte scientifique et le contexte technologique ont suffisamment de différences pour exiger des types de situations distinctes. Afin d'illustrer notre point de vue, examinons deux exemples. La démarche de conception technologique impose parfois un souci d'esthétisme dans la production de l'objet technique (International Technology Education Association, 2000). Ce souci n'est pas partagé en science puisque l'esthétisme n'est que très rarement, voire jamais, considéré. De la même manière, la compréhension d'un objet technique implique souvent le recours au dessin technique qui est une forme de langage propre à la technologie (International Technology Education Association, 2000).

Un des objectifs de cette recherche vise à valider en quelque sorte une partie (Science 1-3) des situations de ce cadre. Les types de situation que nous venons de proposer sont élégants mais ils négligent, pour l'instant, les contraintes liées au contexte interdisciplinaire qui caractérise le champ de la science et de la technologie. Il est hasardeux de ne pas tenir compte du contexte disciplinaire car si ce dernier a fait un impact majeur sur la performance des élèves pour une situation d'évaluation donnée – ce qui serait concordant avec les travaux de Stecher<sup>9</sup> (1998, p. 4) – il faudrait alors vraisemblablement revoir le cadre que nous avons défini afin de mieux préciser les types de situations d'évaluation pour

---

<sup>9</sup> Stecher et al. (citant plusieurs autres études) indiquent que les tâches ouvertes (open-ended tasks) ont généralement des corrélations faibles les unes avec les autres à l'intérieur même d'un domaine d'étude (ex. science).

chacune des quatre catégories (Science 1-3, Techno 1-3, Science 2-3, Techno 2-3) précédemment définies. Les résultats de nos travaux pourraient, par exemple, nous obliger à scinder la première situation d'évaluation en sous-catégories (ex. Science 1-3 (Univers matériel), Science 1-3 (Univers vivant), etc.).

### **1.1.3.3 Un modèle d'évaluation à trois niveaux**

Le modèle que nous proposons et que nous souhaitons expérimenter est inspiré, entre autres, de celui développé par Rey *et al.* (2003) en Belgique. Dans cette section, nous présenterons les grandes lignes de ce modèle.

Traditionnellement, on mesure les apprentissages des élèves québécois dans le contexte d'une évaluation certificative en science au moyen d'un instrument de mesure administré à une seule occasion, soit à la fin d'une année scolaire par exemple. Aussi, tous les élèves se voient administrer la même épreuve. On trouve cependant de légères variantes (ex. des versions différentes dans lesquelles les items sont les mêmes mais qui peuvent apparaître dans un ordre différent pour éviter le plagiat) qui n'affectent pas significativement la qualité de l'information recueillie. Les items qui composent l'instrument de mesure ont généralement un niveau de difficulté moyen et constituent un échantillon représentatif d'un ensemble d'items (univers d'items) qu'il aurait été possible d'administrer. Socialement et politiquement, les résultats à de telles épreuves doivent se situer dans un intervalle « acceptable ». Effectivement, un taux d'échec de 80% serait considéré inacceptable aux yeux des élèves, des parents, des décideurs scolaires et de la population en général. Ainsi, nous pouvons dire que la pression sociale liée à l'atteinte d'un niveau de performance minimal affecte considérablement la nature et les caractéristiques des outils d'évaluation certificative. Or, comme l'indique Rey (2003, p. 103), la compétence de résolution de problème n'est que faiblement maîtrisée par les élèves même au terme d'une longue séquence d'apprentissage :

(...) la majorité des élèves devant résoudre des tâches complexes leur demandant la mise en place de démarches de résolution, sont en réelles difficultés. Ce constat

semble confirmer notre hypothèse : que les compétences exigées dans le référentiel *Socle de compétences* sont peu maîtrisées par les élèves.

On peut résumer la situation de la façon suivante. D'une part, on souhaite mieux développer chez les élèves leurs compétences de résolution de problème mais nous savons que cela est difficile et que les élèves peinent à y parvenir. D'autre part, il plane sur la tête des enseignants et des élèves le spectre de l'évaluation certificative car il faut néanmoins continuer à certifier les apprentissages des élèves au terme d'une ou de plusieurs séquences d'apprentissage. De plus, cette évaluation certificative doit nécessairement permettre d'obtenir des résultats socialement satisfaisants. Compte tenu de l'ensemble des contraintes qui marquent cette situation, une solution facile – et fréquemment employée – consiste à ajuster les outils d'évaluation en élaguant l'évaluation des habiletés supérieures. Nous pensons qu'une telle situation peut être évitée en modifiant la méthodologie utilisée. Nous pensons qu'un modèle d'évaluation à trois niveaux – tel que décrit subséquemment – pourrait permettre d'améliorer cette situation.

Le modèle à trois niveaux consiste à développer, pour une situation d'évaluation donnée, trois « versions » (la section 3.1.2 est consacrée exclusivement à la description et aux caractéristiques de chacun de ces niveaux) de celle-ci comme l'illustre le tableau qui suit. Dans le cadre de cette recherche, nous avons développé trois situations d'évaluation : SE1, SE2, SE3.

**Tableau I**  
**Synthèse des situations d'évaluation qui ont été expérimentées.**

Situation d'évaluation (SE)	Niveau 1 (Compétence)	Niveau 2 (Compétence dirigée)	Niveau 3 (Habilité)
SE1 : La tasse de chocolat chaud (Univers matériel)	SE1-1	SE1-2	SE1-3
SE2 : La balle de tennis (Démarche pure)	SE2-1	SE2-2	SE2-3
SE3 : Les lombrics (Univers vivant)	SE3-1	SE3-2	SE3-3

Une première version (ex. SE1-1<sup>10</sup>) consiste à placer l'élève dans un contexte et des conditions lui permettant de faire réellement la démonstration de ses compétences. Cette situation est la plus exigeante et permet de recueillir des informations sur la capacité de l'élève à mobiliser ses ressources en contexte de résolution de problème scientifique en ne fournissant aucune aide à celui-ci. Par exemple, le cahier de l'élève est épuré au maximum et ne comporte pratiquement aucun indice sur la façon de procéder pour résoudre le problème. L'élève doit donc prendre en charge de façon autonome ses démarches de résolution de problème. La seconde version vise également à mesurer le niveau d'atteinte de la compétence évaluée mais cette fois-ci en donnant quelques indices<sup>11</sup> à l'élève. Il peut s'agir d'indices méthodologiques ou de références documentaires afin de l'aider dans les démarches qu'il met de l'avant. La troisième version est, quant à elle, une situation qui vise essentiellement à mesurer des habiletés et non plus des compétences. À l'intérieur d'une telle situation, l'élève reconnaît généralement une démarche qu'il a déjà faite et il peut prévoir assez facilement les résultats qu'il obtiendra au terme de sa démarche. La figure 5

<sup>10</sup> Dans ce texte, l'expression SE1-1 signifie Situation d'Évaluation 1, version 1 (compétence), l'expression SE1-2 signifie Situation d'Évaluation 1, version 2 (compétence dirigée), l'expression SE1-3 signifie Situation d'Évaluation 1, version 3 (habileté), etc.

<sup>11</sup> Au chapitre 3, nous présenterons les caractéristiques marquant les nuances entre les différents niveaux du modèle que nous proposons.

qui suit résume les séquences possibles d'administration des différentes versions de chacune des situations d'évaluation.

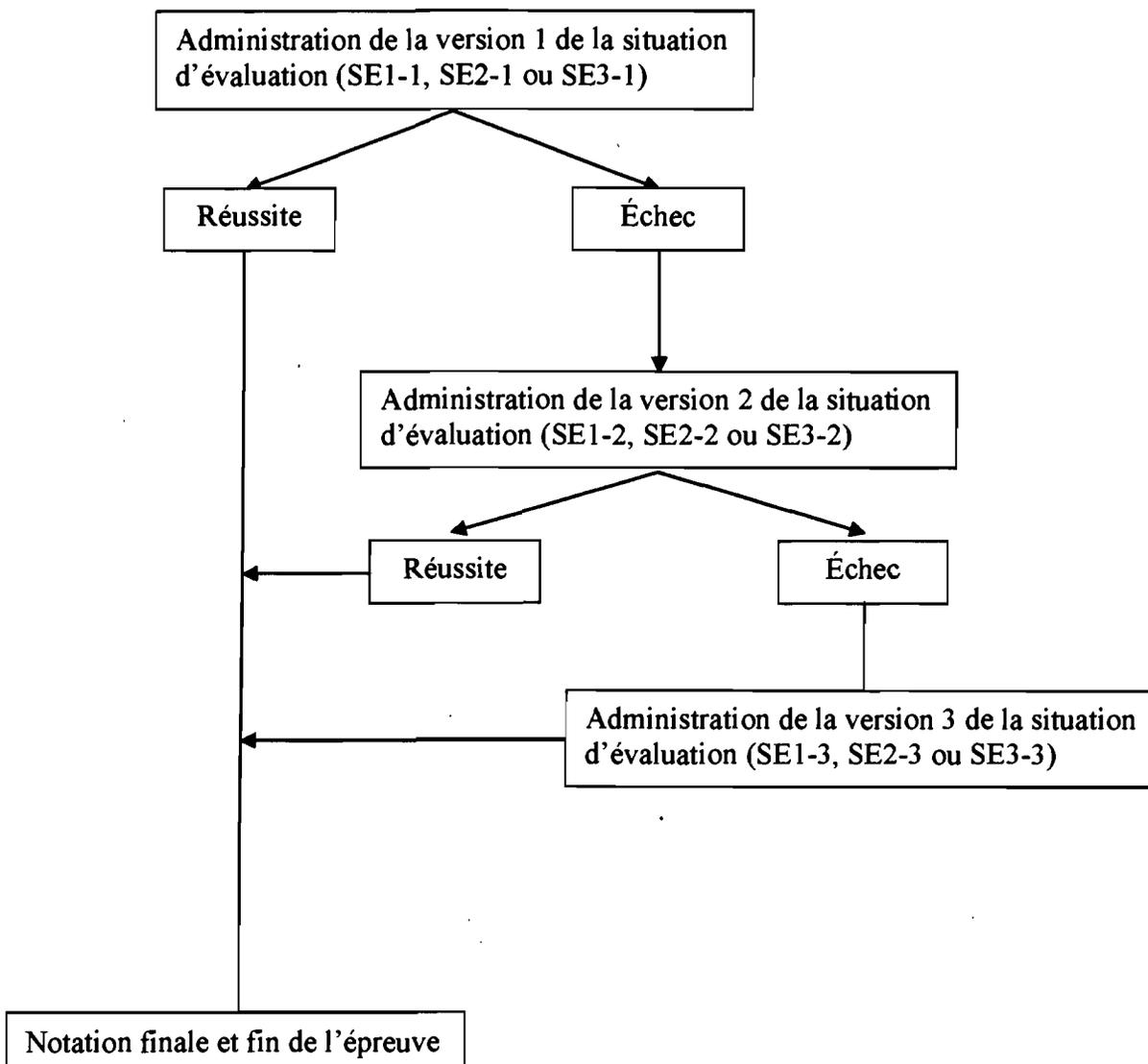


Figure 5

Séquences possibles d'administration des situations d'évaluation.

Ce modèle possède l'avantage de placer l'élève dans un contexte réel de résolution de problème scientifique tout en tenant compte de la pression sociale liée au rendement des élèves. En effet, un élève qui est incapable de réussir la version 1 se voit administrer la

version 2 afin de mieux cibler ses difficultés. Par exemple, un élève peut être incapable de cerner le problème qui lui est dévolu s'il ne reçoit aucune aide mais il peut être tout à fait capable d'y parvenir avec quelques indices. Un élève manifestant toujours des difficultés à réussir la version 2 se verra alors administrer la version 3 afin de déterminer, par exemple, les habiletés qu'il maîtrise. Nous pensons que ce modèle permet une plus grande latitude pour les enseignants car un taux d'échec fulgurant lors de la passation de la version 1 n'est pas catastrophique étant donné qu'il est alors possible d'aller chercher d'autres informations aux moyens des deux autres versions de la situation d'évaluation. Ceci représente, sans nul doute, le principal avantage de ce modèle par rapport au modèle à un niveau (à passation unique). Qui plus est, la possibilité de pouvoir recourir à d'autres instruments de cueillette de données peut éviter à l'enseignant d'être confronté à une situation où certains élèves ont été atteints du « syndrome de la page blanche » n'ayant rien inscrit comme information pertinente quant à la situation qui leur était dévolue.

Nous avons opté pour un modèle à trois niveaux pour deux raisons principales. D'abord, Baxter, Elder et Glaser (1996, p. 138) ont montré qu'en général les élèves démontrent une activité cognitive qualitative qui peut être regroupée en trois catégories : nettement élevée, nettement faible et intermédiaire. Au regard de ces résultats, un modèle à trois niveaux semble le plus approprié pour discriminer le niveau de maîtrise des élèves. Ensuite, lors d'une recherche semblable à la nôtre, Rey, Carette, Defrance et Kahn (2003) ont aussi utilisé un modèle à trois niveaux. Puisque nous espérons croiser nos résultats avec ceux obtenus par ces auteurs, un modèle à trois niveaux s'imposait.

## ***1.2 Objectifs de la recherche***

Les objectifs de la recherche que nous proposons sont de deux ordres. Premièrement, nous espérons que les résultats de notre recherche pourront éclairer avantageusement les enseignants de science et technologie ainsi que les différents acteurs du milieu scolaire (conseillers pédagogiques, administrateurs, chercheurs universitaires, etc.) sur une des méthodologies liées à l'évaluation des compétences en science. Deuxièmement, nous

espérons que les résultats de cette recherche, qui concernent particulièrement le champ de science et technologie, pourront éclairer les autres champs scolaires concernant les éléments méthodologiques liés à l'évaluation des compétences que nous expérimentons dans le cadre de cette recherche.

### *1.3 Questions de recherche*

En science, il n'est pas rare que certains élèves souffrent du « syndrome de la page blanche » lors d'une situation d'évaluation qui vise à mesurer la façon dont ils résolvent des problèmes scientifiques. Contrairement, par exemple, aux disciplines où la composante linguistique est importante, les disciplines scientifiques offrent peu d'échappatoires aux élèves quand le problème est, par exemple, trop difficile ou mal compris. Cette situation pose alors un triple problème aux enseignants. D'abord, un problème lié à la validité de l'inférence : en n'ayant aucune information en main, que pouvons-nous conclure au sujet du niveau de maîtrise de l'élève ? Ensuite, les enseignants font face à un problème pédagogique : comment se fait-il qu'un élève manifeste si peu de maîtrise au terme d'une séquence d'enseignement ? Enfin, ils se heurtent à un problème de généralité (échantillonnage et fidélité) abondamment discuté par plusieurs auteurs (McMillan, 2001; Popham, 1999a, p. 164; Shavelson, Baxter, & Pine, 1991). En effet, est-il raisonnable d'attribuer une note nulle alors qu'on se doute qu'un élève est en mesure de résoudre un certain nombre de problèmes semblables ? Ou à l'inverse, que pouvons-nous conclure si un élève réussit exceptionnellement bien à une situation d'évaluation alors que des doutes raisonnables nous laissent croire que l'élève en question ne maîtrise pas l'ensemble des apprentissages visés ? En ce sens, un modèle d'évaluation à passation unique nous apparaît plutôt limitatif compte tenu de la situation que nous venons de décrire.

Un autre aspect qui peut causer problème dans un contexte d'évaluation certificative – et qui est relié au syndrome de la page blanche – concerne le soutien qu'il est possible de donner à l'élève. En situation d'apprentissage, il est tout à fait concevable et normal d'aider les élèves en difficulté en leur donnant des pistes ou en leur suggérant des éléments de

démarche afin qu'ils puissent progresser. En situation d'évaluation certificative, le contexte est tout autre et des aspects éthiques entrent irrémédiablement en jeu. À ce sujet, plusieurs questions se posent : quelle est la nature de l'aide à apporter ? Quelle est l'ampleur de cette aide ? Quelles sont les règles d'équité impliquées ? Etc. Comme l'indique Gott (1987), l'évaluation doit nous renseigner sur ce que les élèves sont en mesure de faire et non pas sur le niveau d'aide devant leur être donné afin de résoudre la tâche. Ajoutons en terminant que bien qu'il soit certainement possible de considérer tous ces aspects, il faut également que les conditions soient facilement gérables et applicables en contexte de classe ordinaire c'est-à-dire avec environ 32 élèves par groupe.

Bien qu'il soit difficile d'obtenir un consensus afin de définir ce qu'est ou ce que n'est pas une compétence, on peut néanmoins affirmer, sans peur de se tromper, que l'approche par compétences exige beaucoup plus de la part des élèves. En effet, non seulement doivent-ils acquérir, entre autres, des connaissances, des habiletés, des attitudes, des techniques, des savoir-faire, mais ils doivent, de plus, les utiliser à bon escient et dans le bon contexte. En ce sens, nous rejoignons la proposition de Tardif (2006, p. 22) quand il énonce qu'une compétence est « un savoir-agir complexe prenant appui sur la mobilisation et la combinaison efficaces d'une variété de ressources internes et externes à l'intérieur d'une famille de situations ». Dans une certaine mesure, ces visées pédagogiques – bien que formulées par objectifs – rejoignent celles que l'on retrouve dans les programmes de science actuels. En effet, l'ensemble des programmes de science du deuxième cycle du secondaire (Ministère de l'Éducation du Québec, 1990, 1992a, 1992b, 1992c) visent déjà à développer des habiletés de résolution de problème de haut niveau chez les élèves. Pour plusieurs, ces visées sont partiellement atteintes et ce, même au terme des études secondaires. Cela n'est pas nécessairement étonnant puisque ce sont des habiletés fort complexes à développer. En ce sens, il serait étonnant que des élèves du premier cycle, et à plus forte raison de la première secondaire, soient en mesure de très bien réussir dans des situations d'évaluation de compétences. Autrement dit, on constate que les niveaux de compétence doivent être bien définis afin de connaître le niveau auquel on peut s'attendre pour une majorité d'élèves. À défaut d'obtenir des attentes réalistes, il est fort possible que

l'on assiste à une mise en œuvre insatisfaisante des changements pédagogiques comme nous l'avons constaté pour certains cours de science du deuxième cycle (Dionne, 1999).

Dans le cadre de cette recherche, nous souhaitons expérimenter un modèle d'évaluation certificative en science et technologie qui vise à surmonter les difficultés auxquelles nous venons de faire référence. Nous pensons que ce modèle à trois niveaux pourrait permettre : (1) d'obtenir plus d'informations sur la compétence d'un élève à réaliser une investigation scientifique en laboratoire et (2) de mieux baliser l'aide à offrir aux élèves en difficulté. De façon plus formelle et, pour les fins de notre recherche, nous posons la première question qui suit :

#### Première question

**Quels sont les avantages à recourir à un modèle d'évaluation à trois niveaux afin de rendre compte du niveau de développement d'une compétence d'investigation scientifique en laboratoire ?**

Dans le contexte d'une approche interdisciplinaire en science, nous pensons qu'il est possible d'observer des différences selon la majeure disciplinaire qui caractérise la situation d'évaluation. Pour l'instant, nous n'avons aucune donnée sur l'impact de la nature du contenu disciplinaire sur le rendement des élèves dans des situations d'investigation scientifique. Au chapitre suivant, nous présenterons certains résultats de recherche qui tendent à démontrer que le contenu disciplinaire scientifique semble jouer un rôle concernant la réussite de telles situations. Aussi, s'il s'avérait que la nature du contenu disciplinaire influence considérablement les inférences qu'il est possible de réaliser au regard de la compétence, il faudrait alors soit ajuster le modèle que nous proposons ou considérer sérieusement cette limite s'il n'est pas possible concrètement d'ajuster le modèle<sup>12</sup>. Nous retenons donc la seconde question qui suit et qui s'intéresse à cet aspect.

---

<sup>12</sup> Nous savons que le nombre de situations d'évaluation qu'il est possible d'administrer en fin de cycle est limité. Il ne serait pas possible, par exemple, d'administrer 6 ou 7 situations d'évaluation de l'investigation scientifique qui nécessiterait chacune 4 ou 5 périodes de 75 minutes.

Deuxième question

**Est-ce que le modèle d'évaluation à trois niveaux se comporte de la même façon selon le contexte disciplinaire privilégié ?**

Enfin, nous souhaitons investiguer les effets possibles de trois variables contextuelles qui pourraient affecter le modèle que nous proposons. Nous avons divisé cette troisième question en trois sous-questions. La première sous-question s'intéresse aux différences possibles entre les garçons et les filles. La deuxième sous-question vise à chercher des liens potentiels entre la formation des enseignants et leur capacité à développer des compétences d'investigation scientifique chez leurs élèves et le modèle que nous expérimenterons. Nous pensons en effet qu'il devrait exister un lien entre la formation des enseignants et la réussite des élèves. Enfin, la dernière sous-question vise à identifier quelle catégorie d'élèves réussissent à obtenir les meilleurs résultats à chacun des niveaux (1, 2, 3) du modèle.

Troisième question

**Dans quelle mesure un modèle d'évaluation à trois niveaux est-il sensible à certains facteurs contextuels liés aux caractéristiques des élèves et des enseignants ?**

Première sous question

**Est-ce que le modèle permet de voir des différences entre les garçons et les filles ?**

Deuxième sous question

**Est-ce que le modèle permet de distinguer les enseignants qui sont expérimentés en enseignement des sciences ?**

Troisième sous question

**Est-ce que le modèle permet de distinguer les enseignants qui sont formés adéquatement pour développer chez leurs élèves l'investigation scientifique ?**

#### ***1.4 Pertinence de la recherche***

Nous croyons que cette recherche mérite d'être menée pour deux raisons principales. Premièrement, nous considérons qu'il existe présentement un vide méthodologique très important concernant la problématique que nous abordons. En effet, nous savons très peu de chose sur la méthodologie à employer et l'information qu'il est possible d'obtenir dans une approche par compétences basée sur une évaluation de performance. En ce sens, nous espérons que cette recherche visera à apporter des pistes de solution pertinentes à un besoin méthodologique manifeste. Deuxièmement, nous considérons qu'il existe un besoin social à répondre aux questions que nous soulevons. En effet, les intervenants du milieu scolaire québécois sont présentement à la recherche de modèles afin de mieux rendre compte du niveau de compétence des élèves. La mise en œuvre des nouveaux programmes d'études en science et technologie représente une belle occasion d'apporter une contribution non seulement théorique mais aussi pratique aux problèmes que rencontrent les différents intervenants en enseignement de la science et de la technologie. En ce sens, notre recherche se situe dans la lignée des travaux de Mislavy (2004) qui s'intéresse aux modalités d'évaluation de l'investigation scientifique (*science inquiry*) en se situant dans le contexte des propositions de l'*American Association for the Advancement of Science* (AAAS) aux États-Unis.

J'ai vu plus loin que les autres parce que je  
me suis juché sur les épaules de géants.  
Isaac Newton

## **Chapitre 2. Recension des écrits**

---

Ce chapitre est divisé en quatre. Dans un premier temps, nous présenterons les nouvelles tendances en évaluation. Dans un second temps, nous présenterons les caractéristiques de l'évaluation basée sur la performance en mettant l'accent sur les avantages et les limites de cette dernière. Dans un troisième temps, nous discuterons de l'investigation scientifique de type *hands-on*. Enfin, nous terminerons ce chapitre en exposant l'état de la question concernant les variables investiguées dans cette recherche.

### ***2.1 Nouvelles tendances en évaluation***

Comme nous l'avons mentionné au premier chapitre, l'approche par compétences nécessite de repenser les outils d'évaluation (situation d'évaluation, outils de notation, outils de consignation des résultats, etc.) qui permettront de collecter l'information afin de baliser le jugement de l'enseignant. Une des caractéristiques de l'approche par compétences est de tenir compte non seulement du résultat d'apprentissage mais également du processus par lequel l'élève est passé afin de réaliser l'apprentissage en question. Dans ce contexte, les outils d'évaluation basés sur la sélection de bonnes réponses (ex. items à correction objective) ne répondent pas aux nouvelles exigences (Lasnier, 2000). En effet, ce genre d'items permet difficilement de retracer le cheminement intellectuel de l'élève et de plus, comme le souligne Gitomer (1993, p. 242) – citant Frederiksen (1984) – et Ward & Lee (2002) ce format d'items a des limites importantes en ce qui a trait à l'évaluation de certains niveaux d'habiletés. Afin de palier à ces difficultés, un courant axé sur des formes dites alternatives d'évaluation a pris de l'ampleur vers la fin des années mille-neuf-cent-quatre-vingt. L'évaluation de la performance (*performance-based assessment*), l'évaluation authentique (*authentic assessment*), le portfolio et les démonstrations représentent quelques exemples d'approches alternatives en évaluation (McMillan & Workman, 1998, p. 7) auxquelles on a de plus en plus recours en milieu scolaire. Selon ces auteurs, les nouvelles

approches en évaluation doivent permettre de mieux rendre compte du niveau des élèves au regard des habiletés de résolution de problème et des habiletés cognitives supérieures. Le tableau II qui suit présente d'ailleurs les nouvelles tendances en évaluation des apprentissages.

**Tableau II**  
Nouvelles tendances en évaluation des apprentissages à l'échelle de la classe.<sup>13</sup>

De	Vers
Accent sur les résultats d'apprentissage	Accent sur le processus d'évaluation
Évaluation d'habiletés isolées	Évaluation d'habiletés intégrées
Évaluation de faits ou de concepts pris isolément	Évaluation de l'application des concepts ou des connaissances
Tâche de type papier-crayon	Tâches authentiques
Généralement une seule bonne réponse à fournir	Généralement plusieurs bonnes réponses possibles
Tâches décontextualisées	Tâches contextualisées
Critères de correction non nécessairement connus par les élèves	Critères de correction annoncés à l'avance aux élèves
Évaluation qui vient au terme d'une séquence d'enseignement	Évaluation qui s'effectue en cours d'apprentissage
Peu de rétroaction	Beaucoup de rétroaction
Recours à des tests à correction objective	Recours à des situations de performance
Évaluation sporadique	Évaluation continue

À la lecture du tableau précédent, on remarque que plusieurs des caractéristiques associées aux évaluations alternatives sont en étroite relation avec les caractéristiques de situations permettant de développer des compétences. À titre d'exemple, on remarque que les

<sup>13</sup> Traduit et adapté de McMillan et Workman (1998, p. 7)

tendances actuelles mettent l'accent sur des tâches contextualisées et complexes pour lesquelles plusieurs réponses sont plausibles et qui nécessitent, de la part de l'élève, une structuration et une organisation de ses connaissances afin de résoudre la problématique qui lui est présentée. Cette description concorde avec les qualités recherchées quant aux situations d'apprentissage et d'évaluation (Ministère de l'Éducation du Québec, 2003c, p. 272).

Au Québec, ce courant vers des formes d'évaluation dites alternatives en science n'est pas nouveau. En effet, les épreuves de laboratoire en sciences physiques 416-430 ainsi qu'en chimie 534 et en physique 534 – introduites au début des années 90 – ont, à plusieurs égards, des caractéristiques qui les associent à des évaluations de performance.

Aux États-Unis, Shavelson, Baxter et Pine (1991), ont identifié trois raisons qui justifient de recourir à des méthodes d'évaluation alternatives : (1) les avancés en recherche et particulièrement en science cognitive, (2) la réforme des programmes scolaires et (3) un désenchantement face aux méthodes de testing traditionnelles (ex. items à choix multiple). Dans le cadre du renouvellement des programmes d'études en science, ces aspects sont tous présents et contribuent au mouvement vers l'évaluation de la performance.

## ***2.2 Caractéristiques et définition de l'évaluation basée sur la performance (EBP)***

Avant de présenter un bref état de la question concernant l'évaluation basée sur la performance, nous souhaitons au préalable clarifier deux points.

D'abord, mentionnons que l'on retrouve dans la littérature un nombre important de définitions se rapportant à l'expression « évaluation basée sur la performance ». Cependant, on note rapidement – comme le soulignent aussi certains auteurs (Popham, 1999a; Scallon, 2004; Simon & Forgette-Giroux, 2001) - qu'il ne semble pas y avoir de consensus apparent sur le sens à donner à cette expression. À défaut de donner une définition univoque, la plupart des auteurs essaient de donner les caractéristiques de ce qu'est et/ou ce que n'est pas l'évaluation basée sur la performance. Nous n'agissons pas différemment de ces auteurs

et nous adopterons aussi cette stratégie afin de préciser le sens de l'EBP. Ainsi, dans un premier temps nous identifierons les caractéristiques les plus largement mentionnées au regard de l'EBP. Dans un second temps, nous présenterons la conception que nous avons retenue compte tenu des caractéristiques associées à l'évaluation basée sur la performance.

Ensuite, la clarification du concept d'EBP nous apparaît être une opération particulièrement importante à réaliser. En effet, l'analyse de recherches s'apparentant à la nôtre nous indique que les caractéristiques des situations administrées aux élèves peuvent varier considérablement d'une étude à l'autre même si chacune d'elles s'appuie sur un même modèle d'évaluation basée sur la performance. Cette situation est plutôt ennuyeuse car il devient alors difficile de comparer des résultats provenant de différentes recherches même si elles prétendent toutes s'appuyer sur une approche d'EBP.

Examinons maintenant les six principales caractéristiques de l'EBP que nous avons retenues.

### **1. Une évaluation basée sur la performance est significative et contextualisée.**

La plupart des auteurs (Herman, Aschbacher, & Winters, 1992; Hibbard, 1996; Scallon, 2004; Simon & Forgette-Giroux, 2001; Solano-Flores & Shavelson, 1997; Wiggins, 1993a, 1993b) s'entendent sur l'importance d'une contextualisation adéquate afin de rendre les situations les plus significatives possible aux yeux des élèves. Cependant, on retrouve peu d'indications sur le niveau attendu de contextualisation. En ce qui nous concerne, nous considérons que cette dernière peut être exprimée sur un continuum sur lequel on retrouverait, à une extrémité, les situations scolaires totalement décontextualisées et à l'autre extrémité le courant de l'évaluation authentique qui, lui, trouve appui sur un fort réalisme des situations présentées aux élèves. Une situation contextualisée qui répond aux exigences de l'EBP pourrait se situer entre ces deux extrêmes.

## **2. Une évaluation basée sur la performance fait appel au jugement du correcteur pour assigner une valeur à la production de l'élève.**

Le jugement professionnel des enseignants est particulièrement sollicité afin de juger de la valeur de la performance d'un élève (Gitomer, 1993; McMillan, 2001; Popham, 1999a; Stiggins, 1987). Une évaluation basée sur la performance vise essentiellement à rendre compte du niveau de maîtrise d'un élève face à une situation complexe comme par exemple une résolution de problème en science. Ainsi, il est possible que le problème comporte plusieurs solutions et que chacun des élèves ait à réaliser un certain nombre de choix dans le cadre de sa démarche. Plusieurs solutions et plusieurs démarches peuvent alors être acceptables. Dans ce cas, les grilles de correction critériée sont des outils qui peuvent grandement aider le ou les correcteurs à baliser le jugement qui est, en définitive, toujours subjectif (Arter & McTighe, 2001). Ajoutons que cette caractéristique associée aux situations de performance permet de les distinguer des items à réponse choisie où le jugement du correcteur, lors de la correction, est alors négligeable.

En ce qui concerne le jugement du correcteur, Popham (1999a) ajoute que la complexité des performances attendues fait également en sorte qu'il faut définir de multiples critères afin de juger adéquatement d'une performance.

## **3. Une évaluation basée sur la performance fait en sorte que l'élève crée un produit, construit une réponse ou fait quelque chose qui fait appel et qui démontre ses connaissances, ses habiletés et/ou ses compétences.**

Un aspect important lié à l'EBP concerne la réalisation d'une production de la part de l'élève qui peut prendre une variété de formes différentes (Kober, 1993; McMillan, 2001; Scallon, 2004; Simon & Forgette-Giroux, 2001; Stiggins, 1987). Les situations de performance exigent des élèves qu'ils soient actifs intellectuellement mais aussi parfois physiquement en mettant, par exemple, la main à la pâte en manipulant du matériel de laboratoire (Shavelson et al., 1991). Aussi, les situations de performance visent à placer les élèves dans un contexte où ils devront utiliser à bon escient les apprentissages qu'ils auront réalisés en classe (Baker, 1997; Hibbard, 1996) tout en défendant leurs réponses et leur(s) démarche(s) (McMillan, 2001; Scallon, 2004). Il ne s'agit donc pas de reproduire une

situation déjà vécue en classe et qui n'aurait été que pédagogiquement « maquillée » comme par exemple un problème où on demanderait aux élèves de refaire une expérience déjà réalisée intégralement dans les semaines précédentes. À ce sujet, Wiggins (1993a) a émis une mise en garde sur les activités *hands-on* en science. Il indique qu'il serait tentant de les associer systématiquement à des situations de performance alors que ce n'est pas nécessairement le cas. Il donne l'exemple d'une tâche en science où on demandait aux élèves de mesurer la température de l'eau en présentant aux élèves un contenant, de l'eau, un thermomètre et divers autres instruments de mesure faisant office de leurre. On constate qu'une telle mise en situation, bien qu'elle oblige les élèves à mettre la main à la pâte, ne répond en rien aux autres conditions définissant l'EBP.

#### **4. Une évaluation basée sur la performance nécessite un temps de passation élevé.**

Une évaluation basée sur la performance vise à mesurer l'atteinte de compétences ou d'habiletés intellectuelles de haut niveau (Aschbacher, 1991; Herman et al., 1992) comme – si on se rapporte à la taxonomie de Bloom (1975) – l'analyse, la synthèse ou l'évaluation. Il va sans dire que la résolution de problèmes complexes exige du temps afin de permettre aux élèves de mettre en œuvre une démarche de résolution de problème itérative pouvant être ponctuée de multiples retours en arrière. De plus, le fait d'administrer aux élèves des situations contextualisées qui exigent de faire la démonstration de ses compétences exige forcément du temps puisque les situations d'évaluation associées ne peuvent être constituées de questions fermées ou de tableaux à remplir (Tardif, 1993). Également, les situations de performance devraient normalement être relativement étrangères aux élèves. Bref, ils ne devraient pas retrouver facilement les indices leur permettant d'associer la situation d'évaluation à d'autres situations identiques à celles vécues en cours d'apprentissage. Ces aspects font en sorte que l'EBP exige un temps de passation beaucoup plus grand qu'une situation d'évaluation conventionnelle (Baker, 1997; McMillan, 2001).

**5. En science, une évaluation basée sur la performance possède des points communs avec les tâches qui sont dévolues normalement aux scientifiques professionnels.**

Solano-Flores et Shavelson (1997) identifient un sixième critère – qui recoupe en partie le quatrième – étant, en quelque sorte, une contextualisation des tâches authentiques telles que formulées par Wiggins (1993b) aux disciplines scientifiques. Selon eux, une situation de performance en science doit permettre de recréer les conditions intellectuelles (raisonnement, réflexion, etc.) dans lesquelles les scientifiques oeuvrent. Ainsi, la sélection de bonnes réponses est rarement une activité qu'un scientifique exécute dans le cadre de ses activités quotidiennes. Il doit plutôt émettre des hypothèses, circonscrire des problématiques, réaliser des expériences, faire des choix théoriques, méthodologiques et épistémologiques, etc. Une situation de performance devrait donc normalement simuler ce genre d'activités.

**6. Une évaluation basée sur la performance requiert des outils de cueillette de données permettant de saisir la compréhension des élèves au regard des concepts et des démarches jugés importants à l'intérieur d'une discipline.**

Comme nous l'avons mentionné précédemment, l'EBP vise à recueillir des données qui seraient plus difficilement accessibles au moyen d'outils de cueillette de données traditionnels (ex. tests avec items à correction objective). Les outils associés à l'EBP permettent généralement d'analyser des réponses ouvertes afin de mieux saisir la compréhension des contenus disciplinaires et les raisonnements complexes (Baker, 1997). Ce type de situation d'évaluation vise à mieux comprendre le niveau de compréhension des aspects pédagogiques jugés réellement importants par les spécialistes de la discipline (Baker, 1997). McMillan (2001) ajoute que la performance est directement observable par le correcteur. Enfin, Shavelson (1991) affirme que l'EBP doit nécessairement être cohérente avec les fondements du curriculum. Ainsi, ce modèle d'évaluation peut être particulièrement intéressant dans le cadre d'un programme qui vise justement à développer des habiletés de haut niveau à l'intérieur d'une discipline scolaire.

### 2.2.1 Les avantages

Popham (1999a, p. 160-161), s'appuyant sur les écrits de Mehrens (1992), identifie trois raisons de recourir à des situations d'évaluation basées sur la performance : l'insatisfaction face aux tests à réponses choisies, l'influence des recherches en psychologie cognitive et l'influence parfois nocive des tests de rendement conventionnels. D'abord, selon Popham, il y a plusieurs auteurs qui mettent en doute la validité de certains tests à réponses choisies afin d'évaluer des habiletés de haut niveau. Selon eux, ce genre de tests permettraient plutôt d'évaluer la capacité de l'élève à reconnaître une bonne réponse. Aussi, dans ce genre d'épreuve il est difficile de retracer la démarche intellectuelle de l'élève puisque celui-ci ne fait que choisir une réponse et cela représente généralement la seule information vérifiable sur laquelle il est possible de se baser pour porter un jugement. Ensuite, il mentionne que les travaux en psychologie cognitive mettent l'accent sur les connaissances déclaratives et procédurales. Ces dernières connaissances seraient particulièrement difficiles à évaluer par le biais d'items à réponse choisie. Enfin, toujours selon Popham, le format des items (souvent à choix multiple) présentés dans les tests de rendement à grande échelle pourrait avoir un impact sur l'enseignement. Il donne l'exemple d'un enseignant qui pourrait être tenté de faire converger son enseignement principalement sur les aspects qui seront mesurés par cet examen. De fait, cet enseignant pourrait volontairement écarter des activités pédagogiquement pertinentes afin de mettre l'accent sur ce qui fera l'objet de l'évaluation. En postulant que les items à choix multiple ne mesurent que difficilement des habiletés de haut niveau, cela voudrait dire que cet enseignant développe majoritairement des habiletés de base ou inférieures.

Quant à lui, McMillan (2001) identifie au moins quatre avantages à recourir à des situations de performance. Premièrement, l'EBP permettrait une meilleure congruence apprentissage-évaluation puisque la situation d'évaluation est aussi une situation d'apprentissage pour l'élève. Puisque ce qui est évalué est congruent avec ce qui est enseigné, il est plus facile de faire bénéficier l'enseignant des résultats d'évaluation (par exemple afin de corriger ses séquences d'enseignement-apprentissage) ce qui permet alors d'obtenir de meilleurs résultats aux évaluations. Deuxièmement, l'apprentissage repose de plus en plus sur une

vision constructiviste qui met l'accent sur le raisonnement et sur l'intégration des matières scolaires. L'EBP cadrerait mieux avec une vision constructiviste que les items à réponses choisis. Troisièmement, les situations de performance permettraient un meilleur engagement (meilleure motivation) de la part des élèves puisqu'ils accompliraient quelque chose de plus concret que de répondre à des tests de mémorisation. Quatrièmement, ce type d'évaluation forcerait les enseignants à identifier de multiples critères pour juger du succès de l'élève. Ces critères sont présentés aux élèves avant qu'ils n'accomplissent la tâche, ce qui permet aux élèves d'apprendre à s'auto-évaluer.

Accongio et Doran (1993) considèrent que les nouvelles formes d'évaluation, dont l'évaluation authentique, permettent de recueillir une information plus pertinente concernant ce que les élèves peuvent accomplir en science par rapport aux tests standardisés traditionnels de type papier-crayon. Ils indiquent que les modes d'évaluation traditionnels peuvent faire en sorte qu'un élève réussisse un cours de science tout en étant incapable de réaliser des démarches scientifiques complètes :

*Acceptance of standardized tests – driven by a desire to satisfy critics of public education, and the need to reduce complex intellectual behaviours to simple numbers – continue to define instruction. Within the prevailing framework for assessment, it is possible for a student to pass science courses and still remain scientifically illiterate. (Accongio & Doran, 1993, p. 27)*

Dans la même foulée, Jackson (2001) considère qu'il peut être injuste pour l'élève de le confronter à des situations de laboratoire et ensuite de l'évaluer au seul moyen de tests à correction objective. En ce sens, elle considère que l'évaluation, les programmes et les situations d'apprentissage doivent être cohérents entre eux.

### **2.2.2 Les limites**

L'évaluation basée sur la performance a, comme tout autre modèle d'évaluation, ses limites. Dans cette section, nous présentons les principales limites liées à l'EBP que nous avons regroupées en trois grandes catégories : (1) les limites liées au temps de rédaction et

de passation, (2) les limites liées à la formation des enseignants et des correcteurs, et (3) les limites liées à la qualité des informations recueillies et des inférences pouvant être réalisées.

D'abord, il est particulièrement difficile de bâtir des situations de performance (Baker, 1997). De plus, le temps de rédaction et de passation de telles épreuves est particulièrement long (Kober, 1993; O'Neil, 1992; Popham, 1999a; Shavelson & Baxter, 1992; Shavelson et al., 1991). Au sujet des contraintes de temps, McMillan (2001) ajoute qu'au-delà du temps de conception et d'administration de situations de performance, il est difficile pour un enseignant de prendre le temps d'observer et d'interagir avec les élèves lorsqu'ils réalisent la situation. Lors d'une première passation, il est aussi très difficile d'estimer le temps qui sera nécessaire aux élèves pour accomplir la tâche. Une autre limite, liée au temps de passation relativement élevé, tient à l'échantillonnage des performances pour chacun des élèves (McMillan, 2001; Popham, 1999a). En effet, puisqu'une situation de performance exige un temps de passation élevé, il n'est pas possible d'en administrer plusieurs; par conséquent, la somme d'information recueillie afin de porter un jugement sur une compétence est parfois limitée.

Ensuite, le recours à un tel modèle implique de devoir former adéquatement les enseignants ce qui entraîne des coûts particulièrement importants (Baker, 1997; Kober, 1993; Popham, 1993). Dans le cadre d'une évaluation certificative, on doit aussi considérer les coûts liés à la formation des enseignants concernant tout ce qui touche à la correction et à la notation. Il faut, par exemple, s'assurer d'une certaine homogénéité dans la compréhension et l'application des critères de correction afin d'obtenir un niveau d'accord inter juges acceptable. O'Neil (1992) rapporte que plusieurs experts évaluent que le coût d'administration de situations de performance est de deux à trois fois plus élevé que pour celles composées d'items à correction objective. Quant à elle, Aschbacher (1991) rapporte que pour certains états américains le coût d'administration d'une situation de performance varie entre trois et dix dollars américains, alors que ce coût s'élève à seulement 1,50 dollar américain pour un test standardisé classique. Enfin, Stecher et Klein (1997, p. 11) rapportent que ce coût passe à 102 dollars américains lorsqu'on souhaite obtenir un score

dont la fidélité est élevée (0,80) à une telle épreuve administrée à large échelle. Outre les coûts liés à la correction, ces derniers relèvent cinq autres aspects qui peuvent s'avérer problématiques dans une approche basée sur la performance :

- (1) accroissement de la charge de travail des enseignants et des directeurs d'école;
- (2) réduction du temps d'enseignement;
- (3) accroissement de l'espace disponible pour conserver les productions des élèves;
- (4) augmentation des ressources pour l'achat du matériel;
- (5) diminution de la clarté de l'interprétation des scores.

Enfin, Stiggins (1994) indique que certains enseignants abandonnent le recours à un modèle d'évaluation basée sur la performance à cause de mauvaises expériences passées qui n'ont pas donné les résultats escomptés. Ainsi, certaines tentatives plus ou moins heureuses de la part d'enseignants peuvent être à ce point décevantes que ces derniers ne risquent plus de nouvelles tentatives. On peut penser que la formation des enseignants et l'accompagnement de ces derniers dans le processus de changement de pratiques évaluatives pourraient pallier aux difficultés soulevées par Stiggins. Quant à elle, Baker (1997) mentionne que l'évaluation basée sur la performance peut souffrir d'un problème de validité apparente face aux parents des élèves. Selon elle, plusieurs parents peuvent penser que l'EBP est moins rigoureuse que les tests à choix multiple pour évaluer leur(s) enfant(s).

### ***2.3 L'investigation scientifique en laboratoire de type hands-on***

En examinant différents résultats de recherche, nous avons pu remarquer que l'expression « investigation scientifique *hands-on* » était souvent employée pour désigner des situations d'apprentissage ou d'évaluation fort différentes. Dans les sections qui suivent, nous présenterons les définitions les plus couramment associées à cette expression. Nous exposerons aussi les caractéristiques liées à l'investigation scientifique. Enfin, nous terminerons en identifiant les types de situations *hands-on* qu'il est possible de retrouver dans les écrits.

### 2.3.1 Activités *hands-on* : définitions et caractéristiques

En consultant la littérature, on peut se rendre compte que l'expression « *hands-on* » ne possède pas d'acception unique. Comme le soulignent Haury et Rillero (1994), l'expression « *hands-on* » est souvent employée pour représenter une panoplie d'activités réalisées en laboratoire. Ces activités empruntent parfois des expressions fort diverses telles : laboratoire, expérimentation, investigation, tâche *hands-on*, laboratoire recette, travaux pratiques, activité de démarche scientifique, laboratoire de vérification, etc. Aussi, on peut constater que l'on définit souvent cette expression au moyen de deux dimensions à savoir, d'une part, les activités réalisées en classe et, d'autre part, les conceptions entretenues face à l'activité scientifique. Dans le premier cas, on présente une liste d'activités qui peuvent être regroupées sous le thème d'activités *hands-on* comme les activités mentionnées plus haut dans ce paragraphe. Dans le second cas, on caractérise les activités *hands-on* comme étant des activités qui s'apparentent à des activités scientifiques professionnelles (Solano-Flores & Shavelson, 1997, p. 18). Cette situation reflète la difficulté à standardiser une définition liée aux activités ou à l'approche *hands-on*.

Dans le cadre de cette recherche, nous adoptons une définition qui tient compte des avis de deux groupes d'auteurs. En premier lieu, Solano-Flores, Shavelson, et Bachman (1999, p. 295) considèrent qu'une situation *hands-on* est une activité où « les étudiants sont appelés à utiliser l'équipement mis à leur disposition afin de réaliser une expérience leur permettant de résoudre le problème ou de vérifier l'hypothèse émise » (traduction libre<sup>14</sup>). Cette définition nous semble intéressante car elle met en lumière trois aspects qui nous paraissent fondamentaux dans le cadre d'une investigation scientifique en laboratoire. Le premier aspect concerne la manipulation d'équipement de laboratoire. Les élèves doivent obligatoirement manipuler du matériel (qui peut être très simple). Le second aspect est lié au processus de résolution de problème que doit mettre en branle l'élève. Ce second aspect vient baliser le premier car le matériel de laboratoire doit être manipulé afin de répondre à une question scientifique. Enfin, le troisième aspect touche l'émission et la vérification d'une hypothèse. Ce dernier aspect souligne l'importance de l'ingéniosité dans une

---

<sup>14</sup> La citation originale anglaise est la suivante : « *student are asked to use the equipment provided and conduct an experiment to solve a problem or test a hypothesis* ».

démarche scientifique d'investigation. Nous retenons également les propositions de Watson, Goldsworthy et Wood-Robinson (1999) qui, au cours d'une enquête nationale touchant environ mille enseignants anglais de sciences aux niveaux *key stage 2 et 3*<sup>15</sup>, concluent que les aspects suivants devraient être associés à l'investigation scientifique :

*In investigative work pupils have to make their own decisions either individually or in groups: they are given some autonomy in how the investigation is carried out.*

*An investigation must involve pupils in using investigational procedures such as planning, measuring, observing, analysing data and evaluating procedures.*

On peut extraire de ces descriptions quelques dimensions associées à l'investigation scientifique et que nous considérerons dans le cadre de la rédaction des situations d'évaluation. D'abord, l'élève doit être placé en situation d'autonomie lui permettant de faire des choix sur la façon dont il mènera ses investigations. Il est donc question ici de la démarche de l'élève et non de la démarche que l'enseignant suggère. Ensuite, l'élève doit être en mesure d'exercer certaines habiletés scientifiques (mesurer, observer, analyser les données, etc.) de façon à résoudre un problème à sa mesure.

### **2.3.2 Types de situations *hands-on***

On remarque dans la littérature qu'il est possible de classifier les situations *hands-on* de multiples façons. Nous avons choisi de présenter trois classifications qui nous apparaissent intéressantes.

Dans un premier temps, Watson et al. (1999) suggèrent une classification des situations d'investigation en six catégories : (1) classification et identification, (2) relations entre variables et contrôle de variables, (3) recherche d'une structure, (4) exploration, (5) investigation d'un modèle et (6) utilisation du savoir scientifique en technologie. Le tableau III qui suit résume chacune de ces catégories.

---

<sup>15</sup> Ces niveaux sont équivalents à la fin du primaire et le début du secondaire dans le système éducatif québécois.

**Tableau III**  
Description des catégories d'investigation (Watson et al., 1999).

Catégorie d'investigation	Description
Classification	Processus de réarrangement de phénomènes, d'objets ou d'événements en catégories.
Identification	Processus de reconnaissance d'objets et d'événements appartenant à une catégorie.
Relations entre variables et contrôle de variables	Processus d'observation et d'exploration entre les variables d'un problème. Identification d'une ou plusieurs variables indépendantes et des facteurs pouvant être contrôlés.
Recherche d'une structure	Cette catégorie est similaire à la précédente mais à trois différences près : (1) les variables peuvent être plus difficilement contrôlables. (2) l'investigation est généralement présentée de façon plus large et (3) l'échantillonnage représentatif est parfois plus difficile à obtenir.
Exploration	Fait référence à l'observation d'objets ou d'événements ou à une série d'observations d'un phénomène naturel sur une période de temps relativement longue. L'accent est mis sur la nature de ce qui est observé et la fréquence des observations.
Investigation d'un modèle	Un modèle est développé afin d'expliquer un phénomène scientifique. L'élève décide des données à recueillir afin de mettre à l'épreuve le modèle.
Utilisation du savoir scientifique en technologie	L'élève a recours aux concepts scientifiques afin de résoudre un problème où la technologie est sollicitée afin de répondre à un besoin.

Dans un second temps, Ruby (2001) classe les activités de type *hands-on* en cinq catégories : (1) vérification/démonstration, (2) découverte, (3) exploration, (4) investigation et (5) habiletés. En situation de vérification/démonstration, l'enseignant présente aux élèves, de façon théorique, une situation à caractère scientifique. Par la suite, les élèves sont invités à réaliser une expérimentation ou assister à une démonstration visant à « prouver » les allégations scientifiques théoriques discutées préalablement. Ces situations de

laboratoire sont généralement très directives et visent à répondre à une question précise. En contexte de découverte, on présente alors aux élèves le matériel auquel ils ont accès tout en leur fournissant peu ou pas d'indications sur ce qu'il faut faire ou chercher. Ruby précise que ce type de situations *hands-on* vise à développer deux aspects : la découverte de phénomènes scientifiques et les habiletés à mettre en pratique pour conduire adéquatement une investigation scientifique. Ce type d'activité est cependant souvent très (voire trop) difficile pour les élèves et certains concepts peuvent difficilement être découverts par les élèves (ex. théorie atomique). Une activité de type exploration vise, dans un premier temps, à ce que les élèves réalisent des découvertes en manipulant le matériel. S'ils n'y parviennent pas, on les guide alors pour leur faire découvrir ce que l'enseignant souhaite aborder au niveau conceptuel. Une situation d'investigation ressemble à une situation de découverte sauf que le rôle de l'enseignant est d'agir davantage comme un guide et de poser des questions pertinentes afin de mettre les élèves sur des pistes de recherche ou de solution. Ce type de situation possède cependant les mêmes limites ou difficultés que l'approche par découverte. Enfin, le développement d'habiletés scientifiques met l'accent sur des aspects interdisciplinaires tels la mesure, l'histoire des sciences, l'épistémologie et la nature de la science dans le cadre d'activités pratiques.

Enfin, Solano-Flores et al. (1997) ont proposé une classification des activités scientifiques de type *hands-on* en quatre catégories : (1) expérimentation comparative, (2) expérimentation d'identification, (3) activités de classification et (4) activités d'observation. La première catégorie vise à ce que les élèves réalisent une comparaison en s'appuyant sur des critères ou des attributs. La situation classique « les papiers absorbants » qui consiste à comparer différents papiers essuie-tout en vue de déterminer celui qui possède la meilleure absorbance en est un bon exemple. Les situations d'identification consistent à découvrir un aspect inconnu dans une situation comme dans le cas des boîtes mystères en électricité où il s'agit de reconstituer un circuit électrique qui est caché à l'intérieur d'une boîte qu'il n'est pas possible d'ouvrir. Les activités de classification ont pour objet de permettre à l'élève d'inventer un système logique lui permettant regrouper des objets selon des attributs spécifiques. La classification des minéraux et la création d'un système de classification de ces minéraux sont des exemples classiques appartenant à ce

type de tâche. Enfin, les activités d'observation permettent aux élèves de modéliser des phénomènes qu'ils ne peuvent recréer en laboratoire. Les auteurs donnent l'exemple de la création d'une maquette permettant de recréer la trajectoire du soleil du lever au coucher.

Ces classifications mettent en lumière l'importance de bien déterminer les caractéristiques (le squelette) des situations d'apprentissage et, à plus forte raison, des situations d'évaluation qui peuvent être associées à l'investigation scientifique de type *hands-on*. C'est d'ailleurs un aspect que nous déplorons dans plusieurs recherches que nous avons consultées. Par exemple, certains chercheurs ont construit des situations d'évaluation sans nécessairement documenter le processus de construction et les caractéristiques sur lesquelles ils se sont basés pour les construire. Ce faisant, il devient difficile de comparer des résultats de recherche qui apparaissent, à première vue, tout à fait comparables mais qui ne le sont pas nécessairement.

#### ***2.4 État de la question concernant les variables investiguées***

Dans cette section, nous présenterons les principaux résultats de recherche concernant les variables que nous allons investiguer dans le cadre de cette recherche. À notre grand étonnement, nous avons déniché très peu de recherches s'étant intéressées à l'évaluation de l'investigation scientifique en laboratoire de type *hands-on* à l'échelle de la classe. Ce constat nous a laissé perplexe à bien des égards. En effet, l'investigation scientifique n'est certes pas une nouveauté dans le paysage de l'enseignement des sciences et nous nous attendions à trouver une kyrielle de recherches traitant non seulement de celle-ci mais également des modalités d'évaluation qui pouvaient lui être associées. La revue de la littérature que nous avons effectuée nous a permis de constater que la plupart des études qui se sont intéressées à l'évaluation de l'investigation scientifique se sont déroulées à partir du début des années 90. On peut penser que ce domaine de recherche est, somme toute, relativement nouveau. On peut émettre quelques hypothèses pour expliquer ce constat. La première hypothèse que nous retenons est que les chercheurs qui s'intéressent à l'évaluation ne le font pas dans le seul champ des sciences mais dans plusieurs autres disciplines scolaires. Aussi, plusieurs d'entre eux n'ont pas une formation scientifique. Par

exemple, nous présenterons dans les deux prochaines sections les travaux de Richard J. Shavelson qui a produit plusieurs recherches intéressantes sur l'évaluation en science. Par contre, ce chercheur s'intéresse aussi à d'autres sujets (entre autres, la théorie de la généralisabilité) qui ne sont pas liés à l'évaluation en science. La seconde hypothèse que nous pouvons formuler est à l'effet que l'investigation scientifique est un sujet abordé principalement par les didacticiens. Nous avons effectivement identifié plusieurs recherches qui traitaient de l'investigation scientifique mais dans une perspective didactique et non d'évaluation.

#### **2.4.1 Liens entre le modèle d'évaluation, la forme de l'épreuve, le contenu disciplinaire et les résultats des élèves en science**

En Belgique, Rey, Carette, Defrance et Kahn (2003) ont proposé un modèle visant à évaluer des compétences. Ils ont en effet réalisé une étude auprès de 1200 élèves du primaire. Les épreuves ont été administrées à des enfants de troisième année (enfants d'environ 8 ans) et de dernière année du primaire (enfants d'environ 12 ans). Le projet consistait à expérimenter un modèle d'évaluation diagnostique en trois phases qui partage plusieurs points communs avec les niveaux que nous avons présentés au premier chapitre. Par contre, les chercheurs ont choisi d'administrer les trois versions de leurs épreuves à chacun des élèves. Pour établir leur modèle, ils ont procédé en trois étapes :

1. ils ont réalisé une étude des différents sens de la notion de compétence;
2. ils ont ensuite identifié les aspects les plus formateurs qui ont ensuite servi à construire les épreuves;
3. ils ont analysé les données sous l'angle de l'évaluation et de la didactique au terme de l'expérimentation<sup>16</sup>.

Les six épreuves qu'ils ont produites visaient à évaluer des compétences disciplinaires en langue et en mathématiques ainsi que des compétences transversales (ex. recherche

---

<sup>16</sup> La recherche a duré deux ans. La première année a permis surtout de construire les épreuves et d'assurer une première validation. La seconde année a permis de recueillir des données plus appuyées.

d'information). À titre d'exemple, l'une des tâches intitulée « le jardinage » consistait à planifier la plantation de légumes dans le jardin de l'école. Pour y arriver, les élèves devaient accomplir trois tâches :

1. choisir les trois graines qui seraient semées afin de récolter les légumes dans les prochains quatre mois;
2. réaliser un schéma du jardin selon des consignes données;
3. construire un panneau représentant les légumes cultivés.

Concernant la nature des épreuves construites et expérimentées, les auteurs de cette recherche indiquent (Rey et al., 2003, p. 98) :

Toutes ces situations complexes font appel à la mobilisation et l'organisation de procédures. Les élèves, pour les résoudre en phase 1, sont obligés de construire leur propre démarche. Par ailleurs ces situations complexes sont construites dans une visée interdisciplinaire. D'une manière plus précise, les différentes tâches proposées peuvent se différencier suivant plusieurs critères. En effet, nous avons des tâches orientées vers la production d'écrits ou des tâches orientées vers la résolution de problèmes faisant appel à des procédures mathématiques. Parmi les différentes tâches, nous en trouvons qui font appel à la recherche d'informations dans des documents inédits et d'autres où les informations font partie de la consigne. En phase 2, les élèves reçoivent la même tâche mais cette fois-ci, ils sont aidés dans leur démarche soit par un questionnement sur la tâche et/ou par la présentation d'un modèle (...). En phase 3, nous proposons des exercices systématiques, décontextualisés évaluant des procédures automatisées et nécessaires à la réussite des phases 1 et 2.

Les résultats obtenus indiquent que les élèves maîtrisent très peu les compétences puisque seulement 2,95% des élèves ont réussi (à 80% ou plus) l'ensemble des tâches lors de la première année d'expérimentation et cette valeur a grimpé à 4,65% lors de la deuxième cueillette de données une année plus tard. Avec un seuil de passage fixé à 60%, les données de la première année d'expérimentation nous indiquent que 8,85% des élèves réussissent la phase 1, 48,44% réussissent la phase 2 et 78,85% réussissent la phase 3. Lors de la deuxième année, ces valeurs passent respectivement à 25,18%, 52,00% et 66,49%. Selon ces chercheurs, il existerait effectivement un lien important entre les phases. Le calcul du coefficient de corrélation de Pearson pour les données recueillies lors de la deuxième année

indique que le lien entre la phase 1 et 2 se situe à 0,46, entre la phase 1 et 3 la valeur se situe à 0,49 et entre la phase 2 et 3 elle se situe à 0,63. L'analyse des données permet aussi de constater que les élèves « forts » réussissent très bien chacune des phases. Les élèves « faibles » sont en difficulté dans chacune des phases. En ce qui concerne les élèves « moyens », ils ont tendance à obtenir des résultats moyens pour chacune des phases. Ces résultats sont d'ailleurs concordants avec ceux obtenus par Lawrenz, Huffman et Welch (2001) qui sont arrivés aux mêmes constats.

La forme de l'instrument de mesure aurait, selon Shavelson, Baxter et Pine (1991) un effet sur la nature des informations recueillies lors de l'évaluation de l'investigation scientifique chez les élèves. Ces auteurs ont en effet comparé les résultats d'élèves (n=300) de cinquième et sixième année au moyen de trois situations d'évaluation (les papiers absorbants, les insectes et les circuits électriques mystères) pour lesquels ils avaient élaboré quatre formes différentes d'instruments de mesure : une tâche de type *hands-on*, une tâche de simulation par ordinateur, une tâche ouverte de type papier-crayon et un test à choix multiples. Les résultats montrent que les tâches de type *hands-on* mesurent des aspects différents concernant la démarche scientifique en comparaison avec une approche basée sur un test à choix multiples. Ils ont calculé que la corrélation moyenne entre ces deux formes d'épreuve était d'environ 0,45. La forme des instruments de mesure ne serait donc pas interchangeable. Selon eux, les situations qui permettent de recueillir une information semblable à celle obtenue avec des situations de type *hands-on* sont les simulations informatiques. Une recherche plus récente, menée par Lawrenz, Huffman et Welch (2001), qui touchait 3500 élèves de la neuvième année arrive à des résultats similaires à ceux précédemment présentés. Ces auteurs ont en effet comparé les résultats des élèves au moyen de quatre formes d'épreuves distinctes : test à choix multiples, questions ouvertes, test d'habileté *hands-on* et investigation *hands-on*<sup>17</sup>. De façon similaire à Shavelson *et al.*, ils ont constaté que les résultats des élèves varient selon la forme de l'épreuve. La forme *hands-on* donne des résultats significativement différents de ceux des formes « choix

---

<sup>17</sup> L'article que nous avons consulté ne présentait pas les situations utilisées ni les caractéristiques de ces dernières. Selon notre compréhension, les situations qualifiées « habiletés *hands-on* » seraient équivalentes à nos situations de niveau 3 (SE1-3, SE2-3 et SE3-3) et les situations qualifiées « investigation *hands-on* » seraient équivalentes à nos situations de niveau 1 (SE1-1, SE2-1, SE3-1).

multiples » et « questions ouvertes ». Les formes donnant les résultats les plus semblables sont les formes « choix multiples » et « questions ouvertes ». Cette recherche a aussi permis de montrer que les élèves ont obtenus les meilleurs résultats avec la forme d'épreuve portant sur l'investigation *hands-on*. Les chercheurs expliquent ceci en disant que ces tâches sont peut-être plus motivantes pour les élèves plus faibles. Ces situations sont aussi peut-être moins abstraites, ce qui favoriserait ces élèves. Par contre, les élèves n'ont pas pu réaliser de multiples tâches *hands-on*, ce qui représente une limite importante concernant la généralisation de ces résultats.

Selon les recherches que nous avons consultées, la nature des concepts disciplinaires aurait un rôle important à jouer concernant la performance des élèves à des situations d'investigation scientifique de type *hands-on*. En reprenant les données issues de la recherche décrite au paragraphe précédent (Shavelson, Baxter et Pine (1991)), Shavelson et Baxter (1992) ont remarqué que les situations faisant appel à des concepts disciplinaires précis (ex. les circuits électriques mystères) ont tendance à donner des résultats différents par rapport aux situations qui mettent l'accent uniquement sur l'investigation (ex. les papiers absorbants).

Quant à eux, Stecher et son équipe (2000) se sont aussi intéressés à l'influence du contenu disciplinaire, à la forme de la situation d'évaluation et au niveau d'investigation scientifique mesuré dans le cadre de situations ouvertes à caractère scientifique. Pour ce faire, ils ont bâti et administré huit tâches ouvertes auprès d'environ 1200 élèves, ces tâches se distinguaient selon :

1. la forme de la situation (*hands-on* vs papier-crayon);
2. le niveau d'investigation mesuré (guidé vs non guidé);
3. le contenu (« acides et bases », « force et mouvement » et « propriétés des matériaux »).

L'hypothèse à l'effet que les tâches qui étaient similaires au niveau du contenu, du niveau d'investigation et du format (*hands-on* vs papier crayon) allaient être corrélées entre elles s'est avérée une piste fructueuse. Les valeurs des corrélations obtenues, pour l'ensemble

des mesures prises, variaient en effet entre 0,50 et 0,66. De même, l'ordre de présentation des tâches n'a pas eu d'effet sur le score des élèves. Par exemple, les scores aux tâches « non guidées » n'ont pas été plus élevés quand ces tâches venaient après une tâche « guidée ». Cette recherche a aussi permis aux chercheurs de constater que les situations administrées aux élèves n'étaient pas aussi identiques – en terme de format, de contenu disciplinaire et de niveau d'investigation scientifique – que ce qu'ils avaient d'abord pensé lors de la passation initiale soulevant l'importance de concevoir des situations d'évaluation selon un cadre théorique rigoureux.

Ayala, Yin, Schultz et Shavelson (2002) se sont intéressés, entre autres, à la comparabilité des scores lors de l'administration de trois situations de performance en science. Pour ce faire, ils ont utilisé et adaptés trois situations à caractère scientifique : (1) les circuits mystères (physique), (2) la durée du jour (astronomie) et (3) simulation d'une explosion à bord d'un sous-marin (*aquacraft*) (chimie). Chacune de ces situations ciblait trois dimensions différentes qui étaient respectivement : (1) connaissance et raisonnement, (2) habiletés spatio-mécanique et (3) maîtrise des aspects quantitatifs. Pour assurer la notation, ils ont défini quatre dimensions dans le cadre de la première situation (n=6) et ils ont obtenu les résultats qui suivent : schématisation (moy. = 8,55, é.-t.<sup>18</sup> = 3,84), observation (moy = 7,77, é.-t. = 3,14), inférence (moy. = 9,49, é.-t. = 3,10) et explication (moy. = 2,39, é.-t. = 2,29). Les sujets ont obtenu des résultats satisfaisants sauf en ce qui concerne l'émission d'explications justifiées et pertinentes. En ce qui concerne la deuxième situation (n=6), trois dimensions ont été évaluées : résultats (moy. = 14,45, é.-t. = 5,35), modélisation (moy. = 5,36, é.-t. = 2,32) et explication (moy. = 7,29, é.-t. = 3,50). À la lumière de ces résultats, on constate que les résultats des élèves varient considérablement à l'intérieur même d'une situation. Enfin, la troisième situation étaient notées en se basant sur les dimensions qui suivent : réaction chimique (moy. = 5,27, é.-t. = 2,43), identification du gaz (moy. = 2,07, é.-t. = 1,34), balancement d'équations chimiques (moy. = 6,16, é.-t. = 2,78) et calcul de l'énergie impliquée (moy. = 2,20, é.-t. = 2,53). On peut remarquer des similitudes entre les résultats obtenus à cette situation et ceux obtenus aux deux précédentes situations. En comparant les résultats, les auteurs ont pu remarquer que certains

---

<sup>18</sup> e.t. signifie écart-type

élèves obtenaient de bien meilleurs résultats à une des trois situations. Nous pouvons penser que ces résultats soutiennent le fait que le contenu disciplinaire pourrait avoir un impact sur les résultats des élèves.

#### **2.4.2 Liens entre le sexe de l'élève, les caractéristiques de l'enseignant et les résultats des élèves en science**

Globalement, la plupart des recherches indiquent qu'il y a peu de différences entre les scores obtenus par les filles et les garçons dans des situations d'évaluation de l'investigation scientifique de type *hands-on*. C'est la conclusion à laquelle sont arrivés Lawrenz *et al.* (2001). Par exemple, les scores moyens obtenus par les garçons et les filles aux tâches d'investigation ont été respectivement de 69,15% (e.t. = 3,03) et de 71,88 (e.t. = 3,09). Jovanovic, Solano-Flores et Shavelson (1994) se sont intéressés à cette problématique dans le cadre d'une étude réalisée auprès d'environ 2400 élèves issus de la cinquième, sixième et neuvième année. Ils ont conçu et expérimenté une batterie de tâches qui différaient selon la forme (items à choix multiples, items à choix multiples avec justification et questions ouvertes provenant du CLAS (California Learning Assessment System), items du ITBS (Iowa Tests of Basic Skills), tâches d'investigation du CLAS et tâches d'investigation produites par les chercheurs), les habiletés ou compétences visées (analyse, design, classification) et les concepts disciplinaires en cause (biologie, physique, géologie). Les résultats indiquent qu'il n'existe que peu de différence entre les garçons et les filles. Les résultats aux situations de performance ont montré que les filles obtenaient de meilleurs résultats en sixième année mais les différences n'étaient pas statistiquement significatives en cinquième et en neuvième année. Quant à eux, Jovanovic, Solano-Flores et Shavelson (1994) ont réalisé trois études en utilisant les données issues de cinq recherches<sup>19</sup>. Dans la première étude, ils ont comparé les résultats d'élèves (n=600) de sixième année du primaire à cinq situations *hands-on* issues de cinq domaines disciplinaires différents (conductivité électrique, classification de feuilles, identification de roches, identification du pH d'une solution et mesure). Les résultats ont

---

<sup>19</sup> Les recherches en questions sont celles de : Shavelson, Baxter et Gao (1993), Gao (1992), Baxter (1991), Shavelson, Baxter et Pine (1992) et Solano-Flores, Jovanovic et Shavelson (1994)

démontré que les filles obtenaient de meilleurs résultats à trois de ces situations (classification de feuilles, identification de roches, identification du pH d'une solution et mesure). Les tâches liées à la mesure n'ont indiqué aucune différence significative alors que les garçons ont obtenu de meilleurs scores que les filles à l'activité concernant l'électricité. La seconde étude visait à comparer la performance des élèves à trois situations : boîtes mystères (physique), préférence des insectes à un milieu donné (biologie), identification du meilleur papier absorbant (situation d'investigation pure). Les analyses ont permis de constater qu'il y avait un effet « genre » car les résultats des garçons étaient supérieurs encore une fois à ceux des filles pour la situation impliquant des concepts en électricité. La troisième étude concernait spécifiquement l'étude de cet effet en soumettant aux élèves deux situations de physique (plan incliné et friction). Dans le cas de cette étude, les analyses de variance (ANOVA) n'ont permis de découvrir aucune différence significative entre les garçons et les filles. Selon les auteurs de cette étude, les expériences antérieures pourraient expliquer ces résultats. Les travaux en didactique des sciences indiquent en effet que les garçons sont généralement plus motivés à manipuler du matériel électrique (fils, ampoules, piles, etc.) ce qui pourrait leur donner un avantage lorsque les situations font appel à ces concepts.

La revue de la littérature que nous avons réalisée ne nous a pas permis de trouver beaucoup de résultats de recherche concernant le lien entre les résultats des élèves à des situations de performance en science et les caractéristiques associées aux enseignants ou aux méthodes pédagogiques qu'ils préconisent. Cependant, Rey *et al.* (2003) se sont intéressés à ce lien. Ils ont comparé les résultats des élèves en se basant sur une variable « dynamisme pédagogique ». Pour caractériser cette variable, ils s'en sont remis à la connaissance de sept inspecteurs qui connaissaient les enseignants qui participaient à la recherche. Selon des critères subjectifs et discutables (de l'aveu même des auteurs) ils ont classé les écoles en deux catégories : écoles dynamiques et écoles moins dynamiques. Les résultats obtenus n'indiquent aucune différence quant au succès des élèves selon qu'ils appartiennent à une école dynamique ou moins dynamique. Par exemple, la moyenne des scores obtenus aux phases 1,2 et 3 a été respectivement pour les écoles dynamiques de 46,27%, 55,16% et de 63,69%. Pour les écoles moins dynamiques, ces scores ont été de

46,24%, 58,52% et de 68,78%. Les auteurs proposent quatre hypothèses pouvant expliquer ces résultats : (1) certains élèves pourraient ne pas avoir pris les évaluations au sérieux, (2) les situations pourraient ne pas avoir été suffisamment motivantes aux yeux des élèves, (3) il n'y aurait, effectivement, aucun lien entre le résultat des élèves et le dynamisme de l'école et (4) certaines pédagogues plus ouvertes pourraient moins considérer le développement d'automatismes qui étaient à la base des situations de phase 3 qu'ils ont produites. Nous nous permettons d'ajouter à ceci deux hypothèses que nous considérerons dans le cadre de cette étude. La première concerne la difficulté à bien cerner le « dynamisme pédagogique » des écoles. Idéalement, nous pensons qu'il faudrait réaliser des entrevues auprès des enseignants et des visites en classe afin de mieux cerner ce dynamisme et les méthodes d'enseignement privilégiées par les enseignants en cours d'apprentissage. La seconde hypothèse porte sur les consignes et les conditions de passation des épreuves. Nous pensons que la façon de présenter, par exemple, les situations aux élèves ou les règles de passation d'une phase à l'autre peuvent affecter considérablement les résultats des élèves. Nous pensons qu'il est plausible que de telles interactions aient pu masquer le rôle du dynamisme pédagogique dans le cadre de leur étude. Quoiqu'il en soit, nous avons essayé de tenir compte de ces résultats dans le cadre de cette recherche.

Baxter, Shavelson, Goldman et Pine (1992) se sont aussi intéressés au lien entre l'enseignement et les résultats des élèves à des situations d'évaluation de la performance en investigation scientifique. Ils ont en effet comparé les résultats des élèves (n=96) de cinquième année en analysant les effets d'un certain nombre de variables telles la forme de l'épreuve (test à choix multiple, évaluation basée sur une performance et un test d'habileté en science<sup>20</sup>), la façon de noter les élèves (observation directe vs cahier de l'élève) et le type d'enseignement offert aux élèves (élèves réalisant des investigations scientifiques de type *hands-on* vs élèves n'en réalisant pas). Ces chercheurs ont montré que les élèves ayant déjà réalisé des tâches de type *hands-on* ont obtenu de meilleurs résultats que les élèves n'en ayant jamais réalisé. Il faut cependant noter que cette tendance s'est vérifiée aussi lors de l'administration des épreuves à choix multiples.

---

<sup>20</sup> Le test en question est le CTBS (*Comprehensive Test of Basic Skills*).

À notre connaissance, il n'y a aucune étude qui s'est intéressée spécifiquement à l'effet de variables « enseignant » (ex. expérience et formation des enseignants) sur les résultats des élèves à des situations de performance en science au Québec.

On ne peut se passer d'une méthode pour se  
mettre en quête de la vérité des choses.  
René Descartes

## **Chapitre 3. Méthodologie**

---

Ce chapitre aborde les éléments de méthode sur lesquels nous nous sommes appuyés tout au long de ce processus de recherche. Nous avons divisé ce chapitre en plusieurs sections pour plus de clarté. D'abord, nous présentons les situations d'évaluation que nous avons expérimentées ainsi que leurs principales caractéristiques. Par la suite, nous nous attardons au processus que nous avons mis en œuvre afin de valider les dites situations. Nous enchaînerons en présentant les variables investiguées. Enfin, nous terminerons en précisant les aspects liés au traitement et à l'analyse des données.

### ***3.1 Les situations d'évaluation expérimentées***

Il existe plusieurs méthodes permettant de recueillir de l'information quand les élèves sont confrontés à une situation d'évaluation basée sur la performance. McColskey et O'Sullivan (1993, p. 15) suggèrent trois méthodes : (1) observation des élèves, (2) dialogue avec les élèves et (3) observation de la production des élèves. La première méthode consiste à recueillir des données alors que les élèves sont dans l'action. Cette observation est qualifiée d'informelle quand elle est recueillie de façon spontanée sans but précis, et de structurée quand elle vise à recueillir des données sur des dimensions identifiées à l'avance. L'observation peut également être colligée de façon exhaustive au moyen d'un texte quand l'enseignant veut rendre compte d'un phénomène complexe. La seconde méthode vise à interroger les élèves afin d'analyser, par exemple, leur compréhension d'un concept particulièrement difficile à saisir. Les auteurs suggèrent deux façons de procéder soit par interview ou par auto-évaluation. Enfin, la production des élèves peut être analysée au moyen de questions ouvertes, de situations de performance, de portfolios ou du produit de la démarche de l'élève (ex. projet, maquette, etc.).

Dans le cadre de cette recherche, nous avons opté pour la troisième méthode de collecte de l'information. Nous considérons que les deux premières méthodes s'opérationnalisent difficilement dans un contexte d'évaluation certificative. Avec des groupes d'élèves nombreux (32-34 élèves), il est en effet difficile de trouver du temps pour interviewer les élèves ou noter de façon systématique leurs actions en classe-laboratoire. La troisième méthode permet, quant à elle, de recueillir l'information et ensuite les enseignants ont un temps raisonnable pour l'analyser adéquatement. Puisque nous souhaitons expérimenter un modèle d'évaluation viable dans le contexte actuel marquant la discipline de science et technologie au secondaire, les limites que nous venons tout juste d'évoquer nous ont poussé à n'utiliser que des données qu'il est possible d'analyser après leur cueillette.

### **3.1.1 La conception des situations d'évaluation**

De l'avis de plusieurs auteurs (Baker, 1997; Scallon, 2004; Solano-Flores et al., 1999; Solano-Flores & Shavelson, 1997), la conception de situations d'évaluation de performance n'est pas une sinécure. L'absence de méthodologies éprouvées pour évaluer des compétences et l'absence de situations d'évaluation expérimentées sont quelques facteurs plausibles pouvant expliquer ce constat. Dans cette section, nous présenterons la démarche que nous avons mise de l'avant pour la conception des situations d'évaluation et qui représente une synthèse de trois propositions (Doran & Hejaily, 1992; Herman et al., 1992, p. 34; Moskal, 2003) schématisée à la figure 6. Nous sommes conscients que les situations d'évaluation que nous avons rédigées peuvent être critiquées comme le serait d'ailleurs n'importe quelle épreuve. En ce qui nous concerne, nous avons été dans l'obligation, à plusieurs reprises, de prendre des décisions. Certaines ont été plus déchirantes que d'autres. Dans tous les cas, nous avons tenté de ne pas perdre de vue les intentions qui nous habitaient dans le cadre de cette recherche à savoir expérimenter des situations d'évaluation qui nous informent sur le développement d'une compétence d'investigation scientifique dans un contexte réaliste d'enseignement au secondaire. Dans les sections qui suivent, nous identifierons les choix que nous avons dû réaliser et les raisons qui expliquent ces choix en raison des contraintes auxquelles nous avons dû faire face.

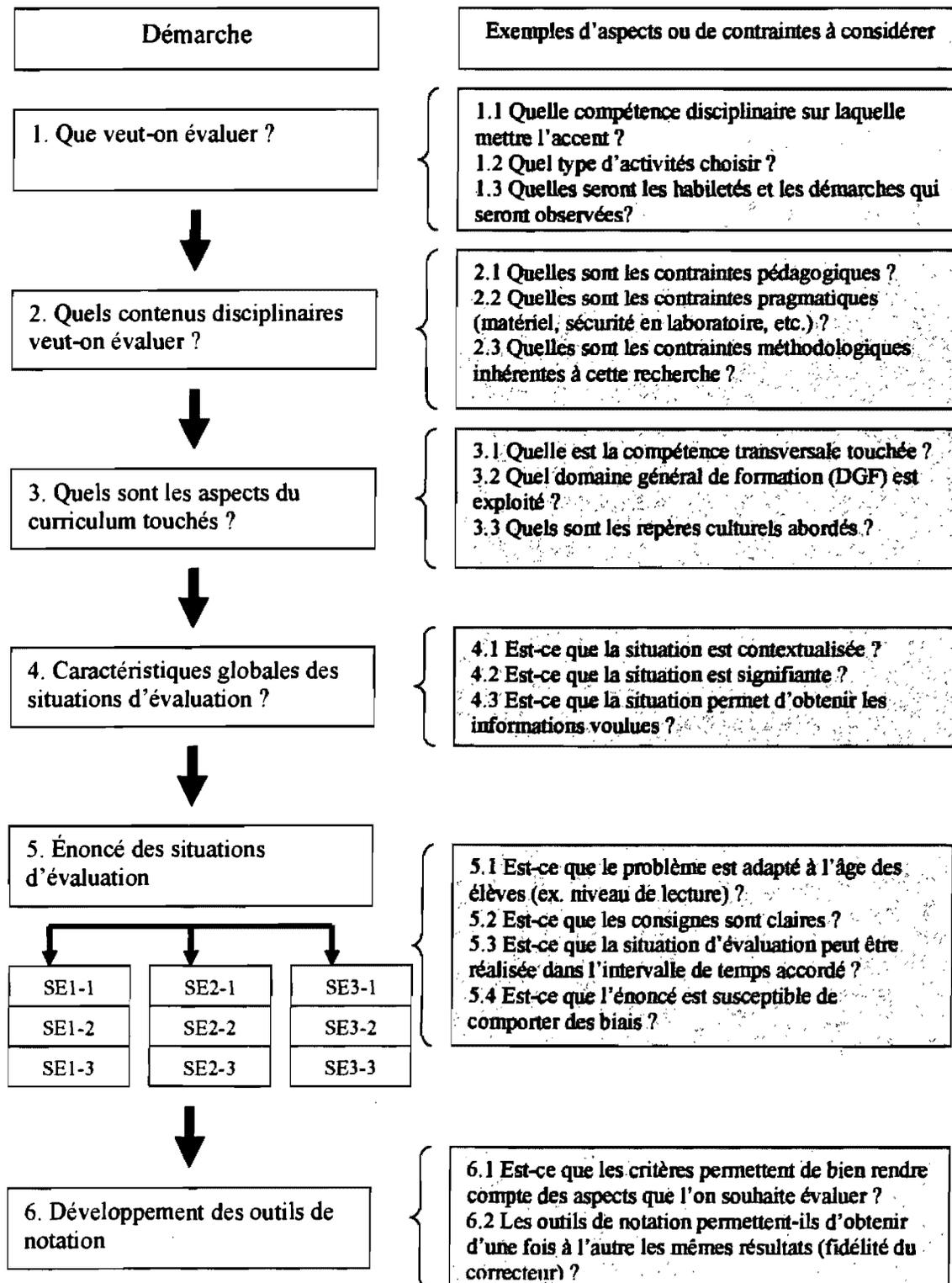


Figure 6  
Démarche pour la conception initiale des situations d'évaluation-bilan.

La première étape – le choix de la compétence disciplinaire à évaluer – allait de soi puisque nous avons d’ores et déjà décidé de mettre l’accent sur la démarche d’investigation en science. Par la suite, nous avons choisi d’axer les situations d’évaluation sur des démarches de résolution de problème dans un contexte d’investigation scientifique où les élèves doivent corroborer un fait scientifique au moyen d’observations, de mesure et d’expérimentation.

La seconde étape – consistant à déterminer les concepts, les techniques et les habiletés<sup>21</sup> présents dans chaque situation d’évaluation – a donné lieu à plusieurs difficultés. Ces difficultés étaient d’ailleurs prévisibles puisque le choix des contenus est tributaire de plusieurs contraintes pédagogiques, pragmatiques et aussi expérimentales. D’un point de vue pédagogique, plusieurs concepts généraux du Programme de formation se prêtaient difficilement à une expérimentation en laboratoire comme par exemple, le concept général de « perpétuation des espèces » dans l’univers vivant ou le concept général « d’organisation » dans l’univers matériel. Dans ces cas, nous avons choisi de ne pas inclure ces contenus dans les situations d’évaluation. Certains contenus étaient aussi difficilement accessibles en situation d’évaluation-bilan en raison de contraintes pragmatiques liées au travail en laboratoire. Par exemple, une situation d’évaluation qui aurait exigé la croissance de végétaux aurait été à toute fin pratique impossible à réaliser en raison, entre autres, du temps de germination des graines, des conditions d’ensoleillement nécessaires et de l’espace nécessaire à l’entreposage des montages expérimentaux. Ainsi, tous les concepts qui présentaient des contraintes pragmatiques majeures ont été mis de côté pour la conception des situations d’évaluation. Enfin, une dernière catégorie de contraintes était liée au plan expérimental de cette recherche. Les situations d’évaluation produites dans le cadre de cette recherche devaient pouvoir être expérimentées dans plusieurs écoles de plusieurs commissions scolaires. Ainsi, chaque école a généralement des contraintes spécifiques concernant le matériel de laboratoire (incluant la disponibilité de classes-laboratoires) et les ressources humaines (ex. la disponibilité des techniciens de laboratoire). Les situations d’évaluation ne devaient donc pas comporter des contraintes

---

<sup>21</sup> Pour des raisons de clarté, nous nous référerons à l’expression « contenu » pour faire référence aux concepts, techniques, attitudes, habiletés prescrits dans le Programme de formation.

trop rigides pouvant remettre en question les conditions de leur passation. Lorsque des contenus étaient susceptibles de générer de telles contraintes, nous les avons exclus des situations d'évaluation.

La troisième étape avait pour but de s'assurer de la meilleure adéquation possible entre les situations d'évaluation et les fondements généraux du Programme de formation de l'école québécoise et de la Politique d'évaluation des apprentissages. Comme nous l'avons souligné au chapitre 1, il faut s'assurer que le triangle curriculum-apprentissage-évaluation soit le plus cohérent possible afin de maximiser la validité de construit des épreuves évaluatives. À ce sujet, nous avons identifié une compétence transversale<sup>22</sup> qui devrait normalement être mobilisée pour chacune des situations mais qui ne serait pas évaluée formellement puisque cela aurait dépassé le cadre de notre recherche. Lorsque cela était pertinent, nous avons également identifié un domaine général de formation<sup>23</sup> qui devait normalement être exploité par les élèves lors de la résolution du problème ainsi que des repères culturels liés à la tâche.

La quatrième étape consistait à s'assurer que la situation d'évaluation était conforme à une situation de performance comme nous l'avons définie au chapitre 2. La situation devait être, entre autres, contextualisée, signifiante, ouverte et permettre à l'élève de véritablement résoudre un problème ayant un niveau de difficulté bien dosé. La structuration du problème dévolu à l'élève devait également faire en sorte de permettre de recueillir une information à la fois valide et suffisante sur le niveau de développement de la compétence de l'élève.

---

<sup>22</sup> Les neuf compétences transversales répertoriées dans le Programme de formation de l'école québécoise, premier cycle du secondaire (Ministère de l'Éducation du Québec, 2003c) sont : (1) exploiter l'information, (2) résoudre des problèmes, (3) exercer son jugement critique, (4) mettre en œuvre sa pensée créatrice, (5) se donner des méthodes de travail efficaces, (6) exploiter les technologies de l'information et de la communication, (7) actualiser son potentiel, (8) coopérer et (9) communiquer de façon appropriée. Ces compétences peuvent être développées dans toutes les disciplines (mathématique, langue, science et technologie, etc.) et elles sont de la responsabilité de tous les enseignants.

<sup>23</sup> Selon le Ministère (Ministère de l'Éducation du Québec, 2003c), les domaines généraux de formation sont « les problématiques auxquelles les jeunes doivent faire face dans différentes sphères importantes de leur vie ». Les domaines généraux de formation sont regroupés en cinq catégories : (1) santé et bien-être, (2) orientation et entrepreneuriat, (3) environnement et consommation, (4) médias et (5) vivre-ensemble et citoyenneté.

La cinquième étape a consisté à rédiger l'énoncé du problème ainsi que les différentes sections du cahier de l'élève en vue de la pré-expérimentation. Nous avons tâché de tenir compte des principaux aspects qui peuvent affecter la performance des élèves tels le niveau de formulation, l'absence de biais et le respect du temps de passation. À cette étape, nous avons accordé une attention toute spéciale à la formulation du problème que nous allions soumettre aux élèves afin de nous conformer aux exigences ministérielles. Selon le ministère de l'Éducation du Québec, une situation d'évaluation doit être ouverte au sens où « elle présente des données de départ susceptibles de fournir différentes pistes de solution » (Ministère de l'Éducation du Québec, 2003c, p. 272). Sur ce sujet, Arzac, Germain et Mante<sup>24</sup> (1988) sont plus explicites sur les caractéristiques de tels problèmes ou situations. Selon eux, les problèmes ouverts doivent :

- avoir un énoncé court qui n'induit ni la méthode, ni la solution;
- éviter de se réduire à l'utilisation ou l'application immédiate des derniers résultats vus dans les cours précédents;
- se trouver dans un domaine conceptuel avec lequel les élèves ont peu de familiarités.

Quant à lui, Scallon (2004) (citant Fabre (1999)) ajoute que les problèmes ouverts offrent la possibilité de répondre adéquatement au problème de multiples façons. Ainsi, plusieurs solutions peuvent être acceptées pour un problème donné. Ajoutons que nous avons tenu compte de l'ensemble de ces recommandations lors de la rédaction des problèmes.

Enfin, la dernière étape avait pour but de structurer et de construire l'outil de notation que nous comptons utiliser. Nous avons opté pour une grille critériée dans laquelle on retrouve les principales dimensions des compétences 1 et 3.

### **3.1.2 Caractéristiques de chacun des niveaux**

La rédaction d'une seule situation de performance permettant de recueillir des informations sur le développement d'une compétence scientifique est en soi une activité complexe. La

---

<sup>24</sup> cité dans Astolfi (1997, p. 141)

rédaction de plusieurs situations de performance est évidemment une activité encore plus complexe. Un élément de complexité supplémentaire s'ajoute lorsqu'on souhaite que les situations aient des propriétés relativement semblables les unes aux autres à la manière des « tests parallèles classiques » de McDonald (1999) dans le contexte de la théorie classique des tests. On cherche alors à concevoir des tests qui possèdent des caractéristiques semblables (contenus semblables, même type d'items, moyennes semblables, écarts-types semblables, etc.). Même dans le contexte d'une évaluation basée sur la performance, nous souhaitons créer une « coquille » réutilisable ayant suffisamment de balises afin de créer d'autres situations « semblables » ou « parallèles » à la manière de la proposition de Solano-Flores (1999) ou de celle de Klein *et al.* (1997). Nous avons donc identifié cinq dimensions – présentées au tableau IV - qui, à notre avis, permettent de caractériser chacun des trois niveaux (1-2-3) pour chacune des tâches.

- les indices méthodologiques;
- la liste du matériel accessible;
- les choix méthodologiques, conceptuels et pragmatiques possibles;
- la détermination des concepts impliqués;
- les aspects métacognitifs;

Ces dimensions ont été choisies pour deux raisons. D'abord, nous pensons que ces dernières permettent de bien marquer la différence entre chacun des niveaux. Notons que certaines dimensions peuvent parfois se recouper avec d'autres ou pourraient même être facilement fusionnées. Cela pourrait être le cas des deux premières dimensions si on considère, par exemple, que la présence ou l'absence d'une liste de matériel peut être en soi un indice méthodologique. Cependant, nous avons distingué ces dimensions par souci de clarté. Ensuite, ces dimensions ont été choisies car elles caractérisent, en quelque sorte, le niveau de complexité de chacun des niveaux. Par exemple, l'identification d'un élément de méthode peut, pour certains élèves, les aider à poursuivre leur progression jusqu'à l'obtention d'une solution satisfaisante. L'ajout d'un tel indice ou le retrait de ce dernier peut considérablement affecter le niveau de complexité de la situation d'évaluation. Puisque les situations d'évaluation que nous avons construites se caractérisent par un degré de difficulté décroissant, l'identification et la prise en compte de ces dimensions nous apparaissait incontournable. Examinons maintenant chacune de ces dimensions en détail.

Tableau IV  
Principales dimensions caractérisant chacun des niveaux pour une situation d'évaluation donnée.

Dimension	Niveau 1 (Compétence)	Niveau 2 (Compétence dirigée)	Niveau 3 (Habilité)
1. Les indices méthodologiques permettant de réaliser l'investigation.	Présence d'un cahier de l'élève comportant les étapes génériques qu'il devra réaliser.	Présence d'un cahier de l'élève comportant les étapes génériques qu'il devra réaliser ainsi que quelques questions permettant de baliser la démarche de l'élève.	Présence d'un cahier de l'élève avec les étapes détaillées qu'il devra réaliser.
2. Le matériel fourni aux élèves.	Aucune liste de matériel n'est fournie dans le cahier de l'élève. Le matériel disponible est simplement rendu disponible à chaque poste de travail. En plus des éléments indispensables à la réalisation de l'expérience, on retrouve aussi du matériel superflu.	Présence d'une liste comportant les éléments indispensables dans le cahier de l'élève afin de réaliser au moins une démarche satisfaisante. Le matériel superflu n'est pas présenté à l'élève.	Présence d'une liste complète permettant d'arriver directement à une solution (induite dans le problème soumis à l'élève).
3. La latitude laissée aux élèves concernant les choix à exercer pour réaliser l'investigation.	L'élève peut exercer des choix sur le matériel employé, la méthode pour réaliser l'investigation et les variables étudiées en respectant certaines balises.	L'élève peut exercer des choix parmi ceux disponibles concernant le matériel employé, la méthode pour réaliser l'investigation et les variables étudiées. Les balises sont plus serrées qu'au niveau précédent.	L'élève a très peu de choix à réaliser concernant le matériel, la méthode, les variables investiguées. Tout lui est fourni et il doit simplement suivre la procédure.

4. Les indices conceptuels offerts.	L'élève doit mobiliser lui-même tous les concepts nécessaires à l'investigation.	Les principaux concepts sont rappelés à l'élève mais sans les associer à l'une ou l'autre des parties de la démarche d'investigation.	Les concepts sont rappelés à l'élève et ils sont associés correctement à chacune des parties de la démarche d'investigation.
5. Les aspects métacognitifs liés à la démarche scientifique de l'élève.	L'élève doit lui-même assurer la régulation de sa démarche et du produit obtenu.	Des indices (ex. questions à la fin de chacune des grandes étapes de la démarche) rappellent à l'élève d'assurer une régulation au niveau de sa démarche et du produit obtenu.	À chaque étape de la démarche, des questions systématiques sont posées à l'élève afin d'assurer une régulation au niveau de la démarche et du produit obtenu.

---

### Première dimension : Les indices méthodologiques

Au regard de sa démarche, un élève compétent devrait normalement être en mesure de la concevoir, de la réaliser et d'en rendre compte. Aussi, il devrait être capable de réaliser ceci en contexte d'autonomie entière. À ce sujet, un élève compétent ne devrait pas avoir besoin d'indices méthodologiques afin d'accomplir la tâche qui lui est dévolue. De fait, la situation de compétence de niveau 1 n'induit aucun indice méthodologique à l'élève et vise à vérifier si celui-ci est capable seul de réaliser une investigation scientifique. La situation de compétence dirigée de niveau 2, quant à elle, présente les grandes étapes par lesquelles l'élève devrait normalement passer pour réaliser la situation sans en induire une de façon particulière. Par exemple, on identifiera les sous-étapes liées à chacune des grandes étapes de la démarche. Concrètement, dans le cas de la première étape (cerner le problème), on pourra fournir quelques questions génériques visant à l'aider à cerner le problème (ex. Quel est le but de l'expérimentation ? Quelles sont les données importantes du problème ? Etc.). Dans le cas de la situation de niveau 3, qui vise à recueillir des informations sur les habiletés des élèves, une démarche méthodologique sera privilégiée par le concepteur de l'épreuve et une série de questions judicieusement choisies guideront de façon précise la démarche de l'élève. Pour tous les niveaux, nous avons volontairement omis de donner des indices inutiles ou qui rendaient la situation ambiguë.

### Deuxième dimension : La liste du matériel accessible

Le matériel de laboratoire rendu disponible aux élèves a généralement un impact majeur sur la difficulté de la tâche à accomplir dans le cadre d'une investigation scientifique de type *hands-on*. L'ajout de matériel superflu ou l'omission de matériel indispensable peut considérablement influencer la démarche de résolution de problème que l'élève mettra de l'avant. Ainsi, un élève compétent devrait normalement être en mesure d'identifier le matériel utile parmi tout ce qui est mis à sa disposition. C'est pourquoi les situations de niveau 1 présentent une liste de matériel incomplète ou comportant des éléments superflus. Idéalement, nous aurions souhaité ne donner aucun indice quant à la liste de matériel mais des contraintes techniques nous obligent à réaliser un échantillonnage du matériel de

laboratoire *a priori*. En ce qui concerne les situations de niveau 2, elles comportent une liste de matériel constituée exclusivement de matériels indispensables afin de réaliser une des démarches possibles pour la résolution du problème. Les situations de niveau 3 présentent, quant à elles, une liste exhaustive permettant de réaliser normalement une seule démarche et d'arriver à une seule solution.

### Troisième dimension : Les choix méthodologiques, conceptuels et pragmatiques possibles

Cette dimension concerne la notion de « choix » qui caractérise les problèmes ouverts. Compte tenu de la problématique de l'évaluation des apprentissages en première secondaire (âge des élèves, contraintes techniques, etc.), nous avons convenu de laisser aux élèves la possibilité de faire des choix sur trois aspects bien précis. D'abord, au niveau 1 l'élève peut faire certains choix balisés en ce qui concerne les variables investiguées, la méthode employée pour réaliser l'investigation et le matériel qui servira à la réaliser. Au niveau 2, l'élève peut encore exercer certains choix mais cette fois-ci à l'intérieur de balises beaucoup plus serrées. Par exemple, la variété des instruments de mesure serait beaucoup moins importante pour les tâches de niveau 2 que celles de niveau 3. En ce qui concerne les situations de niveau 3, l'élève n'a que très peu de choix à réaliser. Dans le cadre de l'activité qui lui est soumise, il n'y a généralement qu'une seule démarche ne comportant qu'une seule solution vraisemblable qu'il est possible d'obtenir au moyen d'un seul montage. En ce sens, les choix sont inexistants et inutiles pour l'obtention d'une solution satisfaisante.

### Quatrième dimension : La détermination des concepts impliqués

Un élève compétent devrait non seulement maîtriser les contenus disciplinaires liés à l'investigation scientifique mais également être en mesure de les mobiliser à bon escient dans le cadre d'une problématique qu'on lui soumet. C'est pourquoi les situations de niveau 1 à 3 fournissent de plus en plus d'indices sur tous les aspects conceptuels pertinents liés à la résolution du problème. Ainsi, les situations de niveau 1 ne fournissent pratiquement aucun indice alors que les situations de niveau 3 rappellent aux élèves les

contenus importants qui doivent être traités et elles fournissent également certains indices sur la façon de les traiter en vue de résoudre le problème.

#### Cinquième dimension : Les aspects métacognitifs

Le caractère itératif lié aux démarches caractérisant la compétence 1 devrait normalement être observé chez les élèves compétents. Spontanément, ils devraient réaliser – lorsque cela est pertinent – des retours en arrière afin de faire des vérifications ou des contre vérifications au sujet autant de leur démarche que de leurs résultats. À cet égard, ces élèves n'ont donc pas besoin de se voir proposer des indices. Pour les élèves manifestant un degré de maîtrise moins élevé, il devient alors pertinent de les aiguiller en leur proposant des rappels au sujet de leur démarche et des résultats qu'ils obtiennent. Les situations de niveau 2 et 3 proposent progressivement des rappels de ce genre. Dans le cadre des situations de niveau 2, il s'agit de questions générales soumises aux endroits névralgiques de leur démarche. En ce qui concerne les situations de niveau 3, les questions sont à la fois plus spécifiques et plus fréquentes.

#### **3.1.3 Description des situations d'évaluation**

Nous présentons ici une description sommaire de chacune des situations d'évaluation que nous avons construites. L'appendice D présente en détail chacune des neuf situations (SE1-1, SE1-2, SE1-3, SE2-1, SE2-2, SE2-3, SE3-1, SE3-2, SE3-3).

La première situation (SE1) s'intitule « La tasse de chocolat chaud ». Cette situation est une adaptation d'une idée de situation d'apprentissage recueillie dans le document de mise en œuvre du programme d'études « Science de la nature, secondaire 1 » du Manitoba (Éducation Formation professionnelle et jeunesse Manitoba, 2001, p. 0.08). Cette situation fait appel aux concepts de l'univers matériel et consiste, en résumé, à choisir un contenant de remplacement aux traditionnels verres en polystyrène et d'isoler ce contenant afin de minimiser les pertes thermiques.

La seconde situation (SE2) se nomme « La balle de tennis ». Cette dernière est aussi une adaptation d'un problème répertorié, entre autres, par Gott (1987) et par le *Maine Department of Education (2001)*. Pour cette tâche, il s'agit d'identifier une variable qui peut affecter le rebond d'une balle de tennis et de concevoir une expérience permettant de mesurer l'effet de la variable choisie sur le rebondissement.

La troisième situation (SE3) est une situation originale qui s'appelle « les lombrics ». Cette dernière se rapporte à l'univers vivant du Programme de formation. Dans le cas de cette dernière, il s'agit d'identifier une variable qui peut affecter la présence des lombrics dans un milieu donné et d'observer si cette variable affecte ou non la présence de ces derniers.

### ***3.2 Validation des situations d'évaluation***

Le processus de validation que nous avons réalisé est composé de deux étapes. D'abord, nous avons effectué une validation de contenu auprès d'experts<sup>25</sup> de différents domaines (mesure et évaluation, didactique des sciences et de la technologie, conseillers pédagogiques, rédacteurs du programme de science et technologie). Par la suite, nous avons procédé à une pré expérimentation en situation réelle. Les sections qui suivent présentent les détails liés à chacune de ces étapes.

#### **3.2.1 Validation de contenu des situations d'évaluation**

Parvenir à déterminer si une situation pédagogique est conforme aux fondements d'un programme d'études est une opération pour le moins difficile. Pour certains, une situation d'évaluation permettra de rendre compte de telle compétence alors que pour d'autres, la même situation permettra de rendre compte de telles habiletés et non d'une compétence. Or, pour les fins de cette recherche, nous devons nous assurer que les situations d'évaluation produites permettaient véritablement aux élèves de faire la démonstration de leur niveau de compétence et non uniquement de leur niveau d'habiletés ou de

---

<sup>25</sup> La liste des experts est présentée à l'appendice A.

connaissances. Pour les besoins de la cause, nous nous sommes référés à cinq experts qui avaient à émettre leur avis sur la qualité des situations d'évaluation produites à la lumière des treize critères suivants :

1. La mise en situation est signifiante
2. La mise en situation présente un défi réaliste et des contraintes
3. La situation fait appel à différentes disciplines
4. La situation permet à l'élève de faire des choix à différentes étapes de la démarche
5. La situation permet une découverte de la part de l'élève
6. La situation permet le développement d'une ou de plusieurs compétences
7. La situation peut-être facilement enrichie afin de favoriser la différenciation des apprentissages
8. La situation exige que l'élève fasse montre de son autonomie
9. La situation nécessite que l'élève soit intellectuellement actif
10. La situation est complexe
11. La situation offre des attentes claires
12. La situation est bien formulée et intelligible pour un élève de première secondaire
13. Les critères d'évaluation sont pertinents

Nous avons également demandé à chacun des experts de se prononcer par écrit sur chacun des critères au moyen d'une échelle de Likert en 4 points (totalement en désaccord, en désaccord, d'accord, totalement en accord) présentée à l'appendice B. Les experts ont été également invités à émettre leurs commentaires sur chacun des critères et, au besoin, de proposer des suggestions permettant d'améliorer chacune des situations d'évaluation.

Ces critères correspondaient à une adaptation et une bonification de la proposition de Millman et Greene (1993, p. 354) au sujet de l'évaluation d'items servant à bâtir un test de rendement scolaire. Bien que le contexte soit passablement différent (items classiques vs situations de performance), leur proposition concernant les critères demeure intéressante pour juger de la qualité des situations d'évaluation produites. C'est pourquoi nous avons retenu ces critères et nous en avons ajouté un certain nombre afin de compléter et de bonifier la qualité du jugement au regard des situations d'évaluation produites.

En terminant, mentionnons les principaux aspects concernant les modalités pratiques liées à cette validation de contenu. Dans un premier temps, nous avons fait parvenir tous les documents pertinents (situations d'évaluation, canevas de la SE, outil d'évaluation, etc.)

aux experts par courrier électronique. Par la suite, nous avons convenu d'un rendez-vous avec chacun d'eux afin de leur expliquer les objectifs et les grandes lignes de cette recherche pour qu'ils soient en mesure de bien situer les instruments de mesure dans leur contexte. Les experts ont ensuite eu un temps suffisant pour discuter de chacune des situations d'évaluation au moyen des outils que nous avons mis à leur disposition. Quand les experts ont eu terminé leur travail, nous les avons tous consulté, sur une base individuelle et pendant environ une heure, afin de recueillir leurs commentaires et leurs impressions et ainsi nous assurer de bien saisir leurs points de vue.

### 3.2.2 Pré-expérimentation

Le processus de validation que nous avons mis en place s'appuyait sur les suggestions de Solano-Flores et Shavelson (1997, p. 23) telles qu'illustrées à la figure 7 qui suit.

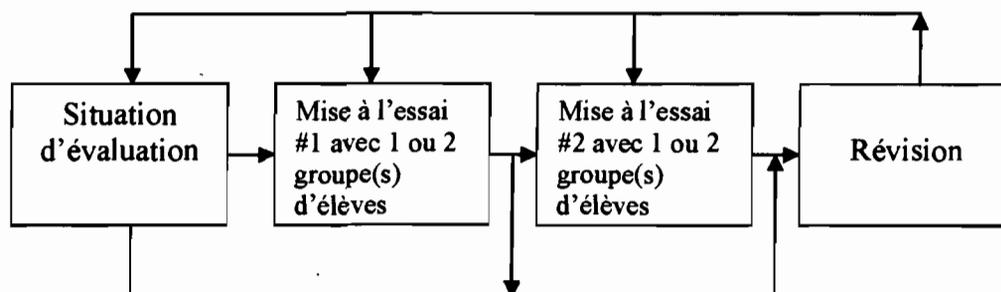


Figure 7

Processus de développement d'une situation d'évaluation de performance (Solano-Flores & Shavelson, 1997).

Ces derniers suggèrent de procéder à deux mises à l'essai avec un ou deux élèves lors de la première mise à l'essai et 15 à 20 lors de la seconde mise à l'essai. En ce qui nous concerne, nous souhaitons augmenter la taille de ces échantillons compte tenu du contexte dans lequel nous devons œuvrer (nouveaux programmes, situations d'évaluation nouvellement créées et jamais expérimentées, etc.). Ainsi, les mises à l'essai ont été effectuées auprès de quatre groupes d'élèves (n=102). Cet échantillon nous a permis de recueillir des données qui ont eu pour effet d'améliorer considérablement les situations

d'évaluation et le contexte général entourant la passation de ces dernières (consignes aux élèves et aux enseignants, choix du matériel de laboratoire, etc.)

La première mise à l'essai visait à corriger les aspects problématiques liés à une première passation et à obtenir des données sur le déroulement de chacune des situations. Lors de cette première mise à l'essai, nous avons essayé de recueillir des informations sur des aspects qui étaient pour nous encore nébuleux. Voici une liste présentant quelques exemples d'aspects que nous avons surveillés avec attention lors de cette première mise à l'essai :

- Est-ce que la formulation des mandats (des problèmes) est adéquate pour des élèves de première secondaire ?
- Quelles sont les difficultés auxquelles les enseignants pourraient être confrontés concernant le matériel de laboratoire ?
- Quel est le temps minimum, maximum et moyen pour la passation de chacune des situations d'évaluation ?
- Quelles sont les meilleures consignes à donner aux élèves à différents moments névralgiques de chacune des SE ?
- Est-ce que les règles de sécurité que nous présentons aux élèves sont exhaustives et pertinentes ?
- Etc.

Un enseignant s'était déjà porté volontaire afin de réaliser cette première mise à l'essai auprès de trois à quatre groupes d'élèves de première secondaire. Cette mise à l'essai s'est déroulée en avril 2005 en présence du chercheur qui a alors effectué des visites en classes afin de réaliser des observations sur le déroulement de la passation des épreuves. Fort de l'expérience acquise lors de la première mise à l'essai, nous avons grandement ajusté les situations d'évaluation afin de les bonifier en enlevant les irritants et les ambiguïtés. Nous avons alors procédé à une seconde validation avec un autre enseignant de la même école afin de tester les versions finales des situations d'évaluation. Cette seconde validation a eu lieu avec deux groupes d'élèves (n=41) sans la présence du chercheur. Les résultats obtenus, autant au niveau des modalités de passation qu'au niveau des résultats des élèves, nous sont apparus suffisamment prometteurs pour procéder à l'essai avec notre échantillon expérimental.

### **3.2.3 Modalités d'expérimentation et échantillonnage**

Dans cette section, nous exposerons les aspects pratiques associés à l'expérimentation que nous avons réalisée. Dans un premier temps, nous décrirons les détails liés à la collecte des données. Dans un second temps, nous développerons sur l'échantillonnage des sujets qui ont participé à cette étude. Enfin, nous terminerons en discutant des consignes que nous avons données aux enseignants afin que l'expérimentation se déroule de la façon la plus uniforme possible dans chacune des écoles.

#### **3.2.3.1 Collecte des données**

La collecte des données s'est déroulée durant les mois de mai et juin 2006 et elle s'est échelonnée sur une période de 3 à 4 semaines (de 5 à 8 périodes de 75 minutes). Préalablement, nous avons rencontré les enseignants à deux reprises (l'équivalent de deux demi-journées). La première rencontre avait pour objectif d'informer les enseignants à propos de la recherche et de susciter leur engagement à y participer. La seconde rencontre avait pour but, quant à elle, de présenter les documents pertinents à l'administration des épreuves et de répondre aux questions. Lors de cette rencontre, nous avons invité les enseignants, les techniciens de laboratoire ainsi que les conseillers pédagogiques. Toutes ces personnes étaient invitées à communiquer avec le chercheur si la moindre question surgissait. Nous avons également produit une vidéo (disponible à l'appendice E) dans laquelle nous présentons l'ensemble des consignes liées à la recherche ainsi que certaines modalités pratiques associées à l'administration des situations d'évaluation. Toujours dans le but d'assurer le maximum d'uniformité quant à la passation des situations d'évaluation, nous avons également construit un site web dans lequel nous avons colligé des informations plus « pratiques » (ex. échancier, procédure pour la remise des cahiers d'élève, etc.) afin de faciliter la récupération des données de recherche. Les conseillers pédagogiques de chacune des commissions scolaires ont été particulièrement mis à contribution lors de cette recherche. Nous les avons impliqués dès le début du processus d'échantillonnage afin qu'ils deviennent la courroie de transmission entre le chercheur principal et les enseignants participants. En cas de difficulté, les enseignants étaient invités

à communiquer avec le chercheur ou le conseiller pédagogique afin de trouver des solutions aux problèmes rencontrés. En ce sens, la contribution des conseillers pédagogiques a été fort importante afin d'assurer la cohésion entre les enseignants qui ont administré nos épreuves.

### **3.2.3.2 Échantillonnage**

Millman et Greene (1993, p. 357) considèrent que la taille de l'échantillon expérimental en éducation devrait être déterminée à la lumière des trois facteurs suivants : (1) la façon dont les sujets sont choisis, (2) la quantité d'information et la précision attendue au regard des analyses et (3) le choix de la méthode d'analyse des items. On peut résumer la proposition de ces auteurs en indiquant que généralement, on doit recourir à plus de sujets si :

- ces derniers ne sont pas suffisamment représentatifs de la population;
- on souhaite réaliser des analyses à l'intérieur de plusieurs sous-échantillons;
- on désire obtenir la meilleure précision possible;
- le modèle statistique utilisé est sophistiqué (ex. analyses paramétriques ou analyses non paramétriques).

Puisque notre recherche était à caractère exploratoire, plusieurs des aspects précédents nous encourageaient à expérimenter les situations d'évaluation sur un nombre relativement important de sujets. En effet, nous souhaitions expérimenter notre modèle auprès de plusieurs clientèles d'élèves (élèves appartenant à un programme particulier (ex. programme d'éducation internationale), élèves appartenant à des classes dites « ordinaires », élèves en difficulté, etc.) afin d'avoir un portrait d'ensemble relativement représentatif. De plus, nous souhaitions également analyser les données au regard de plusieurs variables contextuelles ce qui supposait un échantillon de taille importante. Ainsi, à la lumière de la proposition précédente nous avons administré nos situations d'évaluation à un échantillon de 560 élèves issus de 49 groupes-classes. Un total de 22 enseignants, en provenance de cinq commissions scolaires et de 14 écoles différentes, ont procédé à l'administration des situations d'évaluation. La taille de l'échantillon d'élèves est également suffisamment importante pour nous permettre de réaliser des analyses

statistiques (ex. corrélations, analyses de variance, test du khi-carré, etc.) de base. La taille de notre échantillon est comparable à la taille d'échantillon utilisée dans des recherches semblables à la nôtre (Rey et al., 2003; Shavelson et al., 1991).

### **3.2.3.3 Les consignes données aux enseignants lors de l'expérimentation**

Comme le souligne avec justesse Scallon (2004, p. 158), le contexte entourant l'évaluation certificative implique un resserrement des différentes modalités liées à l'évaluation. Certains auteurs comme Legendre (2001) affirment que les situations d'évaluation peuvent se confondre aux situations d'apprentissage. À ce sujet, deux remarques. D'abord, dans le cadre de cette recherche, nous souscrivons d'emblée aux remarques de Scallon. En effet, les enjeux liés à l'évaluation certificative sont suffisamment importants pour s'assurer de mettre en œuvre toutes les conditions afin de maximiser la validité des inférences qui seront réalisées. Ensuite, comme l'indique Lasnier (2000, p. 222) « les conditions de l'évaluation certificative de fin de cycle ne doivent pas présenter une rupture avec l'évaluation en cours d'apprentissage ». Dans ce contexte, nous croyons qu'un compromis entre la proposition de Scallon et celle de Lasnier peut être une solution intéressante. Dans le cadre de cette recherche, nous avons tenté de réunir le maximum des conditions d'évaluation « en cours d'apprentissage » tout en s'assurant que ces dernières soient suffisamment rigoureuses pour assurer la validité des inférences qui devaient être réalisées.

Cette prise de position avait évidemment un impact sur les conditions de passation des épreuves. Premièrement, puisqu'il fallait porter le jugement sur l'atteinte des compétences pour chacun des élèves, nous avons convenu que chaque épreuve soit administrée individuellement. Bien que cette modalité puisse différer des situations d'apprentissage administrées en cours d'apprentissage, il pouvait difficilement en être autrement dans la mesure où l'on souhaite véritablement porter le jugement sur chacun des élèves. De plus, les cahiers d'élève devaient être remis à l'enseignant à la fin de chaque période afin d'éviter le plus possible les risques de contamination.

Puisque les épreuves visaient à évaluer le degré d'atteinte des compétences 1 et 3, tous les élèves se sont vus administrer, au départ, une situation d'évaluation de niveau 1. Les élèves qui montraient des signes évidents de difficulté se voyaient alors administrer l'épreuve de niveau 2. Puisque l'expression « signes évidents de difficulté » pouvait être arbitraire, nous avons convenu de deux balises pour l'administration d'une épreuve de niveau 2. Notons que ces balises ne servaient pas à remplacer le jugement professionnel des enseignants. Ces derniers demeuraient, sans aucun doute, les plus compétents pour juger si un élève était en difficulté ou pas. La première balise était basée sur un facteur « temps » alors que la seconde s'appuyait sur un facteur « réussite ».

#### Première balise

Un élève se voit administrer une épreuve de niveau inférieur (ex. niveau 1 à niveau 2) si au terme de la première période (75 minutes) il n'arrive pas – ou il arrive mal – à reformuler le problème et à identifier les principales variables importantes liées au problème.

#### Deuxième balise

Un élève se voit administrer une épreuve de niveau 2 si le résultat prévisible est en deçà du seuil de réussite.

De façon analogue, les balises précédentes s'appliquaient pour l'administration des épreuves de niveau 3 pour les élèves en difficulté lors de l'épreuve de niveau 2.

Dans tous les cas litigieux où l'enseignant avait un doute sur la capacité de l'élève à réussir la situation d'évaluation, le mot d'ordre était de laisser la chance maximale à l'élève de faire la démonstration de sa compétence.

Puisqu'il était impossible de prévoir tous les scénarios possibles, nous demandions aux enseignants de colliger dans un carnet les « incidents critiques » pouvant survenir à un moment ou à un autre de l'expérimentation et qui pouvaient affecter la qualité des données. Aussi, nous exigeons des enseignants qu'ils assignent au hasard l'une ou l'autre des trois situations d'évaluation à chacun des élèves.

### ***3.3 Variables investiguées***

Afin de répondre à nos questions de recherche, trois variables principales et trois variables contextuelles ont été analysées.

#### **3.3.1 Les variables principales**

Les deux variables principales que nous souhaitons investiguer concernaient le niveau de maîtrise de la compétence par les élèves et l'effet du contexte disciplinaire sur le niveau de maîtrise de la compétence.

##### **3.3.1.1 Le niveau de maîtrise de la compétence**

La première variable principale concerne le niveau général de maîtrise de la première compétence scientifique en ne tenant pas compte du contexte disciplinaire. Il s'agit d'identifier le nombre d'élèves en mesure de réussir chacun des niveaux (ex. SE1-1, SE1-2, SE1-3) pour chacune des situations d'évaluation proposées. Aussi, nous cherchions à connaître plus précisément le score des élèves à chacun des critères en notant la performance de l'élève sur une échelle de type Likert à cinq niveaux et ce pour les douze critères de correction retenus. Nous présenterons plus en détail cette grille de correction à la section 3.4.1. Enfin, nous tâcherons d'analyser les différentes séquences de passation des épreuves par lesquelles les élèves sont passés et qui sont illustrées à la figure 4 du premier chapitre.

##### **3.3.1.2 L'effet du contexte disciplinaire sur la maîtrise de la compétence**

La seconde variable principale concerne l'effet du contexte disciplinaire dans le cadre d'une investigation scientifique. Nous souhaitons analyser les scores obtenus par les

élèves afin de voir s'il y a ou pas des différences selon que la situation d'évaluation possède une majeure en biologie, en physique ou aucune majeure particulière. Pour ce faire, nous avons comparé les résultats obtenus par les élèves pour chacune des situations d'évaluation.

### **3.3.2 Les variables contextuelles**

Nous avons retenu trois variables contextuelles qui permettaient d'obtenir de l'information sur des aspects intéressants que nous pensons liés à la performance des élèves. Ces variables sont les suivantes : l'expérience de l'enseignant, la formation initiale de l'enseignant et le sexe des élèves.

En ce qui concerne les deux dernières variables, les données étaient accessibles facilement. Cependant, les données liées à la première variable contextuelle étaient beaucoup plus difficiles à obtenir. Compte tenu de la taille de l'échantillon des enseignants qui ont participé, nous avons été en mesure d'obtenir trois catégories d'enseignants. La section suivante présente les aspects liés à chacune de ces variables contextuelles.

#### **3.3.2.1 L'expérience de l'enseignant**

Nous n'avons pas répertorié beaucoup de recherches qui s'intéressaient à l'influence de l'expérience des enseignants sur l'appropriation d'un nouveau modèle d'évaluation. Pourtant, nous pensions que cette variable pouvait avoir une influence sur le rendement des élèves dans un tel contexte. Normalement, des enseignants plus expérimentés devraient être en mesure de mieux gérer les changements que nous proposons au regard de l'évaluation. C'est pourquoi nous avons essayé de tenir compte de cette variable. De fait, nous avons classé les enseignants qui ont participé à cette étude en trois catégories : ceux ayant entre 0 et cinq années d'expérience, ceux ayant entre six et dix années d'expérience et ceux ayant onze années d'expérience et plus.

### **3.3.2.2 La formation initiale de l'enseignant**

On pouvait penser que la formation des enseignants à l'égard du nouveau programme de science et technologie pouvait avoir un impact sur le niveau de réussite des élèves. De fait, un modèle d'évaluation adéquat qui serait expérimenté auprès d'enseignants mal ou peu formés donnerait possiblement de bien piètres résultats. Ainsi, nous avons séparé les résultats des élèves en deux groupes. D'une part, nous avons regroupé les résultats des élèves ayant un enseignant formé en enseignement des sciences. Dans cette catégorie, nous avons inclus les enseignants ayant soit un baccalauréat en enseignement des sciences ou un baccalauréat dans une discipline scientifique (ex. chimie, physique, biologie, etc.) jumelé ou non à une formation en pédagogie (ex. certificat en pédagogie). D'autre part, nous avons réuni les résultats des élèves n'ayant pas un enseignant formé en enseignement des sciences. Nous avons ensuite comparé les deux grandes catégories ainsi produites.

### **3.3.2.3 Le sexe des élèves**

À notre connaissance, aucune étude ne s'est intéressée aux différences possibles entre les garçons et les filles en ce qui concerne le niveau de maîtrise de compétences expérimentales au niveau secondaire. *A priori*, nous ne supposons pas qu'il y aura de différences importantes entre les sexes mais puisque aucune recherche connue ne s'y était intéressée, nous espérons investiguer cette variable dans le contexte de cette recherche.

## **3.4 Traitement et analyse des données**

Comme nous l'avons mentionné précédemment, nous avons recueilli des données sur deux types de variables : les variables principales et les variables contextuelles. Les données concernant les variables principales ont été recueillies directement au moyen des situations d'évaluation. En ce qui concerne les variables contextuelles, nous les avons recueillies à l'aide d'un court questionnaire administré aux enseignants. Ce questionnaire est reproduit à l'appendice F.

Au départ, nous avons envisagé différentes stratégies d'exploration des données. Dans plusieurs cas, nos données nous permettaient de recourir à des outils paramétriques (ex. ANOVA) mais dans d'autres, nos données ne rencontraient pas les conditions d'utilisation de tels outils statistiques comme l'homogénéité des variances, des échantillons de tailles égales ou la condition de normalité. Finalement, compte tenu de la situation que nous venons de décrire et du contexte exploratoire de la recherche nous avons choisi de recourir à des statistiques descriptives et inférentielles de type non paramétrique. Ces tests imposent des conditions d'utilisation moins strictes ce qui nous apparaissait satisfaisant dans les circonstances. Aussi, puisqu'il n'était pas possible de recourir à des statistiques multivariées, nous avons utilisé, entre autres pour la comparaison des résultats entre les situations d'évaluation, des statistiques univariées répétées. Nous sommes conscient que cette stratégie n'est pas idéale puisque le fait de multiplier les comparaisons pourrait requérir un niveau de signification plus élevé par exemple 0,001 plutôt que 0,05. De fait, cette situation peut entraîner une erreur de type II.

À plusieurs occasions, nous avons aussi choisi de présenter nos données à l'aide de la moyenne. Bien que nous soyons conscient des limites de cette statistique, cette mesure de tendance centrale nous apparaissait appropriée afin de comparer les tendances dans les résultats des élèves à chacune des situations d'évaluation. La médiane ou les écarts interquartiles nous semblaient moins utiles pour réaliser ces comparaisons.

Nous avons également fait le choix de traiter les données brutes. Nous avons réalisé un très grand nombre d'analyses avec le modèle *rating scale* de Rasch afin de modéliser nos données brutes pour tenir compte des problèmes de mesure que peut engendrer le traitement des données brutes. Cependant, pour comparer les situations d'évaluation entre elles, il aurait fallu que certains élèves réalisent au moins deux situations d'évaluation afin de se servir de ces données pour réaliser l'ancrage des critères d'évaluation. Dans le cadre de cette recherche, les contraintes liées à la collecte des données ne nous ont pas permis d'administrer plus d'une situation d'évaluation par élève. Bien que nous soyons sensible aux problèmes de mesure liés à l'utilisation des scores bruts, nous avons quand même

choisi de réaliser les analyses avec ces derniers, entre autres, par souci de comparaison des situations d'évaluation entre elles.

#### **3.4.1 La correction des situations d'évaluation**

La correction des épreuves et l'analyse des données ont été réalisées par l'étudiant-chercheur. L'étape de la correction et de l'entrée des données s'est échelonnée sur environ cinq semaines. Les situations d'évaluation ont toutes été corrigées au moyen d'une grille d'appréciation en cinq niveaux dont le seuil est situé au niveau deux. Nous avons donc comptabilisé les scores des élèves sur une échelle dont le score maximum se situait à 4,00. La description de chacun d'eux apparaît dans le tableau ci-dessous :

Tableau V  
Échelle de notation générale des situations d'évaluation.

Critères d'évaluation du programme d'études	Critères d'évaluation généraux de la SE	Critères d'évaluation spécifiques de la SE	Échelle de notation				
			0	1	2	3	4
Représentation adéquate de la situation	Première étape : Je prépare mon expérience	1. Reformulation du problème					
		2. Identification de la variable à investiguer					
3. Élaboration d'une liste de matériel							
4. Formulation d'une hypothèse vérifiable et plausible							
5. Formulation et présentation des manipulations							
Élaboration d'une démarche pertinente pour la situation	Deuxième étape : Je réalise l'expérience	6. Consignation des données expérimentales et des observations					
		7. Identification des causes d'erreurs					
Mise en œuvre adéquate de la démarche	Troisième étape : J'analyse les résultats de mon expérience	8. Traitement des données expérimentales					
		9. Maîtrise des concepts, des règles et des techniques					
	Élaboration de conclusions, d'explications ou de solutions pertinentes	Quatrième étape : Je propose une solution au problème	10. Application des résultats expérimentaux dans la solution proposée				
Respect de la terminologie, des règles et des conventions propres à la science et à la technologie dans la production de messages			11. Présentation générale				
	12. Utilisation des symboles et des termes						

Une échelle à cinq niveaux nous est apparue nettement suffisante compte tenu des analyses que nous comptons réaliser. En effet, le but de cette recherche n'est pas de noter finement chacun des élèves mais plutôt de situer globalement les élèves par rapport à un certain nombre de critères. En ce sens, une échelle à cinq niveaux permet de dégager trois grands groupes d'élèves (les élèves en grande difficulté, les élèves en réussite et les élèves en très grande réussite) ce qui est suffisant pour les analyses que nous comptons réaliser. De plus, nous nous sommes arrimés aux échelles de niveaux de compétence proposées par le MELS (2006a). Ainsi, nous avons retenu une échelle en cinq points dont chacun des échelons correspond au descripteur indiqué au tableau suivant. Les descripteurs sont ceux proposés par le MELS (2006a, p. 7). Le seuil que nous avons retenu (fixé au niveau 2) correspond aussi à celui du MELS.

**Tableau VI**  
Description de chacun des échelons de la grille de notation.

Échelon	Description
0	Nettement en deçà des attentes
1	En deçà des exigences
2 (seuil)	Satisfait minimalement aux exigences
3	Satisfait clairement aux exigences
4	Dépasse les exigences

### **3.4.2 Contrôle de la qualité concernant la fiabilité de la notation**

Bien que les situations d'évaluation aient été corrigées par une seule personne, il demeure nécessaire de s'assurer de la fiabilité du jugement du correcteur. Pour ce faire, nous nous

sommes inspirés des recommandations de Huberman et Miles (1991) sur la méthodologie à employer pour le codage de données qualitatives. Ils définissent la fiabilité comme étant le rapport du nombre d'accords sur le nombre total d'accords et de désaccords. Nous avons procédé à une vérification de la fiabilité sur environ 10% des copies d'élève et nous avons obtenu un niveau d'accord de 90,3%. Cette valeur est concordante avec les suggestions de Huberman et Miles qui la fixe à 90%.

La plupart des études que nous avons consultées indiquent qu'il est possible d'obtenir une fiabilité intra juge ou inter juges satisfaisante lors de la notation de situations de performance en science. Shavelson et Baxter (1992) ont étudié l'accord inter juges lorsque les enseignants devaient noter les élèves alors que ces derniers accomplissaient la tâche (évaluation « live » des élèves). Le taux d'accord (0,80) a été satisfaisant. Shavelson a également entrepris une étude plus formelle en ayant recours à la théorie de la généralisabilité pour mieux cerner les facettes ayant un impact sur l'erreur de mesure. Ils ont réalisé plusieurs analyses (ex.  $p \times r \times t \times o$ ,  $p \times r \times t$ ,  $p \times t \times m$ <sup>26</sup>) et les valeurs des interactions personnes-correcteur ( $p \times r$ ) et tâche-correcteur ( $t \times r$ ) sont nulles suggérant que l'effet correcteur est négligeable dans l'erreur de mesure. Selon ces auteurs, une seule personne bien formée serait suffisante pour assurer une correction adéquate dans des situations basées sur la performance en science et aussi en mathématique.

Klein *et al.* (1998) se sont aussi intéressés à la problématique de la fidélité de la correction de situations de performance. Dans l'une de leurs études, ils ont comparé l'accord inter juge, la fiabilité des scores et le temps de notation à trois situations de performance en science comportant trois dimensions (compréhension des concepts, performance en science et application des concepts en contexte). Les données ont été recueillies auprès de 168 élèves de cinquième année, 98 élèves de huitième année et 102 élèves de dixième année. Le but de cette étude était de comparer les résultats obtenus à l'aide de deux outils : une échelle analytique et une échelle holistique. Les résultats indiquent que les correcteurs obtiennent des résultats semblables pour une même méthode et ce, pour les trois échantillons analysés (élèves de cinquième, huitième et dixième année). Les corrélations

---

<sup>26</sup> p : person (sujet), r : rater (correcteur), t task (tâche), o : occasion (occasion), m : method (méthode)

inter correcteurs sont plus élevées pour la méthode analytique que la méthode holistique. Une étude de généralisabilité a permis de constater que de recourir à deux ou plusieurs juges ne ferait que très peu augmenter la fiabilité des scores.

### **3.4.3 Logiciels utilisés pour l'analyse et la présentation des données**

Les statistiques ont été obtenues avec l'aide du logiciel SPSS version 15 (2006). Les logiciels Excel (2003) et SPSS ont été utilisés pour effectuer les représentations graphiques.

Il ne suffit pas de « voir » un objet jusque-là invisible pour le transformer en objet d'analyse. Il faut encore qu'une théorie soit prête à l'accueillir.

François Jacob

Extrait de la « Logique du vivant »

## **Chapitre 4. Analyse des données**

---

### ***Introduction***

Ce chapitre est divisé en trois grandes parties. Dans un premier temps, nous présentons la description de notre échantillon. Dans un second temps, nous présentons les données issues des deux premières questions de recherche et qui concernent spécifiquement la performance du modèle d'évaluation que nous avons expérimenté. Dans un troisième temps, nous discutons de la robustesse du modèle au regard de certaines variables contextuelles qui sont associées à la troisième question de recherche et de ses sous questions. C'est dans le chapitre suivant que nous discuterons des résultats de l'analyse des données. Cependant, nous terminons chacune des sections du présent chapitre en dégageant quelques pistes de discussion. Ces pistes seront poursuivies au chapitre 5 alors que nous discuterons à fond des principaux résultats obtenus dans le cadre de cette recherche.

### ***Description de l'échantillon***

Afin de présenter de façon exhaustive l'échantillon retenu, nous discuterons, dans un premier temps, de l'échantillon relatif aux élèves. Dans un second temps, nous présenterons l'échantillon des enseignants qui ont participé à la collecte des données. Enfin, nous terminerons en décrivant la répartition des élèves ayant réalisé l'une ou l'autre des neuf versions des situations d'évaluation.

À la lecture du tableau VII, on constate que le groupe des 560 élèves ayant accompli les situations d'évaluation est composé à 44,1% de garçons et à 55,9% de filles. Ce résultat est étonnant puisque nous nous attendions à obtenir une proportion à peu près égale de

garçons et de filles. En effet, selon le MELS (2006b) , on retrouvait un total de 212 870 (49,1%) filles au secondaire et 220 731 (50,9%) garçons pour le secteur francophone des jeunes. En ce qui nous concerne cette disparité s'explique par le retour des formulaires de consentement à participer à l'étude.

La très grande majorité des élèves de notre échantillon (93,7%) ont entre 12 et 13 ans et n'ont donc pas de retard scolaire notable. Sur l'échantillon total (N=560), seulement 35 élèves sont âgés de 14 ou de 15 ans. Compte tenu de leur âge, ces élèves ont un retard scolaire d'au moins un an sur le cheminement « normal » des élèves de première secondaire.

**Tableau VII**  
Répartition des élèves en fonction de l'âge et du sexe.

Âge	Masculin		Féminin	
	Effectif	Pourcentage du total	Effectif	Pourcentage du total
12	69	12,3%	96	17,1%
13	164	29,3%	196	35,0%
14	12	2,1%	19	3,4%
15	2	0,4%	2	0,4%
Total	247	44,1%	313	55,9%

En survolant le tableau VIII, on apprend que plus de 65% des élèves (65,7%) proviennent d'une école dite « régulière ». Nous entendons par « régulière » une école qui ne réalise aucune opération de sélection à l'entrée et qui n'offre aucun cheminement de type élitiste. Dans notre échantillon, environ un élève sur dix (9,6%) est inscrit au programme d'éducation internationale (PEI) qui est, lui, un programme où les élèves sont sélectionnés. Une partie de notre échantillon (6,3%) est également constituée d'élèves inscrits dans un programme spécifique pour élèves en difficulté (volet alternatif). Une autre partie (12,7%) regroupe les élèves engagés dans un cheminement axé sur

l'autonomie des apprentissages (enseignement modulaire). Enfin, une trentaine d'élèves (5,7%) sont rattachés à un programme qui met l'accent sur les sciences et les techniques (PISICUI). Cependant, il n'existe aucune opération de sélection des élèves pour adhérer à ce programme particulier. Tous les élèves sont invités à s'y inscrire sur une base purement volontaire.

**Tableau VIII**

Répartition des élèves en fonction de la vocation attirée à leur école.

Vocation de l'école	Fréquence	Pourcentage
Aucune	368	65,7%
Programme d'éducation international (PEI)	54	9,6%
Volet alternatif	35	6,3%
Enseignement modulaire	71	12,7%
Programme d'intégration des sciences et de l'informatique en collaboration avec l'université et l'industrie (PISICUI)	32	5,7%
<b>Total</b>	<b>560</b>	<b>100,0%</b>

Quant à eux, les 22 enseignants ayant collaboré à la collecte des données étaient majoritairement des femmes (71,4%). En prenant connaissance du tableau IX qui suit, on constate que les enseignants peuvent être classés en trois grandes catégories d'égale effectif (33,3%) : ceux qui possèdent moins de 5 années d'expérience, ceux ayant entre 6 et 10 années d'expérience et finalement ceux ayant plus de 11 ans d'expérience. Il est à noter que les enseignants les plus expérimentés (11-35 ans) sont tous de sexe féminin.

Tableau IX  
Répartition des enseignants en fonction du sexe et de l'expérience.

Expérience (année)	Masculin		Féminin	
	Effectif	Pourcentage du total	Effectif	Pourcentage du total
0-5	3	14,3%	4	19,0%
6-10	3	14,3%	4	19,0%
11-15	0	0,0%	3	14,3%
16-35	0	0,0%	4	19,0%
<b>Total</b>	<b>6</b>	<b>28,6%</b>	<b>15</b>	<b>71,3%</b>

Notre échantillon n'est pas représentatif de la population au regard du sexe des enseignants. Selon les données du MELS (2005), on retrouvait 3138 enseignants en mathématique et science au secondaire en 2005 ce qui représente 47,5% des effectifs qui sont de sexe masculin. Par ailleurs, on dénombrait 3474 femmes qui enseignaient les mathématiques et les sciences ce qui représentait 52,5% des effectifs totaux.

Le tableau X nous indique également que près du tiers (31,8%) des enseignants possèdent une formation initiale disciplinaire en science (B.Sc.), un peu plus d'un enseignant sur cinq (22,7%) une formation en enseignement des sciences (B.Ed.) alors qu'un seul enseignant combine ces deux types de formation à savoir un baccalauréat en enseignement des sciences jumelé à un baccalauréat en science. On peut noter que près du tiers (31,8%) des enseignants n'ont de formation initiale ni en science ni en enseignement des sciences. Ce constat est particulièrement important à souligner. Il faut préciser que cette recherche s'est déroulée à une période où la pénurie d'enseignant qualifié en science était particulièrement importante. Malgré nos efforts, nous avons tenté d'obtenir des informations sur le nombre d'enseignants non légalement qualifiés en science et technologie mais en vain. Bien que cette situation ait été indépendante de notre volonté, nous devons quand même tenir compte de cet aspect dans nos résultats.

**Tableau X**  
**Répartition des enseignants en fonction de leur formation initiale.**

Formation initiale	Fréquence	Pourcentage
Baccalauréat en science (B.Sc.)	7	31,8%
Baccalauréat en enseignement des sciences (B.Ed.)	5	22,7%
Baccalauréat en science et en enseignement des sciences (B.Sc. + B.Ed.)	1	4,5%
Autre formation	7	31,8%
Valeur manquante	2	9,1%
<b>Total</b>	<b>22</b>	<b>100,0%</b>

Le tableau XI nous informe que près de 60% (59,1%) des enseignants ont reçu, sur une base volontaire, une ou plusieurs séance(s) de formation sur le renouveau pédagogique et sur l'approche par compétences en science et technologie. À la statistique précédente, il s'ajoute 13,6% des enseignants qui ont suivi non seulement les formations offertes par la commission scolaire mais qui se sont également investis dans divers projets de recherche et développement avec le ministère de l'Éducation, des loisirs et du sport. Dans notre échantillon, deux enseignants (9,1%) étaient inscrits dans un programme universitaire de deuxième cycle en didactique des sciences et de la technologie. Enfin, trois enseignants (13,6%) n'ont reçu aucune formation particulière sur le renouveau pédagogique et les nouveaux programmes formulés par compétences.

**Tableau XI**  
Répartition des enseignants en fonction de leur profil de formation continue.

Type de formation continue	Fréquence	Pourcentage
Participation volontaire à des formations offertes par la commission scolaire	13	59,1%
Participation volontaire à des formations offertes par la commission scolaire et participation à des projets de développement avec le MELS	3	13,6%
Inscription à un programme universitaire de deuxième cycle en didactique des sciences et de la technologie	2	9,1%
Aucune formation	3	13,6%
Valeur manquante	1	4,5%
<b>Total</b>	<b>22</b>	<b>100,0%</b>

Le tableau XII présente les différents profils de passation des épreuves évaluatives. Avant de poursuivre, une parenthèse s'impose afin de rappeler au lecteur la méthodologie liée à la passation des épreuves. On demandait aux enseignants d'administrer d'abord le niveau 1 à tous les élèves. Pour les élèves qui manifestaient des difficultés, il était indiqué de leur administrer alors le niveau 2. Les élèves qui étaient toujours en difficulté au niveau 2 se voyaient alors donner le niveau 3. Dans les faits, nous nous attendions à voir seulement trois profils de passation à savoir : niveau 1 seulement, niveaux 1 et 2 et niveaux 1, 2 et 3. Le tableau XII présente les profils de passation que nous avons identifiés lors de l'analyse des données. On remarque trois profils distincts de passation qui se démarquent particulièrement. Le premier profil, qui regroupe le plus d'effectifs (45,0%), est celui dont les élèves ont réalisé uniquement la situation de niveau 1 (SE1 : 92, SE2 : 77 et SE3 : 83). Le deuxième profil (20,4%) implique les élèves ayant réalisés les trois niveaux (SE1 : 24, SE2 : 59 et SE3 : 31). Enfin, le troisième profil de passation (28,6%) se rapporte aux élèves s'ayant vus offrir les situations de niveau 1 et 3 seulement (SE1 : 54, SE2 : 40 et SE3 : 66). Les profils de passation niveau 1 et 2 seulement (5,9%) et niveau 3

(0,2%) sont, quant à eux, négligeables. En comparaison avec ce qui était prévu, le profil #3 (N1N3) nous est apparu comme une surprise puisqu'il n'était pas suggéré aux enseignants. Plusieurs enseignants nous ont dit avoir voulu « comparer » les résultats de leurs élèves à une épreuve dite ouverte (niveau 1) avec une épreuve plus traditionnelle et fermée (niveau 3). C'est la principale raison que nous avons retenue pour expliquer ce constat. Après mûres réflexions, nous avons choisi de considérer ces résultats dans le cadre de notre étude. Nous avons considéré que l'hypothèse émise par les enseignants méritait d'être investiguée. Par ailleurs, on constate que la répartition des élèves pour les trois situations d'évaluation est pratiquement égale (SE1 : 189 élèves, SE2 : 186 élèves et SE3 : 185 élèves).

Tableau XII

Répartition des élèves ayant réalisé chacun des profils de passation des situations d'évaluation.

Situation d'évaluation réalisée (SE)	1 N (%)	1-2 N (%)	1-2-3 N (%)	1-3 N (%)	3 N (%)	Total N (%)
1	92 (16,4%)	19 (3,4%)	24 (4,3%)	54 (9,6%)	0 (0,0%)	189 (33,8%)
2	77 (13,8%)	10 (1,8%)	59 (10,5%)	40 (7,1%)	0 (0,0%)	186 (33,2%)
3	83 (14,8%)	4 (0,7%)	31 (5,5%)	66 (11,8%)	1 (0,2%)	185 (33,0%)
Total	252 (45,0%)	33 (5,9%)	114 (20,4%)	160 (28,6%)	1 (0,2%)	560 (100,0%)

De façon légèrement différente à ce qui est présenté au tableau précédent, le tableau XIII présente la répartition des élèves ayant réalisé l'un ou l'autre des niveaux pour chacune des situations d'évaluation et ce, nonobstant le profil réalisé. Ce tableau permet de découvrir que plus de la moitié des élèves (57,2%) ont fait le niveau 1. La proportion des élèves ayant réalisé le niveau 2 se situe à 14,9%. Enfin, 27,9% des élèves ont, quant à

eux, eu recours au niveau 3. On peut voir que la distribution des élèves pour les trois situations d'évaluation est uniforme (entre 31,7% et 36,0%).

**Tableau XIII**

Répartition des élèves ayant réalisé chacune des neuf versions des situations d'évaluation.

Situation d'évaluation réalisée (SE)	1 N (%)	2 N (%)	3 N (%)	Total N (%)
1	189 (19,3%)	43 (4,4%)	78 (8,0%)	310 (31,7%)
2	186 (19,0%)	68 (6,9%)	98 (10,0%)	352 (36,0%)
3	185 (19,0%)	35 (3,6%)	97 (9,9%)	317 (32,4%)
<b>Total</b>	<b>560 (57,2%)</b>	<b>146 (14,9%)</b>	<b>273 (27,9%)</b>	<b>979 (100,0%)</b>

## *Première partie : Analyse des scores aux différentes situations d'évaluation*

### **Introduction**

Dans cette première partie consacrée à la présentation des données, nous commencerons par présenter globalement les résultats obtenus par les élèves aux trois situations d'évaluation. Par la suite, nous enchaînerons en décrivant les résultats obtenus pour chacune des composantes de la compétence et pour chacun des critères d'évaluation et ce pour toutes les situations d'évaluation expérimentées. La section suivante sera, quant à elle, consacrée à la présentation des résultats pour chacun des quatre profils de passation que nous avons repérés lors de nos investigations. Lors de l'analyse des données, les profils que nous avons détectés permettent de bien saisir l'efficacité du modèle. Aussi, nous profiterons de l'occasion pour discuter des résultats particuliers qui concernent le profil 4 c'est-à-dire les enseignants qui ont administré les situations de niveau 1 et de niveau 3. Enfin, nous terminerons la première partie en présentant les résultats par composantes et critères de notation en mettant en lumière les nuances existant pour chacun des niveaux d'une même situation d'évaluation.

Le lecteur aura tôt fait de remarquer que les différentes sections de la première partie de l'analyse obéissent à la même structure. Nous débutons par une analyse descriptive des données afin de représenter globalement l'allure des distributions obtenues. Par la suite, nous effectuons une analyse comparative des différentes distributions que nous avons alors présentées. Enfin, nous terminons chacune des sections en présentant, de façon sommaire, nos principales interprétations. Au chapitre 5, nous reprendrons ces interprétations en les détaillant et en les expliquant en profondeur. Nous avons opté pour cette structure car il nous est apparu plus facile pour le lecteur de débiter l'interprétation des données en pouvant se référer facilement et rapidement aux graphiques et aux tableaux de données qui apparaissent au chapitre 4.

## **1. Vue d'ensemble des trois situations d'évaluation**

Dans cette section, nous présentons les données générales en lien avec chacune des situations d'évaluation. Les données plus pertinentes concernent les profils de passation mais nous ne pouvons faire l'économie de la présente section qui présente les données de façon plus générale.

### **1.1 Analyse descriptive des données**

Le tableau XIV présente les principales statistiques descriptives issues de l'analyse des scores obtenus par le biais des trois situations d'évaluation (SE1, SE2 et SE3). Pour chacune des situations d'évaluation, nous avons calculé le score moyen obtenu de la part des élèves ayant réalisés chacun des niveaux. D'abord, on constate une tendance dans les trois SE en ce qui concerne la moyenne des scores. Le troisième niveau de chaque situation d'évaluation (SE1N3, SE2N3 et SE3N3) est toujours le mieux réussi et le deuxième niveau de chaque situation d'évaluation (SE1N2, SE2N2 et SE3N2) est toujours le moins bien réussi. La figure 8 permet de visualiser ce constat. À l'aide de la figure, on peut également voir que les scores moyens pour les niveaux 1 et 2 se situent sous le seuil de passation fixé à 2,00.

Tableau XIV

Statistiques descriptives des scores bruts pour chacune des situations d'évaluation (par niveau).

	SE1 N1 N=189	SE1 N2 N=43	SE1 N3 N=78	SE2 N1 N=186	SE2 N2 N=68	SE2 N3 N=98	SE3 N1 N=184	SE3 N2 N=35	SE3 N3 N=97	
Moyenne	1,89	1,16	2,56	1,60	1,13	2,49	1,76	0,99	2,28	
Écart-type	1,07	0,86	0,82	1,05	0,79	0,81	0,89	0,86	0,71	
Variance	1,15	0,74	0,68	1,09	0,62	0,65	0,79	0,75	0,51	
Médiane	2,12	0,94	2,67	1,80	1,02	2,55	1,85	0,58	2,24	
Quartiles	25	0,96	0,41	2,15	0,58	0,49	2,11	1,27	0,28	1,94
	50	2,12	0,94	2,67	1,80	1,02	2,55	1,85	0,58	2,24
	75	2,68	1,90	3,07	2,42	1,58	3,01	2,34	1,63	2,76

Note. Les valeurs des quartiles et de la médiane sont calculées à partir de données groupées en 4 catégories. La moyenne est calculée sur 4,00.

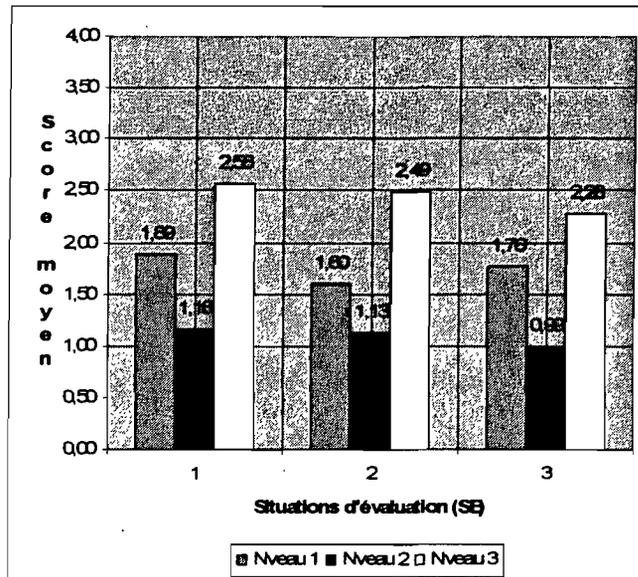


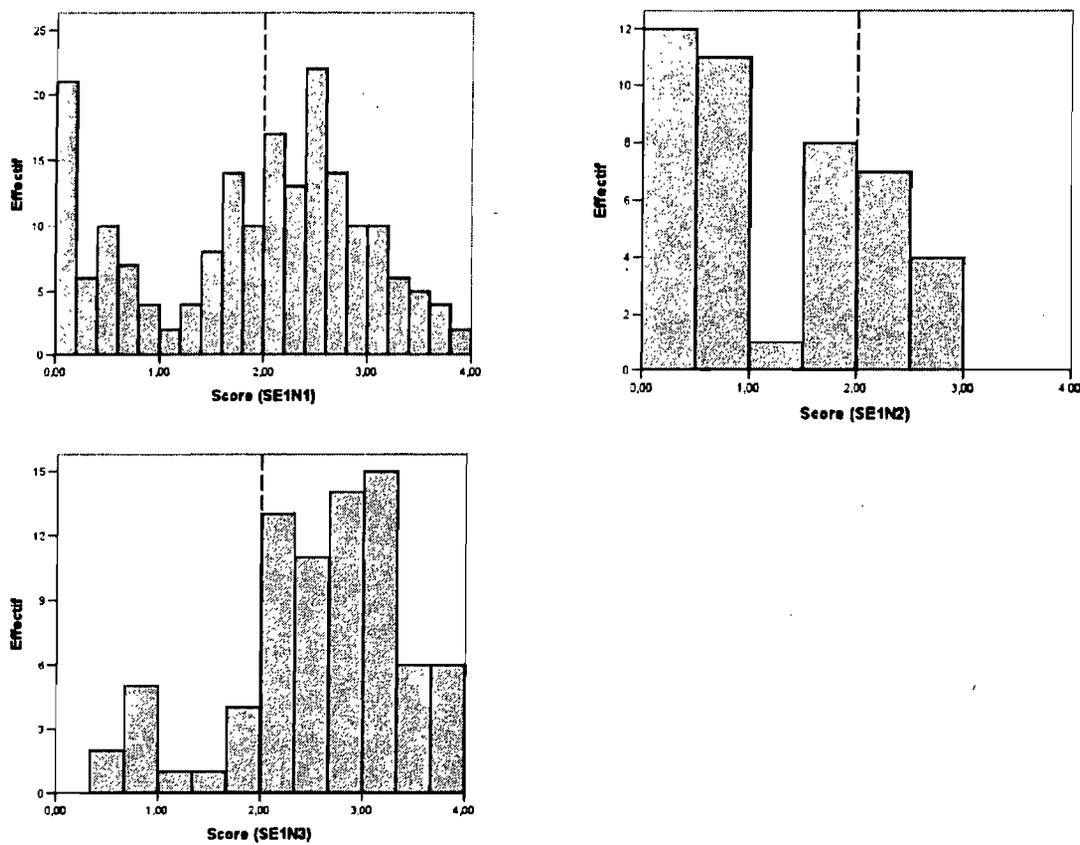
Figure 8

Score moyen par niveau pour chacune des situations d'évaluation (SE).

Aussi, nous nous attendions à observer une gradation dans la moyenne des scores en passant successivement du niveau 1 au niveau 2 et enfin au niveau 3. La figure 8 nous témoigne d'une gradation entre les niveaux 1 et 3 mais, de toute évidence, cela n'est pas le cas entre les niveaux 1 et 2 et entre les niveaux 2 et 3 pour aucune des situations d'évaluation. Nous expliquerons ce constat quand nous présenterons les résultats par profil de passation.

### **1.1.1 Analyse descriptive de la distribution des scores bruts pour la situation d'évaluation 1 (SE1)**

La figure 9 présente la distribution des scores bruts issus de la passation de la situation d'évaluation 1. Le seuil de passation est fixé à 2,00 et est représenté par un trait pointillé sur chacun des histogrammes. Examinons chacune des distributions pour les niveaux 1, 2 et 3. La distribution des scores pour la SE1N1 est bimodale. Le premier mode se situe tout près du score nul (0,00) et le second près du seuil de passation (2,50). En ce qui concerne le niveau 2 (SE1N2), on constate que les scores sont majoritairement regroupés dans l'intervalle compris entre 0,00 et 1,00. La distribution est clairement bimodale et on peut remarquer l'absence de sujets dans l'intervalle compris entre 3,00 et 4,00. Globalement, la majorité des élèves se situent en deçà du seuil de passation. Pour le niveau 3 (SE1N3), on note que la distribution se caractérise par une asymétrie négative. La majorité des sujets se retrouvent au-dessus du seuil de passation (à droite du trait pointillé). Bien que certains élèves soient encore en sérieuse difficulté (score inférieur à 1,00), il n'y a aucun élève qui a un score complètement nul comme nous pouvons le noter dans les deux précédentes distributions.



**Figure 9**  
Distribution des scores pour la situation d'évaluation 1 (SE1) pour chacun des niveaux (N1, N2 et N3).

La figure 10 présente les mêmes distributions qu'à la figure 9 mais sous une forme différente. Cette représentation, dite diagramme en forme de boîte<sup>27</sup> (*boxplot*), permet de mettre en évidence l'étendue et la variance des scores (Howell, 1998). Ainsi, la distribution SE1N1 s'étend sur l'ensemble des scores de l'échelle soit de 0,00 à 4,00. La situation de niveau 2 s'étend, quant à elle, de 0,00 à environ 3,00 ce qui laisse un vide à la droite de l'échelle (entre 3,00 et 4,00). Pour ce qui est de la situation de niveau 3, les scores s'étendent d'environ 1,00 à 4,00. Dans le cas de cette dernière distribution, peu d'élèves sont en sévère difficulté comme nous l'avons d'ailleurs noté précédemment.

<sup>27</sup> Cette représentation est également appelée diagramme en forme de boîte avec moustaches. Dans le cadre de ce texte, nous considérons les deux expressions de façon équivalente.

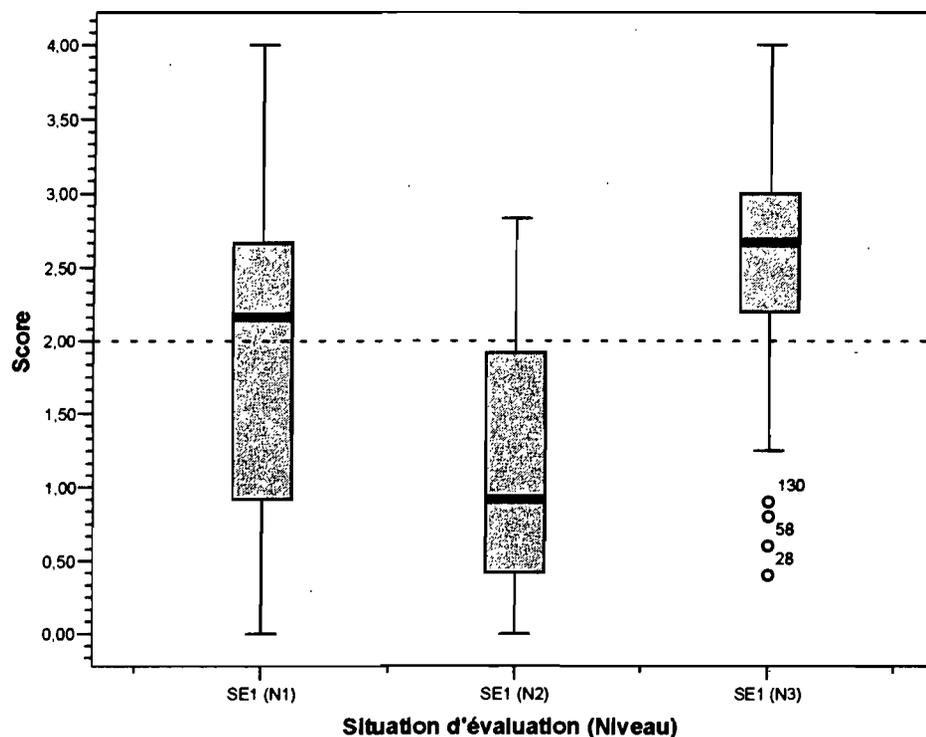


Figure 10  
Distribution des scores bruts pour la situation d'évaluation 1 (SE1) pour chacun des niveaux (N1, N2 et N3).

### 1.1.2 Analyse descriptive de la distribution des scores bruts pour la situation d'évaluation 2 (SE2)

La distribution des scores du niveau 1 pour la seconde situation d'évaluation possède deux modes (figure 11). Le premier mode se situe dans l'intervalle  $[0,00 ; 0,40]$  alors que le second est dans l'intervalle  $[2,00 ; 2,40]$ . La distribution des scores pour SE2N2 est marquée par une asymétrie positive. En effet, peu d'élèves ont obtenu des scores supérieurs à 2,00. Ainsi, la grande majorité des élèves qui ont fait cette version de la situation d'évaluation sont en échec. En ce qui concerne le niveau 3, la distribution des scores est marquée par une asymétrie négative. La plupart des élèves ont obtenu un score moyen supérieur au seuil de passation (fixé à 2,00). Le mode est situé entre les scores 2,00 et 3,00.

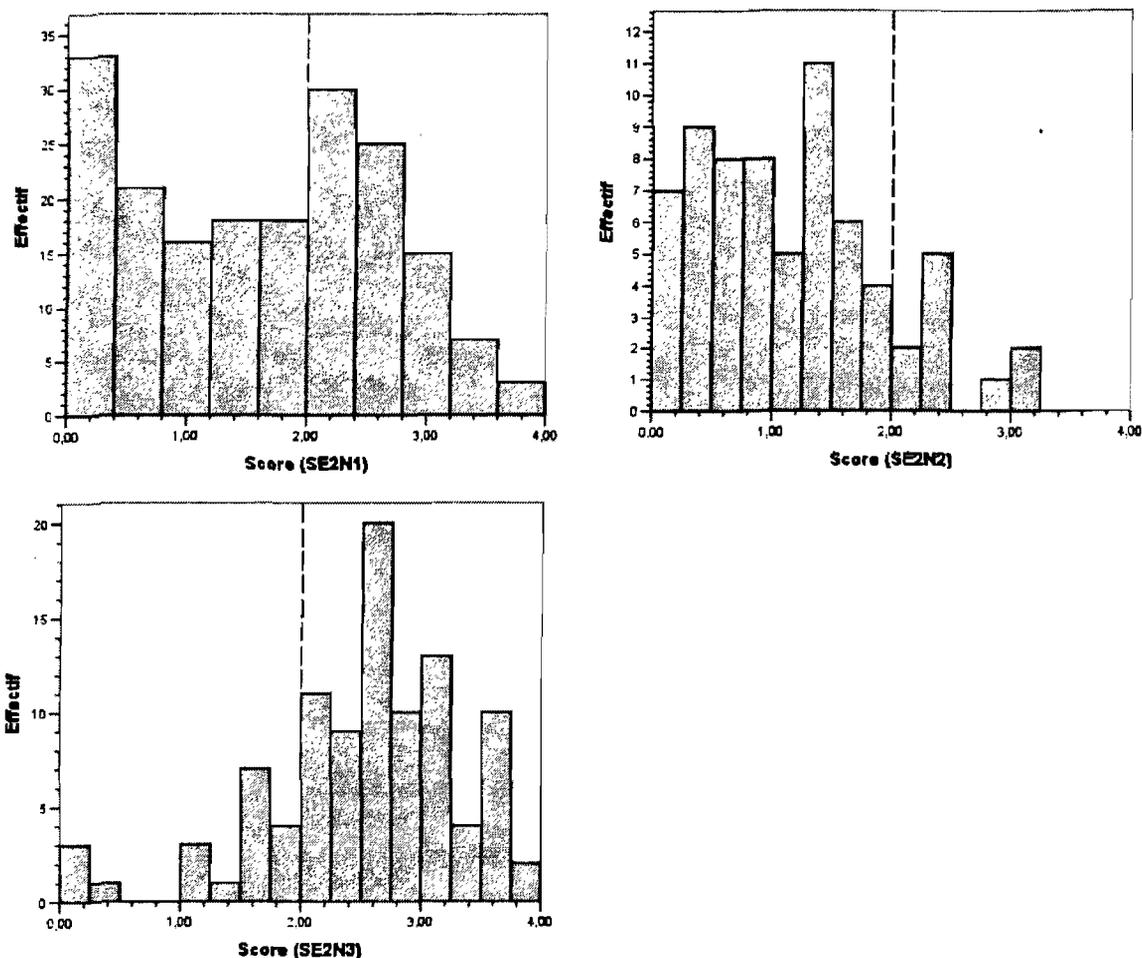


Figure 11  
Distribution des scores pour la situation d'évaluation 2 (SE2) pour chacun des niveaux (N1, N2 et N3).

La figure 12 nous permet de constater que la distribution des scores du niveau 1 couvre l'ensemble de l'échelle soit de 0,00 à 4,00. Cette distribution possède d'ailleurs la plus grande variance (1,09) en comparaison avec celle du niveau 2 (0,62) et du niveau 3 (0,65). On peut également observer que le score médian (1,80) se situe juste sous le seuil de passage. La distribution des scores du niveau 2 s'étend de 0,00 à 3,00. Le score médian (1,02) est le plus faible, et de loin, des trois distributions comparées. Enfin, la distribution des scores du niveau 3 se situe de 1,00 à 4,00. La variance est relativement faible ce qui nous amène à constater que 50% des scores se situent dans l'intervalle [2,00 ; 3,00].

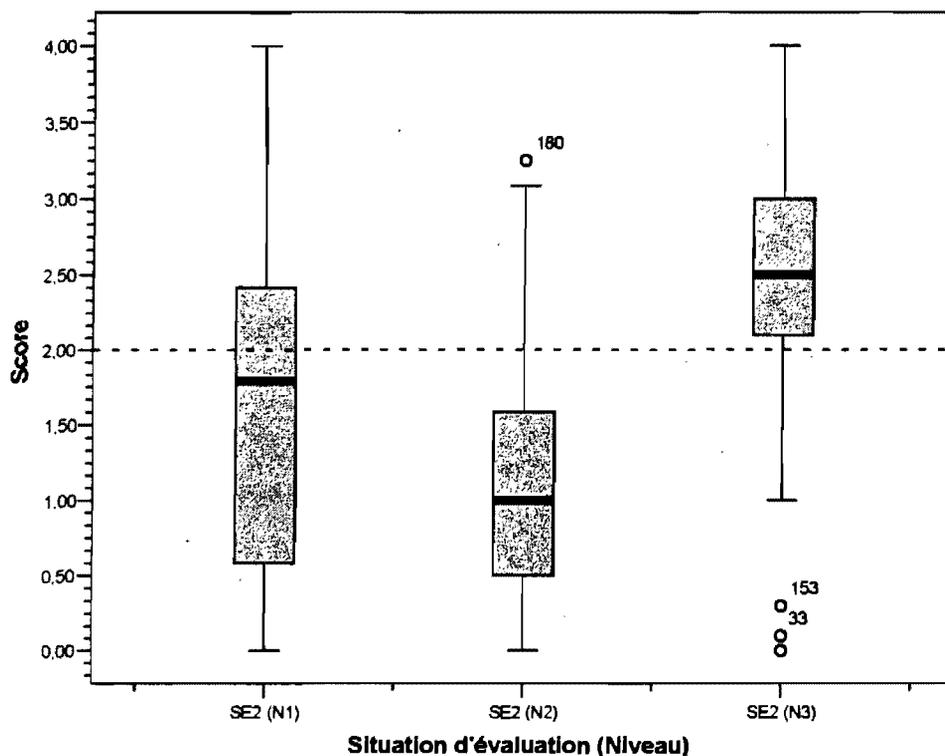


Figure 12

Distribution des scores bruts pour la situation d'évaluation 2 (SE2) pour chacun des niveaux (N1, N2 et N3).

### 1.1.3 Analyse descriptive de la distribution des scores bruts pour la situation d'évaluation 3 (SE3)

En jetant un coup d'œil à la figure 13, on constate que la distribution des scores du niveau 1 en ce qui concerne la situation d'évaluation 3 tend vers une distribution normale. En comparant cette distribution avec une distribution normale, on obtient une valeur au test de Kolmogorov-Smirnov de 1,208 ( $p = 0,108$ ). La valeur de cette statistique ne nous permet pas de rejeter l'hypothèse nulle qui stipule que la distribution est normale. Cependant, on peut remarquer une concentration de sujets à la gauche de l'échelle (entre 0,00 et 0,40). Au sujet de la distribution des scores pour le niveau 2, cette dernière possède plusieurs ressemblances avec celles de SE1N2 et SE2N2. La distribution possède une asymétrie positive et aucun sujet n'a obtenu un score supérieur à 3,00. On peut facilement remarquer que les sujets se trouvent majoritairement sous le seuil de passation. La distribution des scores du niveau 3 a également une allure de courbe normale. Bien

que nous ne nous attendions pas nécessairement à une distribution normale, la statistique de Kolmogorov-Smirnov se situe à 1,208 ( $p = 0,108$ ) et confirme la normalité de la distribution. Le mode se situe dans l'intervalle [2,00 ; 2,40].

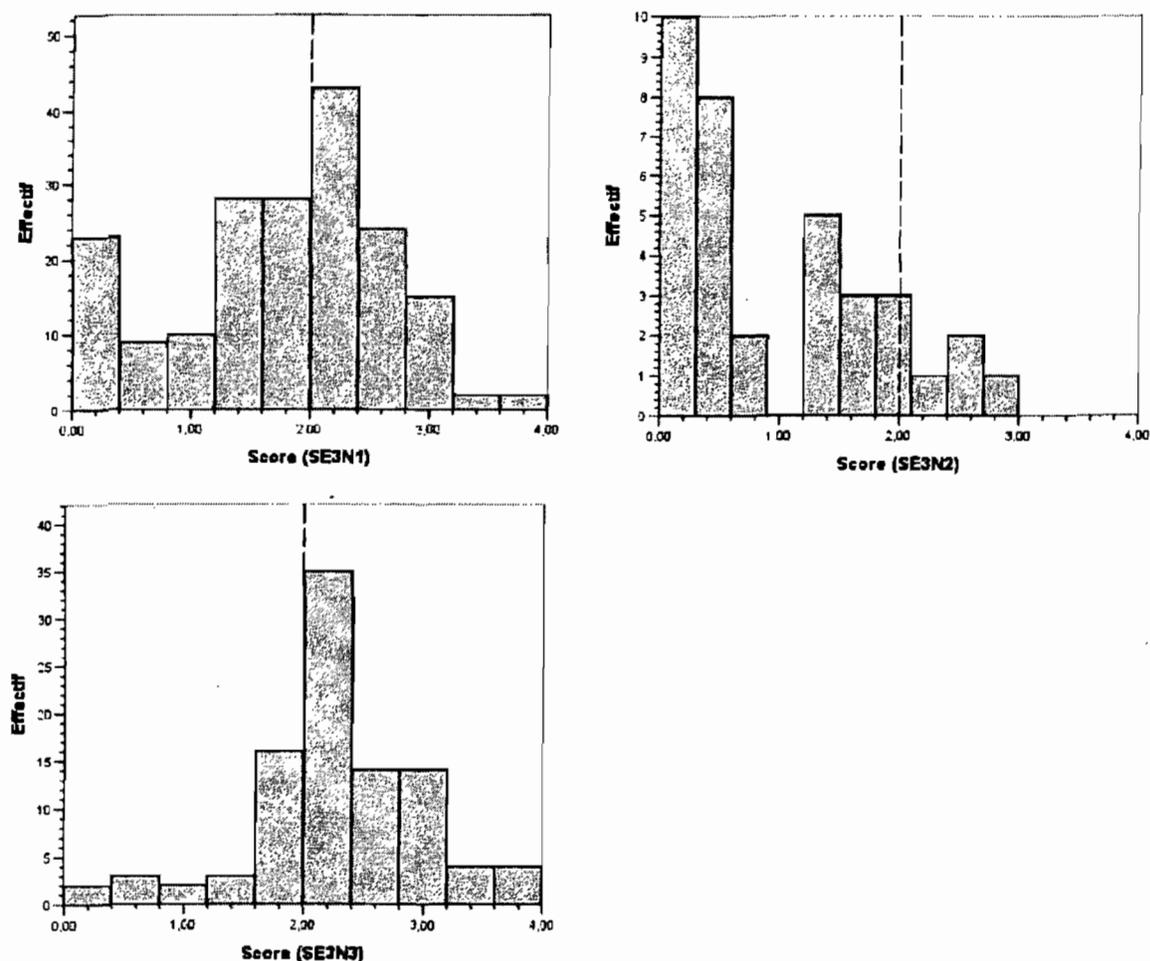


Figure 13

Distribution des scores pour la situation d'évaluation 3 (SE3) pour chacun des niveaux (N1, N2 et N3).

Les boîtes à moustaches de la figure 14 nous montrent que la distribution des scores de l'épreuve SE3N1 s'étend de 0,00 à 4,00. Le score médian (1,85) se situe sous le seuil de passation. La distribution des scores du niveau 2 s'étend de 0,00 à 3,00. On remarque que plus de 75% des scores se situent sous le seuil de passation (fixé à 2,00). Les scores du niveau 3, quant à eux, se distribuent généralement entre les échelons 1,00 et 4,00. Le

score médian est situé à 2,24. Le mode est situé dans l'intervalle compris entre les scores 2,00 et 2,40.

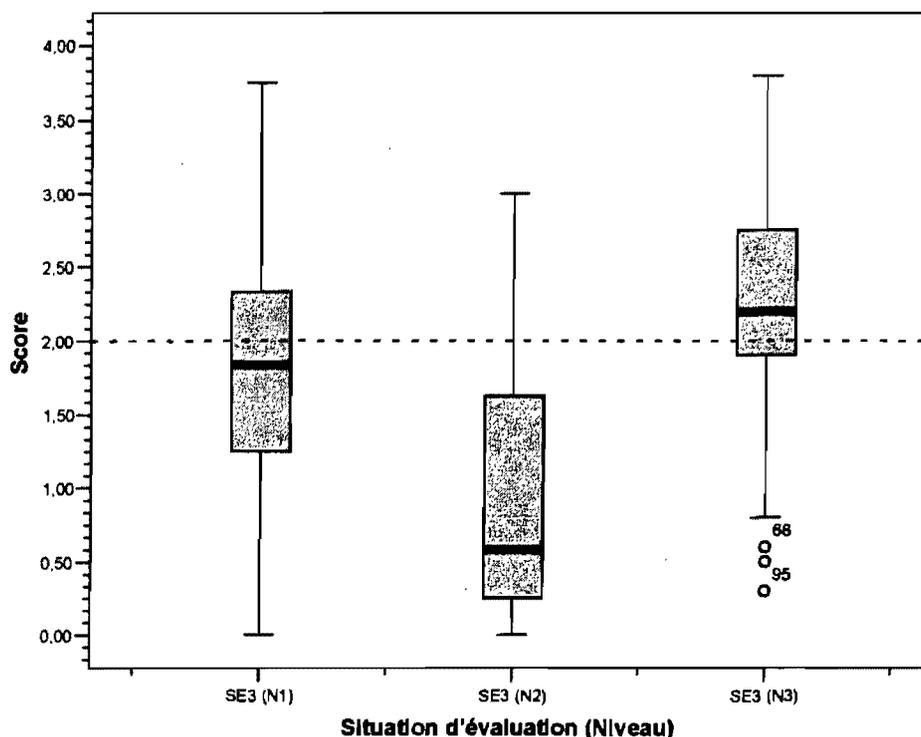


Figure 14

Distribution des scores bruts pour la situation d'évaluation 3 (SE3) pour chacun des niveaux (N1, N2 et N3).

#### 1.1.4 Analyse comparative de la distribution des scores bruts concernant les trois situations d'évaluation

Dans cette section, nous nous attarderons aux différences entre les différents niveaux pour les trois situations d'évaluation. L'analyse de la figure 15 nous apprend que la distribution des scores moyens est relativement semblable d'une situation d'évaluation à l'autre. Dans chaque cas, le score médian à la situation de niveau 1 est plus élevé qu'à celui du niveau 2. Le score médian au niveau 3 est, quant à lui, toujours supérieur aux scores obtenus aux deux niveaux précédents. Autre fait à noter, la distribution des scores pour le niveau 1 s'étend généralement de 0,00 à 4,00 (sauf dans le cas de la SE3 où le score maximum est d'environ 3,75). Dans le cas du niveau 2, les scores s'étendent de

0,00 à environ 3,00. Pour ce qui est du niveau 3, les scores se situent entre environ 1,00 et 4,00.

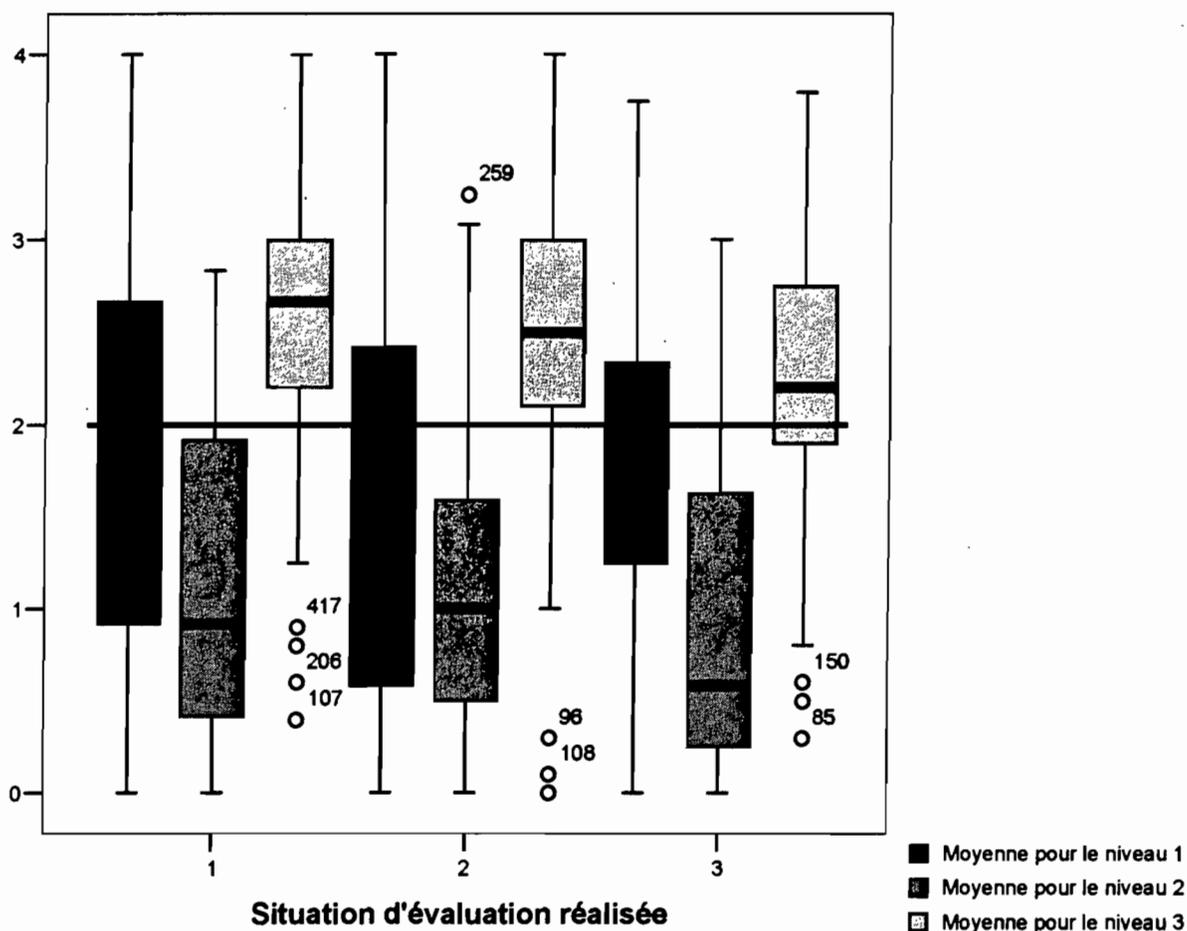


Figure 15

Diagramme en moustaches présentant la distribution des scores moyens pour les trois situations d'évaluation pour chacun des niveaux.

Le tableau XV présente une comparaison des scores moyens pour les trois niveaux de chacune des situations d'évaluation. On peut voir que la plupart des comparaisons deux à deux indiquent qu'il n'y a pas de différence entre les moyennes. Cependant, tel n'est pas le cas pour les trois paires ci-contre : SE1N1-SE2N1, SE1N3-SE3N3 et SE2N3-SE3N3. Ainsi, sur les neuf paires comparées, uniquement trois d'entre elles montraient des différences significatives à  $p=0,01$ .

Tableau XV  
 Comparaison deux à deux des moyennes pour toutes les situations d'évaluation et par niveau.

SE	Statistique standardisée associée au test U de Mann-Whitney
SE1N1- SE2N1	-2,647‡
SE1N1- SE3N1	-1,902
SE2N1- SE3N1	-1,277
SE1N2- SE2N2	-0,142
SE1N2- SE3N2	-1,097
SE2N2- SE3N2	-1,098
SE1N3- SE2N3	-0,644
SE1N3- SE3N3	-3,023‡
SE2N3- SE3N3	-2,580‡

Note. ‡  $p \leq 0,01$

En somme, on constate qu'il existe plusieurs similitudes dans les scores obtenus à chacune des situations d'évaluation qui comportent toutes un contexte disciplinaire très différent.

En terminant, jetons un coup d'œil au tableau XVI qui présente certaines statistiques concernant les élèves qui manifestent le « syndrome de la page blanche ». Ces élèves se reconnaissant au fait qu'ils n'arrivent pas à manifester la moindre démarche ou parcelle de solution quant au problème qui leur est proposé. Pour le niveau 1, on remarque que cette proportion varie entre 4,89% et 8,06%. C'est pour ce niveau que le taux de vecteur réponse nulle<sup>28</sup> est le plus élevé. La passation du niveau 2 demeure problématique pour plusieurs élèves. Pour la situation d'évaluation 1 on note que ce taux grimpe à 6,98% et à

<sup>28</sup> Lorsqu'un élève obtient un score nul à tous les critères, on parle alors d'un vecteur réponse nulle.

11,43% dans le cas de la SE3. Au niveau 3, le nombre d'élèves ayant un vecteur réponse nul chute de façon considérable pour tendre vers zéro.

Tableau XVI

Proportion des élèves ayant obtenu un vecteur réponse nul pour chacune des situations d'évaluation (SE) et chacun des niveaux (N).

Niveaux	SE 1	SE 2	SE 3
N1	5,29%	8,06%	4,89%
N2	6,98%	2,94%	11,43%
N3	0,00%	1,16%	0,00%

## 1.2 Interprétation des données

Il y a beaucoup à dire sur les données qui viennent d'être présentées à la section précédente (section 1). En guise d'introduction à l'interprétation des données, qui sera réalisée au chapitre suivant, nous discuterons pour l'instant uniquement des données plus factuelles. Nous aborderons brièvement les aspects liés à la passation des situations d'évaluation, à la performance des élèves aux situations de niveau 2 et aux scores globaux pour les situations d'évaluation 1, 2 et 3

### Passation des situations d'évaluation

Dans le protocole de recherche, on demandait aux enseignants d'administrer d'abord la situation d'évaluation de niveau 1 et ensuite, au besoin, administrer les niveaux 2 et 3 aux élèves qui éprouvaient des difficultés manifestes. On se serait alors attendu à voir plus d'élèves réaliser chacune des situations du niveau 2 que du niveau 3. Or, dans tous les cas, les effectifs associés au niveau 3 sont toujours supérieurs aux effectifs du niveau 2 (tableau xiv). Plusieurs enseignants nous ont dit qu'ils avaient administré le niveau 3 même lorsque cela n'était pas nécessaire afin (1) d'observer s'il existait des différences entre les scores obtenus à une situation ouverte (niveau 1) et fermée (niveau 3) ou (2)

d'occuper les élèves qui avaient terminé le niveau 1 bien avant les autres élèves du groupe. Ce résultat inattendu sera discuté plus longuement au chapitre suivant.

### La performance des élèves aux situations de niveau 2

L'analyse des scores moyens nous permet de constater que les élèves ont rencontré passablement de difficulté à la situation de niveau 2 et ce, pour chacune des situations d'évaluation. Comme nous avons pu le constater à la figure 8, les scores associés au niveau 2 sont, et de loin, nettement inférieurs à ceux du niveau 1 ou 3. Nous pensons que les consignes données au niveau 2 étaient peut-être trop générales et floues aux yeux des élèves. Aussi, tous les indices concernaient les aspects méthodologiques associés à la compétence à démontrer. Lors de la conception des épreuves, nous avons choisi d'axer davantage nos interventions sur les aspects de méthode en accordant moins d'importance aux indices conceptuels. Les résultats nous donnent à penser qu'il serait préférable d'ajouter plus d'indices conceptuels afin de mieux guider les élèves au niveau 2. Comme par exemple, nous pourrions identifier les concepts clés de chacune des situations et inviter les élèves à les circonscrire afin de mieux se représenter le problème. Bref, rendre les consignes plus concrètes.

### Les situations d'évaluation 1-2-3

L'analyse descriptive des scores de chacune des situations d'évaluation nous apprend que les distributions ne sont pas normales. De plus, en consultant le tableau XIV, on peut voir que les effectifs pour chacune des situations expérimentées ne sont pas égaux. Ce constat nous a obligé à faire des choix méthodologiques pour la suite de l'analyse des données. Par conséquent, nous avons choisi d'avoir recours à des outils statistiques non paramétriques pour la suite de nos analyses. La nature de nos données nous a incité à choisir une méthode d'analyse non paramétrique qui n'oblige pas à rencontrer des contraintes rigides comme dans le cas de l'utilisation de méthodes paramétriques. À ce sujet, nous souscrivons aux réflexions de Capéraà et Van Cutsem (1988, p. 4) à ce sujet : « Ainsi, à la différence d'un modèle paramétrique qui spécifie la loi de chaque

observation par la valeur d'un paramètre, un modèle non paramétrique laisse beaucoup plus de souplesse quant à la forme et à la nature possible des lois d'observation ». L'échelle que nous avons utilisée (échelle d'appréciation) justifie également notre choix d'avoir opté pour des tests non paramétriques.

## **2. Analyse détaillée des résultats par composantes et critères de notation**

### **2.1 Analyse descriptive des données**

Dans cette section, nous nous intéressons plus spécifiquement aux résultats obtenus à chacun des douze critères de notation ainsi qu'aux résultats à chacune des composantes de la compétence. Le lecteur remarquera que les valeurs associées aux critères 3 et 5 du niveau 3 sont absents autant dans les tableaux de données (ex. tableau XVIII) que dans les graphiques (ex. figure 16). Ceci s'explique par le fait que le matériel (critère 3) et les manipulations (critère 5) étaient fournis aux élèves dans le cadre du niveau 3.

La correspondance entre les critères de notation et les composantes de la compétence 1 du Programme de formation de l'école québécoise, discipline science et technologie, apparaît dans le tableau XVII ci-dessous. Le lecteur prendra note que les critères 11 et 12 concernent les aspects liés à la communication. Ces aspects font l'objet d'une compétence spécifique dans le Programme de formation de l'école québécoise, enseignement secondaire. Dans le cadre de cette recherche, nous considérons les aspects liés à la communication comme des composantes à part entière de la compétence d'investigation scientifique.

Une analyse des coefficients de corrélations entre les différents critères de correction a été effectuée en guise d'exploration. Une telle analyse a été conduite afin de vérifier s'il pouvait exister, par exemple, un lien entre les critères appartenant à une même composante. Si tel est le cas, cela viendrait appuyer la validité associée à chacune des composantes. Aussi, l'étude des corrélations peut nous donner des pistes supplémentaires

au sujet des éléments de la démarche d'investigation scientifique pour lesquels les élèves ont eu de la facilité ou des difficultés.

### Tableau XVII

Correspondance entre les critères de notation et les composantes de la compétence d'investigation scientifique telles que formulées dans le Programme de formation de l'école québécoise, enseignement secondaire (1<sup>er</sup> cycle), discipline science et technologie.

Critère	Composante
1. Reformulation du problème 2. Identification de la variable à investiguer	1. Cerner un problème
3. Élaboration d'une liste de matériel 4. Formulation d'une hypothèse vérifiable et plausible 5. Formulation et présentation des manipulations	2. Choisir un scénario d'investigation
6. Consignation des données expérimentales et des observations 7. Identification des causes d'erreurs	3. Concrétiser sa démarche
8. Traitement des données expérimentales 9. Maîtrise des concepts, des règles et des techniques 10. Application des résultats expérimentaux dans la solution proposée	4. Analyser ses résultats ou sa solution
11. Présentation générale 12. Utilisation des symboles et des termes	5. Communiquer à l'aide des langages utilisés en science <sup>29</sup>

<sup>29</sup> Cet énoncé n'est pas une composante de la compétence d'investigation scientifique mais plutôt l'énoncé de la compétence de communication scientifique.

### **2.1.1 Analyse descriptive de la distribution des scores bruts moyens pour la situation d'évaluation 1 (SE1) pour chacun des critères de notation pour les trois niveaux**

La figure 16 (A) illustre la distribution des scores moyens pour chacun des douze critères de notation. Dans un premier temps, on constate que les trois courbes ont, *in globo*, la même allure. Par exemple, le critère 1 est celui parmi les douze qui est généralement le mieux réussi à chacun des niveaux. Le critère 7 est constamment le moins bien réussi. Au niveau 3, qui est le niveau le mieux réussi, le seul critère dont le score moyen (1,74) se situe sous le seuil de passation est le critère 7.

Les scores obtenus aux quatre derniers critères sont généralement assez semblables et ce pour tous les niveaux. À ce sujet, l'étude des coefficients de corrélation de Spearman nous apprend que les quatre derniers critères sont fortement corrélés entre eux. Les tableaux XIX à XXI présentent les valeurs des coefficients de corrélations bivariés de Spearman des critères de notation en provenance de chacun des trois niveaux. Les coefficients inscrits sur fond noir représentent des valeurs de corrélation qui montrent une forte liaison positive ( $\geq 0,700$ ). À la lecture de chacun de ces tableaux, on peut apercevoir que les valeurs de corrélation sont fortes entre les critères 8, 9, 10, 11 et 12. Au niveau 3, seul le critère 10 montre une corrélation un peu moins forte avec les critères 11 et 12.

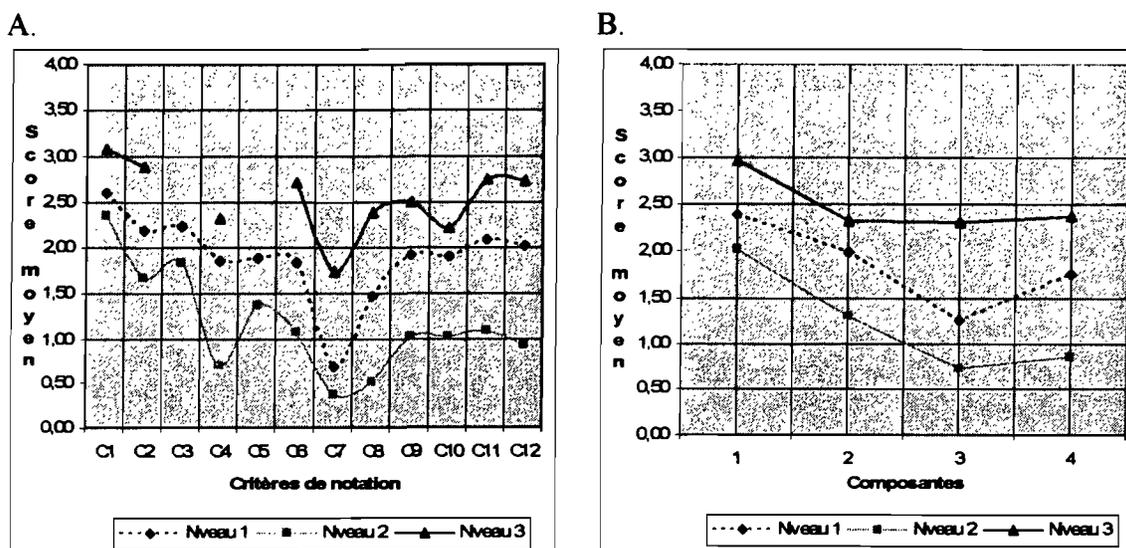


Figure 16  
Distribution des scores moyens par critère de notation (A) et par composante (B) pour chacun des niveaux de la SE1.

Autre fait intéressant à souligner, on peut voir sur la figure 16 (A) que l'ordre des courbes est inversé quant aux niveaux. En effet, les scores moyens pour le niveau 2 se situent en deçà de ceux obtenus au niveau 1. Au chapitre suivant, dédié à l'interprétation des données, nous émettrons quelques hypothèses pour expliquer cette inversion des courbes.

La figure 16 (B) présente, quant à elle, les scores moyens pour certains regroupements de critères de notation. On peut remarquer que les deux premiers critères sont généralement assez bien réussis pour les niveaux 1 et 3 puisque le score moyen est supérieur à 2,00. Aux niveaux 1 et 2, on note une baisse des scores moyens par la suite pour les critères associés aux composantes 2, 3 et 4. La composante la moins bien réussie pour ces niveaux est la troisième (celle qui regroupe le critère 7). En ce qui concerne le niveau 3, on remarque que les scores sont relativement constants pour les composantes 2, 3 et 4.

Tableau XVIII

Scores moyens pour chacun des critères de notation pour chacun des niveaux de la SE1.

Critère	N1	N2	N3
1	2,60	2,35	3,08
2	2,19	1,67	2,87
3	2,23	1,84	NSP
4	1,85	0,7	2,32
5	1,89	1,37	NSP
6	1,83	1,07	2,71
7	0,68	0,37	1,74
8	1,47	0,51	2,38
9	1,91	1,02	2,49
10	1,90	1,02	2,21
11	2,09	1,09	2,75
12	2,02	0,93	2,73

Note. NSP : Ne s'applique pas.

Tableau XIX

Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE1N1.

Critère	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
C 1	1,000	.712	,457	,470	,431	,529	,154	,467	,553	,485	,518	,532
C 2		1,000	,454	,483	,462	,575	,214	,541	,660	,596	,590	,615
C 3			1,000	,446	,554	,527	,136*	,543	,541	,462	,551	,538
C 4				1,000	,537	,531	,221	,579	,592	,577	,570	,581
C 5					1,000	,656	,276	,636	,674	,635	,682	,651
C 6						1,000	,474	.749	.845	.790	.813	.762
C 7							1,000	,341	,449	,410	,363	,315
C 8								1,000	.821	.761	.766	.741
C 9									1,000	.886	.841	.825
C 10										1,000	.822	.791
C 11											1,000	.849
C 12												1,000

Note. \*Seule valeur dont la corrélation n'est pas statistiquement significative au seuil de 5%.

Tableau XX

Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE1N2.

Critère	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
C 1	1,000	,475	,290	,338†	-,12	,143	,115	,120	,149	,095	,089	,150
C 2		1,000	.632‡	,298	,417‡	,579‡	,365†	,483‡	,548‡	,502‡	,517‡	,588‡
C 3			1,000	,368†	,52‡6	,547‡	,247	,276	,402‡	,396‡	,430‡	,456‡
C 4				1,000	,066	,197	,018	,105	,135	,060	,084	,112
C 5					1,000	.644	,287	,470‡	,581‡	,555‡	,608‡	,584‡
C 6						1,000	,507‡	.800‡	.898‡	.854‡	.847‡	.855‡
C 7							1,000	,345†	,565‡	.656‡	,566‡	,619‡
C 8								1,000	.89‡1	.823‡	.792‡	.802‡
C 9									1,000	.956‡	.940‡	.946‡
C 10										1,000	.910‡	.914‡
C 11											1,000	.973‡
C 12												1,000

Note. †  $p \leq 0,05$ . ‡  $p \leq 0,01$ .

Tableau XXI

Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE1N3.

Critère	C1	C2	C4	C6	C7	C8	C9	C10	C11	C12
C1	1,000	,439‡	,320‡	,362‡	,068	,330†	,256	,186	,299†	,309†
C2		1,000	,422‡	,444‡	,038	,476‡	,335†	,305†	,369‡	,320†
C4			1,000	,596‡	,335†	,522‡	,560‡	,455‡	,573‡	,393‡
C6				1,000	,509‡	.836‡	.739‡	,583‡	.792‡	.791‡
C7					1,000	,495‡	,574‡	,463‡	,385‡	,41‡
C8						1,000	.815‡	.714‡	.746‡	.775‡
C9							1,000	.761‡	.756‡	.750‡
C10								1,000	,645‡	,552‡
C11									1,000	.845‡
C12										1,000

Note. †  $p \leq 0,05$ . ‡  $p \leq 0,01$ .

### 2.1.2 Analyse descriptive de la distribution des scores bruts moyens pour la situation d'évaluation 2 (SE2) pour chacun des critères de notation pour les trois niveaux

La figure 17 (A) illustre la distribution des scores pour chacun des critères de la situation d'évaluation 2. Encore une fois, on constate que les courbes du niveau 1 et 2 sont inversées. Logiquement, les scores obtenus au niveau 2 devraient être supérieurs à ceux du niveau 1 mais tel n'est pas le cas. Dans la section 1.3, nous examinerons les résultats à la lumière des profils de passation et nous expliquerons alors pourquoi nous observons une telle inversion dans les scores moyens globaux. Les premiers critères (critères 1 à 3) sont, pour chacun des niveaux, généralement les mieux réussis. En ce qui concerne spécifiquement les niveaux 1 et 2, on constate que les courbes ont de multiples similitudes. On constate également que l'écart entre les courbes est relativement faible. Ainsi les critères semblent se comporter de la même façon pour chacun de ces deux niveaux. On constate aussi que le critère 7 est problématique pour chacun des niveaux. Au niveau 3, le critère 7 est encore le seul (1,94) dont le score moyen se situe, légèrement cependant, sous le seuil de passation.

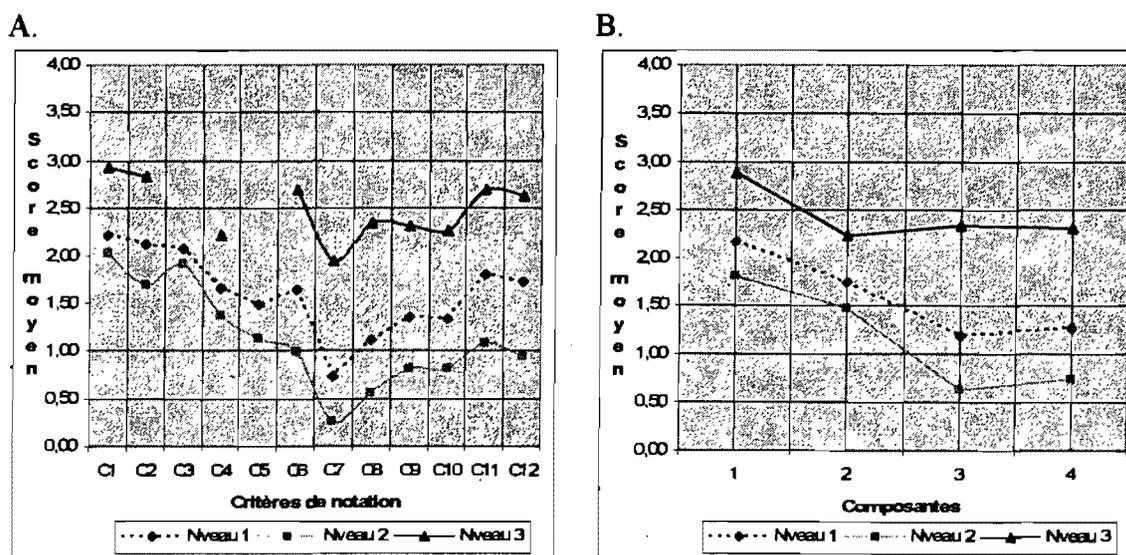


Figure 17  
Distribution des scores moyens par critère de notation (A) et par composante (B) pour chacun des niveaux de la SE2.

La figure 17 (B) nous offre une représentation des scores pour chacune des composantes de la compétence d'investigation scientifique telles que définit dans le programme de science et technologie, premier cycle du secondaire. On peut voir plus clairement le comportement des premiers critères associés à la composante 1. Globalement, cette dernière est assez bien réussie. Par contre, et cela est d'autant plus remarquable pour les niveaux 1 et 2, on remarque une baisse abrupte des scores pour les composantes 2 et 3. Cette baisse est visible également dans le cas du niveau 3 mais l'intensité est beaucoup moindre. De fait, les scores moyens sont généralement constants pour les composantes 2 à 4 dans le cas du niveau 3.

Tableau XXII

Scores moyens des critères de notation pour chacun des niveaux de la SE2.

Critère	N1	N2	N3
1	2,22	2,03	2,94
2	2,12	1,69	2,83
3	2,07	1,91	NSP
4	1,66	1,37	2,22
5	1,49	1,13	NSP
6	1,64	0,99	2,70
7	0,73	0,25	1,94
8	1,11	0,56	2,34
9	1,36	0,82	2,31
10	1,34	0,81	2,27
11	1,80	1,09	2,70
12	1,72	0,94	2,63

Note. NSP : Ne s'applique pas.

Tableau XXIII  
Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE2N1.

Critère	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
C 1	1,000	,580	,371	,397	,279	,399	,272	,317	,408	,366	,403	,402
C 2		1,000	,443	,512	,397	,440	,253	,378	,480	,400	,436	,458
C 3			1,000	,299	,637	,584	,380	,534	,580	,579	,604	,579
C 4				1,000	,390	,430	,206	,392	,420	,370	,387	,409
C 5					1,000	.707	,496	,628	.693	,664	.703	.676
C 6						1,000	,630	.826	.855	.820	.862	.861
C 7							1,000	,552	,599	,638	,567	,527
C 8								1,000	.825	.786	.801	.800
C 9									1,000	.854	.861	.858
C 10										1,000	.859	.817
C 11											1,000	.900
C 12												1,000

Note. Toutes les corrélations sont significatives à  $p \leq 0,05$ .

Tableau XXIV  
Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE2N2.

Critère	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
C 1	1,000	,507†	,265†	,203	,031	,047	,128	-,034	,071	,079	,024	,021
C 2		1,000	,206	,351‡	,142	,220	,130	,068	,168	,078	,051	,081
C 3			1,000	,049	,445‡	,401‡	,334‡	,393‡	,357‡	,490‡	,424‡	,342‡
C 4				1,000	,243†	,190	,113	,107	,207	,059	,120	,128
C 5					1,000	,473‡	,329‡	,535‡	,546‡	,513‡	,540‡	,522‡
C 6						1,000	,452‡	.708‡	.847‡	.760‡	.827‡	.835‡
C 7							1,000	,458‡	,555‡	,574‡	,508‡	,442‡
C 8								1,000	.759‡	.774‡	.831‡	.788‡
C 9									1,000	.857‡	.921‡	.917‡
C 10										1,000	.903‡	.841‡
C 11											1,000	.962‡
C 12												1,000

Note. †  $p \leq 0,05$ . ‡  $p \leq 0,01$ .

Tableau XXV

Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE2N3.

Critère	C1	C2	C4	C6	C7	C8	C9	C10	C11	C12
C1	1,000	,488†	,177	,351‡	,158	,356‡	,209	,221†	,476‡	,388‡
C2		1,000	,380‡	,418‡	,279‡	,417‡	,273†	,312‡	,398‡	,492‡
C4			1,000	,371‡	,189†	,360‡	,316‡	,230†	,320‡	,387‡
C6				1,000	,576‡	<b>.736‡</b>	,582‡	,597‡	<b>.666‡</b>	,621‡
C7					1,000	,607‡	,442‡	,574‡	,552‡	,393‡
C8						1,000	<b>.791‡</b>	<b>.695‡</b>	<b>.715‡</b>	<b>.721‡</b>
C9							1,000	<b>.680‡</b>	,572‡	,605‡
C10								1,000	<b>.727‡</b>	,571‡
C11									1,000	<b>.701‡</b>
C12										1,000

Note. †  $p \leq 0,05$ . ‡  $p \leq 0,01$ .

L'analyse des coefficients de corrélation de Spearman (tableaux XXIII, XXIV et XXV) révèle que les derniers critères (6-8-9-10-11-12) sont très fortement corrélés entre eux sauf dans le cas du critère 7 qui semble se comporter différemment des autres.

### **2.1.3 Analyse descriptive de la distribution des scores bruts moyens pour la situation d'évaluation 3 (SE3) pour chacun des critères de notation pour les trois niveaux**

La figure 18 (A) illustre la distribution des scores dans le cas de la troisième situation d'évaluation. De façon générale, on constate que les scores sont plus élevés en ce qui concerne les premiers critères et plus faibles pour les derniers critères. On remarque cependant une hausse des scores pour les deux derniers critères (C11 et C12) – et ce pour tous les niveaux – mais il faut se rappeler que ces critères sont associés davantage aux aspects liés à la communication des résultats qu'à la démarche scientifique. Si on jette un coup d'œil aux critères du niveau 1, on peut remarquer que les critères 5 à 10 se situent sous le seuil de passation et sont, par conséquent, des aspects non maîtrisés de la part des élèves. Le critère 7 est encore une fois fortement problématique puisque le score global se situe à seulement 0,61 (tableau XXVI). En consultant le tableau XXVII, on peut remarquer que les critères 6, 8, 9, 10, 11 et 12 sont fortement corrélés entre eux. Pour le niveau 2, le tableau XXVI nous informe que toutes les valeurs moyennes se situent sous le seuil de passation fixé, rappelons-le, à 2,00. On peut même remarquer que trois critères (7, 8 et 10) ont des scores qui se situent en deçà de 0,50. Sur la figure 17 (A), on remarque – encore une fois - que les trois premiers critères sont les mieux réussis. Par la suite, les scores tendent à diminuer de façon dramatique. Quant à eux, les scores associés au niveau 3 sont passablement fluctuants. En fait, les scores se retrouvent dans l'intervalle [1,26 ; 3,08]. Parmi eux, quatre scores moyens (C6, C7, C8, C10) demeurent sous le seuil de passation. Outre les deux derniers critères, on remarque que la valeur des scores tend à diminuer en passant d'un critère à l'autre. En consultant le tableau XXIX, il appert que certains critères d'évaluation sont faiblement corrélés entre eux. On remarque des valeurs corrélationnelles élevées et statistiquement significatives entre les critères 6 et 8, 6 et 9, 7 et 9, 8 et 9 et 9 et 11 seulement. Ces résultats sont concordants avec les valeurs fluctuantes que nous avons observées et décrites précédemment.

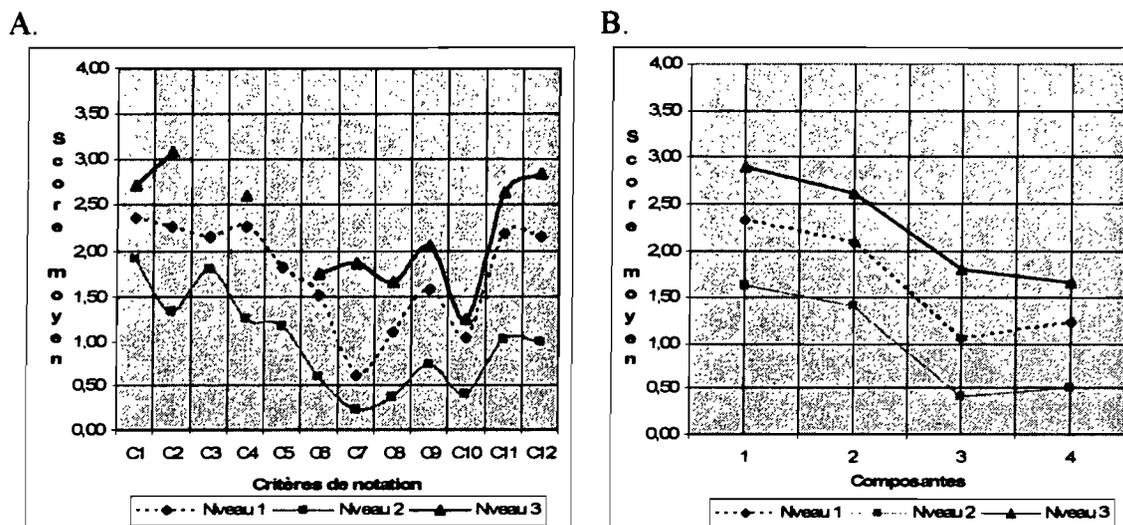


Figure 18

Distribution des scores moyens par critère de notation (A) et par composante (B) pour chacun des niveaux de la SE3.

La figure 18 (B) illustre la variation des scores pour les critères associés à chacune des quatre composantes de la compétence d'investigation scientifique dans le cadre de la troisième situation d'évaluation. Ce graphique permet de bien visualiser la tendance qu'ont les scores à diminuer au fur et à mesure que les élèves avancent dans la résolution du problème qui leur est dévolu. Pour tous les niveaux, on observe à peu près la même chose. Les résultats aux deux premières composantes sont généralement les meilleurs. Par la suite, les scores tendent à diminuer de façon marquée.

**Tableau XXVI**  
**Scores moyens des critères de notation et de chacune des composantes pour chacun des niveaux de la SE3.**

Critère	N1	N2	N3
1	2,36	1,91	2,71
2	2,27	1,34	3,08
3	2,16	1,80	NSP
4	2,27	1,26	2,61
5	1,82	1,17	NSP
6	1,51	0,60	1,75
7	0,61	0,23	1,87
8	1,11	0,37	1,66
9	1,58	0,74	2,06
10	1,04	0,40	1,26
11	2,19	1,03	2,64
12	2,16	1,00	2,84

Note. NSP : Ne s'applique pas.

Tableau XXVII  
Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE3N1.

Critère	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
C 1	1,000	,538‡	,298‡	,360‡	,238‡	,312‡	,105	,328‡	,347‡	,243‡	,375‡	,311‡
C 2		1,000	,299‡	,510‡	,259‡	,322‡	,287‡	,346‡	,470‡	,304‡	,426‡	,451‡
C 3			1,000	,473‡	,472‡	,424‡	,177†	,498‡	,370‡	,433‡	,542‡	,539‡
C 4				1,000	,454‡	,465‡	,223‡	,494‡	,451‡	,345‡	,488‡	,481‡
C 5					1,000	,582‡	,284‡	,519‡	,540‡	,473‡	,559‡	,577‡
C 6						1,000	,493‡	.758‡	.785‡	.678‡	.719‡	.727‡
C 7							1,000	,501‡	,531‡	,393‡	,316‡	,320‡
C 8								1,000	.749‡	.678‡	.675‡	.719‡
C 9									1,000	,583‡	.689‡	.704‡
C 10										1,000	.674‡	.689‡
C 11											1,000	.867‡
C 12												1,000

Note. †  $p \leq 0,05$ . ‡  $p \leq 0,01$ .

Tableau XXVIII

Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE3N2.

Critère	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
C 1	1,000	.673‡	,432‡	,745‡	,310	,382†	,277	,485‡	,450‡	,389†	,432‡	,381†
C 2		1,000	,296	.630‡	,447‡	,482‡	,417†	,574‡	,561‡	,353†	,536‡	,486‡
C 3			1,000	,437‡	,487‡	,553‡	,141	,475‡	,585‡	,445‡	,557‡	,568‡
C 4				1,000	,314	,407†	,065	,616‡	,502‡	,470‡	,477‡	,455‡
C 5					1,000	,560‡	,224	,392†	,602‡	,413†	,553‡	,562‡
C 6						1,000	,375†	.658‡	.795‡	,539‡	.862‡	.870‡
C 7							1,000	,267	,506‡	,339†	,453‡	,387†
C 8								1,000	.715‡	.782‡	.821‡	.829‡
C 9									1,000	.752‡	.892‡	.890‡
C 10										1,000	.720‡	.730‡
C 11											1,000	.989‡
C 12												1,000

Note. †  $p \leq 0,05$ . ‡  $p \leq 0,01$ .

Tableau XXIX  
Valeurs des coefficients de corrélation de Spearman pour les critères de notation de SE3N3.

Critère	C1	C2	C4	C6	C7	C8	C9	C10	C11	C12
C1	1,000	,185	,227†	,271‡	,293‡	,395‡	,395‡	,118	,249†	,172
C2		1,000	,395‡	,301‡	,316‡	,312‡	,401‡	,219†	,417‡	,383‡
C4			1,000	,296‡	,093	,342‡	,285‡	,162	,241†	,309‡
C6				1,000	,496‡	,824‡	,747‡	,415‡	,587‡	,453‡
C7					1,000	,545‡	,642‡	,407‡	,464‡	,368‡
C8						1,000	,749‡	,435‡	,590‡	,428‡
C9							1,000	,547‡	,637‡	,542‡
C10								1,000	,470‡	,525‡
C11									1,000	,581‡
C12										1,000

Note. †  $p \leq 0,05$ . ‡  $p \leq 0,01$ .

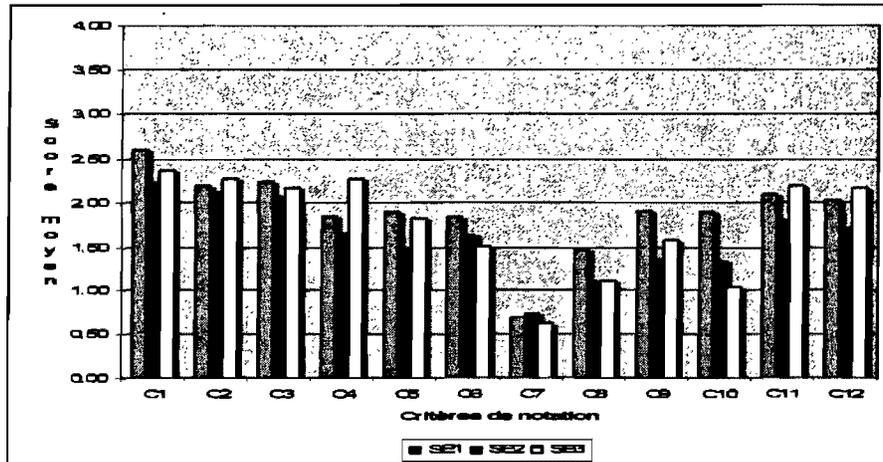
#### **2.1.4 Analyse comparative de la distribution des scores bruts moyens concernant les trois situations d'évaluation pour chacun des critères de notation**

Cette partie est consacrée aux comparaisons entre les scores obtenus aux critères de notation et aux composantes de la compétence d'investigation scientifique pour chacune des trois situations d'évaluation. La figure 19 présente trois histogrammes illustrant chacun la distribution des scores pour les douze critères de notation par situation d'évaluation et par niveau. Dans le cadre de nos analyses, nous avons comparé les scores deux à deux afin de détecter les différences statistiquement significatives. Compte tenu des caractéristiques de nos données (échantillons de taille inégale, variance inégale entre les groupes comparés), nous avons utilisé un test non paramétrique (statistique U de Mann-Whitney) avec un seuil de signification fixé à 5%. Le tableau XX présente les résultats moyens pour chacun des critères ainsi qu'un symbole distinctif afin de repérer les différences statistiquement significatives. Au total, sur les 102<sup>30</sup> paires de moyennes comparées, 29 (28,4%) d'entre elles montraient des différences significatives ( $p = 0,05$ ). Par conséquent, plus de 70% (71,6%) des comparaisons ne montraient pas de différences notables entre les moyennes comparées.

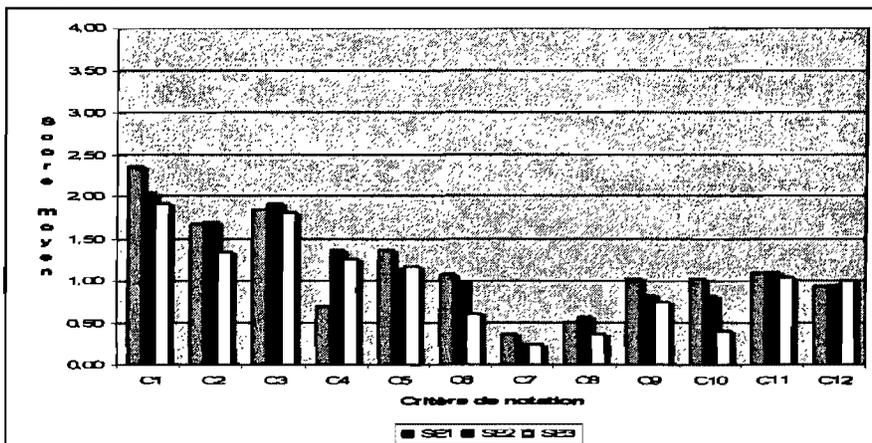
---

<sup>30</sup> Nous avons comparé 36 paires de moyenne pour le niveau 1, 36 paires de moyenne pour le niveau 2 et seulement 30 paires de moyenne pour le niveau 3. Cela s'explique par le fait que les critères 3 et 5 sont caduques pour le niveau 3 puisque la structure de l'épreuve faisait en sorte de fournir aux élèves la liste du matériel (critère 3) et le protocole d'investigation (critère 5).

## A. Niveau 1



## B. Niveau 2



## C. Niveau 3

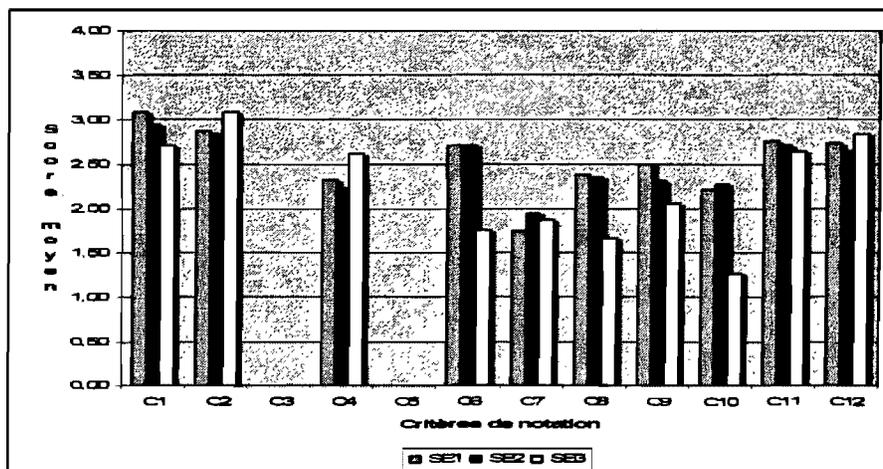


Figure 19

Distribution des scores moyens pour chacun des critères de notation de chacune des situations d'évaluation (par niveau).

Notons que les différences entre les moyennes sont plus marquées au niveau 1. À titre indicatif, on retrouve 16 paires de moyennes ayant des différences statistiquement significatives sur les 48 moyennes comparées. Au niveau 2, il n'existe que très peu de différence. Nous avons relevé seulement 3 paires de moyennes ayant des différences notables. En ce qui concerne le niveau 3, les analyses nous indiquent que 10 paires de moyennes possèdent des différences statistiquement significatives. Examinons chacun de ces histogrammes un par un afin de mieux saisir les détails de la situation globale que nous venons de décrire.

Le premier (A) présente les résultats pour le niveau 1. Débutons notre analyse en comparant les différences notables entre la situation d'évaluation 1 et la situation d'évaluation 2. En observant le graphique, on peut voir qu'il semble exister un écart important entre les scores moyens pour les critères 1, 5, 8, 9, 10 et 12. En se référant également au tableau XXX, on peut remarquer qu'il existe effectivement des différences statistiquement significatives<sup>31</sup> entre les scores obtenus à chacun de ces six critères pour le niveau 1. Pour ces critères, les scores obtenus à la SE1 sont supérieurs à ceux obtenus à la SE2. La comparaison des situations 2 et 3 révèle que cinq critères (4, 5, 10, 11, 12) possèdent des moyennes différentes. De façon générale, les scores obtenus à la SE2 sont inférieurs à ceux obtenus à la SE3 sauf dans le cas du critère 10. Dans le cas de la comparaison des scores issus de la SE1 et de la SE3, on remarque que cinq critères possèdent aussi des moyennes différentes. Il s'agit des scores moyens associés aux critères 4, 6, 8, 9 et 10. Majoritairement, les scores obtenus à la SE1 sont supérieurs à ceux obtenus à la SE3. La seule exception concerne les scores associés au critère 4 où la valeur moyenne est plus élevée dans le cas de la SE3.

Le second histogramme (B) présente la distribution des scores moyens pour les trois situations d'évaluation dans le cas du niveau 2. En comparant les situations d'évaluation 1 et 2, on remarque que seules les valeurs associées au critère 4 montrent des différences importantes. Ainsi, le score moyen pour la SE1 (0,70) est inférieur au score moyen pour

---

<sup>31</sup> Les différences ont été calculées au moyen de la statistique U de Mann-Whitney dans le cas de la comparaison de deux échantillons indépendants à un seuil de 5%.

la SE2 (1,37). En effectuant le même exercice mais - dans ce cas-ci - pour les situations 2 et 3, on peut voir que le score du critère 10 est plus élevé pour la SE2 que pour la SE3. La comparaison des moyennes pour les tâches évaluatives 1 et 3 présente sensiblement les mêmes résultats. Le score moyen pour le critère 10 est plus élevé (1,02) dans le cas de la SE1 que dans le cas de la SE3 (0,40).

Le troisième histogramme (C) présente les résultats comparatifs pour les trois situations d'évaluation pour le niveau 3. On constate d'abord que les scores sont équivalents entre la SE1 et la SE2. En jetant un coup d'œil au tableau XXX, on constate – toujours à titre indicatif – qu'il n'existe aucune différence significative pour les 10 paires de moyennes comparées. En ce qui concerne la comparaison SE2-SE3, on peut voir que la moitié (5) des critères montre une différence significative à savoir les critères 1, 4, 6, 8 et 10. Le constat est sensiblement le même en comparant les moyennes pour les situations d'évaluation 1 et 3. La moitié des critères montre également une différence significative (critères 1, 6, 8, 9 et 10).

**Tableau XXX**  
Scores moyens à chacun des critères de notation pour les trois situations d'évaluation (regroupement par niveau).

Critère	N1			N2			N3		
	SE1	SE2	SE3	SE1	SE2	SE3	SE1	SE2	SE3
C1	2,60†	2,22	2,36	2,35	2,03	1,91	3,08	2,94‡	2,71£
C2	2,19	2,12	2,27	1,67	1,69	1,34	2,87	2,83	3,08
C3	2,23	2,07	2,16	1,84	1,91	1,80	NSP	NSP	NSP
C4	1,85	1,66‡	2,27£	0,70†	1,37	1,26	2,32	2,22‡	2,61
C5	1,89†	1,49‡	1,82	1,37	1,13	1,17	NSP	NSP	NSP
C6	1,83	1,64	1,51£	1,07	0,99	0,60	2,71	2,70‡	1,75£
C7	0,68	0,73	0,61	0,37	0,25	0,23	1,74	1,94	1,87
C8	1,47†	1,11	1,11£	0,51	0,56	0,37	2,38	2,34‡	1,66£
C9	1,91†	1,36	1,58£	1,02	0,82	0,74	2,49	2,31	2,06£
C10	1,90†	1,34‡	1,04£	1,02	0,81‡	0,40£	2,21	2,27‡	1,26£
C11	2,09	1,80‡	2,19	1,09	1,09	1,03	2,75	2,70	2,64
C12	2,02†	1,72‡	2,16	0,93	0,94	1,00	2,73	2,63	2,84

Note. † :  $p \leq 0,05$  (Comparaison entre SE1 et SE2). ‡ :  $p \leq 0,05$  (Comparaison entre SE2 et SE3). £ :  $p \leq 0,05$  (Comparaison entre SE1 et SE3).

## 2.2 Interprétation des données

### L'inversion des courbes (niveau 1 et niveau 2)

Lors de la présentation des données, nous soulignons le fait que les courbes du niveau 1 et du niveau 2 étaient inversées dans le cas des trois situations d'évaluation. De fait, on constate que les scores moyens pour chacun des critères de notation sont plus faibles au

niveau 2 qu'au niveau 1. Cette situation peut s'expliquer par le fait que les scores globaux associés au niveau 1 comprennent les résultats de tous les élèves (forts, moyens et faibles). Au niveau 2, on retrouve uniquement les résultats des élèves en difficulté. Il n'est donc pas étonnant que les scores obtenus au niveau 1 soient plus élevés que ceux obtenus au niveau 2. La cohérence du modèle se voit lorsqu'on scinde les résultats globaux selon les profils d'élèves. Ces résultats sont présentés à la section 3 (Analyse des résultats pour chacun des profils de passation).

### La performance des élèves aux premiers critères

Lors de l'analyse du comportement des 12 critères d'évaluation, nous avons pu remarquer que les deux premiers critères étaient relativement bien réussis (surtout pour les niveaux 1 et 2) et ce, dans le cas des trois situations d'évaluation. Ces deux critères sont liés à la composante « cerner le problème ». Dans le cadre de cette composante, il s'agit pour l'élève de relever les principales dimensions du problème à résoudre et de débiter la structuration de la solution. Ce résultat est étonnant car on remarque souvent que les élèves éprouvent souvent des difficultés dans la structuration des éléments du problème qui leur est soumis. D'ailleurs, Dumas-Carré, Goffard et Gill (1992) identifient plusieurs difficultés liées au processus de résolution de problème dont l'une d'entre elles est justement la difficulté des élèves à se représenter le problème et à émettre des hypothèses. Dans le cadre des trois situations d'évaluation qui ont été proposées aux élèves, ces derniers ont réussi à bien circonscrire le problème.

### Les critères plus problématiques

Sauf dans un cas bien particulier (SE3N3), le critère 7 a été le critère le moins bien réussi de la part des élèves. Ce critère est associé à l'identification des causes d'erreurs lors d'une expérimentation scientifique. En se référant au document Cadre commun de résultats d'apprentissage en sciences de la nature (Conseil des ministres de l'éducation du Canada, 2003) - qui propose des attentes en ce qui concerne les apprentissages des élèves en science – on peut apprendre que cet apprentissage est à développer entre la septième et

la neuvième année (secondaire 1 à 3). On peut supposer que l'identification des causes d'erreurs n'est pas une pratique couramment développée par les enseignants du premier cycle en science. La difficulté pour les élèves (des plus faibles aux plus forts) à obtenir des résultats satisfaisants à ces critères nous permet de penser que ces aspects de la compétence n'ont pas été abordés par les enseignants. Qui plus est, les valeurs associées aux corrélations indiquent un faible lien entre le critère 7 et les onze autres critères étudiés.

#### Les valeurs corrélationnelles associées aux derniers critères

Pour pratiquement l'ensemble des neuf situations d'évaluation, on a pu remarquer que les scores pour les critères 6-8-9-10-11-12 étaient assez fortement corrélés entre eux. Pour les cinq premiers critères, nous n'avons pas observé des corrélations aussi élevées. Ces résultats nous laissent à penser que les élèves qui ont réussi à comprendre et à élaborer une démarche pertinente sont également arrivés à produire une démarche d'ensemble satisfaisante. Les premières embûches étant traversées, il leur était alors possible de répondre adéquatement à l'ensemble du problème qui leur était proposé. D'ailleurs, les critères 8, 9 et 10 sont en lien avec l'élaboration de la solution tandis que les critères 11 et 12 sont associés à la production d'un message scientifique adéquat pour rendre compte de la solution proposée. En appliquant un raisonnement similaire, on peut penser que les élèves qui maîtrisaient moins bien la compétence ont également éprouvé des difficultés dans la construction, la structuration et la communication d'une démarche et d'une solution pertinentes. Ceci expliquerait les valeurs associées aux corrélations des critères 6 à 12 sauf le critère 7.

#### L'effet du contexte disciplinaire

Quand on observe les résultats globaux aux trois situations d'évaluation par niveaux, on remarque qu'il existe peu de différences significatives entre les scores obtenues. Comme on a pu le constater lors de la présentation des données, le contexte disciplinaire qui marquait chacune des situations d'évaluation (SE1, SE2, SE3) n'a pas semblé avoir un

effet sur les résultats des élèves. Ce résultat nous apparaît étonnant puisque nous nous attendions à observer des différences notables entre chacune des situations d'évaluation. À titre d'exemple, et comme nous l'avait souligné un des experts lors de la validation, la situation d'évaluation 2 (la balle de tennis) semblait *a priori* plus difficile (principalement pour le niveau 1) que les deux autres. Puisque cette dernière ne comporte pas de contexte disciplinaire précis et prescrit dans le cadre du programme de science suivi par les élèves, nous nous attendions à voir les élèves obtenir des résultats inférieurs pour cette situation. Or, les résultats obtenus à la SE2 sont comparables aux résultats des élèves aux SE1 et SE3. En ce qui concerne les situations de niveau 1, les résultats de cette recherche nous laissent croire que les situations d'évaluation administrées mesurent essentiellement la compétence d'investigation scientifique peu importe le contexte disciplinaire qui lui est associé. Au niveau 2, nous n'avons remarqué pratiquement aucune différence entre les situations d'évaluation. Il faut préciser que les scores globaux ne nous permettent pas de voir l'effet véritable du niveau 2 auprès des élèves en difficulté. Globalement, les résultats aux situations de niveau 2 ont été très faibles pour toutes les situations ce qui explique le peu de différences. En ce qui concerne le niveau 3, on observe quelques différences significatives surtout entre les situations 2 et 3 et entre les situations 1 et 3. Ces différences nous indiquent que les résultats à la situation 3 (les lombrics) sont inférieurs à ceux obtenus aux situations 1 et 2. Ce résultat peut paraître étonnant car de l'avis de plusieurs membres du comité d'experts (qui se sont prononcés lors de la validation des situations d'évaluation) la situation d'évaluation SE3N3 semblait la plus facile parmi les neuf proposées. Lors de la correction de ces situations d'évaluation (SE3N3) nous avons pu remarquer que les élèves étaient moins rigoureux pour rendre compte de leur démarche pour cette tâche. Lors de nos visites en classe, nous avons aussi remarqué que les élèves qui accomplissaient cette situation démontraient moins de rigueur tout en exprimant verbalement les bonnes « réponses ». Par exemple, la plupart des élèves savaient que les lombrics préféraient les lieux humides et sombres. Par ailleurs, la planification de leur expérience était souvent bâclée et les résultats obtenus ne permettaient pas toujours de tirer les bonnes conclusions. Cependant, cela n'empêchait pas les élèves de tirer les conclusions adéquates quitte à ne pas s'appuyer sur les données expérimentales qu'ils avaient obtenues... Les résultats que nous obtenons nous indiquent

qu'une situation d'évaluation jugée facile par son concepteur, un expert ou un évaluateur peut donner des résultats qui ne rendent pas compte de cette « facilité ».

### **3 Analyse des résultats pour chacun des profils de passation**

#### **3.1 Analyse descriptive des données**

Lors de l'expérimentation, il y a eu quatre profils distincts de passation des situations d'évaluation. Le premier profil regroupe les élèves qui ont réalisé uniquement la situation de niveau 1 (N1) peu importe le score obtenu à ce dernier. Le second profil rassemble les élèves ayant réalisé les niveaux 1 et 2 (N1+N2). Le troisième profil réunit les élèves ayant accompli les trois niveaux (N1+N2+N3) de la situation d'évaluation qui leur était proposée. Enfin, le dernier profil est composé des élèves ayant exécutés les niveaux 1 et 3 (N1+N3). Dans les sections qui suivent, nous allons présenter les résultats aux différentes situations d'évaluation en mettant l'accent sur chacun des quatre profils de passation que nous venons tout juste de décrire.

##### **3.1.1 Analyse descriptive de la distribution des scores bruts pour la situation d'évaluation 1 (SE1) pour chacun des profils de passation**

Le tableau XXXI présente une synthèse des différentes statistiques descriptives associées aux quatre profils de passation dans le cas de la situation d'évaluation 1. On remarque que les élèves qui ont réalisé uniquement le niveau 1 (profil #1) ont obtenu un score moyen de 2,24 qui se situe au-dessus du seuil de passation. Les élèves issus du profil #2 ont rencontré, quant à eux, beaucoup plus de difficulté au niveau 1 en obtenant un score moyen de 0,36. Cependant, cette valeur augmente à 1,72<sup>32</sup> lors de l'administration du niveau 2. Le diagramme à moustaches de la figure 20 montre bien le déplacement de la distribution des scores quand ces élèves se voient administrer le niveau 2. À titre d'exemple, le score médian passe de 0,25 à 1,75 du niveau 1 au niveau 2. Dans le cas du

---

<sup>32</sup> Les moyennes pour N1 et N2 pour le profil #2 sont statistiquement différentes ( $p < 0,01$ ; test de la somme des rangs de Wilcoxon).

troisième profil, on remarque sensiblement la même tendance entre les trois niveaux. La moyenne des scores passe successivement de 0,49 pour le niveau 1 à 0,72 pour le niveau 2 à 2,08 pour le niveau 3. La figure 20 (profil #4) illustre cette progression. On peut remarquer facilement que le score médian augmente (0,33-0,54-2,32). L'amélioration la plus notable se situe lors de la réalisation du niveau 3 mais les moyennes obtenues au niveau 1 et au niveau 2 sont néanmoins statistiquement différentes ( $Z=-2,807$ ,  $p=0,005$ ) selon les résultats obtenus au test de la somme des rangs de Wilcoxon. Par ailleurs, les élèves ayant réalisé les niveaux 1 et 3 (profil #4) ont obtenu les meilleurs résultats. La moyenne pour le niveau 1 est de 2,46 et cette dernière statistique, statistiquement significative, augmente jusqu'à 2,78 lors de la passation du niveau 3.

Tableau XXXI

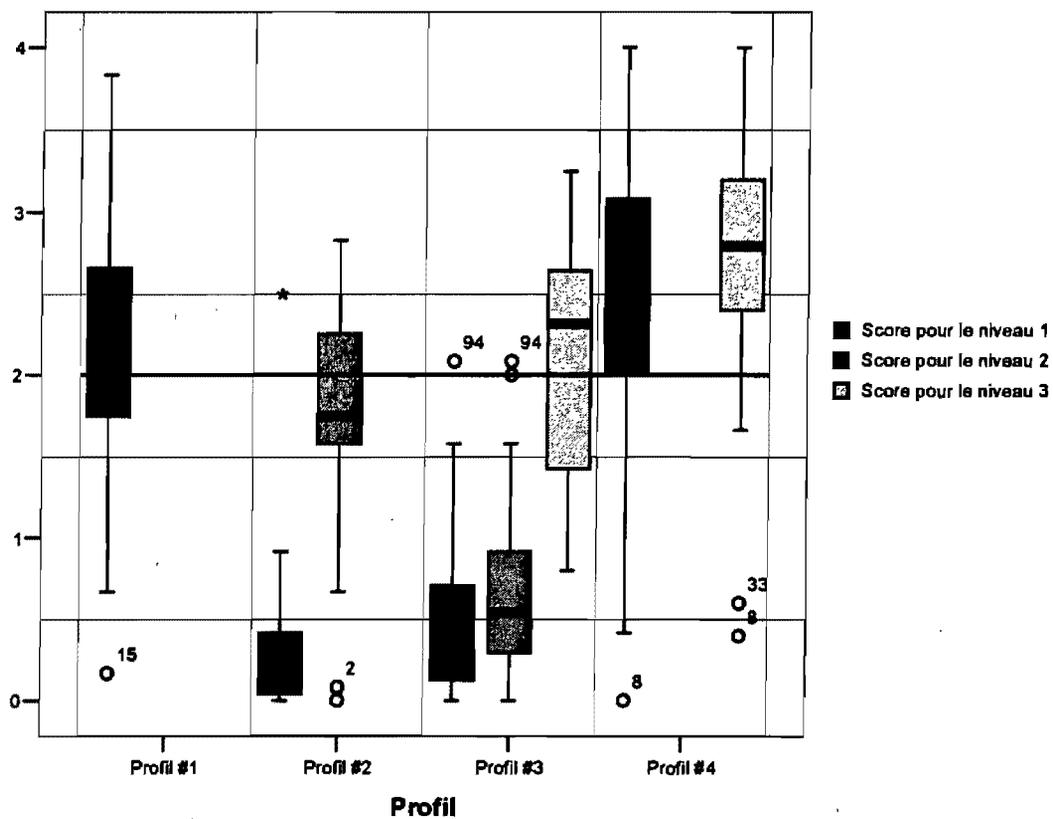
Statistiques descriptives des scores bruts pour chacun des profils de passation pour la situation d'évaluation 1 (SE1).

	Profil 1 (N=92)		Profil 2 (N=19)		Profil 3 (N=24)			Profil 4 (N=54)	
	N1	N1	N2	N1	N2	N3	N1	N3	
Moyenne	2,24	0,36†‡	1,72†‡	0,49*‡	0,72*‡	2,08*‡	2,46†‡	2,78†‡	
Écart-type	0,66	0,57	0,80	0,52	0,63	0,80	0,91	0,77	
Variance	0,44	0,33	0,65	0,27	0,39	0,64	0,83	0,56	
Médiane	2,27	0,25	1,75	0,33	0,54	2,32	2,58	2,80	
Quartiles	25	1,75	0,04	1,58	0,13	0,29	1,43	2,00	2,40
	50	2,27	0,25	1,75	0,33	0,54	2,32	2,58	2,80
	75	2,67	0,42	2,25	0,71	0,92	2,65	3,08	3,20

Note. † Test de la somme des rangs de Wilcoxon.

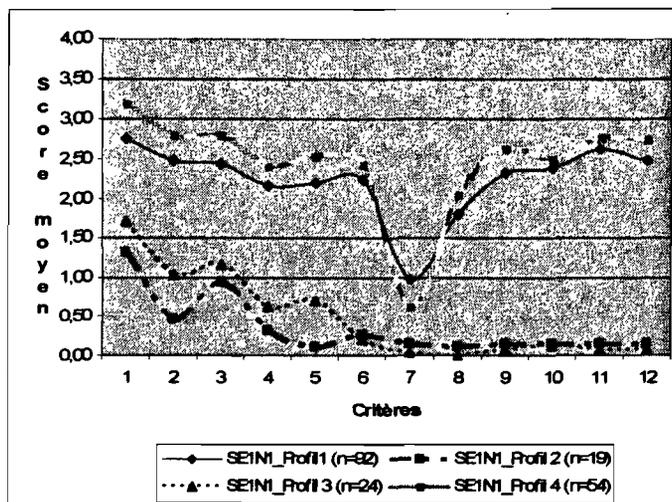
\* Test des rangs de Friedman pour k échantillons non indépendants.

‡  $p \leq 0,001$

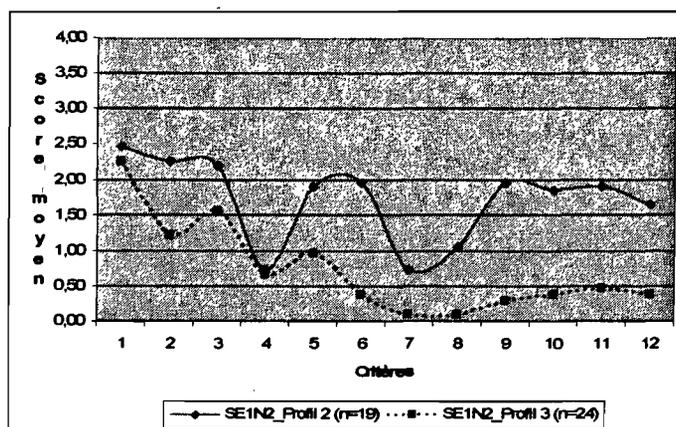


**Figure 20**  
 Distribution des scores pour la situation d'évaluation 1 (SE1) pour les niveaux 1, 2, et 3 (N1, N2 et N3) par profil de passation.

## Niveau 1



## Niveau 2



## Niveau 3

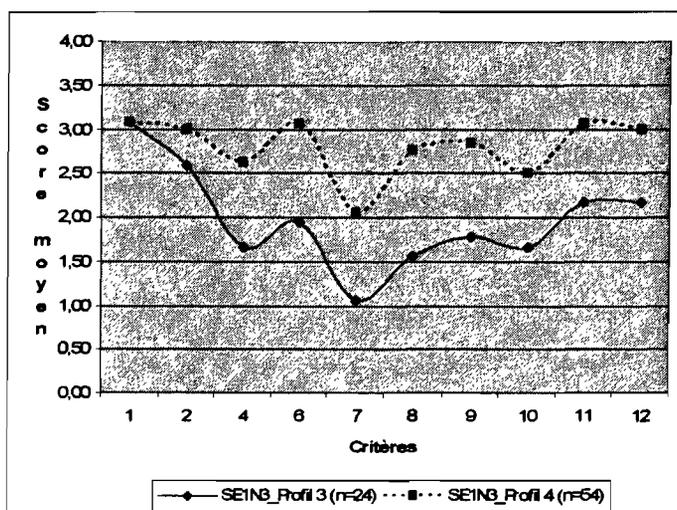


Figure 21  
Distribution des scores moyens par profil de passation et par niveau pour SE1.

La figure 22 qui suit présente les scores moyens pour chacun des 24 élèves qui ont réalisé les trois niveaux de la situation d'évaluation 1. Ce graphique permet de bien visualiser le comportement du modèle lorsque les élèves doivent recourir à tous les niveaux. On peut clairement remarquer la hiérarchie entre les niveaux puisque les scores sont généralement croissants en passant successivement du niveau 1 au niveau 2 et enfin au niveau 3. On peut également noter que les scores obtenus au niveau 3 sont nettement supérieurs à ceux obtenus aux niveaux 1 et 2. En ce qui concerne le niveau 2, on peut voir que son effet est somme toute assez faible puisque les élèves parviennent difficilement à obtenir un score moyen supérieur au seuil de réussite en se voyant octroyer ce niveau.

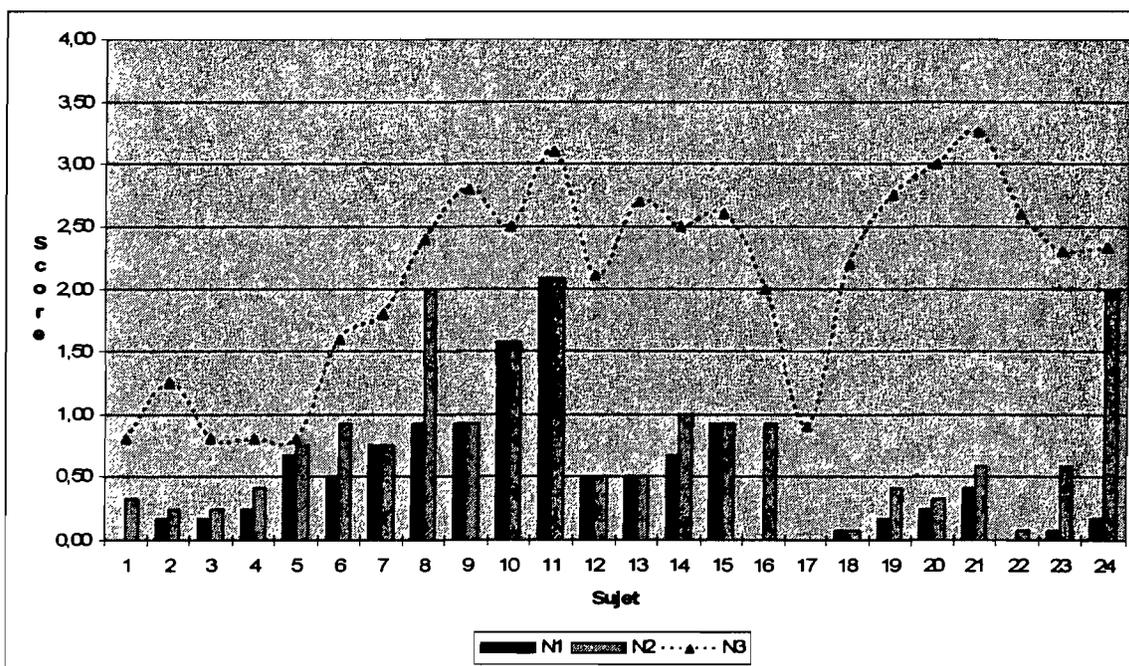


Figure 22

Distribution par niveau des scores moyens pour les élèves appartenant au profil 3 de la situation d'évaluation 1.

Outre les scores moyens, on peut également jeter un coup d'œil au taux de réussite des élèves pour chacun des profils de passation. Le tableau XXXII présente les effectifs liés à l'échec et à la réussite des élèves pour la situation d'évaluation 1. Quant à elle, la figure 23 présente le taux d'échec et de réussite pour chacun des niveaux selon une appréciation

dichotomique<sup>33</sup>. Les élèves ayant réalisés uniquement le niveau 1 (profil #1) ont obtenu un taux de réussite de 65,2% (60/92). Il en est autrement avec les élèves ayant réalisé les profils #2 et #3. Le taux de réussite pour chacun d'eux est respectivement de 5,3% (1/19) et de 4,2% (1/24). Les élèves issus du profil #4 ont eu un taux de réussite relativement semblable à ceux du profil #1 avec un taux de 75,9% (41/54). Ainsi, au terme de la passation du niveau 1, peu importe le profil de passation des épreuves, le taux de réussite se situe à 54,5% (103/189). Pour les élèves rencontrant des difficultés majeures avec le niveau 1, on constate que le niveau 2 les aide, somme toute, très peu. En effet, la passation du niveau 2 permet à 9 élèves d'atteindre le seuil de réussite tout en laissant 32 élèves toujours en situation d'échec. Le taux de réussite au terme de la passation du niveau 2 passe de 54,5% à 59,3% (112/189). En ce qui concerne le troisième niveau, plus du deux tiers des élèves (67,6%, 23/34) qui sont toujours en échec après deux niveaux arrivent alors à obtenir le seuil de passage minimum. Enfin, le taux de réussite après le troisième niveau atteint 71,4% (135/189). Cette statistique serait sans doute plus élevée si nous avions davantage d'informations pour les élèves du profil #1 qui sont en échec et qui ne se sont pas vu offrir les niveaux 2 ou 3. Ainsi, plus de la moitié (57,1%, 16,9%/28,6%) du taux d'échec global (28,6%) est composée des élèves du profil 1.

---

<sup>33</sup> Un élève est réputé être en échec si son score moyen est inférieur à 2,00 et en réussite si son score moyen est égal ou supérieur à 2,00.

Tableau XXXII

Nombre d'élèves en réussite et en échec pour chacun des profils de passation pour chacun des niveaux dans le cadre de la SE1.

Niveaux	Profil 1 (N=92)		Profil 2 (N=19)		Profil 3 (N=24)		Profil 4 (N=54)	
	Échec	Succès	Échec	Succès	Échec	Succès	Échec	Succès
N1	32	60	18	1	23	1	13	41
N2			11	7	21	2		
N3					8	13	3	10

Précisons que ces résultats seront amplement discutés au chapitre suivant car ils apportent un éclairage sur le modèle que nous proposons.

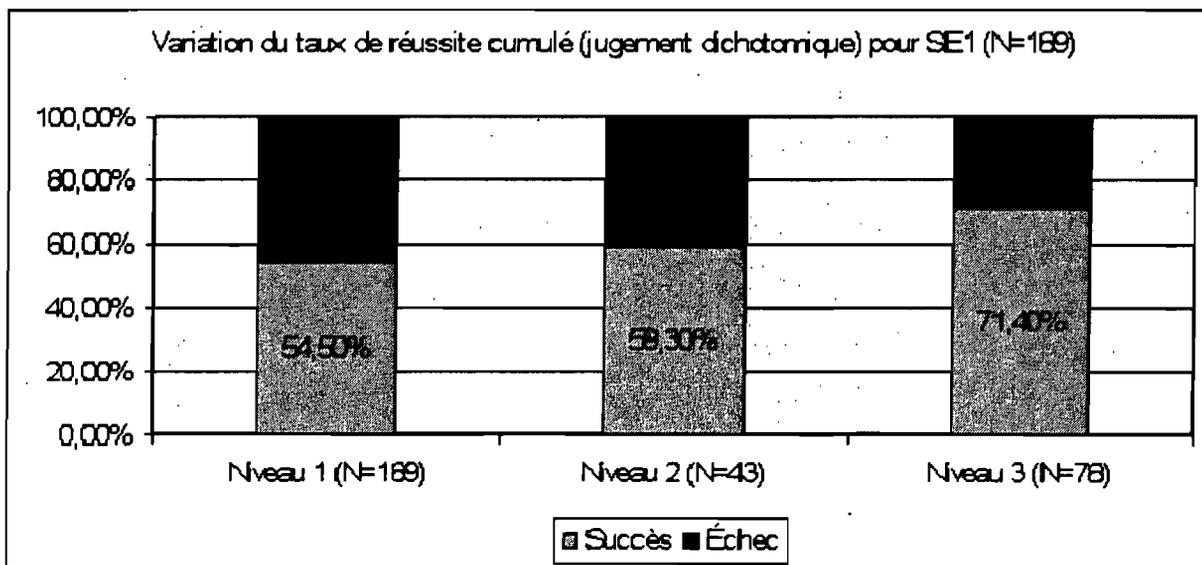


Figure 23

Variation des taux de réussite (et d'échec) pour chacun des niveaux de la situation d'évaluation 1 (SE1).

### **3.1.2 Analyse descriptive de la distribution des scores bruts pour la situation d'évaluation 2 (SE2) pour chacun des profils de passation**

En se référant au tableau XXXIII, on constate que les élèves qui ont réalisé uniquement la situation SE2N1 ont obtenu un score moyen de 2,14. Les élèves appartenant au profil #2 ont obtenu, quant à eux, des moyennes de 0,98 et 1,36 pour les niveaux 1 et 2. Il est à noter que la différence entre ces deux moyennes est non statistiquement significative ( $p = 0,123$ ). Les scores moyens colligés pour les élèves du profil #3 démontrent une tendance croissante d'un niveau à l'autre. Ces élèves ont obtenus, en moyenne, 0,51 au niveau 1, 1,10 au niveau 2 et 2,44 au niveau 3. Les différences sont statistiquement significatives à  $p \leq 0,001$ . La figure 24 permet de visualiser cette progression qui est, à nouveau, plus importante lors du passage du niveau 2 à 3 que lors du passage du niveau 1 à 2. Les scores des élèves formant le quatrième profil sont les meilleurs pour le niveau 1. Le score moyen se situe à 2,35 et le fait de réaliser le niveau 3 n'a rien apporté de vraiment nouveau puisque les moyennes obtenues aux niveaux 1 et 3 sont non statistiquement significatives ( $p = 0,062$ ).

Tableau XXXIII

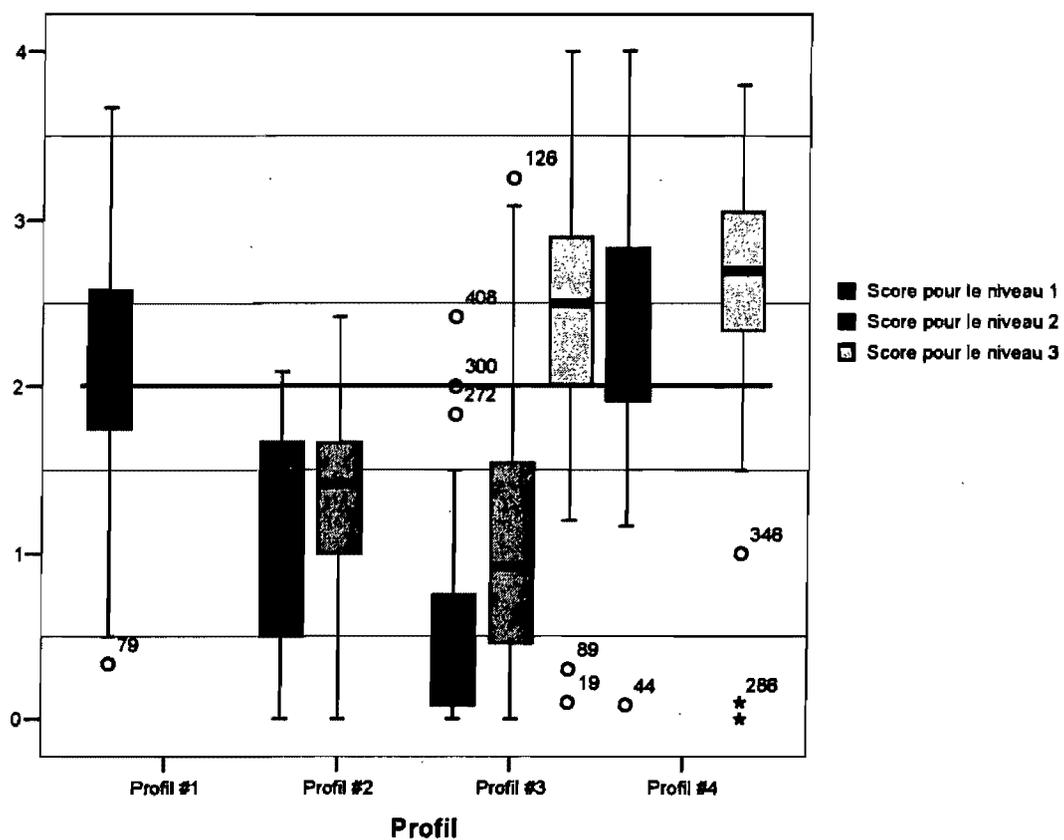
Statistiques descriptives des scores bruts pour chacun des profils de passation pour la situation d'évaluation 2 (SE2).

	Profil 1 (N=77)		Profil 2 (N=10)		Profil 3 (N=59)			Profil 4 (N=40)	
	N1	N1	N2	N1	N2	N3	N1	N3	
Moyenne	2,14	0,98†	1,36†	0,51*‡	1,10*‡	2,44*‡	2,35†	2,57†	
Écart-type	0,71	0,74	0,68	0,54	0,80	0,79	0,78	0,84	
Variance	0,50	0,55	0,46	0,29	0,64	0,62	0,61	0,71	
Médiane	2,17	0,88	1,42	0,42	0,92	2,50	2,42	2,70	
Quartiles	25	1,75	0,50	1,00	0,08	0,46	2,00	1,92	2,33
	50	2,17	0,88	1,42	0,42	0,92	2,50	2,42	2,70
	75	2,58	1,67	1,67	0,75	1,54	2,90	2,83	3,05

Note. † Test de la somme des rangs de Wilcoxon.

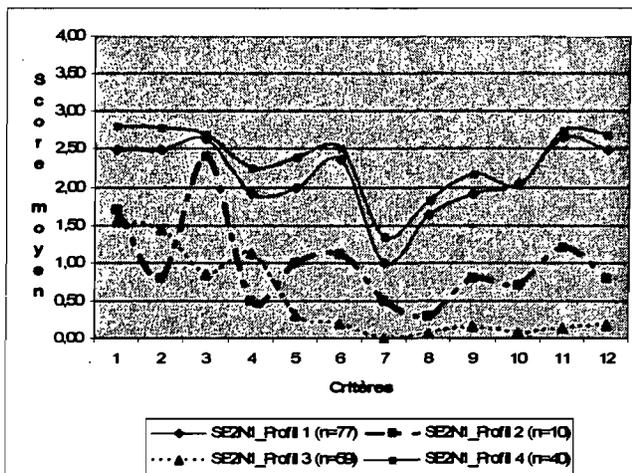
\* Test des rangs de Friedman pour k échantillons non indépendants.

‡  $p \leq 0,001$ .

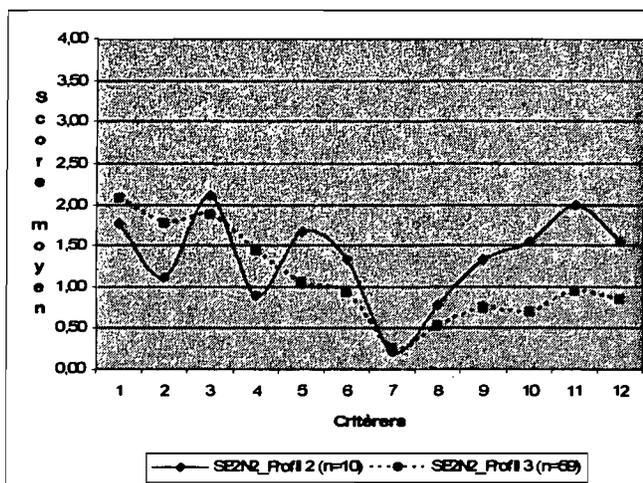


**Figure 24**  
Distribution des scores pour la situation d'évaluation 2 (SE2) pour les niveaux 1, 2, et 3 (N1, N2 et N3) par profil de passation.

Niveau 1



Niveau 2



Niveau 3

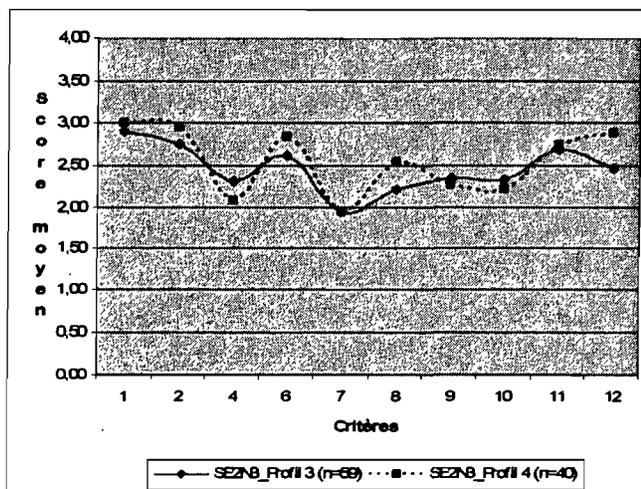


Figure 25  
Distribution des scores moyens par profil de passation et par niveau pour SE2.

La figure 26 présente la distribution des scores moyens obtenus par les élèves (n=59) ayant réalisé les trois niveaux de la deuxième situation d'évaluation. On constate à nouveau la cohérence du modèle puisqu'on observe une amélioration des scores quand les élèves passent du niveau 1 au niveau 2 et du niveau 2 au niveau 3. Dans le cas des premiers élèves (sujets 1 à 12), on peut remarquer que le niveau 2 a permis à ces derniers de réaliser une bonne partie de la situation qui leur était proposée. Pour les autres élèves, l'effet du niveau 2 demeure somme toute relativement négligeable. Lorsque les élèves sont confrontés au niveau 3, on remarque que les résultats sont très satisfaisants puisqu'ils se situent globalement au dessus du seuil de passation fixé à 2,00. Les scores pour le niveau 3 sont généralement supérieurs à ceux du niveau 2 mais on observe quelques exceptions (ex. sujets 35 ou 44).

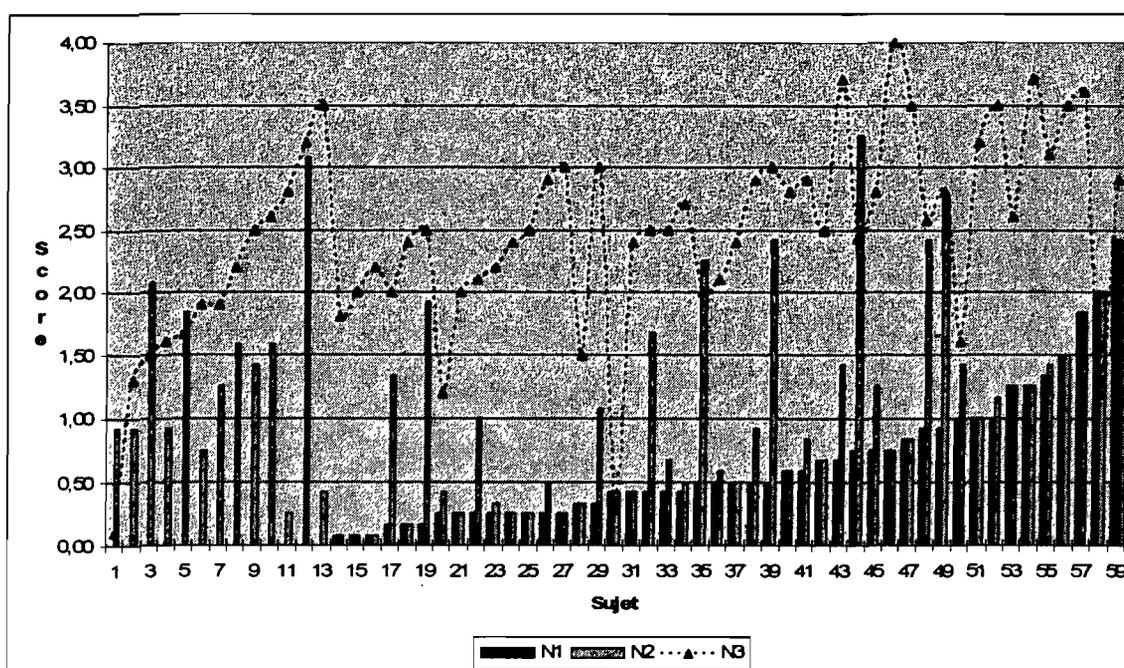


Figure 26  
Distribution par niveau des scores moyens pour les élèves appartenant au profil 3 de la situation d'évaluation 2.

En ce qui concerne le taux de réussite des élèves pour la SE2, on constate à la lecture du tableau XXXIV que 64,9% (50/77) des élèves ont réussi le niveau 1 lorsqu'on leur soumettait uniquement celui-ci. Pour les élèves ayant fait les deux premiers niveaux, le

taux de réussite chute à 10,0% (1/10). Les élèves qui ont accompli les trois niveaux obtiennent un taux de réussite de seulement 3,40%. Ceux qui ont entrepris les niveaux 1 et 3 ont obtenus, quant à eux, un taux de réussite de 67,5% (27/40). Au total, seulement 43,0% des élèves sont en situation de réussite après la passation du niveau 1 seulement. Au terme de l'administration du niveau 2, le taux de réussite passe à 47,3% (88/186) pour atteindre ultimement 72,6% après le niveau 3. La figure 27 présente la progression du taux de réussite en passant du niveau 1 jusqu'au niveau 3. À nouveau, notons que nous n'avons aucune information complémentaire pour les élèves du profil #1 qui sont en échec et qu'ils représentent, à eux seuls, un peu plus de la moitié du taux d'échec global.

Tableau XXXIV

Nombre d'élèves en réussite et en échec pour chacun des profils de passation pour chacun des niveaux dans le cadre de la SE2.

Niveaux	Profil 1 (N=77)		Profil 2 (N=10)		Profil 3 (N=59)		Profil 4 (N=40)	
	Échec	Succès	Échec	Succès	Échec	Succès	Échec	Succès
N1	27	50	9	1	57	2	13	27
N2			8	1	50	7		
N3					11	39	4	8*

Note. \*1 valeur manquante.

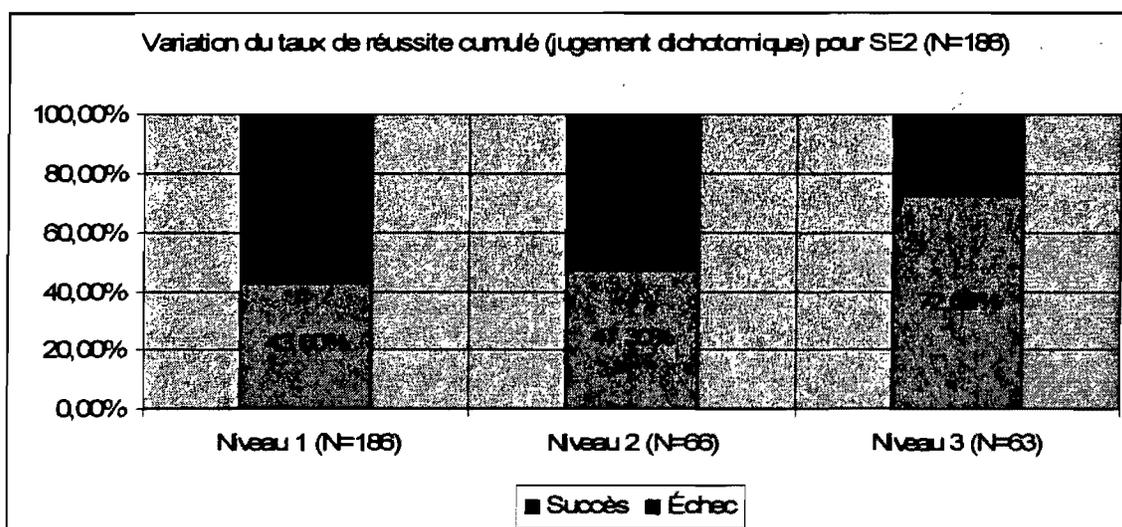


Figure 27

Variation des taux de réussite (et d'échec) pour chacun des niveaux de la situation d'évaluation 2 (SE2).

### 3.1.3 Analyse descriptive de la distribution des scores bruts pour la situation d'évaluation 3 (SE3) pour chacun des profils de passation

En prenant connaissance du tableau XXXV, on remarque que les élèves qui appartiennent au profil #1 ont obtenu une moyenne de 2,04 au niveau 1. Il est difficile de se prononcer sur les scores des élèves du profil #2 en raison de la taille de l'échantillon (N=4). Sans surprise, on remarque que la différence des moyennes est non statistiquement significative ( $p = 0,141$ ) au test non paramétrique de la somme des rangs de Wilcoxon. En ce qui concerne les élèves qui ont réalisé tous les niveaux (profil #3), on constate une tendance à la hausse des scores moyens en passant successivement du niveau 1 (0,54), au niveau 2 (0,95) au niveau 3 (2,06). Notons que la différence entre ces moyennes est statistiquement significative au seuil 0,001. La figure 28 permet de voir que les scores du niveau 3 sont beaucoup plus élevés que ceux des niveaux 1 et 2. En effet, le score médian pour le niveau 3 (2,10) se situe au-delà du seuil de passation. Pour les élèves ayant fait les niveaux 1 et 3, on peut noter que les moyennes sont élevées (N1 : 2,07 et N3 : 2,38). Bien que l'écart entre les deux moyennes soit faible, la différence est néanmoins statistiquement significative.

Tableau XXXV

Statistiques descriptives des scores bruts pour chacun des profils de passation pour la situation d'évaluation 3 (SE3).

	Profil 1 N=83)	Profil 2 (N=4)		Profil 3 (N=31)			Profil 4 (N=66)		
	N1	N1	N2	N1	N2	N3	N1	N3	
Moyenne	2,04	0,25†	1,27†	0,54*‡	0,95*‡	2,06*‡	2,07†‡	2,38†‡	
Écart-type	0,62	0,22	0,72	0,57	0,88	0,69	0,74	0,71	
Variance	0,39	0,05	0,52	0,32	0,78	0,48	0,54	0,50	
Médiane	2,08	0,25	1,33	0,33	0,50	2,10	2,17	2,30	
Quartiles	25	1,58	0,08	0,79	0,17	0,25	1,90	1,75	2,00
	50	2,08	0,25	1,33	0,25	0,50	2,10	2,17	2,30
	75	2,50	0,42	1,75	1,90	1,63	2,50	2,50	2,90

Note. † Test de la somme des rangs de Wilcoxon.

\* Test des rangs de Friedman pour k échantillons non indépendants.

‡  $p \leq 0,001$ .

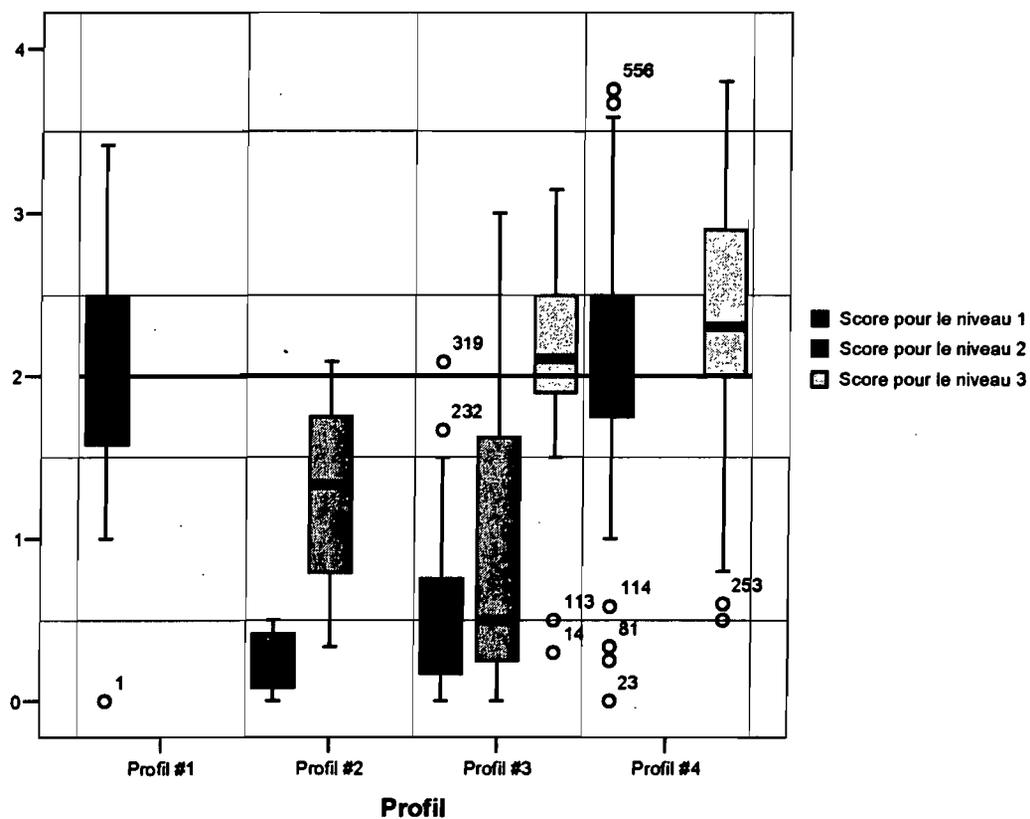
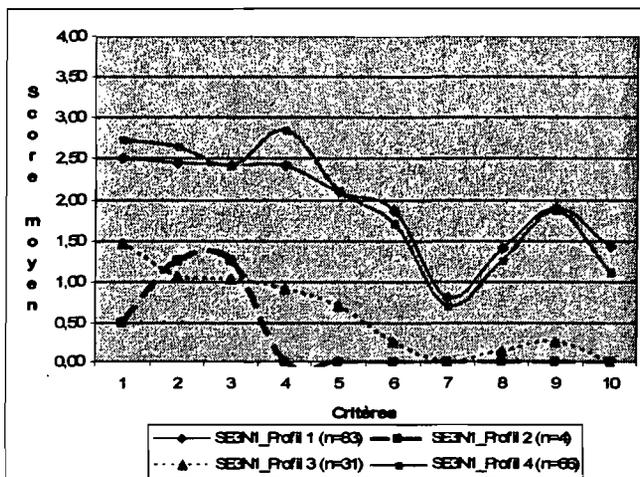
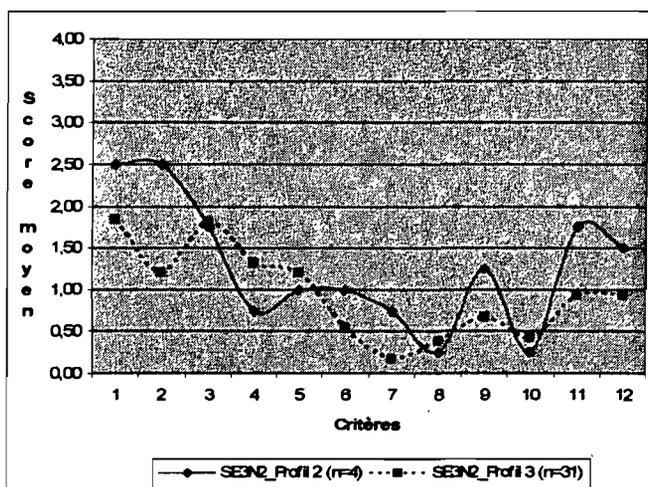


Figure 28  
Distribution des scores pour la situation d'évaluation 3 (SE3) pour les niveaux 1, 2, et 3 (N1, N2 et N3) par profil de passation.

## Niveau 1



## Niveau 2



## Niveau 3

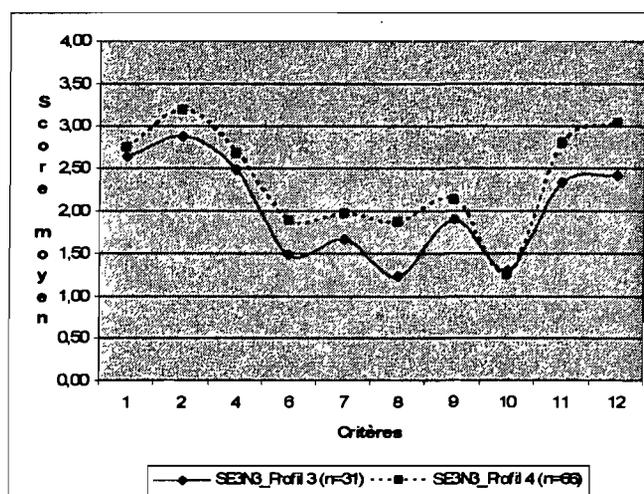


Figure 29  
Distribution des scores moyens par profil de passation et par niveau pour SE3.

La figure 30 montre la gradation des scores pour les élèves qui ont été dans l'obligation de réaliser les trois niveaux de la situation d'évaluation 3. Une fois de plus, on remarque que les scores sont généralement croissants en passant du niveau 1 au niveau 3 en passant par le niveau 2. L'effet du niveau demeure modeste pour la plupart des élèves mais il permet à certains élèves (ex. 4, 6, 8) de véritablement progresser au point où ces derniers obtiennent un score supérieur au seuil de passation. De façon analogue à ce que nous avons constaté précédemment dans le cas des situations 1 et 2, on peut remarquer que les élèves qui réalisent le niveau 3 obtiennent des résultats nettement supérieurs à ce niveau qu'aux deux niveaux précédents. Dans certains cas (ex. 4, 6, 22), le score au niveau 3 est inférieur au score obtenu au niveau 2. Ces cas demeurent somme toute rares.

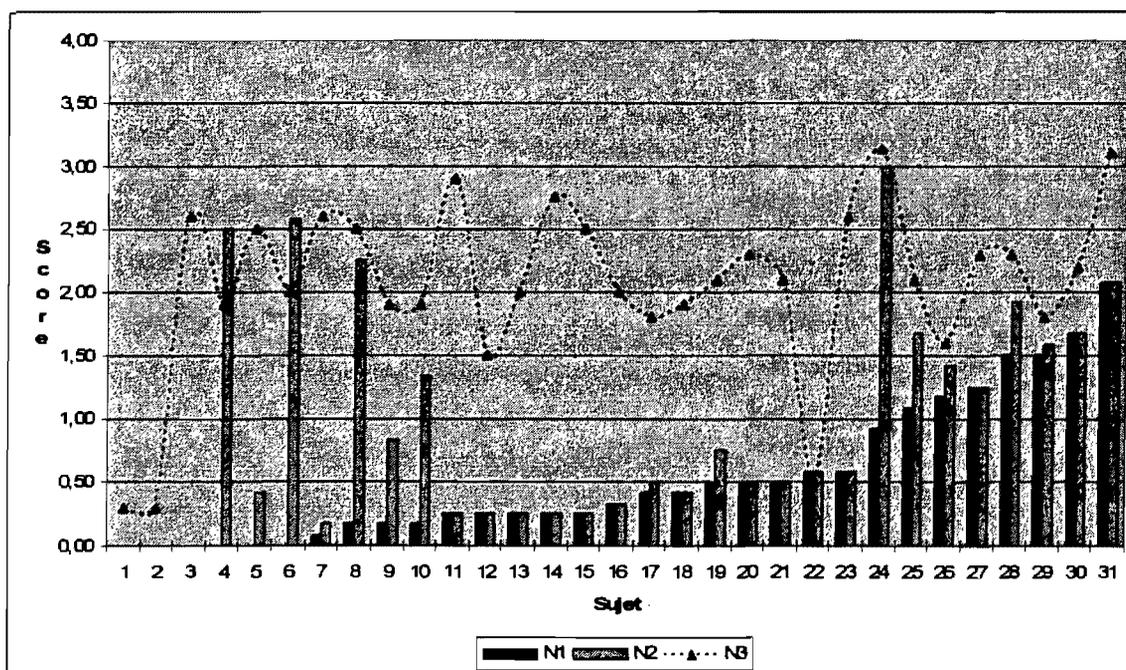


Figure 30

Distribution par niveau des scores moyens pour les élèves appartenant au profil 3 de la situation d'évaluation 3.

Le taux de réussite des élèves qui ont effectué uniquement le niveau 1 se situe à 54,2% (45/83). Le tableau XXXVI nous apprend également que les élèves qui correspondent aux profils 2 ou 3 ont obtenu un taux de réussite particulièrement faible à 2,9% (1/35). Par contre, les élèves qui appartiennent au profil #4 obtiennent un bien meilleur résultat à

60,6% (40/66) pour le niveau 1. Au terme de la passation du niveau 1, le taux de réussite global se situe à 46,5% (86/185). Le niveau 2 ne semble pas permettre d'améliorer le taux de réussite des élèves puisque seuls 5 élèves parviennent à obtenir la note de passage sur les 34 élèves éligibles pour un taux de réussite de 14,7%. Ainsi, le taux de réussite global est majoré d'à peine 2,70% en passant du niveau 1 au niveau 2. L'impact du niveau 3 est cependant plus important à cet égard. En effet, 30 élèves parviennent à obtenir le score minimal sur les 55 qui sont toujours en échec. Ce faisant, le taux de réussite atteint 61,1% à la fin de l'épreuve. La figure 31 illustre la progression du taux de réussite lors du passage d'un niveau à l'autre. On remarque que l'impact du niveau 3 sur le taux de réussite est passablement important.

Tableau XXXVI

Nombre d'élèves en réussite et en échec pour chacun des profils de passation pour chacun des niveaux dans le cadre de la SE3.

Niveaux	Profil 1 (N=83)		Profil 2 (N=4)		Profil 3 (N=31)		Profil 4 (N=66)	
	Échec	Succès	Échec	Succès	Échec	Succès	Échec	Succès
N1	38	45	4	0	30	1	26	40
N2			3	1	26	4		
N3					10	16	12	14

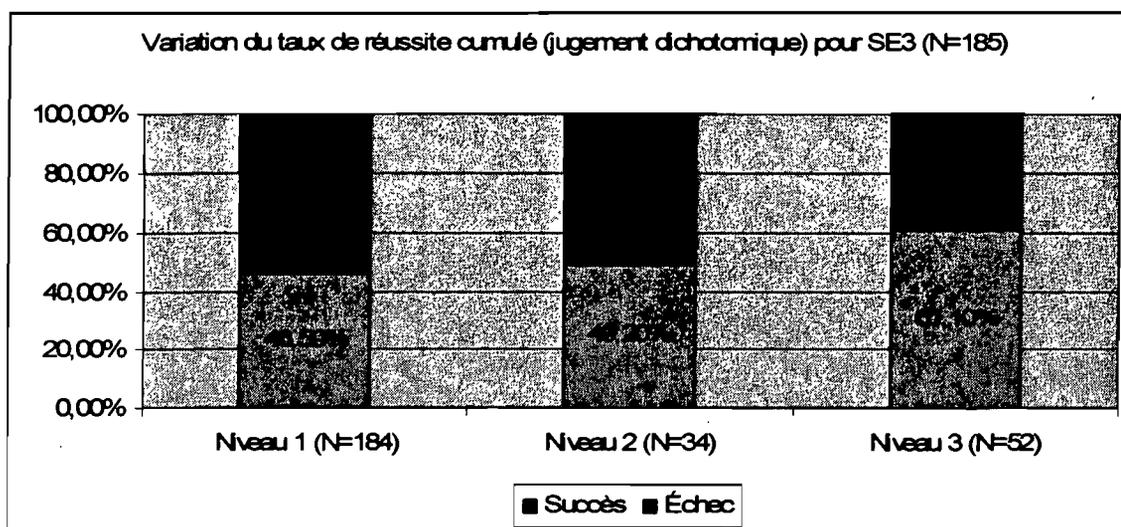


Figure 31  
Variation des taux de réussite (et d'échec) pour chacun des niveaux de la situation d'évaluation 3 (SE3).

### 3.1.4 Analyse comparative de la distribution des scores bruts pour les trois situations d'évaluation pour chacun des profils de passation

Dans la présente section, nous porterons notre attention sur la comparaison des scores bruts moyens entre chacun des quatre profils de passation. Pour ce faire, nous analyserons à nouveau les données des sections 1.3.1 à 1.3.3 en jetant, cette fois-ci, un regard transversal à travers chacun des profils de passation.

#### Profil 1

La figure 32 présente la distribution des scores concernant le niveau 1 de chacune des situations d'évaluation dans le cas spécifique du profil 1 c'est-à-dire les élèves ayant réalisé uniquement ce niveau. Au premier coup d'œil, on peut constater que les distributions sont très semblables. L'analyse des statistiques descriptives (tableau XXXVII) nous indique que la valeur associée à chacune des moyennes est relativement semblable. Les statistiques associées au test de Kruskal-Wallis nous indiquent d'ailleurs qu'il n'existe que peu de différence statistiquement significative entre chacune de ces

valeurs moyennes ( $\chi^2 = 5,114$ , ddl=2, p=0,078). Les scores médians sont également très semblables ainsi que les valeurs associées à la variance de chacune des distributions.

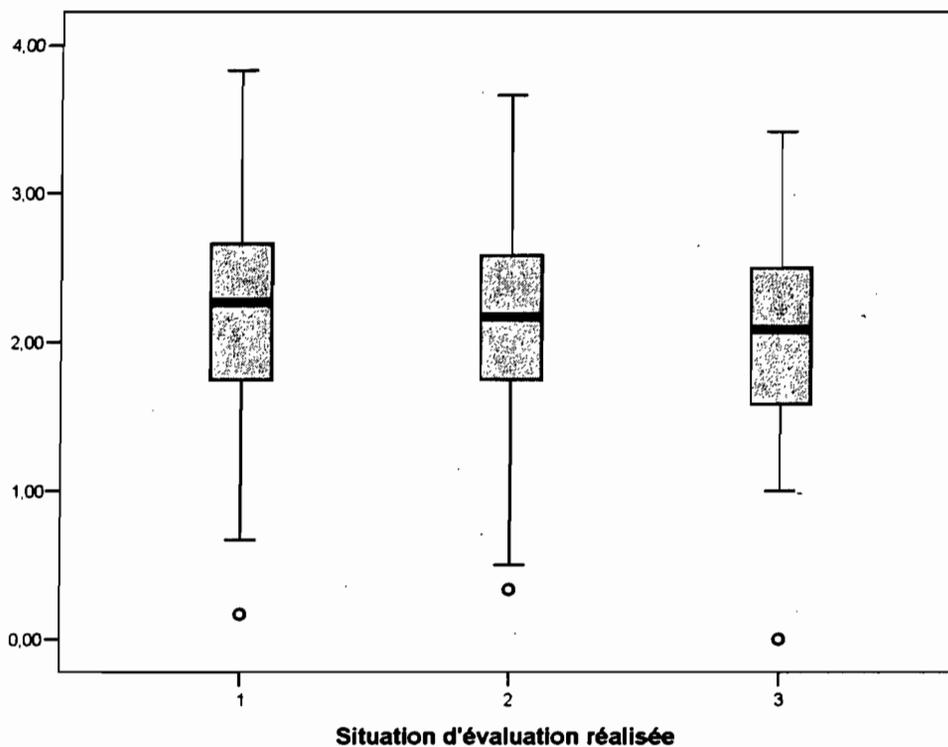


Figure 32  
Distribution des scores moyens pour chacune des situations d'évaluation (profil 1).

Tableau XXXVII

Statistiques descriptives des scores bruts pour le profil de passation #1 pour chacune des situations d'évaluation.

	SE1 (n=92)	SE2 (n=77)	SE3 (n=83)
Moyenne	2,24	2,14	2,04
Écart-type	0,66	0,71	0,62
Variance	0,44	0,50	0,39
Médiane	2,27	2,17	2,08
Quartiles	25	1,75	1,58
	50	2,27	2,08
	75	2,67	2,50

### Profil 2

En ce qui concerne le profil 2, on peut voir en regardant le tableau XXXVIII que les effectifs sont faibles pour chacune des SE. Malgré cela, il est néanmoins possible de recourir à un test non paramétrique pour comparer les moyennes (test de Mann-Whitney). En comparant la première situation d'évaluation avec la deuxième, on constate que la différence entre les moyennes pour le niveau 1 est significative ( $Z=-2,268$ ,  $p=0,024$ ). Cependant, les moyennes sont comparables dans le cas du niveau 2 ( $Z=-1,577$ ,  $p=0,117$ ). En ce qui concerne les situations 1 et 3, les valeurs associées à la statistique de Mann-Whitney nous indiquent que les moyennes sont semblables. En effet, dans le cas du niveau 1, la valeur de la statistique est de  $-0,82$  ce qui correspond à une probabilité de  $0,969$ . Pour le niveau 2, la statistique correspond à  $-1,301$ . Cette statistique est non significative ( $0,218$ ) au seuil de  $5\%$ . Enfin, la comparaison statistique des moyennes entre les SE 2 et 3 montre qu'il n'existe que peu de différence entre les moyennes. Pour le niveau 1, la statistique associée ( $Z=-1,778$ ) est non significative ( $p=0,076$ ) de même que pour le niveau 2 ( $Z=-0,232$ ,  $p=0,825$ ).

Tableau XXXVIII

Statistiques descriptives des scores bruts pour le profil de passation #2 pour chacune des situations d'évaluation.

	SE1		SE2		SE3		
	(n=19)		(n=10)		(n=4)		
	N1	N2	N1	N2	N1	N2	
Moyenne	0,36	1,72	0,98	1,36	0,25	1,27	
Écart-type	0,57	0,80	0,74	0,68	0,22	0,72	
Variance	0,33	0,65	0,55	0,46	0,05	0,52	
Médiane	0,25	1,75	0,88	1,42	0,25	1,33	
Quartiles	25	0,04	1,58	0,50	1,00	0,08	0,79
	50	0,25	1,75	0,88	1,42	0,25	1,33
	75	0,42	2,25	1,67	1,67	0,42	1,75

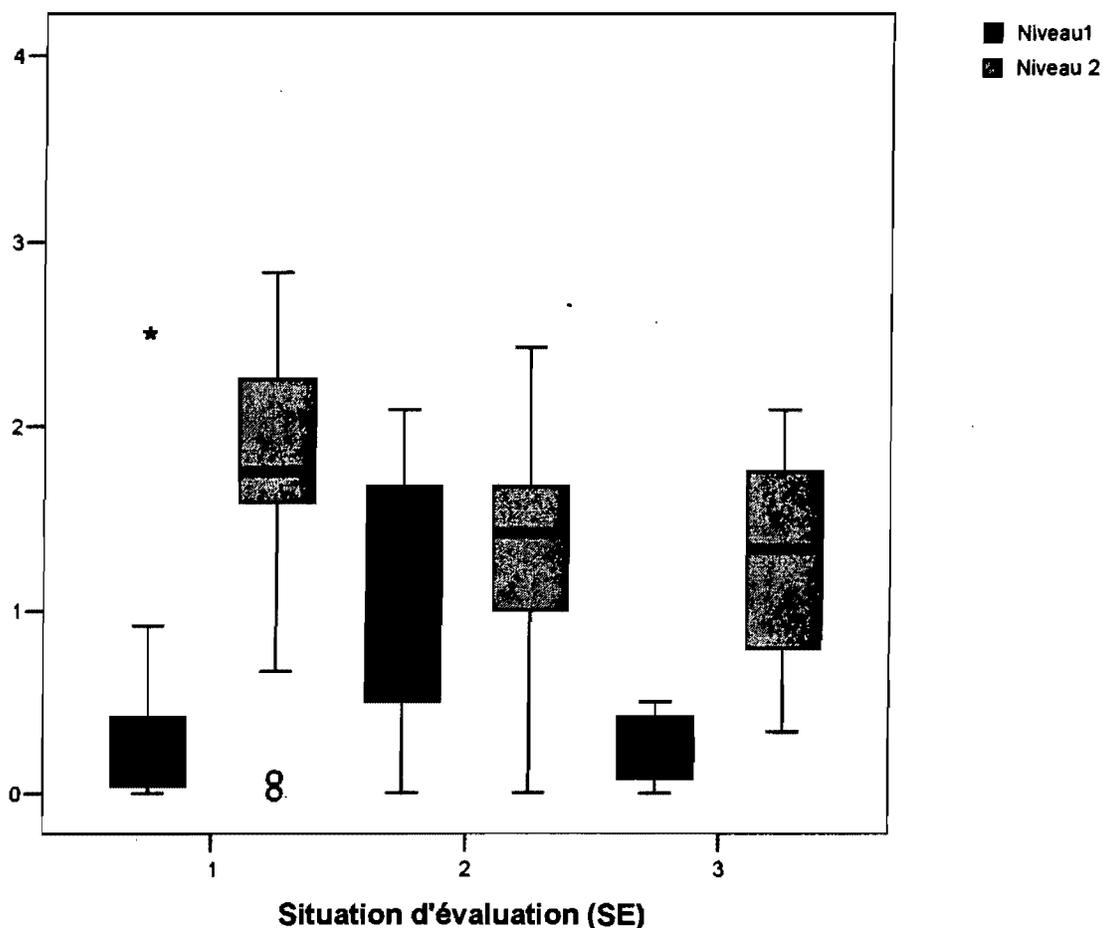


Figure 33  
Distribution des scores moyens pour chacune des situations d'évaluation (profil 2).

### Profil 3

La figure 34 illustre la distribution des données pour chacun des trois niveaux de chacune des situations d'évaluation pour les élèves qui ont réalisé les trois niveaux (profil 3). On peut constater que la distribution des scores pour les niveaux 1 à 3 est très similaire pour chacune des SE. L'analyse statistique des moyennes nous révèle qu'il existe bien peu de différence entre les moyennes pour les niveaux 1 et 2. Le test de Kruskal-Wallis sur l'égalité des moyennes, appliqué aux données du niveau 1, nous indique une valeur de khi-deux de 0,049 avec 2 degrés de liberté. Cette statistique est non significative ( $p=0,976$ ) au seuil de 5%. On peut donc considérer les valeurs moyennes comme étant

équivalentes entre elles. Il en est de même avec les données du niveau 2 puisque la valeur du khi-carré vaut 4,054 avec 2 degrés de liberté. Cette statistique est également non significative ( $p=0,132$ ) au seuil de 5%. Pour le niveau 3, l'écart entre les moyennes semble plus important. Pour le niveau 3, la valeur de la statistique ( $\chi^2=6,279$ , 2 ddl) est significative ( $p=0,043$ ) suggérant ainsi qu'il existe vraiment une différence entre les moyennes.

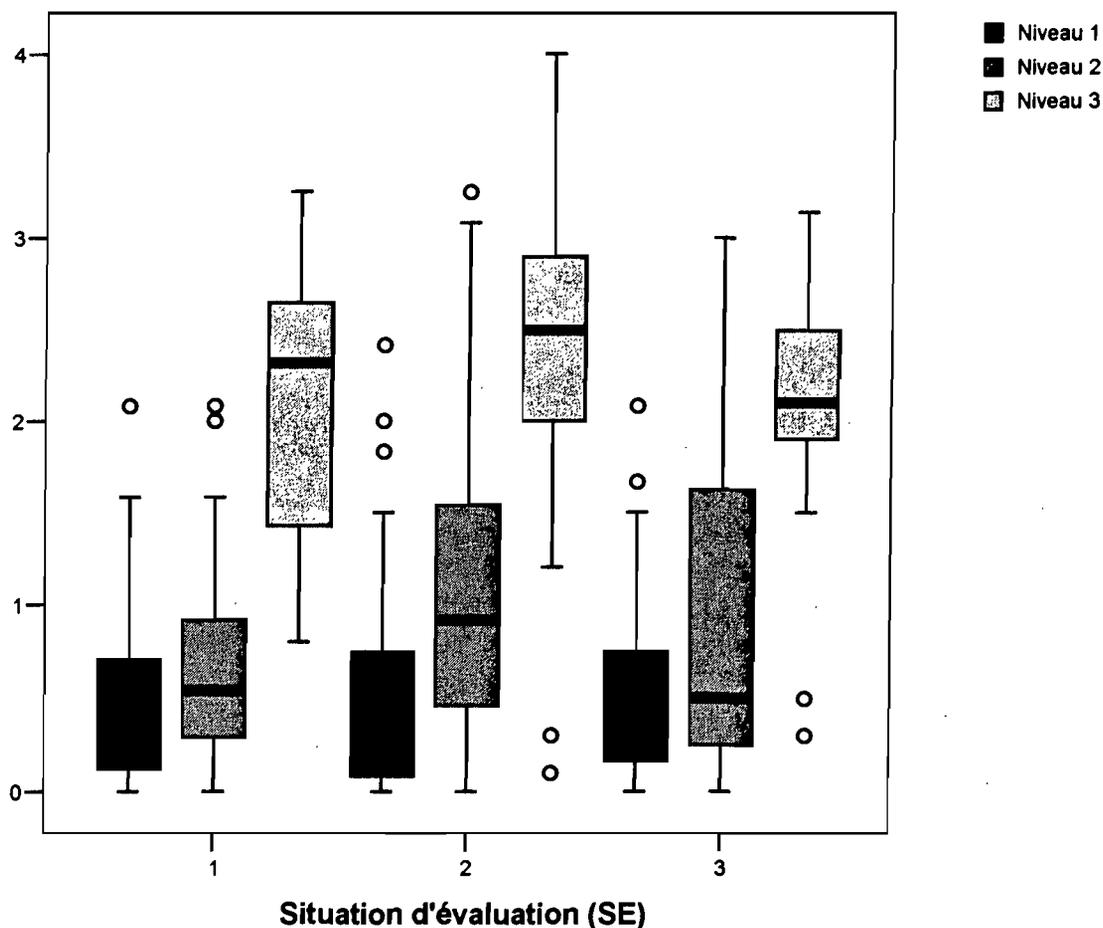


Figure 34  
Distribution des scores moyens pour chacune des situations d'évaluation (profil 3).

Tableau XXXIX

Statistiques descriptives des scores bruts pour le profil de passation #3 pour chacune des situations d'évaluation.

	SE1 (n=24)			SE2 (n=59)			SE3 (n=31)			
	N1	N2	N3	N1	N2	N3	N1	N2	N3	
Moyenne	0,49	0,72	2,08	0,51	1,10	2,44	0,54	0,95	2,06	
Écart-type	0,52	0,63	0,80	0,54	0,80	0,79	0,57	0,88	0,69	
Variance	0,27	0,39	0,64	0,29	0,64	0,62	0,32	0,78	0,48	
Médiane	0,33	0,54	2,32	0,42	0,92	2,50	0,33	0,50	2,10	
Quartiles	25	0,13	0,29	1,43	0,08	0,46	2,00	0,17	0,25	1,90
	50	0,33	0,54	2,32	0,42	0,92	2,50	0,25	0,50	2,10
	75	0,71	0,92	2,65	0,75	1,54	2,90	1,90	1,63	2,50

#### Profil 4

Le profil 4 correspond aux élèves ayant réalisé les niveaux 1 et 3 seulement. La figure 35 présente la distribution des scores moyens pour les niveaux 1 et 3 et ce, pour chacune des situations d'évaluation. L'allure générale des boîtes à moustaches est assez semblable malgré certaines différences visibles à l'œil nu. Par exemple, on peut remarquer que la position des diagrammes en forme de moustaches de la SE3 se situe à un niveau beaucoup plus bas sur l'échelle des ordonnées que la SE1. À partir des valeurs moyennes du tableau XL, nous avons calculé les statistiques U de Mann-Whitney afin de comparer ces valeurs. En ce qui concerne le niveau 1, les différences entre les situations 1 et 2 sont non statistiquement significatives ( $Z=-0,907$ ,  $p=0,364$ ). Par contre, la différence entre les moyennes pour les SE 2 et 3 ainsi que pour les SE 1 et 3 sont statistiquement significatives (SE2-3,  $Z=-2,039$ ,  $p=0,041$  ; SE1-3,  $Z=-2,952$ ,  $p=0,003$ ). Pour le niveau 3, on constate à peu près la même chose. Les moyennes sont équivalentes entre les SE 1 et 2 ( $Z=-1,037$ ,  $p=0,300$ ) alors qu'elles sont différentes entre les SE 1 et 3 ( $Z=-3,157$ ,  $p=0,002$ ) et entre les SE 2 et 3 ( $Z=-2,075$ ,  $p=0,038$ ).

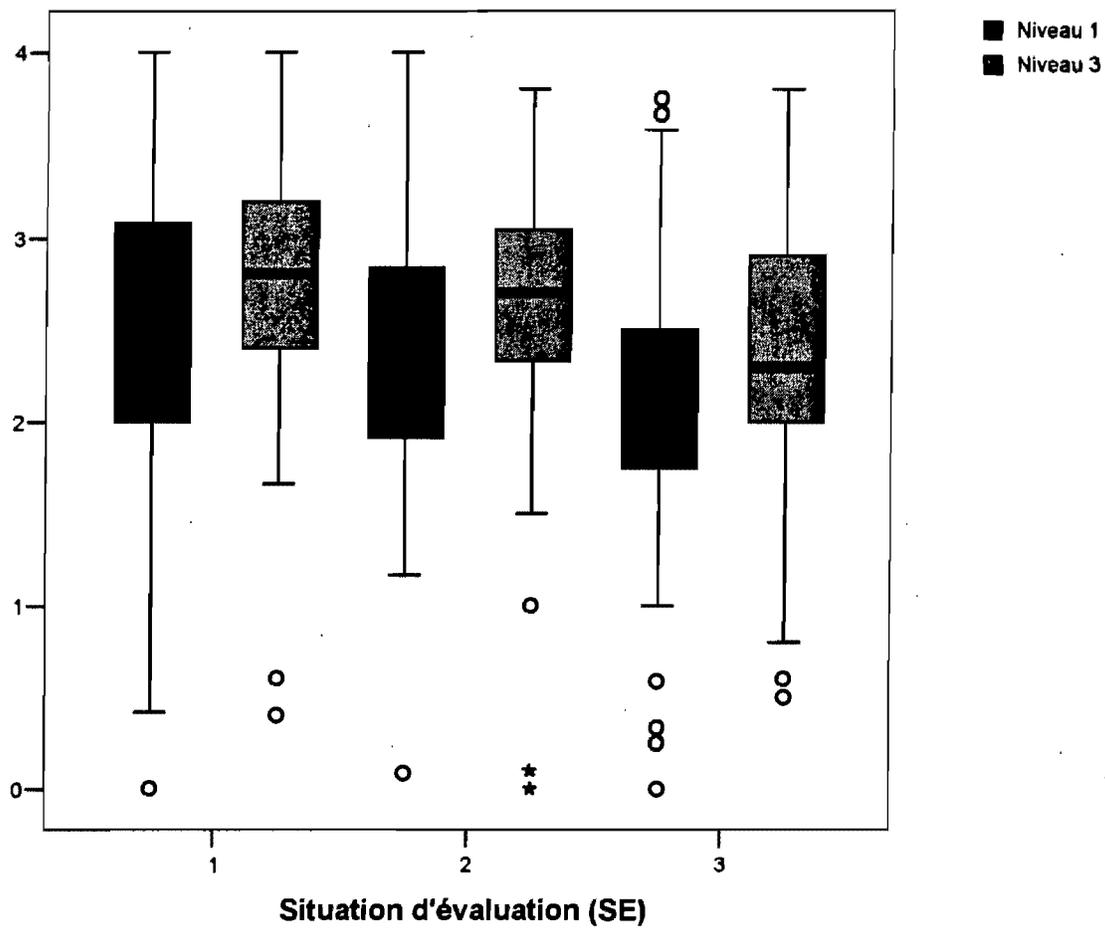


Figure 35  
Distribution des scores moyens pour chacune des situations d'évaluation (profil 4).

Tableau XL

Statistiques descriptives des scores bruts pour le profil de passation #4 pour chacune des situations d'évaluation.

	SE1		SE2		SE3		
	(n=54)		(n=40)		(n=66)		
	N1	N3	N1	N3	N1	N3	
Moyenne	2,46	2,78	2,35	2,57	2,07	2,38	
Écart-type	0,91	0,77	0,78	0,84	0,74	0,71	
Variance	0,83	0,56	0,61	0,71	0,54	0,50	
Médiane	2,58	2,80	2,42	2,70	2,17	2,30	
Quartiles	25	2,00	2,40	1,92	2,33	1,75	2,00
	50	2,58	2,80	2,42	2,70	2,17	2,30
	75	3,08	3,20	2,83	3,05	2,50	2,90

### 3.2 Interprétation des données

#### Les profils de passation et les profils d'élèves associés

Lors de l'analyse des données, nous avons identifié quatre profils distincts de passation. De ces profils, on peut dégager 3 catégories d'élèves associés à ces profils de passation :

- (1) les élèves qui maîtrisent déjà bien les compétences (profils 1 et 4)
- (2) les élèves présentant certaines difficultés mais qui arrivent à démontrer leurs compétences avec un peu d'aide (profil 2)
- (3) les élèves en très grande difficulté qui ont besoin de recourir aux trois niveaux (profil 3).

L'analyse de ces différents profils de passation permet de mettre en lumière l'effet du modèle sur différentes clientèles d'élèves. Examinons chacun de ces profils en discutant des aspects les plus pertinents.

Pour les élèves du profil 4, l'administration des niveaux 2 et 3 s'avère inutile. En effet, certains enseignants ont quand même administré le niveau 3 aux élèves qui avaient réussi le niveau 1 (profil 4). Or, les résultats indiquent que ces élèves ont obtenu des résultats comparables pour le niveau 1 et le niveau 3. Ce résultat est particulièrement intéressant au point de vue de la validité. On peut penser que les résultats obtenus au niveau 1 se comparent à ceux obtenus au niveau 3 ce qui rend superflue l'administration de ce dernier. Rappelons que plusieurs enseignants ont administré le niveau 3 aux élèves ayant réussi le niveau 1 pour voir si ces derniers obtiendraient des résultats équivalents en réalisant une situation d'évaluation plus ouverte et une situation d'évaluation plutôt fermée.

Les résultats obtenus par les élèves du profil 2 montrent que le niveau 2 ne fournit pas les informations nécessaires pour aider les élèves à réussir convenablement la situation qui leur est présentée. Peu importe le contexte disciplinaire, les élèves de ce profil n'ont pas réussi à obtenir un résultat moyen se situant au-delà du seuil de passation. La situation est identique pour les élèves du profil 3 quand on leur a administré le niveau 2. Par rapport à ce constat, nous entrevoyons deux explications possibles. D'abord, nous pensons que l'aide fournie aux élèves dans le cadre du niveau 2 était, peut-être, soit insuffisante ou soit mal adaptée à cette clientèle. Les indices donnés aux élèves concernaient surtout les aspects méthodologiques liés à la compétence d'investigation scientifique. Quoi qu'il en soit, il serait fort pertinent d'ajuster les indices fournis au niveau 2 et examiner si la situation s'améliore. Ensuite, il est possible que le niveau de compétence des élèves soit si faible qu'il faille alors les guider pas-à-pas dans la réalisation d'une démarche d'investigation scientifique comme dans le cas des situations de niveau 3.

Dans le cas des trois situations d'évaluation, nous avons pu observer que l'utilité du niveau 3 est indéniable pour les élèves montrant de grandes difficultés. Dans tous les cas, les élèves ont obtenu des résultats moyens qui se situaient au-dessus du seuil de passation. Contrairement au niveau 2, le niveau 3 permet également de considérablement diminuer le taux d'échec.

En terminant, la distribution des scores illustrée à la figure 33 permet de constater que les trois situations d'évaluation, dans le cas du profil 1, ressemblent à des versions parallèles (moyenne et variance semblables) d'un même instrument de mesure. On peut penser que des élèves maîtrisant généralement bien la compétence pourraient obtenir des résultats semblables à ce genre d'évaluation peu importe le contexte disciplinaire. Nous reviendrons sur ce résultat au chapitre 5.

#### **4. Analyse détaillée des résultats par composantes et critères de notation pour chacun des niveaux : comparaison des élèves en réussite et des élèves en difficulté**

##### **4.1 Analyse descriptive des données**

Dans les prochaines sections, nous nous attarderons aux résultats obtenus à chacun des critères de notation. Cet exercice nous semble incontournable afin de vérifier à quels endroits les élèves rencontrent le plus de difficultés et voir si l'évaluation à niveaux multiples a permis d'apporter des informations pertinentes. Pour ce faire, nous avons scindé les résultats des élèves en deux catégories : celle regroupant les élèves en réussite à chacune des situations d'évaluation et celle formée des élèves en échec (ayant obtenu un score moyen inférieur à 2,00). Par la suite, nous avons comparé les deux groupes afin de détecter les similitudes et les différences.

##### **4.1.1 Analyse descriptive de la distribution des scores bruts moyens pour la situation d'évaluation 1 (SE1) par composantes et critères de notation pour chacun des niveaux**

Le tableau XLI qui suit présente les scores moyens des élèves en réussite et en échec pour chacun des critères de notation pour les trois niveaux de la SE1. La figure 36 illustre l'évolution des scores moyens pour chacun des critères. Une première constatation concerne le critère 7 (identification des causes d'erreurs). On remarque que même les élèves en réussite éprouvent des difficultés à obtenir un score satisfaisant. Ainsi, aux niveaux 1 et 2 les élèves en réussite n'obtiennent pas la note de passage (N1 : 1,19 et N2 :

1,27) alors qu'au niveau 3, ils l'obtiennent de justesse (N3 : 2,06). De façon générale, autant les élèves en échec que ceux en réussite rencontrent des difficultés de plus en plus importantes en passant du critère 1 au critère 7. À la lecture de la figure 36, on constate que la courbe des scores descend du critère 1 au critère 7 et remonte, par la suite, du critère 8 au critère 12. Notons que les élèves qui sont globalement en situation d'échec le sont également pour, à toute fin pratique, tous les critères de notation. Sur la figure 36, on peut voir que la courbe des scores moyens pour les élèves en échec se situe presque toujours sous le seuil de passation (2,00). Seuls les critères 1 (N2 et N3) et 2 (N3) y font exception.

Tableau XLI

Scores moyens pour chacun des critères d'évaluation pour les élèves en réussite et en échec à chacun des niveaux de la SE1.

Critères	SE1N1 (N=189)			SE1N2 (N=43)			SE1N3 (N=78)		
	Réussite (N=103)	Échec (N=86)	Écart	Réussite (N=11)	Échec (N=32)	Écart	Réussite (N=65)	Échec (N=13)	Écart
C1	3,26	1,80	1,46	3,00	2,13	0,87	3,15	2,69	0,46
C2	2,97	1,24	1,73	2,82	1,28	1,54	3,00	2,23	0,77
C3	2,83	1,51	1,32	2,73	1,53	1,20	NSP	NSP	NSP
C4	2,60	0,94	1,66	1,27	0,50	0,77	2,54	1,23	1,31
C5	2,56	1,09	1,47	2,27	1,06	1,21	NSP	NSP	NSP
C6	2,77	0,71	2,06	2,36	0,63	1,73	3,17	0,67	2,50
C7	1,19	0,08	1,11	1,27	0,06	1,21	2,06	0,36	1,70
C8	2,38	0,40	1,98	1,27	0,25	1,02	2,89	0,27	2,62
C9	2,97	0,65	2,32	2,55	0,50	2,05	2,98	0,55	2,43
C10	2,90	0,72	2,18	2,55	0,50	2,05	2,62	0,50	2,12
C11	3,08	0,91	2,17	2,82	0,50	2,32	3,17	1,18	1,99
C12	2,99	0,85	2,14	2,64	0,34	2,30	3,19	0,91	2,28

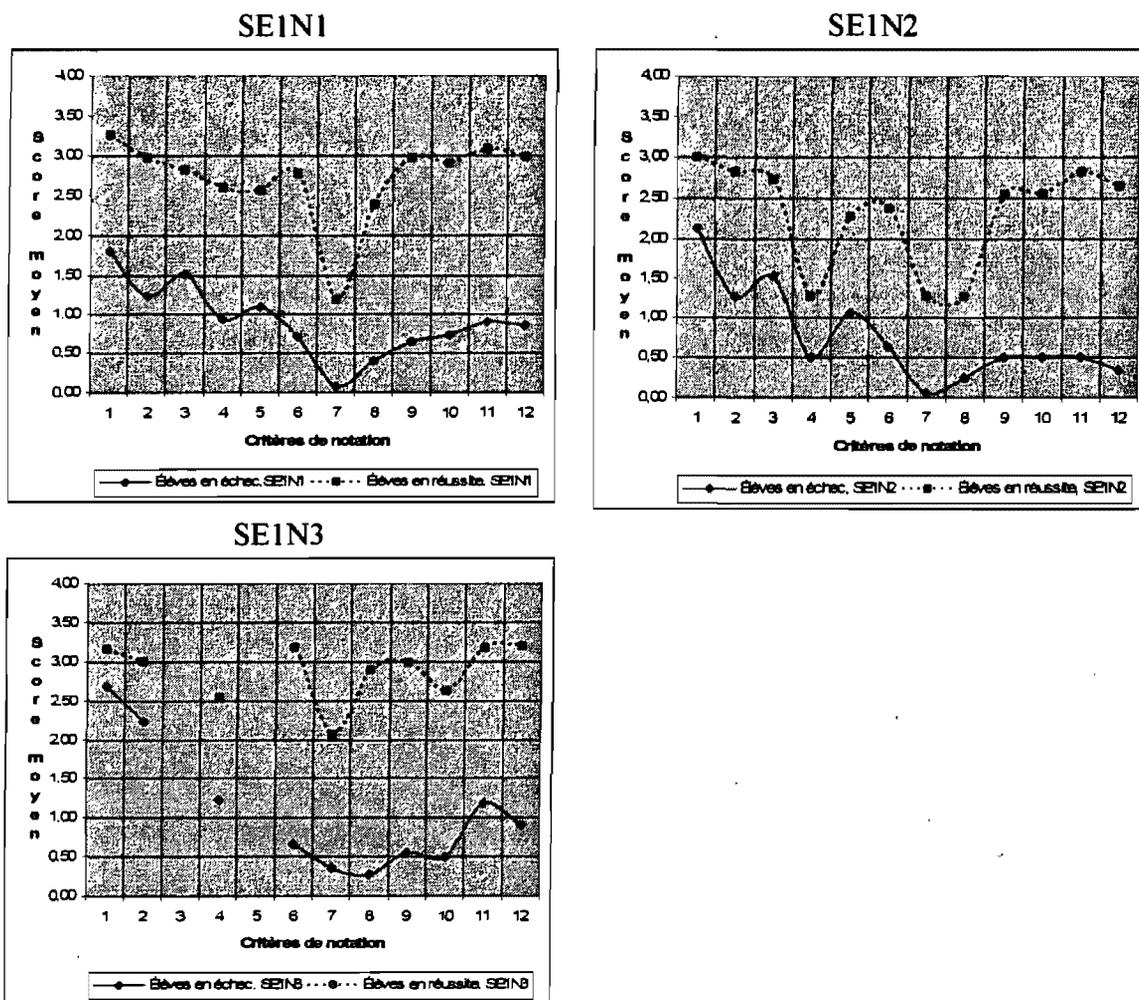


Figure 36

Variation du score moyen des élèves en échec et en réussite pour chacun des niveaux de la situation d'évaluation 1.

Toujours sur la même figure, on peut apercevoir que la distance entre les courbes n'est pas toujours constante pour chacun des niveaux. Afin de mieux visualiser l'écart entre les deux courbes en question, on peut se référer à la figure 37 qui présente la variation de l'écart entre les courbes des scores des élèves en réussite et ceux en échec pour chacune des versions (SE1N1, SE1N2, SE1N3) de la situation d'évaluation 1. Dans le cas de la SE1N1, on peut voir que l'écart entre les courbes est relativement constant pour les critères 1 à 4 et pour les critères 9 à 12. Quant à lui, le critère 7 est celui pour lequel il subsiste le moins de différence entre les élèves en échec et ceux en réussite. En ce qui

concerne la situation SE1N2, l'allure générale de la courbe ressemble à celle de la précédente. Cependant, on note des fluctuations plus importantes entre autres pour les critères 1, 4 et 8. En ce qui concerne le niveau 3, on peut distinguer un accroissement très important de l'écart pour les critères 6 à 12.

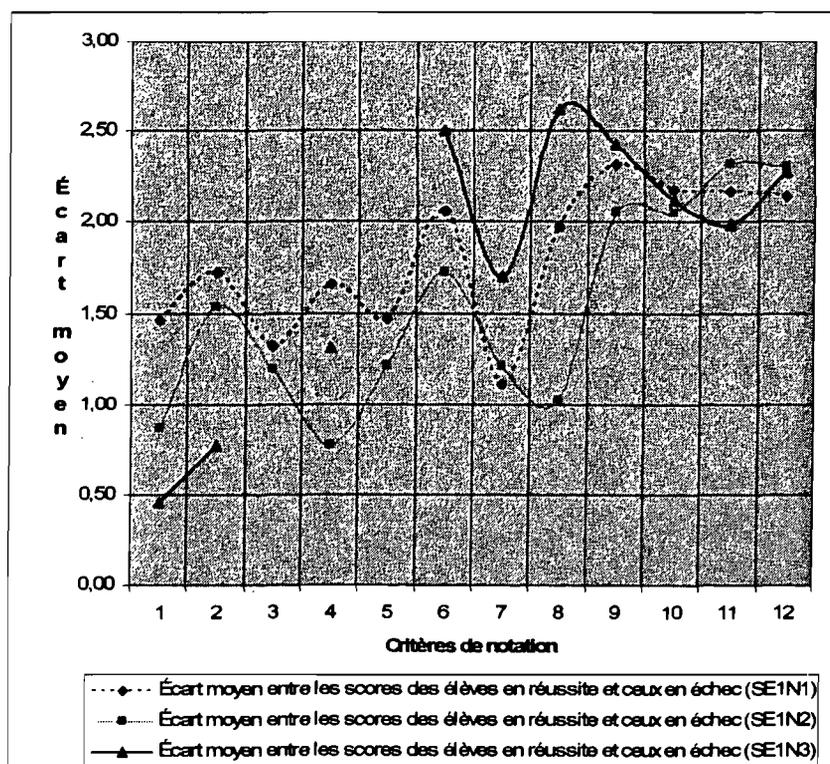


Figure 37

Écart entre les scores moyens des élèves en réussite et ceux en échec pour chacun des critères de notation et pour chacun des niveaux de la SE1.

#### 4.1.2 Analyse descriptive de la distribution des scores bruts moyens pour la situation d'évaluation 2 (SE2) par composantes et critères de notation pour chacun des niveaux

En ce qui concerne la situation d'évaluation 2, on peut remarquer (figure 38) que le critère 7 est le seul qui cause des difficultés aux élèves réputés être en réussite pour les niveaux 1 et 2. Pour ce qui est du niveau 3, tous les élèves en réussite le sont également pour chacun des dix critères de notation. En ce qui concerne les élèves en échec, seuls les

critères 1 et 2 issus du niveau 3 ont été réussis pour ces élèves. Dans tous les autres cas, les élèves en échec le sont pour tous les critères de notation peu importe le niveau auquel ils sont confrontés. En observant les graphiques de la figure 38, on peut également constater que les scores aux critères de notation 1 à 6 tendent à chuter pour les élèves dits en échec. Cette tendance est inversée à compter du critère 8 alors que les scores tendent à augmenter. On peut suspecter qu'une des difficultés des élèves maîtrisant moins bien la compétence est liée à la structuration de la solution (critères 1 à 6) plus qu'à l'identification de la solution en elle-même.

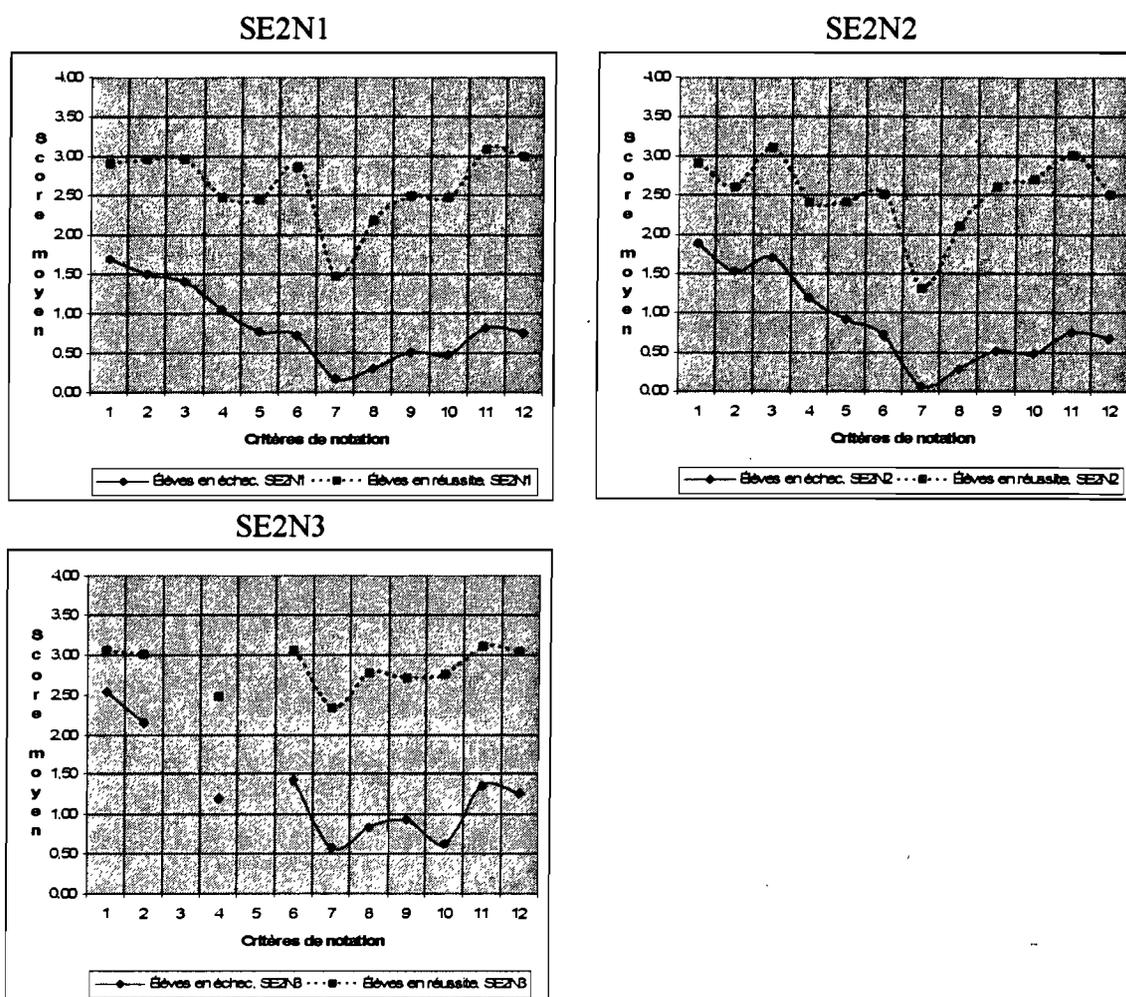


Figure 38  
Variation du score moyen des élèves en échec et en réussite pour chacun des niveaux de la situation d'évaluation 2.

Tableau XLII

Scores moyens pour chacun des critères d'évaluation pour les élèves en réussite et en échec à chacun des niveaux de la SE2.

Critères	SE2N1 (N=186)			SE2N2 (N=68)			SE2N3 (N=98)		
	Réussite (N=106)	Échec (N=80)	Écart	Réussite (N=10)	Échec (N=58)	Écart	Réussite (N=78)	Échec (N=20)	Écart
C1	2,90	1,70	1,20	2,90	1,88	1,02	3,04	2,55	0,49
C2	2,95	1,50	1,45	2,60	1,53	1,07	3,00	2,15	0,85
C3	2,95	1,41	1,54	3,10	1,71	1,39	NSP	NSP	NSP
C4	2,46	1,05	1,42	2,40	1,19	1,21	2,48	1,20	1,28
C5	2,44	0,77	1,66	2,40	0,91	1,49	NSP	NSP	NSP
C6	2,85	0,73	2,12	2,50	0,72	1,78	3,04	1,42	1,62
C7	1,48	0,17	1,31	1,30	0,07	1,23	2,33	0,58	1,75
C8	2,18	0,31	1,86	2,10	0,29	1,81	2,76	0,84	1,92
C9	2,49	0,51	1,98	2,60	0,52	2,08	2,71	0,95	1,76
C10	2,48	0,49	1,98	2,70	0,48	2,22	2,76	0,63	2,13
C11	3,09	0,83	2,26	3,00	0,76	2,24	3,09	1,37	1,72
C12	2,99	0,75	2,23	2,50	0,67	1,83	3,03	1,26	1,77

L'analyse de la figure 39 révèle, de façon sensiblement identique à la SE1, que les écarts entre les scores des élèves en réussite et en échec tendent à être relativement faibles pour les cinq premiers critères. Cet écart tend à augmenter à partir du critère 8 jusqu'au critère 12. De façon générale, on peut distinguer que les écarts, peu importe le niveau, sont somme toute très peu fluctuants surtout entre les critères 8 et 12. Ce résultat nous indique que les différences entre les élèves en échec et ceux en réussite ne sont pas très élevées.

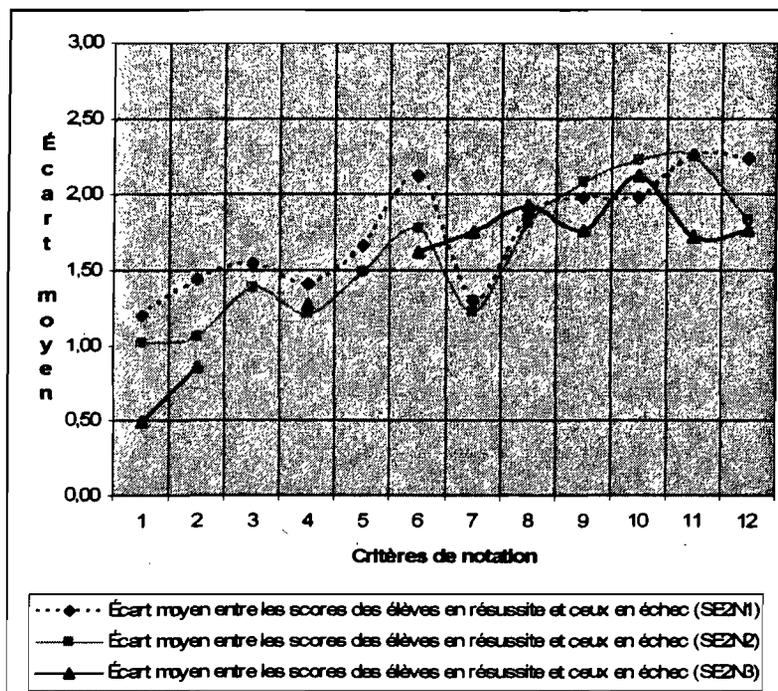


Figure 39

Écart entre les scores moyens des élèves en réussite et ceux en échec pour chacun des critères de notation et pour chacun des niveaux de la SE2.

#### 4.1.3 Analyse descriptive de la distribution des scores bruts moyens pour la situation d'évaluation 3 (SE3) par composantes et critères de notation pour chacun des niveaux

En prenant connaissance des résultats du tableau XLIII, on constate que les élèves en réussite - pour le niveau 1- ont quand même échoué à trois des douze critères à savoir les critères 7, 8 et 10. Au niveau 2, ce résultat est encore pire puisque ces élèves ont alors échoué aux critères 5, 6, 7 et 8. D'ailleurs, sur les graphiques de la figure 40, on peut clairement voir que les résultats chutent de façon dramatique entre les critères 5 et 8 autant pour le niveau 1 que pour le niveau 2. Pour ce qui est du niveau 3, les élèves en réussite obtiennent des résultats plutôt faibles pour les critères 6 à 10. En effet, les résultats à ces derniers oscillent dans un intervalle compris entre 2,00 et 2,50. Pour les niveaux 1 et 2, les élèves en échec obtiennent des résultats sous le seuil de passation pour tous les critères de notation. Par exemple, on remarque qu'au niveau 2 les scores moyens pour les critères 6 à 10 se retrouvent dans l'intervalle compris entre 0,00 et 0,50. Au

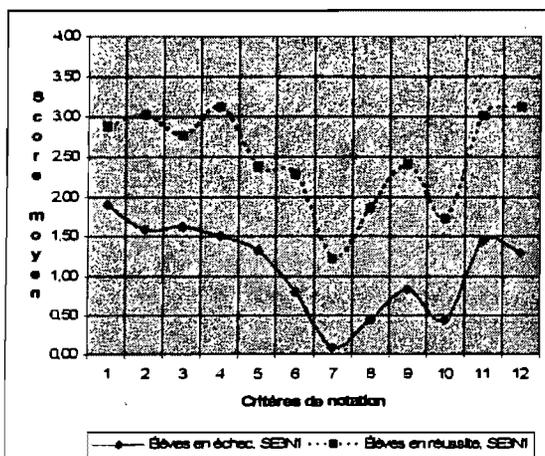
niveau 3, les résultats ne sont guère meilleurs. Les élèves en échec réussissent à obtenir le score de passage uniquement pour les critères 2 (2,35), 4 (2,23) et 12 (2,00).

Tableau XLIII

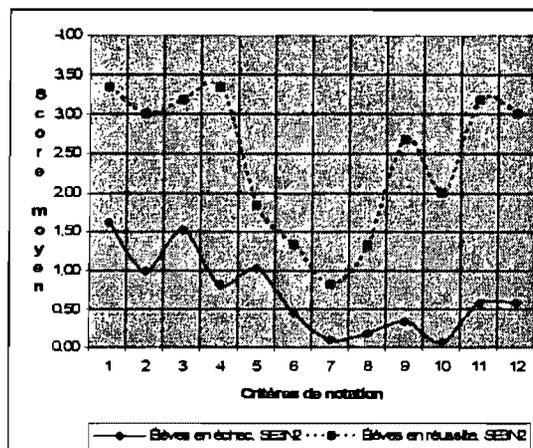
Scores moyens pour chacun des critères d'évaluation pour les élèves en réussite et en échec à chacun des niveaux de la SE3.

Critère	SE3N1 (N=185)			SE3N2 (N=35)			SE3N3 (N=97)		
	Réussite (N=87)	Échec (N=98)	Écart	Réussite (N=6)	Échec (N=29)	Écart	Réussite (N=71)	Échec (N=26)	Écart
C1	2,87	1,89	0,99	3,33	1,62	1,71	2,99	1,96	1,02
C2	3,01	1,58	1,43	3,00	1,00	2,00	3,35	2,35	1,01
C3	2,76	1,61	1,15	3,17	1,52	1,65	NSP	NSP	NSP
C4	3,10	1,50	1,60	3,33	0,83	2,51	2,76	2,23	0,53
C5	2,36	1,33	1,03	1,83	1,03	0,80	NSP	NSP	NSP
C6	2,26	0,82	1,45	1,33	0,45	0,89	2,13	0,77	1,37
C7	1,21	0,08	1,13	0,83	0,10	0,73	2,22	1,00	1,22
C8	1,85	0,44	1,41	1,33	0,17	1,16	2,05	0,69	1,35
C9	2,39	0,84	1,55	2,67	0,34	2,32	2,47	1,04	1,43
C10	1,71	0,44	1,27	2,00	0,07	1,93	1,53	0,60	0,93
C11	3,00	1,45	1,55	3,17	0,59	2,58	2,97	1,84	1,13
C12	3,11	1,30	1,82	3,00	0,59	2,41	3,17	2,00	1,17

SE3N1



SE3N2



SE3N3

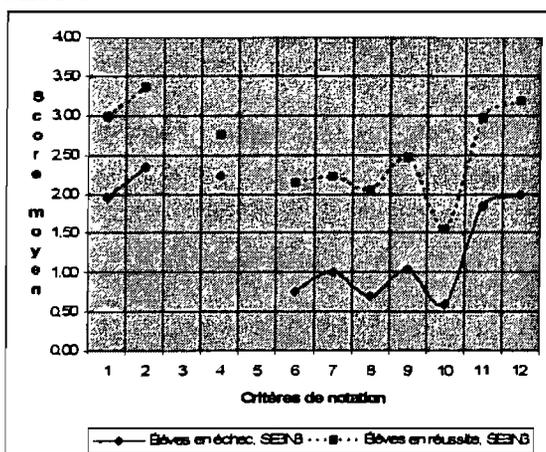


Figure 40

Variation du score moyen des élèves en échec et en réussite pour chacun des niveaux de la situation d'évaluation 3.

Les courbes qui sont illustrées à la figure 41 nous fournissent deux renseignements très pertinents. D'abord, on remarque que les courbes des niveaux 1 et 3 sont caractérisées par de faibles fluctuations. Cela nous indique que les écarts entre les scores moyens pour les élèves en réussite et en échec sont relativement constants pour les douze critères. Par contre, il en est tout autrement pour la courbe du niveau 2. À titre d'exemple, on peut voir que les critères 2, 4, 9, 11 et 12 sont marqués par des écarts de plus de 2,00 échelons entre les élèves en réussite et les élèves en échec.

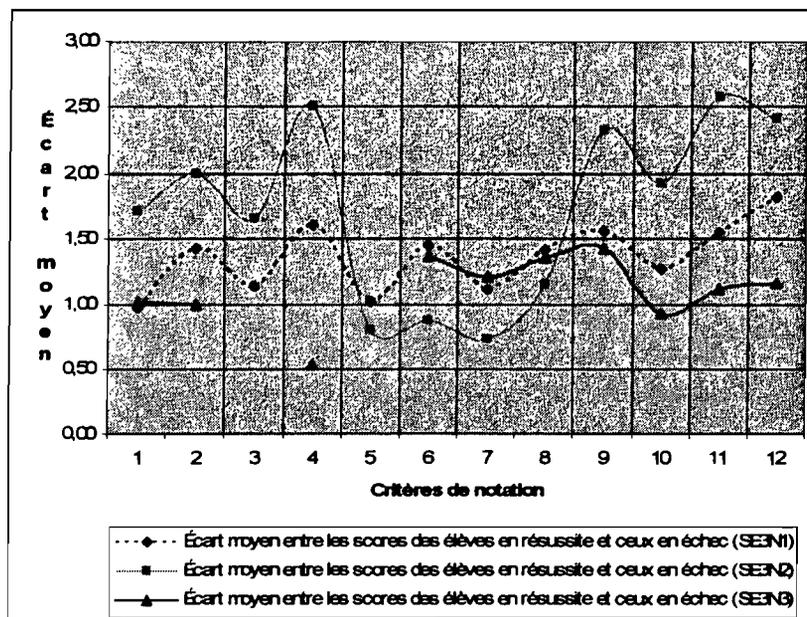


Figure 41

Écart entre les scores moyens des élèves en réussite et ceux en échec pour chacun des critères de notation et pour chacun des niveaux de la SE2.

#### 4.2 Interprétation des données

Les résultats que nous avons obtenus, en comparant les scores des élèves en difficulté avec les élèves en réussite, nous indiquent que les critères retenus pour recueillir des informations sur des éléments de la compétence sont pertinents. Le seul critère problématique a été le septième (identification des causes d'erreurs). Pour ce cas particulier, tous les élèves – en réussite ou pas – ont éprouvé des difficultés à identifier les causes d'erreurs dans leur expérimentation. On peut également noter que la variance des scores des élèves en réussite est relativement faible pour les douze critères (excluant le critère 7). En ce sens, on peut considérer que les élèves en réussite maîtrisent l'ensemble des dimensions de la compétence d'investigation scientifique ce qui inclus également les aspects liés à la communication. Dans le cas des élèves en difficulté, on peut faire un constat semblable. La variance des scores est faible pour l'ensemble des critères (sauf le critère 7). Les élèves en difficulté semblent être en difficulté dans à peu près toutes les dimensions inhérentes à la compétence d'investigation scientifique.

En ce qui concerne les écarts entre les scores des élèves en difficultés et ceux en réussite on note, pour la situation d'évaluation 1, que les quatre derniers critères (C9 : maîtrise des concepts, des règles et des techniques - C10 : application des résultats expérimentaux dans la solution proposée – C11 : présentation générale – C12 : utilisation des symboles et des termes) présentent les écarts les plus importants. On peut penser que ces critères peuvent être des indicateurs intéressants quant à la « sensibilité » des épreuves afin de rendre compte du développement de la compétence. On peut ajouter que les dimensions liées à la structuration et la présentation de la solution sont des aspects qui discriminent bien les élèves qui maîtrisent la compétence des élèves qui la maîtrisent peu ou pas du tout. Éventuellement, il pourrait être intéressant de fractionner l'échantillon des élèves en réussite afin de voir si ces mêmes critères demeurent toujours discriminants.

Pour la situation d'évaluation 2, les écarts – pour les trois niveaux – sont plus faibles que dans le cas de la première situation d'évaluation. Ainsi, il semble y avoir un écart général moins important entre les élèves en réussite et les élèves en difficulté dans le cadre de cette situation. Par contre, on remarque encore que les quatre derniers critères (9-10-11-12) sont ceux pour lesquels les écarts sont les plus grands.

En ce qui concerne la situation d'évaluation 3, nous avons pu remarquer que les écarts étaient faibles pour les niveaux 1 et 3. Par contre, les écarts sont particulièrement importants au niveau 2, entre autres, pour les critères 4 (formulation d'une hypothèse vérifiable et plausible), 9 (maîtrise des concepts, des règles et des techniques), 10 (application des résultats expérimentaux dans la solution proposée), 11 (présentation générale) et 12 (utilisation des symboles et des termes). Nous pensons que ces écarts s'expliquent par le fait que les élèves qui étaient en difficulté, et à qui on a administré le niveau 2, ont alors bénéficié grandement de l'aide accordée. Par contre, les élèves réellement en difficulté ont alors obtenu des résultats particulièrement faibles expliquant les écarts importants pour plusieurs critères.

## ***Deuxième partie : Robustesse du modèle au regard de certaines variables contextuelles***

Cette partie est consacrée aux résultats en lien avec la troisième question de recherche à savoir l'effet de certaines variables contextuelles sur le modèle d'évaluation expérimenté. Dans le cadre de cette recherche, nous nous sommes attardés à trois variables contextuelles : le sexe de l'élève, l'expérience de l'enseignant ainsi que l'effet de sa formation académique. Nous empruntons la même structure que pour la première partie à savoir une analyse descriptive des données pour débiter suivi de l'interprétation des données.

### **1. Les différences entre les garçons et les filles**

#### **1.1 Analyse descriptive des données**

La figure 42 présente les graphiques illustrant la distribution des scores moyens obtenus par les garçons et les filles aux neuf situations d'évaluation expérimentées. En observant les courbes associées aux situations d'évaluation 1 et 2, on constate qu'il semble y avoir très peu de différence dans les scores obtenus par les garçons et les filles. De fait, en jetant un coup d'œil au tableau XLIV, on peut apprendre que la seule différence significative en ce qui concerne la situation 1 est l'écart obtenu au critère 11 pour le niveau 1 à l'avantage des filles. Pour les valeurs de tous les autres critères, les différences entre les garçons et les filles sont non statistiquement significatives. En ce qui concerne la situation d'évaluation 2, les différences notées au critère 1 des niveaux 1 et 2 sont, quant à elles, les seules statistiquement significatives toujours à l'avantage des filles. L'observation des graphiques G, H et I révèle cependant des différences plus notables entre les garçons et les filles. Pour la situation d'évaluation 3, les résultats obtenus par les filles semblent meilleurs que ceux obtenus par les garçons et ce pour les 3 niveaux. En se référant une fois de plus au tableau XLIV, on constate que les filles obtiennent de meilleurs résultats (statistiquement significatifs) aux critères 1, 2, 3, 5, 11 et 12 pour la situation SE1N1 (graphique G). Dans le cas du niveau 2, les filles obtiennent globalement

de meilleurs résultats mais les différences ne sont pas statistiquement significatives à un seuil fixé à 5%. En ce qui concerne la situation SE3N3, les différences sont statistiquement significatives pour les critères 6 et 8 toujours à l'avantage des filles.

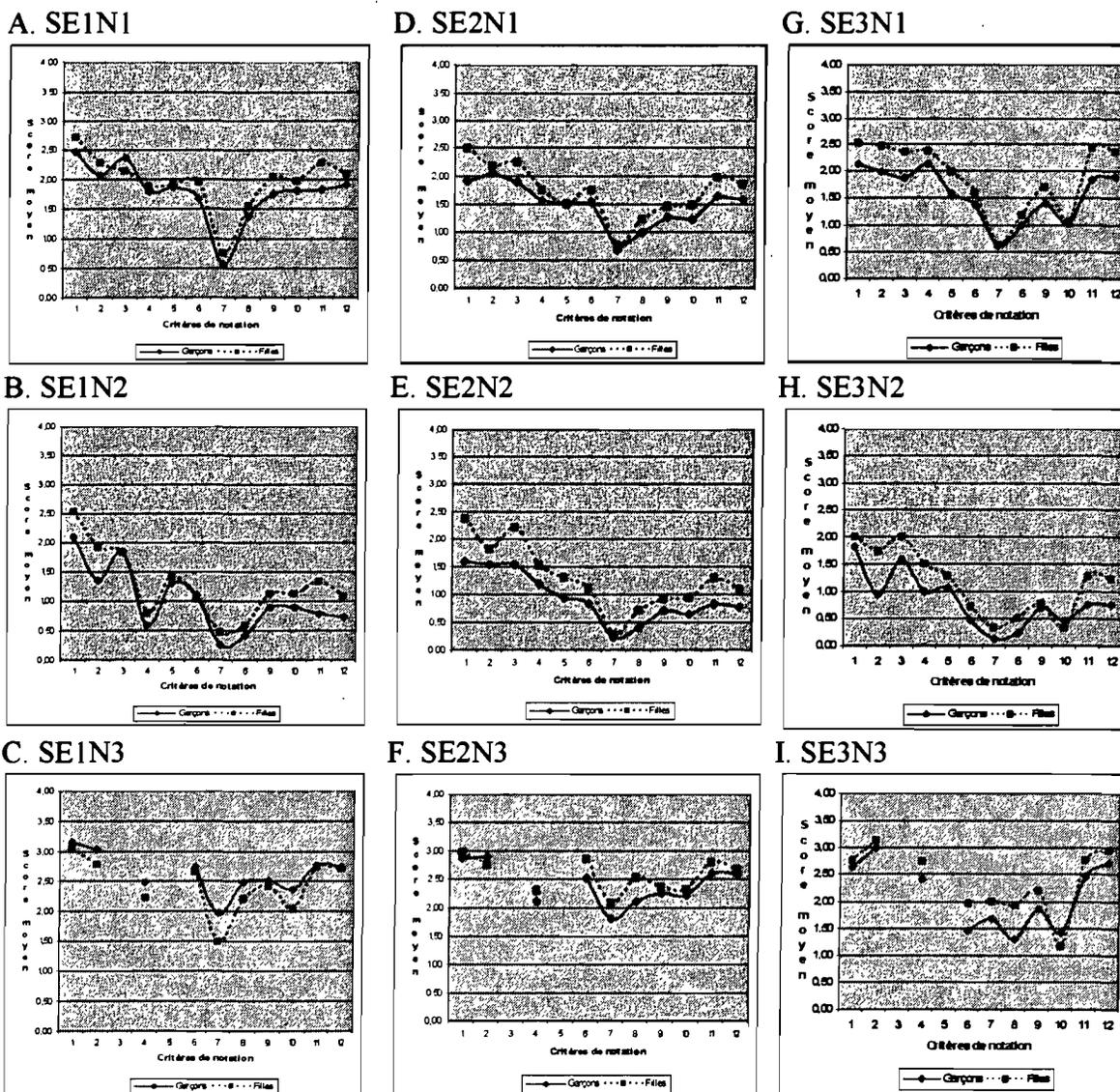


Figure 42  
Distribution des scores obtenus aux neuf situations d'évaluation pour les garçons et les filles.

Tableau XLIV

Résultats du test de Kruskal-Wallis\* pour chacun des critères de notation pour les situations d'évaluation 1, 2 et 3 (comparaison des rangs moyens entre les garçons et les filles).

Critère	SE1			SE2			SE3		
	N1	N2	N3	N1	N2	N3	N1	N2	N3
C1	0,534	0,852	0,337	7,571†	4,858†	0,290	4,023 †	0,077	0,579
C2	1,195	1,238	0,536	0,411	0,653	0,001	4,193 †	2,261	1,053
C3	1,512	0,065	---	3,720	3,118	---	5,880 †	0,383	---
C4	0,089	0,264	0,382	0,657	0,544	0,702	0,977	0,791	2,063
C5	0,153	0,145	---	0,016	1,212	---	6,719 †	0,499	---
C6	1,910	0,006	0,014	1,060	1,141	2,585	1,953	1,316	4,784 †
C7	0,500	0,574	2,043	0,090	0,130	0,863	0,026	0,972	1,487
C8	0,945	0,705	1,280	1,547	1,599	3,234	2,511	1,778	6,868 †
C9	1,953	0,428	0,245	0,698	0,525	0,385	3,104	0,286	2,238
C10	0,681	0,424	1,059	1,463	1,550	0,113	0,127	0,020	1,153
C11	6,001†	1,435	0,046	2,668	1,914	3,130	10,972 †	1,104	2,577
C12	0,408	0,847	0,003	1,541	1,106	0,502	5,256 †	1,074	0,930

Note. \*Statistique du khi-deux (ddl=1) associée au test de Kruskal-Wallis.

† $p \leq 0,05$ .

## 1.2 Interprétation des données

La comparaison des résultats obtenus entre les garçons et les filles a montré qu'il n'existait, en définitive, bien peu de différences. De fait, on constate que seulement 11 critères sur les 132 montrent des résultats statistiquement significatifs. Il faut noter que ces différences sont toutes à l'avantage des filles. Ce résultat nous apparaît très positif puisque plusieurs enseignants avaient manifesté des doutes quant à l'équivalence des

situations d'évaluation pour les garçons et les filles. Par exemple, plusieurs pensaient que la situation d'évaluation concernant la balle de tennis pourrait avantager les garçons. Manifestement, ces doutes ou ces craintes n'étaient pas fondés.

## **2. L'effet de l'expérience des enseignants**

### **2.1 Analyse descriptive des données**

Afin de vérifier l'effet de l'expérience des enseignants, nous avons divisé notre échantillon en trois catégories. La première comprenait les résultats des élèves ayant eu un enseignant comptant cinq années et moins d'expérience en enseignement. La seconde comprenait les résultats des élèves ayant un enseignant comptant entre six et dix années d'expérience en enseignement. Enfin, la troisième partie comprenait les résultats des élèves ayant un enseignant comptant 11 années et plus d'expérience en enseignement. Nous avons choisi ce regroupement car cela faisait en sorte que les effectifs de chacune des catégories étaient plus homogènes. On retrouve en effet le tiers des enseignants qui ont participé à l'étude dans chacune des trois catégories que nous avons définies.

Le tableau XLV présente une synthèse des résultats de la statistique de Kruskal-Wallis dans le cas de la comparaison des scores globaux des élèves (pour chacune des neuf situations d'évaluation) selon les trois catégories d'expérience que nous avons décrites dans le paragraphe précédent. On constate que les moyennes sont équivalentes sauf dans le cas des valeurs associées aux situations SE2N1 et SE3N2.

Tableau XLV

Résultats du test de Kruskal-Wallis\* pour les situations d'évaluation 1, 2 et 3 : comparaison des rangs moyens des scores des élèves pour les enseignants ayant moins de 5 ans d'expérience, entre 5 et 10 ans d'expérience et ceux ayant plus de 10 ans d'expérience.

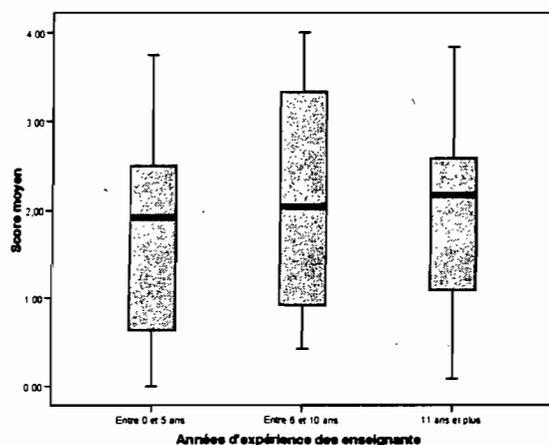
SE	Niveau 1	Niveau 2	Niveau 3
1	3,235	5,093	5,805
2	<b>13,791†</b>	2,397	1,070
3	0,752	<b>10,028†</b>	3,295

Note. \*Statistique du khi-deux (ddl=2) associée au test de Kruskal-Wallis.

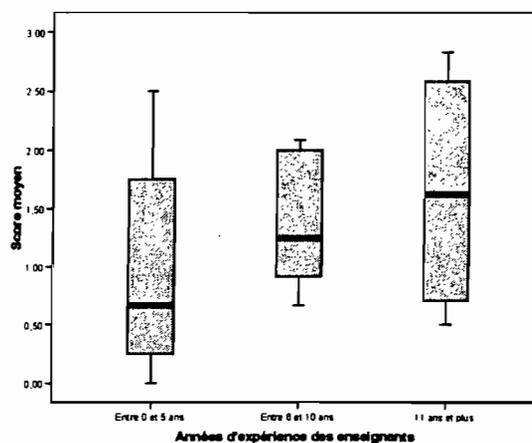
†  $p \leq 0,05$

La figure 43 illustre la distribution des scores pour chacun des niveaux de la première situation d'évaluation. On remarque que les distributions sont équivalentes peu importe l'expérience de l'enseignant. Dans le cas du niveau 2, on peut voir une augmentation croissante du score médian au fur et à mesure que l'expérience de l'enseignant augmente. Par contre, ces différences sont non statistiquement significatives. Au niveau 3, on remarque que la distribution des scores pour les élèves des enseignants ayant 11 ans et plus d'expérience est compacte. On retrouve environ le trois quarts des scores dans l'intervalle [2,00 ; 3,00].

A. Niveau 1



B. Niveau 2



C. Niveau 3

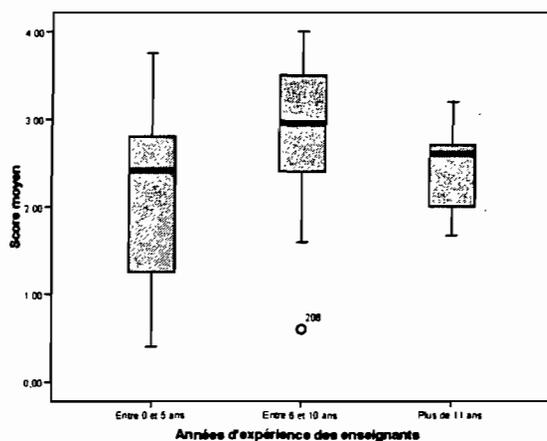
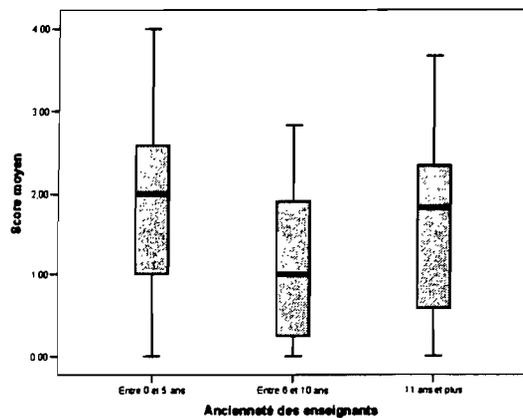


Figure 43

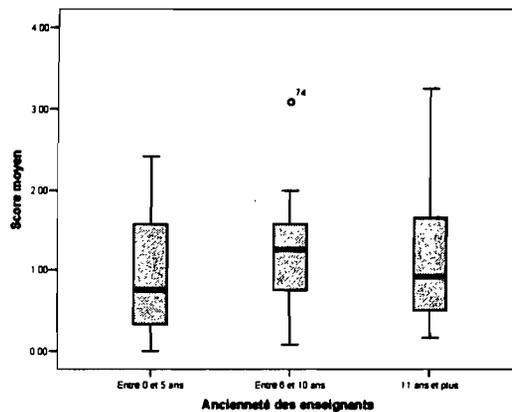
Diagrammes en forme de boîte présentant la distribution des scores moyens pour la situation d'évaluation 1 en fonction du critère d'expérience des enseignants.

La figure 44 présente, quant à elle, la distribution des scores pour la situation 2 toujours en comparant les trois catégories d'enseignants que nous avons décrites précédemment. Au niveau 1, on peut voir que les élèves des enseignants ayant le moins d'expérience (entre 0 et 5 ans) ont obtenu les meilleurs scores. Les résultats les plus faibles ont été obtenus par les élèves des enseignants ayant entre 6 et 10 années d'expérience. Une fois de plus, on remarque que la distribution des scores des élèves ayant un enseignant possédant 11 ans et plus d'expérience montre moins de variance que pour les deux autres catégories d'enseignants dans le cas du niveau 3.

## A. Niveau 1



## B. Niveau 2



## C. Niveau 3

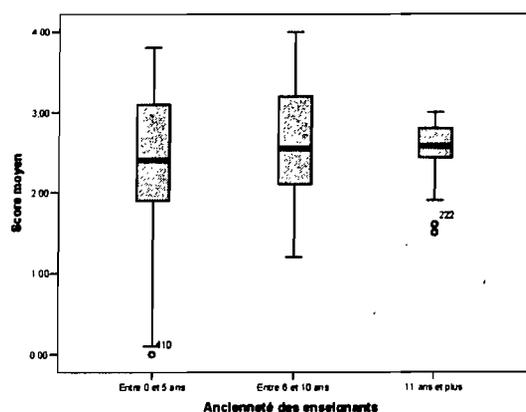


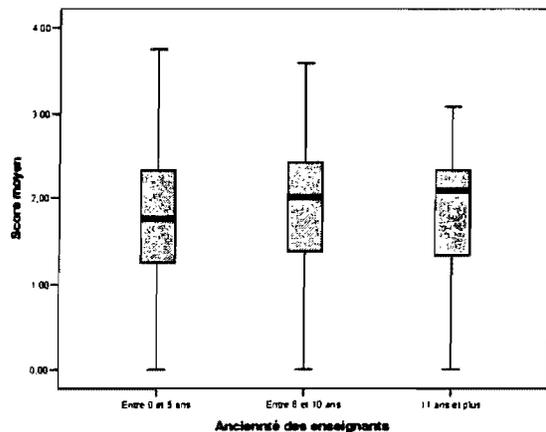
Figure 44

Diagrammes en forme de boîte présentant la distribution des scores moyens pour la situation d'évaluation 2 en fonction du critère d'expérience des enseignants.

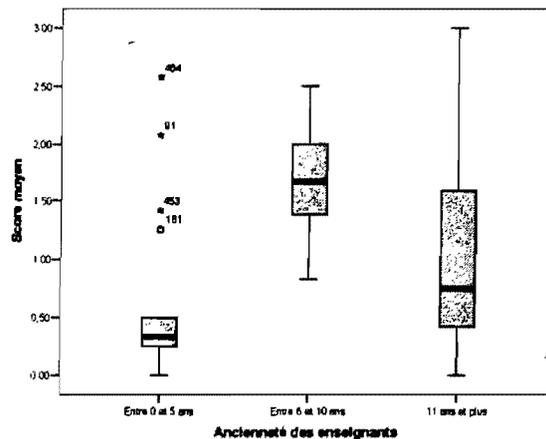
La figure 45 illustre la distribution des scores pour la troisième situation d'évaluation en comparant les résultats des élèves selon l'une ou l'autre des trois catégories d'expérience des enseignants. Au niveau 1, on peut facilement observer que les distributions sont équivalentes. Pour le niveau 2, on remarque aussi facilement que les résultats des élèves ayant un enseignant possédant peu d'expérience (entre 0 et 5 ans) sont nettement plus faibles en comparaison avec les deux autres catégories. Les résultats les plus intéressants sont associés aux élèves ayant un enseignant qui possède entre 6 et 10 ans d'expérience en enseignement. En ce qui concerne le niveau 3, on note que les scores médians sont relativement équivalents d'une distribution à l'autre. À nouveau, on remarque que la

distribution des scores des élèves ayant un enseignant ayant plus de 11 ans d'expérience est plus compacte que les deux autres distributions.

### A. Niveau 1



### B. Niveau 2



### C. Niveau 3

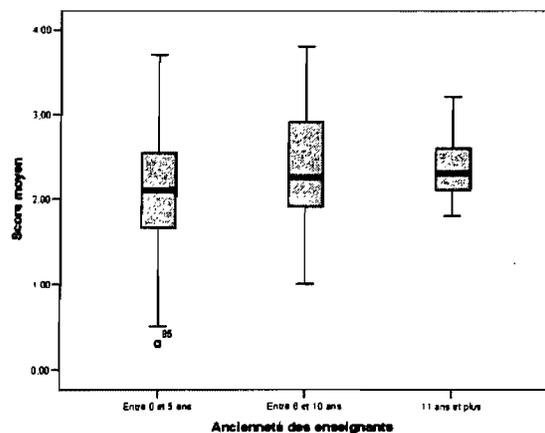


Figure 45

Diagrammes en forme de boîte présentant la distribution des scores moyens pour la situation d'évaluation 3 en fonction du critère d'expérience des enseignants.

## 2.2 Interprétation des données

Il est difficile d'établir des tendances quand on observe les scores en tenant compte de l'expérience des enseignants. Au niveau 1, les scores sont comparables entre les trois catégories d'enseignants sauf dans le cas particulier de la SE2 où les très jeunes enseignants (moins de 5 années d'expérience) semblent voir leurs élèves obtenir de très

bons scores. Au niveau 2, les résultats sont pratiquement identiques à ceux que nous venons de présenter pour le niveau 1 à la différence près que ce sont les enseignants ayant entre 5 et 10 années d'expérience qui obtiennent les meilleurs scores. Au niveau 3, l'expérience des enseignants ne semble pas jouer un rôle déterminant. Par contre, on peut remarquer que les enseignants plus expérimentés voient leurs élèves obtenir des résultats légèrement meilleurs que pour les élèves des enseignants plus jeunes en expérience. On peut penser que les enseignants plus expérimentés sont familiers avec les situations de niveau 3. Ce faisant, ils ont certainement bien préparé leurs élèves à réaliser de telles situations d'évaluation.

Les situations que nous avons proposées étaient nouvelles pour la très grande majorité des enseignants. Aucun enseignant nous a indiqué avoir déjà utilisé de tels outils de cueillette de données. On peut alors raisonnablement penser que, peu importe l'expérience, tous les enseignants se trouvaient dans la même situation face à ces situations d'évaluation sauf dans le cas du niveau 3 qui représente des situations d'évaluation plus « classiques ». Il y a tout lieu de croire, comme nous l'avons mentionné au paragraphe précédent, que les enseignants ayant une expérience plus longue avaient mieux préparé leurs élèves à des situations de niveau 3.

### **3. L'effet de la formation initiale des enseignants**

#### **3.1 Analyse descriptive des données**

Notre échantillon d'enseignants ayant accepté de participer à cette recherche était composé de plusieurs enseignants (7/22) qui n'avaient pas reçu de formation initiale en science ou en enseignement des sciences. Fort de ce constat, nous considérons que cette variable peut influencer positivement ou négativement les résultats au regard du modèle. Dans les pages qui suivent, nous avons analysé les données en essayant de mesurer l'effet de la formation initiale des enseignants sur les résultats des élèves avec le modèle d'évaluation multi niveaux que nous proposons.

Nous avons divisé notre échantillon en deux catégories. La première regroupe les scores des élèves dont l'enseignant possède une formation en science (B.Sc.) ou en enseignement des sciences (B.Ed., en enseignement des sciences). La seconde catégorie circonscrit les résultats des élèves dont l'enseignant ne possède pas de formation en enseignement des sciences ou en science.

Le tableau XLVI présente les valeurs de la statistique standardisée liées au test non paramétrique de Mann-Whitney pour chacun des points de comparaison. Pour chacune des situations d'évaluation, nous avons comparé les valeurs moyennes obtenues par chaque élève selon que l'enseignant soit formé en science ou en enseignement des sciences ou pas. En prenant connaissance de ce tableau, on constate que seules deux valeurs sont statistiquement significatives (SE1N3 et SE3N1). Dans tous les autres cas, il ne semble pas exister de différences statistiquement significatives entre les résultats des élèves ayant eu un enseignement formé en science ou en enseignement des sciences et les résultats des élèves ayant eu un enseignement non formé en science ou en enseignement des sciences. Examinons de plus près les distributions des scores associés à chacune de ces neuf situations d'évaluation.

Tableau XLVI

Résultats du test de Mann-Whitney\* pour les situations d'évaluation 1, 2 et 3 : comparaison des rangs moyens des scores des élèves pour les enseignants ayant une formation initiale en enseignement des sciences ou en science par rapport à ceux qui n'en ont pas.

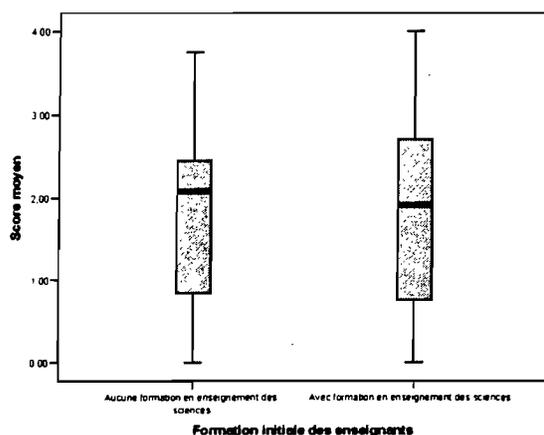
SE	Niveau 1	Niveau 2	Niveau 3
1	-0,355	-0,566	<b>-2,822†</b>
2	-1,578	-1,382	-0,339
3	<b>-2,248†</b>	-1,135	-0,174

Note. Statistique Z associée au test de Mann-Whitney,

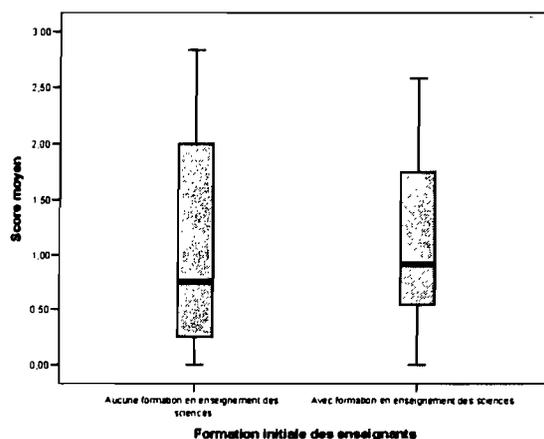
†  $p \leq 0,05$

Débutons notre tour d'horizon avec la figure 46 qui présente les différentes distributions pour la situation d'évaluation 1. Pour les niveaux 1 et 2 (figures A et B), on peut voir facilement que les distributions sont très semblables selon que l'enseignant possède ou pas une formation initiale en science ou en enseignement des sciences. On constate, par exemple, que les enseignants non formés voient leurs élèves obtenir un score médian légèrement supérieur à ceux dont l'enseignant possède une formation. Au niveau 2, on constate le contraire. Ce sont les élèves ayant un enseignant formé en science ou en enseignement des sciences qui obtiennent un score médian plus élevé. Par contre, ces différences ne sont pas statistiquement significatives. Au niveau 3, les résultats sont bien différents. Sur la figure C, on peut voir que les scores obtenus par les élèves ayant un enseignant formé en science sont nettement supérieurs à ceux dont l'enseignant ne possède pas une formation de ce type. Comme nous l'avons souligné précédemment, cette différence est statistiquement significative au seuil de 0,05. De plus, on peut constater que la distribution de droite (enseignants non formés) sur la figure C est beaucoup plus compacte (variance plus faible) que la distribution de gauche associée aux enseignants formés en science ou en enseignement des sciences.

## A. Niveau 1



## B. Niveau 2



## C. Niveau 3

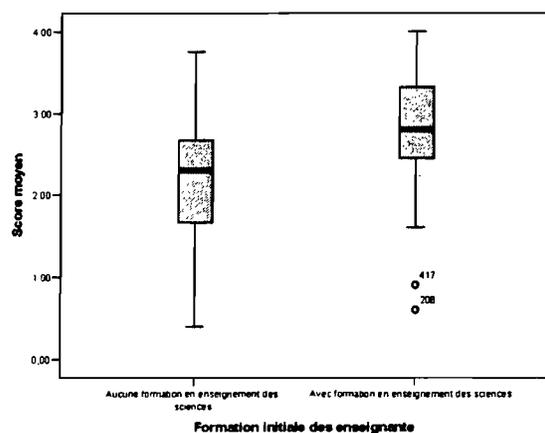


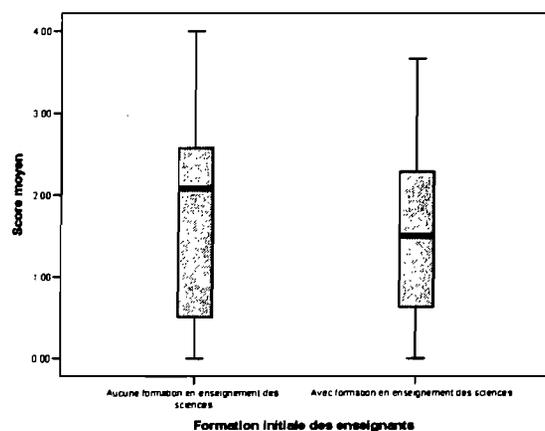
Figure 46

Diagrammes en forme de boîte présentant la distribution des scores moyens pour la situation d'évaluation 1 en fonction de la formation initiale des enseignants.

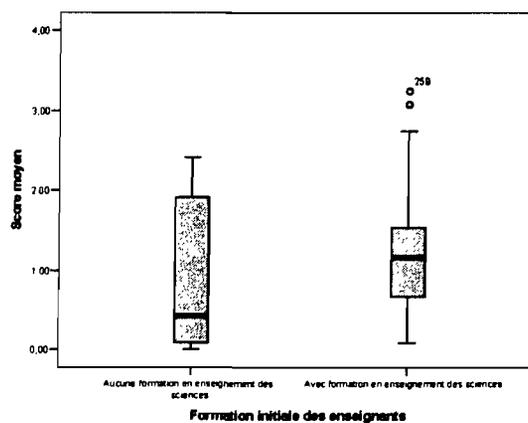
Concernant la situation d'évaluation 2, on constate à la lecture du tableau XLVI qu'il n'existe aucune différence statistiquement significative entre les scores moyens des élèves ayant un enseignant formé en enseignement des sciences ou en science et les scores moyens des élèves ayant un enseignant qui ne possède pas une telle formation. Cependant, on peut noter quelques tendances en consultant la figure 47. Au niveau 1, les élèves des enseignants qui n'ont pas de formation initiale en science obtiennent des résultats légèrement supérieurs. On constate, par exemple, que le score médian de la distribution de gauche (enseignants non formés) est plus élevé que celui de la distribution

de droite (enseignants formés) sur la figure A. Pour les niveaux 2 et 3, on remarque que c'est le contraire puisque les élèves des enseignants formés en enseignement des sciences obtiennent de meilleurs résultats. Par ailleurs, on remarque que les scores médians pour les niveaux 2 et 3 se situent largement en deçà du seuil de passation fixé à 2,00. Ce constat s'applique aux enseignants formés en science ou en enseignement des sciences ainsi qu'aux enseignants possédant une toute autre formation.

A. Niveau 1



B. Niveau 2



C. Niveau 3

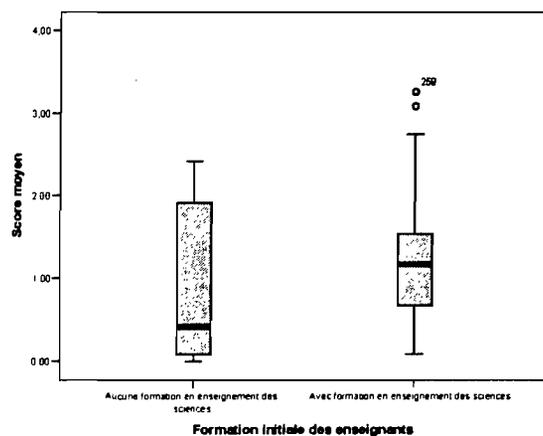
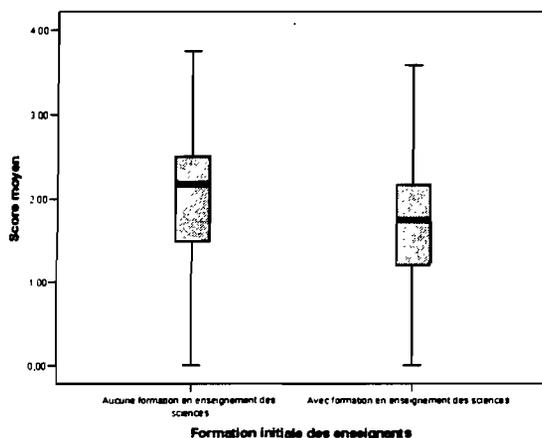


Figure 47

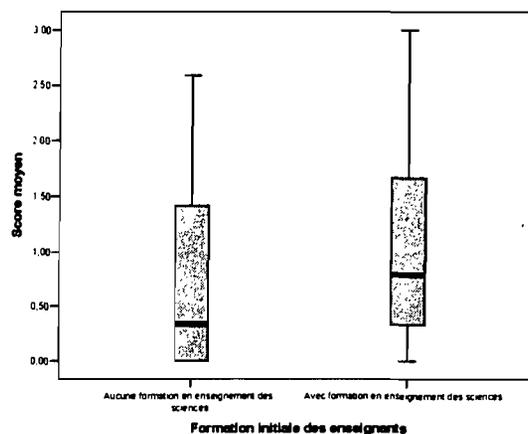
Diagrammes en forme de boîte présentant la distribution des scores moyens pour la situation d'évaluation 2 en fonction de la formation initiale des enseignants.

En ce qui concerne la troisième situation d'évaluation, la figure 48 nous apprend que la formation initiale semble jouer un rôle mineur surtout dans le cas du niveau 1. En se référant au tableau XLVI, on relève que les différences inhérentes au niveau 1 sont statistiquement significatives. Ainsi, les élèves des enseignants qui ne possèdent pas de formation initiale en science ou en enseignement des sciences obtiennent de meilleurs résultats que les élèves ayant des enseignants formés soit en science ou en enseignement des sciences. On remarque également, pour ces derniers, que le score médian se situe sous le seuil de passage fixé à 2,00. Pour les élèves ayant un enseignant formé en science ou en enseignement des sciences, le score médian se situe au-delà du seuil de passage. La tendance que nous venons de décrire est renversée au niveau 2 (figure 47 B). Bien que les résultats ne soient pas statistiquement significatifs, on peut quand même remarquer que le score médian de la distribution de droite est plus élevé que celui de la distribution de gauche. Au niveau 3, les distributions - en provenance des deux échantillons - sont équivalentes. Le score médian est comparable pour les deux distributions.

A. Niveau 1



B. Niveau 2



C. Niveau 3

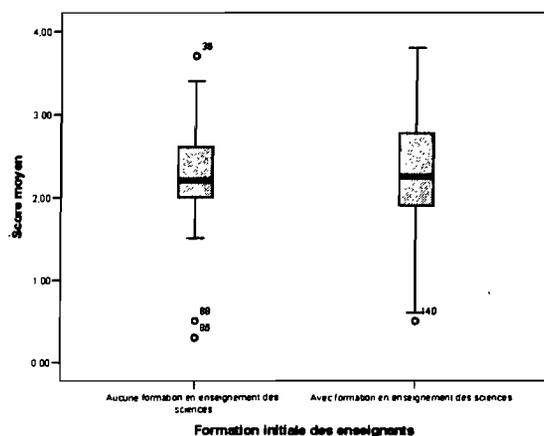


Figure 48

Diagrammes en forme de boîte présentant la distribution des scores moyens pour la situation d'évaluation 3 en fonction de la formation initiale des enseignants.

### 3.2 Interprétation des données

La formation initiale des enseignants ne semble pas avoir joué un rôle prépondérant dans les résultats des élèves. Au total, on repère seulement deux résultats statistiquement significatifs (SE1N3 et SE3N1) sur les neuf analysés. Qui plus est, on ne peut établir de tendance puisque dans un cas ce sont les enseignants ayant une formation en science ou en enseignement des sciences qui ont vu leurs élèves obtenir de meilleurs résultats que les élèves des enseignants n'ayant pas de formation. Dans l'autre cas, nous avons constaté le

contraire. Ce résultat est très étonnant car nous nous serions attendus à voir les enseignants ayant une formation obtenir de bien meilleurs résultats que les enseignants ne s'appuyant sur aucune formation pertinente. Encore une fois, il est vraisemblable de penser que le caractère de nouveauté associé aux SE ait joué un rôle de dénominateur commun en ce sens où même les enseignants formés en science ou en didactique des sciences n'ont pas nécessairement mieux formé leurs élèves au regard de la compétence d'investigation scientifique.

### ***Troisième partie : Commentaires des enseignants sur la passation des situations d'évaluation***

Nous avons demandé aux enseignants de consigner, dans un cahier que nous leur avons remis, les commentaires en lien avec la passation des situations d'évaluation. Il s'agissait, pour les enseignants, de noter tous les événements qui pouvaient – d'une manière ou d'une autre – affecter la qualité et/ou la quantité des données recueillies. Parmi les 22 enseignants participants, 14 d'entre eux ont remis le document en question. Nous avons regroupé les commentaires en trois grandes catégories : (1) gestion de classe et aspects techniques liés à la passation, (2) nature des situations d'évaluation et (3) facteurs liés aux élèves. Les sections qui suivent présentent la synthèse des commentaires des enseignants. Le lecteur remarquera que certains commentaires pourraient appartenir à plus d'une catégorie. Des extraits des commentaires des enseignants viennent appuyer nos observations.

#### **1. Gestion de classe et aspects techniques liés à la passation**

Les sections qui suivent sont dédiées à la gestion de classe et aux aspects plus techniques liés à l'expérimentation de notre modèle d'évaluation. Dans l'ordre, nous aborderons les éléments liés au suivi des élèves, à la gestion de classe et à l'aide apportée par les techniciens de laboratoire.

##### **1.1 Le suivi auprès des élèves**

À chacune des grandes étapes de la résolution de chacune des SE, on demandait aux enseignants de donner leur accord pour la suite des travaux. Il s'agissait pour les enseignants de prendre connaissance du travail de chacun des élèves et de décider si l'élève en question pouvait ou non continuer à réaliser le niveau auquel il était confronté. Si l'enseignant constatait une quelconque difficulté, il pouvait alors décider d'administrer à l'élève un autre niveau (2 ou 3). Cette façon de procéder a été jugée fastidieuse par 6

enseignants sur les 14. De toute évidence, la tâche de l'enseignant s'est alourdie considérablement puisque les enseignants devaient veiller à plusieurs choses à la fois (matériel, gestion disciplinaire de la classe, gestion du suivi des élèves, etc.).

Enseignant numéro 11 :

« Beaucoup de questions, file d'attente à mon bureau. J'ai un mal de tête terrible. Les élèves ont beaucoup de difficultés. »

## **1.2 La gestion de classe**

La gestion de classe a été jugée problématique par quatre enseignants. Le contexte de l'évaluation (individuelle, plusieurs SE, etc.) a semble-t-il contribué à complexifier le travail des enseignants. Les fréquents déplacements en classe ont aussi amené des difficultés concernant le plagiat et la contamination entre les élèves. Nous reviendrons sur cet aspect dans une rubrique qui lui est consacrée.

Enseignant numéro 23 :

« Beaucoup de déplacements, difficile à gérer. Voir à ce que tout se passe bien au niveau de la sécurité. Superviser si les élèves font ce qu'ils ont à faire. S'arranger pour faire taire ceux qui parlent. (...) En général assez bien sauf pour un groupe où ce fut très difficile. Bavardage, ils se lancent des objets, de l'eau, etc. Même en étant trois ce fut presque ingérable!! »

## **1.3 L'aide apportée par le technicien de laboratoire**

Lors de la présentation de la recherche aux enseignants, nous les avons clairement avisés de la nécessité de compter sur la présence du technicien de laboratoire pour réaliser les épreuves. Nous savions que sa présence était une composante essentielle au succès de

l'entreprise. Malheureusement, deux enseignants n'ont pas pu bénéficier de cette aide. Comme nous l'avions prévu, l'absence du technicien a été un facteur nuisant à l'administration des situations d'évaluation. L'un des enseignants a même été dans l'obligation d'annuler l'une des trois situations d'évaluation faute de matériel adéquat.

Enseignant numéro 9 :

« Présentation de deux activités seulement, parce que le matériel pour la troisième n'était pas prêt (chocolat chaud). Il est à noter que le document « À l'intention des techniciens » (sic) avait été remis au « technicien » depuis 2 semaines. »

## **2. Nature des situations d'évaluation**

Dans cette section, nous avons regroupé les commentaires des enseignants qui étaient, de près ou de loin, associés aux situations d'évaluation. Ces commentaires ont été regroupés en quatre catégories : les difficultés liées au passage d'un niveau à l'autre, l'effet de contamination entre les élèves, l'intérêt motivationnel suscité par chacune des situations d'évaluation et l'inadéquation entre les situations d'évaluation et les situations d'apprentissage.

### **2.1 Le passage d'un niveau à l'autre**

La majorité des enseignants (8) ont remarqué et indiqué que certains élèves en difficulté acceptaient difficilement de se voir donner une situation d'évaluation de niveau 2 ou de niveau 3 (changement de version de SE). Plusieurs enseignants ont indiqué que cela était perçu comme un échec de la part des élèves en difficulté. Ces derniers voyaient d'un mauvais œil d'accepter un autre niveau puisque cela pouvait avoir potentiellement un impact sur leur résultat final pour cette épreuve évaluative.

Enseignant numéro 20 :

« Malgré les explications des trois niveaux, les élèves voient beaucoup le passage d'un niveau à l'autre comme un échec??? »

## **2.2 L'effet de contamination**

Comme nous le mentionnions à la section 1, plusieurs enseignants (7) ont remarqué un effet de contamination particulièrement important chez certains élèves. Cet effet s'est manifesté de plusieurs façons. Certains élèves ont, en effet, été influencés par les solutions proposées à la version du niveau 3 (administrées à d'autres élèves de la classe) alors qu'eux-mêmes travaillaient encore sur la version 1. Quelques enseignants ont également remarqué qu'une solution est devenue soudainement fort populaire au deuxième cours alors que les élèves ont eu la chance de se côtoyer entre le premier et le deuxième cours.

Enseignant numéro 8 (compte-rendu du troisième cours) :

« Mes trois élèves qui avaient de la difficulté ont miraculeusement tout compris (je crois qu'il y a eu des fuites entre les périodes). D'autres étudiants ont changé leur façon de faire surtout pour les balles de tennis. J'en ai un qui a eu l'idée de mouiller sa balle et tous les autres ont fait la même chose. »

## **2.3 L'intérêt suscité par chacune des situations d'évaluation**

Lorsque nous avons conçu les situations d'évaluation, nous avons essayé de les rendre intéressantes pour les élèves et aussi pour les enseignants. Un peu plus du tiers des enseignants (5) ont pris la peine d'indiquer que les SE ont suscité un intérêt certain auprès des élèves.

Enseignant numéro 17 :

« J'ai fait un sondage dans mes trois groupes afin de déterminer leurs projets préférés et les moins bien aimés. Sur neuf projets (de janvier à juin), votre SE s'est classée en cinquième position! Pas si mal, hein? »

#### **2.4 L'inadéquation entre les situations d'évaluation et les situations d'apprentissage réalisées durant l'année scolaire**

Une majorité d'enseignants (9) a pris le soin d'indiquer que les situations d'évaluation expérimentées étaient différentes des situations d'apprentissage réalisées en cours d'année. Certains enseignants ont même annexé des exemples de situations d'apprentissage réalisées en cours d'année afin de montrer la distance existant entre ce qui a été fait et ce qui a été expérimenté dans le cadre de cette recherche.

Enseignant numéro 10 :

« Les élèves ont réalisé plusieurs situations d'apprentissage mais aucune d'elles ne ressemblait à celles proposées dans cette recherche. Les situations que j'ai faits faire aux élèves sont en quelque sorte « les classiques ». On donne un maximum d'informations et l'élève à juste à aller chercher les réponses et répondre aux questions. »

### **3. Facteurs liés aux élèves**

Le dernier bloc de commentaires concerne directement les élèves et leur motivation à réussir les situations d'évaluation.

### **3.1 La frustration chez certains élèves**

Une minorité d'enseignants (3) ont remarqué que des élèves manifestaient des signes de frustration à un moment ou à un autre de la passation de l'épreuve. Par exemple, un enseignant mentionnait qu'à un moment donné, l'ampleur de sa tâche était si importante que certains élèves pouvaient avoir la main levée pendant quinze minutes avant de recevoir de l'aide de sa part. Nous pensons que ces frustrations peuvent trouver leur origine dans un ou plusieurs des aspects que nous abordons subséquemment. On peut mentionner le temps pouvant être accordé par l'enseignant à chaque élève en difficulté, la mauvaise compréhension du modèle d'évaluation (le passage d'un niveau à l'autre) ou la nouveauté de la SE par rapport à ce que l'élève a fait durant l'année scolaire.

Enseignant numéro 11 :

« Les élèves insécures (sic) ressortent du groupe.(...) Manque de confiance pour plusieurs élèves. »

### **3.2 La démotivation liée au contexte de passation**

Les épreuves ont été administrées à la toute fin de l'année scolaire 2005-2006. Selon deux enseignants, cela a été un facteur de démotivation chez les élèves et ce, à deux niveaux. D'abord, certains enseignants ont admis avoir manqué de temps pour procéder, par exemple, à l'administration de la situation de niveau 3 pour les élèves qui étaient en difficulté. C'est le cas des enseignants qui ont planifié uniquement quatre périodes pour la passation des épreuves. Ensuite, il y a le fait que ces situations d'évaluation n'étaient pas considérées « sérieuses » de la part des élèves puisqu'il s'agissait d'une passation expérimentale et dont les résultats ne seraient pas comptabilisés au bulletin.

**Enseignant numéro 16 :**

**« Premièrement, j'ai deux groupes très difficiles (05 et 08) au niveau de la gestion de classe. (bavardage, non respect des consignes, non respect du matériel, vol de matériel). Deux élèves ont carrément refusé de faire l'examen. Ces deux groupes n'ont pas pris l'examen au sérieux. Par contre, certains élèves se sont appliqués et ont bien performé. »**

Je ne cherche pas à connaître les réponses, je cherche à comprendre les questions.  
Confucius

## **Chapitre 5. Discussion des résultats**

---

Ce chapitre est consacré à la discussion des données que nous avons décrites au chapitre précédent. Le chapitre a été divisé de façon à interpréter les données relatives à chacune des trois questions principales de recherche et des trois sous-questions associées à la troisième question principale de recherche. Nous terminerons ce chapitre en identifiant les principales limites qui ont marqué la réalisation de cette recherche et les aspects qui nous obligent à une certaine prudence dans l'interprétation et la généralisation des résultats.

### **1. Résultats concernant la première question de recherche**

La première question de recherche se libellait de la façon suivante :

Quels sont les avantages à recourir à un modèle d'évaluation à trois niveaux afin de rendre compte du niveau de développement d'une compétence d'investigation scientifique en laboratoire ?

À notre avis, il y a plusieurs avantages à recourir à un modèle d'évaluation basée sur des différents niveaux de passation. Les sections qui suivent présentent les principaux avantages associés à notre modèle.

#### **1.1 Les informations recueillies**

Une des caractéristiques de l'approche par compétences réside au fait que ces dernière se développent dans le temps. Ainsi, on constate généralement que les élèves ne sont pas tous au même niveau concernant le développement de leur compétence d'investigation scientifique. Un avantage indéniable du modèle réside dans la possibilité de bonifier quantitativement et qualitativement les informations recueillies auprès de chacun des

élèves au regard du développement de la compétence d'investigation scientifique. Lors de l'utilisation d'un modèle plus traditionnel<sup>34</sup>, il faudrait alors recourir à l'un ou l'autre des niveaux (N1, N2 ou N3) pour chacune des situations d'évaluation. Dans le cas où l'on ferait appel uniquement au niveau 1 (ou aux niveaux 1 et 2), il faudrait alors devoir composer avec une qualité d'information réduite puisque plusieurs élèves ont obtenu des scores faibles (ex. élèves des profils 2 et 3) et une proportion non négligeable d'élèves ont également obtenu un vecteur réponse nul pour ce niveau. Cependant, ce niveau est celui qui respecte le plus les exigences de la mesure du développement d'une compétence. Dans le cas où on ferait plutôt appel uniquement au niveau 3, il serait possible pour les élèves de faire la démonstration d'une meilleure performance mais il faudrait alors accepter que les épreuves ne répondent pas aux caractéristiques d'une épreuve visant à mesurer le développement d'une compétence. Dans un cas comme dans l'autre, on constate que le recours à un modèle de passation traditionnel comporte des limites importantes. En ce sens, un modèle à plusieurs niveaux permet de viser l'administration de situations d'évaluation mesurant des compétences tout en offrant des options si jamais les élèves n'y arrivent pas adéquatement. Aussi, les résultats obtenus lors de l'expérimentation nous semblent fort prometteurs dans le contexte d'un changement de pratiques pédagogiques. Le but ultime demeure de recourir uniquement aux situations de niveau 1 qui représentent – à notre avis – les situations qui sont les plus conformes à la nature de la compétence d'investigation scientifique telle que nous l'avons définie au chapitre 1 consacré à la problématique de la recherche. En effet, dans le contexte de changements pédagogiques, il arrive que les enseignants n'aient pas toutes les ressources (formation pédagogique, formation scientifique, ressources matérielles, etc.) afin d'adapter leur enseignement aux changements proposés. Les résultats obtenus lors de cette expérimentation montrent que l'administration seule du niveau 1 entraîne un taux d'échec particulièrement important qui peut être attribuable, entre autres, aux conditions liées aux changements pédagogiques qui avaient cours lors de la présente expérimentation. Dans ce contexte, la présence de niveaux supplémentaires revêt toute son importance. Nous pensons également que notre modèle pourrait s'inscrire dans un

---

<sup>34</sup> Nous entendons par modèle traditionnel un format de passation d'épreuve où on administre une seule version de ladite épreuve qui est commune à tous les élèves.

plan planifié et annoncé de changements pédagogiques qui ferait en sorte de ne plus offrir certains niveaux au fur et à mesure que l'implantation du changement s'opérationnalise. Par exemple, il serait tout à fait vraisemblable de ne plus offrir le niveau 3 (le niveau qui est le plus en opposition avec l'approche par compétences) après deux ou trois années suivant l'implantation du changement pédagogique. Cette façon de procéder permettrait d'avoir en main une kyrielle d'outils de cueillette de données lors des premières années suivant le changement pédagogique souhaité et d'en laisser tomber en cours de route au fur et à mesure que les besoins changent. Concrètement, le modèle expérimenté permet clairement cette possibilité.

## **1.2 Les profils d'élèves**

Le modèle permet d'identifier des profils d'élèves pour qui les niveaux 2 et 3 sont particulièrement utiles. Idéalement, nous souhaitons qu'aucun élève ne soit dans l'obligation de recourir au niveau 2 et au niveau 3. Dans ce cas, cela voudrait dire que tous les élèves ont atteint un niveau de développement de la compétence qui est suffisant. Dans la pratique, cette situation est rarement rencontrée. Comme nous avons pu le constater au chapitre précédent, le niveau 1 est approprié pour pratiquement la moitié des élèves. Pour l'autre moitié des élèves, il demeure nécessaire de les guider davantage et c'est la fonction que remplissent les niveaux 2 et 3. L'identification de ces profils de passation (et par conséquent de profils d'élèves) nous apparaît être d'une grande utilité puisqu'elle permet de savoir qui – et dans quelle proportion – a besoin d'aide supplémentaire pour faire la démonstration de son niveau de compétence.

## **1.3 L'équivalence des situations de niveau 1 et 3 pour les élèves performants**

Dans le protocole de recherche, nous avons demandé aux enseignants d'administrer d'abord le niveau 1 à tous les élèves et de recourir, au besoin, aux niveaux 2 et 3 pour les élèves en difficulté. Nous aurions pu également demander aux enseignants d'administrer les 3 niveaux à chacun des élèves comme l'ont fait Rey et ses collaborateurs (2003). Lors

de l'expérimentation, nous avons constaté que certains enseignants ont administré le niveau 3 aux élèves qui avaient préalablement réussi le niveau 1. Ces résultats ont été regroupés pour former le profil d'élèves 4. Ce résultat était inattendu. Cependant, nous avons considéré que ces données étaient riches et méritaient amplement d'être traitées même si elles n'étaient pas rigoureusement planifiées dans le protocole de recherche. Ces résultats (non prévus) concernant le profil de passation 4 nous indiquent que les élèves qui ont réalisé et réussi de façon satisfaisante le niveau 1 arrivent à obtenir des résultats qui sont aussi bons – et même un peu meilleurs – au niveau 3. Dans ce cas-ci, on peut dire que notre échantillon nous permet de dire que les élèves qui réussissent une situation d'évaluation plus ouverte (niveau 1) réussissent également la même situation dans un contexte plus fermé (niveau 3). En ce sens, les résultats que nous avons obtenus tendent à nous indiquer que les niveaux sont inclusifs de façon unidirectionnelle (du niveau 1 vers le niveau 3). En effet, nous arrivons à ce constat car nous avons expérimenté les situations d'évaluation en commençant par le niveau 1 et en terminant, le cas échéant, par le niveau 3. Ainsi, il n'est pas possible d'affirmer que les situations de niveau 1 et de niveau 3 sont interchangeables en pensant, par exemple, qu'un élève qui réussit la situation de niveau 3 réussirait automatiquement la situation de niveau 1. Pour pouvoir affirmer cela, il faudrait expérimenter à nouveau avec un échantillon d'élèves qui se verraient offrir d'abord la situation de niveau 3 et ensuite la situation de niveau 1. Or, comme l'ont souligné les experts consultés, les situations de niveau 1 sont complexes et ouvertes et il est loin d'être acquis qu'un élève ayant réussi le niveau 3 réussisse par la suite le niveau 1. Une recherche future pourrait être entreprise afin de vérifier cette avenue. Quoi qu'il en soit, l'équivalence des résultats appuie notre choix méthodologique de mettre un terme à l'épreuve aussitôt que l'élève réalise de façon adéquate le niveau le plus exigeant (dans l'ordre : 1-2-3) plutôt que d'administrer chacun des niveaux de façon systématique comme Rey *et al.* (2003) ont choisi de faire.

#### **1.4 L'efficacité relative du niveau 2**

Les résultats obtenus lors de la passation des niveaux 2 (pour l'ensemble des situations d'évaluation) se sont avérés décevants. Comme nous avons pu en prendre note au chapitre précédent, le niveau 2 n'a pas permis aux élèves qui étaient en difficulté au niveau 1 d'atteindre un score médian qui se serait situé au-delà du seuil de passation fixé à l'échelon 2. Dans tous les cas, le taux de réussite a très peu augmenté suite à la passation du niveau 2. Puisque le niveau 3 a permis aux élèves de résoudre la situation qui leur était présentée, nous pensons que le modèle est adéquat mais qu'il est nécessaire de peaufiner le niveau 2 afin de mieux guider les élèves vers des démarches et des solutions adéquates. L'analyse des scores pour chacun des 12 critères de notation nous a révélé que les élèves manifestent généralement des difficultés à organiser leurs idées et leur(s) démarche(s). Par la suite, les élèves arrivent à assez bien structurer leurs données et à les présenter.

Les résultats obtenus pour le niveau 2 de chacune des situations d'évaluation soulignent la difficulté d'ordonner des situations d'évaluation en terme de difficulté. Dans le cadre de cette recherche, nous avons défini cinq dimensions (indices méthodologiques, matériels de laboratoire, autonomie de l'élève, indices conceptuels et aspects métacognitifs) qui nous ont permis de baliser le niveau de difficulté quand on passe d'un niveau à un autre. Nous savions que d'autres dimensions pouvaient également nous aider à fixer le niveau de difficulté. À la lumière de nos résultats, il pourrait être intéressant d'ajouter des dimensions et mieux préciser les dimensions actuelles afin de bonifier l'efficacité du niveau 2.

Compte tenu des résultats que nous avons obtenus pour le niveau 2 et de la lourdeur du mécanisme de passation évoqué par certains enseignants, il faudrait peut-être aussi envisager de simplifier le modèle en proposant seulement deux niveaux. Dans ce cas, nous pourrions envisager un niveau 1 qui propose quelques indices et qui serait une version hybride entre les niveaux 1 et 2 que nous avons expérimentés. La situation de

niveau 2 pourrait, quant à elle, offrir des indices dont la teneur se situerait entre les niveaux 2 et 3 que nous avons également expérimentés.

### **1.5 Le syndrome de la page blanche**

Le modèle que nous avons concocté visait, entre autres, à essayer de pallier au problème du syndrome de la page blanche (aucune information émise par l'élève par rapport au problème qui lui est soumis) quand on confronte les élèves à de véritables investigations scientifiques qui font appel à des stratégies de résolution de problème. Dans le cadre de notre expérimentation, nous avons pu observer ce syndrome surtout lors de la passation des niveaux 1 et 2. Au niveau 3, le nombre de vecteur réponse nul a pratiquement chuté à 0. Un modèle d'évaluation à multiples niveaux comme le nôtre semble donc avoir permis, dans le cadre de cette expérimentation, de réduire considérablement le syndrome de la page blanche. Ce résultat est particulièrement intéressant car notre modèle permet aux élèves qui seraient normalement en difficulté (par exemple, au niveau 1 ou au niveau 2) de fournir des traces de leurs démarches. Ce faisant, il est alors possible de recueillir des données sur le développement de la compétence. Dans le cadre d'une passation avec épreuve simple et unique pour tous (en administrant par exemple uniquement le niveau 1) nous n'aurions alors accès à aucune donnée ou à très peu de données pour appuyer le jugement.

Dans le même ordre d'idée, nous pensons que notre modèle est particulièrement adapté dans la perspective d'une évaluation différenciée. Comme nous l'avons mentionné dans le paragraphe précédent, certains élèves peuvent rencontrer des difficultés particulièrement importantes ce qui nécessite des outils d'évaluation particuliers alors que d'autres sont suffisamment avancés pour faire la démonstration de leurs apprentissages dans un contexte de pleine autonomie. En ce sens, nous pensons que notre modèle permet une meilleure flexibilité afin de s'adapter à des clientèles diversifiées d'étudiantes en terme de niveau de compétence atteint.

## **2. Résultats concernant la deuxième question de recherche**

Lorsque nous avons entrepris cette étude, nous étions convaincus que le contexte disciplinaire pouvait affecter considérablement les scores obtenus par les élèves. À cet égard, nous avons formulé la deuxième question de recherche de la façon qui suit :

Est-ce que le modèle d'évaluation à trois niveaux se comporte de la même façon selon le contexte disciplinaire privilégié ?

À notre grand étonnement, nous avons constaté que le modèle se comportait de façon semblable peu importe le contexte disciplinaire privilégié dans pratiquement chacune des situations d'évaluation. Il existe bien peu de différences entre les résultats des élèves pour l'une ou l'autre des trois situations d'évaluation. Rappelons au lecteur que les trois situations d'évaluation comportaient chacune un contexte disciplinaire distinct. La première situation d'évaluation comportait des aspects liés à la biologie, la seconde situation ne possédait pas de contexte disciplinaire particulier alors que la troisième situation était associée à la physique (thermodynamique). Les résultats que nous avons obtenus nous indiquent que rien ne permet de croire que le contexte disciplinaire joue un rôle avec les situations d'évaluation et les conditions de l'expérimentation. Cependant, nous avons noté que les différences les plus notables liées au contexte disciplinaire se situaient au niveau 1 donc dans un contexte de situation ouverte où les élèves sont confrontés à une situation qui respecte le plus l'approche par compétences. Pour les niveaux subséquents, les nuances sont plus ténues. On peut penser que les indices donnés aux élèves ont le même impact sur ces derniers peu importe le contexte disciplinaire. Si tel est le cas, cela voudrait dire que les élèves ne maîtrisent pas tant les concepts que la démarche d'investigation.

Dans la mesure où on souhaite administrer surtout la situation de niveau 1, les résultats obtenus nous amènent à être prudents dans les interprétations que nous pouvons faire sur l'effet du contexte disciplinaire. Comme nous l'indiquons au chapitre 4, il aurait été intéressant d'administrer plusieurs variantes des situations d'évaluation pour chacun des contextes disciplinaires. De cette façon, il aurait été possible de vérifier l'effet réel du

contexte disciplinaire. Dans le cadre de cette recherche, et compte tenu du contexte pédagogique qui a cours au Québec, il n'a pas été possible de donner suite à cette idée. Malgré tout, nous pouvons quand même dire que des situations d'évaluation bâties selon les règles que nous avons présentées au chapitre 2 ne semblent pas être affectées par le contexte disciplinaire. Par contre, nous ne pouvons nous prononcer sur le rendement du modèle si nous utilisons des situations d'évaluation conçues selon d'autres règles. Dans le contexte d'une évaluation certificative, ce résultat est particulièrement intéressant. En effet, une des difficultés liées à l'évaluation des compétences ou d'apprentissages complexes réside dans l'échantillonnage des situations d'évaluation. En effet, l'évaluation d'une compétence d'investigation scientifique nécessite du temps. Dans notre cas, il a fallu consacrer de 4 à 5 périodes de 75 minutes. Il est donc évident qu'il n'est pas possible d'administrer 4 ou 5 situations d'évaluation différentes dans le contexte actuel de l'enseignement des sciences au Québec. Il faut donc faire des choix qui sont souvent liés au contexte disciplinaire associé à la situation d'évaluation. Si le contexte disciplinaire a un impact majeur sur les résultats des élèves, cela soulève des questions au regard de la validité du processus d'évaluation. Dans ce contexte, quel contexte disciplinaire faut-il privilégier et pourquoi ? Si un élève est moins doué pour un contexte disciplinaire donné, faut-il modifier le contexte pour ce dernier afin de lui donner toutes les chances de montrer ses compétences ? À la lumière de nos résultats, il ne semble pas que le contexte disciplinaire ait une influence marquée quand on utilise des situations d'évaluation qui sont concordantes avec notre modèle. Nous pensons qu'il faut investiguer davantage ce résultat car s'il advenait que ce résultat soit confirmé, dans des recherches ultérieures, cela viendrait amenuiser les écueils que nous venons de décrire.

### **3. Résultats concernant la troisième question de recherche**

La troisième question s'intéressait à la robustesse du modèle d'évaluation au regard de certaines variables contextuelles. La troisième question se formulait de la manière suivante :

Dans quelle mesure un modèle d'évaluation à trois niveaux est-il sensible à certains facteurs contextuels liés aux caractéristiques des élèves et des enseignants ?

De façon opérationnelle, nous avons obtenu des données qui touchaient trois aspects et qui font l'objet des trois sous-questions qui suivent.

### **3.1 Première sous-question**

En abordant cette étude, nous avons choisi d'analyser le rendement du modèle en observant s'il existait des différences dans les scores obtenus par les garçons et les filles. De façon formelle, la première sous-question se présentait comme suit :

Est-ce que le modèle génère des différences entre les garçons et les filles ?

Globalement, nous avons remarqué qu'il n'y avait pas de réelles différences entre les scores obtenus par les garçons et ceux obtenus par les filles. Par contre, les quelques différences statistiquement significatives étaient toutes à l'avantage des filles. Parmi les hypothèses d'explication que nous pouvons évoquer, on peut penser que ces dernières ont peut-être manifestés plus de sérieux dans l'exercice qui leur était demandé. Quoi qu'il en soit, cette situation représente une bonne nouvelle car un résultat contraire nous aurait indiqué la présence d'un biais quelconque qui aurait pu avantager un genre en particulier. Dans tous les cas, mais en particulier dans le contexte d'une évaluation certificative, nous essayons généralement d'éviter ce genre de biais, ce qui semble être le cas à la lumière de nos résultats.

Plusieurs recherches et enquêtes internationales s'intéressent aux différences possibles entre les résultats obtenus par les garçons et les résultats obtenus par les filles en science. Il ne semble pas se dégager de tendances claires sur une différence possible entre les sexes. Examinons les résultats à deux de ces enquêtes. La quatrième enquête internationale sur la mathématique et les sciences (TEIMS 2003) cherche notamment à mesurer les connaissances des élèves dans plusieurs champs disciplinaires (sciences de la

vie, sciences physiques, sciences de la terre, sciences de l'environnement) à l'aide d'un instrument de mesure composé de questions à choix multiples, de questions à réponses courtes ou élaborées et de résolutions de problèmes. Mentionnons que cette épreuve ne comporte aucun volet expérimental. Dans le rapport faisant état des résultats des élèves québécois (Ministère de l'Éducation du Québec, 2004), on remarque que les élèves québécois masculins de deuxième secondaire ont obtenu de meilleurs résultats que les élèves de sexe féminin. De fait, sur les 50 états qui ont participé à l'étude, il y en avait 32 dont les garçons obtenaient de meilleurs résultats que les filles, il y en avait 11 qui ne montraient aucune différence significative et 7 étaient à l'avantage des filles. Une autre enquête bien connue est le Programme international pour le suivi des acquis des élèves (PISA). Une partie de cette enquête s'intéresse au niveau de connaissance et de compétence des élèves de 15 ans. Pour le volet scientifique, cette étude s'intéresse à trois volets : identifier des questions d'ordre scientifique, expliquer des phénomènes de manière scientifique et utiliser des faits scientifiques. Cette enquête ne comporte, tout comme la précédente, aucun volet expérimental. L'édition 2006 de cette enquête (Bussière, Knighton, & Pennock, 2007) révèle que les élèves canadiens obtiennent sensiblement les mêmes résultats qu'ils soient masculins ou féminins. Par contre, on remarque des différences quant aux résultats obtenus à l'un ou l'autre des volets. Par exemple, les garçons obtiennent de meilleurs résultats au volet « expliquer des phénomènes de manière scientifique » alors que les filles obtiennent de meilleurs résultats que les garçons au volet « identifier des questions d'ordre scientifique ».

### **3.2 Deuxième sous-question**

La deuxième sous-question concernait l'effet de l'expérience des enseignants sur les résultats obtenus par les élèves dans le contexte de l'expérimentation du modèle que nous proposons. Pour le bénéfice du lecteur, nous rappelons l'énoncé de cette sous-question qui était :

Est-ce que le modèle permet de distinguer les enseignants qui sont expérimentés en enseignement des sciences ?

Les résultats que nous avons obtenus ne nous permettent pas de supposer un lien direct entre l'expérience et les résultats obtenus par les élèves. En fait, le type d'épreuves développées et les conditions de passation ne permettent pas de distinguer les enseignants d'expérience. Pour le niveau 1, aucun groupe<sup>35</sup> d'enseignant ne s'est démarqué dans le cas des 3 situations d'évaluation. À titre d'exemple, les enseignants ayant entre 6 et 10 années d'expérience ont vu leurs élèves obtenir des résultats inférieurs aux autres élèves dans le cas de la deuxième situation d'évaluation. Par contre, cette différence n'est pas présente dans le cas des 2 autres situations d'évaluation. Dans l'ensemble, la variable liée à l'expérience n'a pas semblé jouer un rôle important sur les scores obtenus par les élèves. De fait, en comparant les neuf situations d'évaluation, on constate que seulement deux d'entre elles montrent des différences statistiquement significatives. Dans la grande majorité des cas, l'effet « expérience » semble tout à fait négligeable. On peut supposer que les nouveautés liées à la fois aux situations d'évaluation et également à la méthodologie proposée ont été suffisamment importantes pour placer tous les enseignants sur le même pied d'égalité.

### **3.3 Troisième sous-question**

La troisième sous-question concernait l'effet de la formation des enseignants. Voici le libellé de cette dernière sous-question :

Est-ce que le modèle permet de distinguer les enseignants qui sont formés adéquatement pour développer chez leurs élèves l'investigation scientifique ?

De façon étonnante, les élèves ayant un enseignant formé en science n'ont pas obtenu nécessairement les meilleurs résultats. Au départ, nous nous attendions, de façon tout à fait logique, à ce que les enseignants possédant une solide formation scientifique soient davantage en mesure de développer la compétence d'investigation scientifique chez leurs

---

<sup>35</sup> Les groupes étaient définis, selon leur expérience, de la façon suivante : entre 0 et 5 ans, entre 6 et 10 ans et plus de 11 ans.

élèves. Nos résultats nous indiquent que tel n'est pas le cas. Ce résultat n'est pas nécessairement étonnant car certaines recherches montrent la difficulté à lier la formation des enseignants aux résultats d'apprentissage de leurs élèves. Dans le cas qui nous occupe, nous pensons que plusieurs raisons pourraient expliquer ce constat.

Premièrement, il est possible que le caractère de nouveauté inhérent à notre modèle ait eu exactement le même impact chez tous les enseignants nonobstant leur formation. Dans la mesure où aucun enseignant n'a travaillé en cours d'apprentissage avec notre modèle, il est possible que les élèves aient été déstabilisés également face à la tâche qui leur était demandée. Par contre, on peut se demander pourquoi un élève ayant développé sa compétence n'arrive pas à la manifester dans un contexte complètement nouveau ? Éventuellement, il serait fort intéressant de mieux contrôler cette variable. Par exemple, nous pourrions recommencer notre expérimentation mais auprès d'enseignants que nous aurions formé à l'utilisation de notre modèle d'évaluation au début de l'année scolaire. De cette façon, nous contrôlerions mieux l'aspect de nouveauté lié au modèle d'évaluation.

Deuxièmement, nous pensons qu'une des raisons expliquant le faible impact de la formation des enseignants concerne la difficulté intrinsèque à développer des compétences chez les élèves. Nous pouvons penser qu'un enseignant qui maîtrise fort bien les aspects disciplinaires (concepts, techniques, démarches, etc.) n'arrive pas nécessairement automatiquement à les développer convenablement chez les élèves. Les résultats que nous obtenons nous indiquent peut-être simplement que les enseignants n'ont pas reçu la formation nécessaire pour développer des compétences chez leurs élèves. De fait, tous les enseignants seraient au même niveau peu importe leur formation initiale (scientifique ou pas).

## Limites de la recherche

*A priori*, nous pouvons identifier plusieurs limites inhérentes à des recherches comme la nôtre : l'échantillonnage, le niveau d'appropriation du nouveau programme, les conditions d'application du nouveau programme, le respect des conditions expérimentales par les enseignants, la qualité de nos instruments de mesure, le niveau d'accord intra juge, l'effet de contamination et les ressources disponibles pour mener à bien cette recherche.

Premièrement, nous aurions vivement souhaité avoir un échantillon aléatoire d'enseignants mais cela n'a pas été possible. Pour le recrutement de notre échantillon, nous avons lancé un « appel à tous » dans plusieurs commissions scolaires en mentionnant que la participation à cette étude était sur une base de volontariat. Nous sommes donc conscient que notre échantillon pourrait être biaisé en comportant, par exemple, davantage d'enseignants qui sont plus favorables à l'expérimentation de nouvelles méthodes d'évaluation. Concernant les échantillons liés aux élèves, le lecteur aura remarqué que plusieurs de ces derniers sont de petites tailles comme par exemple les résultats associés au profil 2 (voir tableau XXXVI). Idéalement, nous aurions espéré obtenir des échantillons contenant plus de données. Les interprétations et les conclusions de cette étude doivent tenir compte de cette réalité.

Deuxièmement, les données ont été recueillies après une seule année d'appropriation du nouveau programme. Lorsqu'un enseignant administre pour une première fois une situation d'apprentissage ou d'évaluation, il acquiert généralement beaucoup d'expérience qu'il s'empresse, la fois suivante, de réinvestir. Compte tenu du contexte de cette recherche, les enseignants auront vécu, tout au long de l'année, une première expérience pour chacune des situations d'apprentissage qu'ils auront fait vivre à leurs élèves. Les inférences que nous avons réalisées – au regard des résultats obtenus par les élèves aux situations d'évaluation – devront tenir compte de cet aspect.

Troisièmement, les conditions d'application du nouveau programme demeurent relativement variables d'une institution scolaire à l'autre. Il nous est difficile de s'assurer que tous les enseignants qui ont participé à cette recherche ont eu les ressources suffisantes afin de bien mener les situations d'évaluation. À titre d'exemple, nous savons que certaines écoles considèrent prioritaire que les techniciens de laboratoire s'occupent d'assurer une présence en classe auprès des groupes d'élèves du deuxième cycle où les risques liés à l'expérimentation en laboratoire sont plus élevés. Là où cette situation prévaut, les enseignants du premier cycle ont alors été désavantagés en n'ayant pas une présence suffisante en classe de la part du technicien de laboratoire.

Quatrièmement, nous avons préparé soigneusement la documentation qui accompagnait les situations d'évaluation en expliquant clairement la marche à suivre. Nous avons fourni des documents aux enseignants et aux techniciens de laboratoire afin de bien leur expliquer le contexte de la recherche, la nature des situations d'évaluation et les modalités de passation. Nous avons également produit un site web ainsi qu'une vidéo explicative afin d'uniformiser le plus possible les conditions de passation. Cependant, nous ne pouvions être totalement à l'abri de mauvaises interprétations de la part des participants. Afin d'éviter cette situation, nous avons tenté de nous assurer que les consignes avaient été bien comprises de la part des enseignants et nous nous sommes rendus disponible afin de répondre à leurs questions lors de l'expérimentation. Nous avons rencontré les enseignants à deux reprises avant l'expérimentation afin de bien les informer sur les différents aspects de cette étude, sur leur rôle et sur les exigences de cette recherche. Malgré tout, il demeure impossible de s'assurer que tous les enseignants ont appliqué uniformément le protocole de recherche que nous leur avons soumis. L'émergence du quatrième profil l'illustre bien. En ce sens, nos résultats doivent être interprétés à la lumière de cette situation.

Cinquièmement, nous avons dû construire nos propres instruments de mesure afin de déterminer, par exemple, le niveau de compétence des élèves. Comme nous l'avons mentionné à plusieurs reprises, la rédaction de situations d'évaluation de type *hands-on* n'est pas une entreprise aisée. Malgré les mécanismes de validation que nous avons mis

en place pour nous assurer de la qualité des données, nous ne sommes pas à l'abri de biais. Pour l'instant, le niveau de connaissance concernant la rédaction de telles situations est encore bien parcellaire comme l'indiquent aussi Stecher, Klein, Solano-Flores, McCaffrey, Robyn, Shavelson et Haertel (2000, p. 155) :

*At our present state of knowledge, test development is as much an art as a science. We suspect that no set of rules, however well documented, will fully capture all of the element of a well-constructed task, and that no set of rules will guarantee that tasks similar on the surface will elicit similar performances from students.*

Sixièmement, nous savons que la correction de situations basées sur la performance s'appuie largement sur le jugement du correcteur. Dans le cadre de cette recherche, seul le chercheur principal a effectué ce travail afin d'éliminer les biais liés à la correction inter juges. Par contre, nous ne sommes pas à l'abri de biais liés à la correction intra juge. Afin de minimiser ce biais, nous avons effectué une contre correction qui nous a donné un taux d'accord supérieur à 90%. En ce sens, l'expérience ne tient pas compte d'une source de variation qui existe toujours dans la pratique : les possibles différences entre les correcteurs.

Septièmement, il faut souligner les effets de contamination qui ont pu avoir un impact sur la qualité des données puisque les situations d'évaluation ont été réalisées sur plusieurs périodes de classe. De fait, les élèves ont ainsi pu se parler entre les périodes où ils devaient réaliser leur situation d'évaluation. Cet aspect soulève des questions au niveau de l'applicabilité de l'administration de telles situations d'évaluation. Là où ont eu lieu nos expérimentations, très peu d'écoles étaient en mesure de bloquer des périodes à l'horaire (par exemple en bloquant un avant-midi ou un après-midi complet) afin de permettre aux élèves de bénéficier de quelques heures continues pour réaliser leur situation d'évaluation. Les contraintes liées à la disponibilité des classes de laboratoire et la disponibilité des techniciens de laboratoire ne sont pas étrangères à cette situation.

Enfin, nous tenons à mentionner que cette recherche n'a pas été subventionnée. Cela n'a pas influencé la qualité des analyses. Par contre, plus de moyens financiers nous aurait

permis une cueillette de données plus substantielle ce qui aurait somme toute bonifié cette thèse.

## Conclusion

---

Pour terminer, nous présenterons un bref résumé de cette recherche en mettant l'accent sur les principaux résultats de recherche que nous avons obtenus. Par la suite, nous discuterons de la pertinence de cette recherche et son originalité. Enfin, nous indiquerons les recherches futures qu'il pourrait être possible de mener en s'appuyant sur les résultats que nous avons obtenus dans le cadre de cette étude.

### Résumé de la recherche

Au Québec, l'enseignement des sciences a vécu au cours des dernières années de profonds changements. En effet, les enseignants ont vu les programmes par objectifs être remplacés graduellement par des programmes par compétences. L'approche par discipline unique (ex. cours distincts de sciences physiques et de biologie) a été délaissée pour une approche par disciplines multiples (chimie, physique, biologie, géologie, astronomie, technologie). Le paradigme dans lequel considérer les apprentissages a aussi été ajusté. En effet, les nouveaux programmes prennent maintenant appui sur des conceptions de l'apprentissage à caractère cognitiviste, constructiviste ou socioconstructiviste. Tous ces choix ont eu des impacts sur les façons d'enseigner et inévitablement sur les façons d'évaluer les apprentissages des élèves. En particulier, la question de l'évaluation des compétences scientifiques pose de nombreux défis, entre autres, méthodologiques. Dans le cadre de cette recherche doctorale, nous nous sommes intéressé précisément aux modalités qui doivent être prises en compte afin d'évaluer une compétence d'investigation scientifique en laboratoire. Ainsi, nous nous sommes inspirés d'une proposition de Rey *et al.* (2003) et nous avons conçu un modèle d'évaluation certificative visant à permettre aux enseignants en exercice de porter un jugement, que nous souhaitons le plus valide possible, sur la compétence de leurs élèves à réaliser des investigations scientifiques en laboratoire. Une des principales particularités de ce modèle est de proposer des outils d'évaluation à niveaux multiples permettant à différentes clientèles d'élèves d'exprimer leur compétence sans dénaturer la valeur du problème qui

leur est soumis. Aussi, le modèle permet d'offrir à tous les élèves en difficulté le même niveau d'aide, ce qui est primordial dans le contexte d'une évaluation à enjeux critiques.

Cette recherche à caractère exploratoire visait à répondre à trois questions de recherche. Dans un premier temps, nous avons cherché à déterminer s'il y avait des avantages à recourir à un modèle d'évaluation à trois niveaux afin de rendre compte du développement d'une compétence d'investigation scientifique en laboratoire. Dans un second temps, nous avons tenté de cerner si le contexte disciplinaire principal pouvait avoir un impact sur les résultats des élèves. Dans un troisième temps, nous nous sommes penché sur l'étude de certaines variables contextuelles et leurs effets possibles sur les résultats des élèves.

Afin de répondre à nos questions de recherche, nous avons rédigé, validé et expérimenté trois situations d'évaluation comportant chacune trois niveaux. Ces situations d'évaluation ont ensuite été administrées auprès de 560 élèves de la première secondaire issus de 5 commissions scolaires. Au total, 22 enseignants ont accepté de participer à cette recherche. Puisque nous souhaitons proposer des solutions aux praticiens, nous avons choisi d'expérimenter les situations d'évaluation en contexte réel auprès d'élèves et d'enseignants représentatifs du milieu scolaire québécois.

Les résultats que nous avons obtenus nous indiquent que le modèle permet d'obtenir des données intéressantes sur le cheminement d'élèves pouvant être regroupés en trois catégories : élèves ayant globalement beaucoup de difficultés, élèves manifestant certaines difficultés et élèves exprimant peu de difficultés dans le déploiement de leur compétence d'investigation scientifique. Nous avons remarqué que le modèle était pertinent puisqu'un nombre important d'élèves a été dans l'obligation de recevoir soit le niveau 2 soit le niveau 3 pour la situation d'évaluation qui leur était assignée. À ce sujet, nous avons également remarqué que le niveau 2 ne semblait pas offrir aux élèves une aide suffisante pour leur permettre de résoudre la tâche qui leur était proposée. Par contre, le niveau 3 a permis à une majorité d'élèves de faire la démonstration de leurs habiletés. Étonnamment, nous n'avons pas observé de différences entre les résultats des élèves pour

l'une ou l'autre des trois situations d'évaluation. Ainsi, les élèves ayant reçu la situation d'évaluation dans le contexte des sciences physiques ont obtenu à peu près les mêmes résultats que ceux ayant reçu la situation dans le contexte de la biologie ou selon un contexte générique. Selon ce que nous avons pu observer, l'effet du contexte disciplinaire ne semble pas avoir joué un très grand rôle en ce qui concerne les résultats des élèves. Dans le cadre de notre étude, nous nous sommes également intéressé à l'effet possible de trois variables contextuelles : le genre des sujets, l'expérience des enseignants ainsi que leur formation initiale. Nous n'avons pas observé de différence entre les résultats des garçons et des filles et ce pour l'ensemble des situations d'évaluation expérimentées. Au sujet des enseignants, leur expérience et leur formation initiale n'ont pas eu un effet déterminant sur les résultats des élèves.

### **Pertinence de la recherche**

À notre connaissance, aucune recherche s'apparentant à la nôtre n'avait encore été menée. Les travaux de Rey *et al.* (2003), sur lesquelles nous nous sommes appuyé pour construire notre modèle, ont été réalisés dans un contexte autre que celui de l'enseignement scientifique. Par conséquent, nous pensons que les données et les résultats issus de cette recherche sont uniques. Aussi, nous avons tenté – du mieux que nous avons pu – de nous arrimer au nouveau programme de science et technologie afin que les connaissances produites par cette recherche soient les plus pertinentes possibles autant pour les chercheurs que les praticiens. Au terme de cette recherche, nous pensons que nos travaux peuvent alimenter la réflexion concernant l'implantation d'un dispositif d'évaluation dans le contexte de l'enseignement secondaire québécois. L'originalité de cette recherche repose également sur le fait d'avoir réussi à obtenir des données réelles obtenues dans un contexte pratique auprès d'élèves engagés dans un programme de science et technologie formulé par compétences. Selon nous, il y a présentement très peu de recherches qui présentent des données sur l'efficacité d'un modèle d'évaluation permettant d'évaluer des compétences.

## **Recherches futures**

Cette recherche en est résolument une à caractère exploratoire. Comme nous le mentionnions au premier chapitre, il n'existe que bien peu de recherches qui se sont intéressées à l'évaluation des compétences et encore moins en ce qui concerne les compétences liées à l'investigation scientifique. Les résultats que nous avons obtenus en menant cette recherche sont suffisamment intéressants pour nous suggérer plusieurs pistes de recherche. Nous pouvons classer ces dernières en deux catégories : (1) celles nécessitant une nouvelle cueillette de données et (2) celles nécessitant le recours à d'autres outils d'analyse afin de traiter les données brutes.

### **Pistes de recherche éventuelle nécessitant une nouvelle expérimentation**

Tel que nous l'avons mentionné au chapitre 5, les situations d'évaluation du niveau 2 ont eu une certaine utilité pour aider certains élèves en difficulté. Cependant, nous pensons qu'il serait possible de les bonifier afin de les rendre plus efficaces auprès de tous les élèves manifestant des difficultés. Nous pensons qu'il pourrait être utile de procéder à une nouvelle expérimentation qui s'appuierait sur des situations de niveau 2 remaniées qui tiendraient mieux compte des besoins des élèves en difficulté. À titre d'exemple, nous pensons qu'il pourrait être pertinent d'offrir plus d'indices conceptuels aux élèves et aussi plus d'indices méthodologiques en lien avec le problème qui leur est dévolu. Ce travail permettrait également d'enrichir notre compréhension des étapes définissant le développement d'une compétence d'investigations scientifique.

Il pourrait être pertinent de reproduire cette expérimentation avec le même échantillon d'enseignants. De cette façon, nous pourrions éliminer l'effet de nouveauté lié au format d'administration des situations d'évaluation. En effet, pour la majorité des enseignants et des élèves, le format des situations d'évaluation était nouveau. Il est possible que cet effet nouveauté ait joué un rôle non négligeable sur la qualité des données bien que nous ne soyons pas en mesure de quantifier l'impact de cet effet. Dans le contexte d'une évaluation certificative à enjeux critique, il serait impératif de contrer cet effet afin

d'éviter des préjudices possibles aux élèves (American educational research association, American psychological association, & National council on measurement in education, 1999). Tous ces arguments nous semblent suffisamment pertinents pour justifier une seconde expérimentation en recourant aux mêmes enseignants qui ont participé à la présente recherche. Non seulement seraient-ils plus à l'aise de procéder à l'administration des situations d'évaluation mais ils auraient également l'occasion de mieux préparer leurs élèves au format des outils de cueillette de données. Une recherche menée de façon longitudinale pourrait être intéressante aussi pour voir dans quelle mesure les enseignants appliquent le nouveau programme de science et technologie. En effet, nous devrions nous attendre à voir diminuer la fréquence d'utilisation des situations de niveau 3 au fur et à mesure que les enseignants appliquent les nouvelles prescriptions ministérielles. Normalement, le même phénomène devrait être observé dans le cas du niveau 2.

Dans le cadre de cette recherche, l'une des variables que nous souhaitions investiguer était celle concernant l'effet du contexte disciplinaire. Nous avons ainsi comparé les résultats de situations issues de trois contextes disciplinaires différents. Par contre, nous avons développé seulement une situation d'évaluation par contexte disciplinaire. Éventuellement, il serait intéressant de créer d'autres situations disciplinaires afin d'obtenir suffisamment de variance afin de se prononcer de façon plus solide sur l'effet disciplinaire. Un plan d'expérimentation cubique de type 3X3 (3 contextes disciplinaires X 3 niveaux par SE X 3 tâches par contexte disciplinaire) pourrait être conçu et expérimenté.

### **Pistes de recherche éventuelle nécessitant le recours à d'autres outils d'analyse**

Les données que nous avons recueillies se révèlent être d'une grande richesse au point de vue de l'évaluation mais également à l'égard de plusieurs aspects plus spécifiques à la mesure. Lors de l'analyse des données, nous avons survolé certaines perspectives concernant le traitement des données brutes. Les essais que nous avons réalisés en ayant recours au modèle de Rasch (1960) se sont avérés particulièrement prometteurs. Les

différentes analyses que nous avons réalisées avec le logiciel *Winsteps* (Linacre, 2006) nous permettent d'entrevoir qu'il serait possible de bonifier à la fois les situations d'évaluation et également la grille de notation que nous avons employées pour établir les scores pour chacun des douze critères de correction. Dans l'éventualité où nous procéderions à une nouvelle expérimentation, il pourrait être fort pertinent de d'abord réaliser ces analyses afin de bonifier davantage les outils de cueillette de données et de correction. Éventuellement, nous souhaitons également explorer les possibilités qu'offre le logiciel *Facets* qui permet de placer différentes dimensions sur une même échelle. Dans le cas qui nous concerne, nous pourrions modéliser avec cinq dimensions qui sont : les critères, les juges, les niveaux, les situations d'évaluation et les élèves. Dès lors, il serait possible d'identifier les aspects qui ne se conforment pas au modèle.

Cette thèse s'inscrit dans le champ de la mesure et de l'évaluation. Nous avons donc traité les données en nous appuyant sur les principes de la mesure et de l'évaluation. Il serait cependant possible d'analyser les données en y jetant un regard plus didactique afin de mieux cibler les aspects qui mériteraient d'être revus et corrigés, par exemple, dans les niveaux 2 et 3. Nous pensons que ce travail pourrait augmenter la validité de construit des épreuves. Aussi, nous sommes très conscient que nous avons été dans l'obligation de faire des choix en ce qui concerne la forme et les contenus des situations d'évaluation. Dans l'avenir, au gré de l'application des nouveaux programmes du MELS, il est possible que les situations d'évaluation de la compétence d'investigation scientifique prennent un autre format avec d'autres contenus que ceux que nous avons définis. Quoi qu'il en soit, nous pensons que notre modèle sera toujours applicable avec de nouvelles situations d'évaluation.

## Références

---

- Accongio, J. L., & Doran, R. L. (1993). *Classroom assessment: Key to reform in secondary science education* (Evaluative/Feasibility).
- American Association for the Advancement of Science. (1993). *Benchmarks for science literacy, project 2061*. New-York: Oxford University Press.
- American educational research association, American psychological association, & National council on measurement in education. (1999). *Standards for educational and psychological testing*. Washington, DC: American educational research association.
- Anderson, L. W. (2002). Curricular alignment: A re-examination. *Theory into practice*, 41(4), 255-260.
- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-18.
- Arsac, G., Germain, G., & Mante, M. (1988). *Problème ouvert et situation-problème*. Villeurbanne: Irem, Université Claude Bernard (Lyon 1).
- Arter, J. A., & McTighe, J. (2001). *Scoring rubrics in the classroom : using performance criteria for assessing and improving student performance*. Thousand oaks, California: Corwin Press.
- Aschbacher, P. R. (1991). Performance assessment : State activity, interest, and concerns. *Applied measurement in education*, 4(4), 275-288.
- Astolfi, J.-P., Darot, É., Ginsburger-Vogel, Y., & Toussaint, J. (1997). *Mots-clés de la didactique des sciences : Repères, définitions, bibliographies*. Bruxelles: DeBoeck Université.

- Auger, R., & Dassa, C. (1992). Les pratiques de mesure et d'évaluation des apprentissages et la validité des tests dans un contexte de démarche évaluative. Dans D. Laveault (Ed.), *Les pratiques d'évaluation en éducation : Textes rédigés en vue de la XV<sup>e</sup> session d'étude de l'Association pour le développement de la mesure et de l'évaluation en éducation* (pp. 151-165). Montréal: Éditions de l'ADMÉE et M éditeur.
- Ayala, C. C., Yin, Y., Schultz, S., & Shavelson, R. (2002). *On Science Achievement from the Perspective of Different Types of Tests: A Multidimensional Approach to Achievement Validation. CSE Technical Report* (Reports - Research No. CSE-TR-572, R305B60002, REC9628293): California Univ., Los Angeles. Center for the Study of Evaluation.; Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
- Bain, D. (1999). De l'évaluation aux compétences : mise en perspective de pratiques émergentes. Dans J. Dolz & E. Ollagnier (Eds.), *L'énigme de la compétence en éducation* (Vol. 1-2, pp. 129-145). Bruxelles: DeBoeck & Larcier.
- Baker, E. L. (1997). Model-Based Performance Assessment. *Theory into Practice*, 36(4), 247-254.
- Baker, E. L. (2004). *Aligning Curriculum, Standards, and Assessments: Fulfilling the Promise of School Reform. CSE Report 645*: Center for Research on Evaluation Standards and Student Testing CRESST.
- Baxter, G. P. (1991). *Exchangeability of science performance assessments*. University of California, Santa Barbera.
- Baxter, G. P., Elder, A. D., & Glaser, R. (1996). Knowledge-based cognition and performance assessment in the science classroom. *Educational psychologist*, 31(2), 133-140.

- Baxter, G. P., Shavelson, R. J., Goldman, S. R., & Pine, J. (1992). Evaluation of procedure-based scoring for hands-on science assessment. *Journal of Educational Measurement, 29*(1), 1-17.
- Beckers, J. (2002). *Développer et évaluer des compétences à l'école : vers plus d'efficacité et d'équité*. Bruxelles: Éditions Labor.
- Bélair, L. M. (1999). *L'évaluation dans l'école : Nouvelles pratiques*. Paris: ESF.
- Bloom, B. S. (1975). *Taxonomie des objectifs pédagogiques, 1 domaine cognitif* (M. Lavallée, Trans.). Québec: Les Presses de l'Université du Québec.
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New-York: McGraw-Hill Book Co.
- Brookhart, S. M. (1994). Teachers' Grading: Practice and Theory. *Applied Measurement in Education, 7*(4), 279-301.
- Brookhart, S. M. (1999). *The Art and Science of Classroom Assessment. The Missing Part of Pedagogy. ASHE-ERIC Higher Education Report, Volume 27, Number 1* (Classroom Use - Teaching Guides No. ED-99-00-0036): Association for the Study of Higher Education.[BBB14212], ERIC Clearinghouse on Higher Education, Washington, DC.[BBB15669], George Washington Univ., Washington, DC. Graduate School of Education and Human Development.[BBB32577].
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice, 22*(4), 5-12.
- Brown, J. H., & Shavelson, R. J. (1996). *Assessing hands-on science: A teacher's guide to performance assessment*. Thousand Oaks, CA: Corwin Press, inc.
- Bussière, P., Knighton, T., & Pennock, D. (2007). *À la hauteur : Résultats canadiens de l'étude PISA de l'OCDE, la performance des jeunes du Canada en sciences, en lecture et en mathématiques, premiers résultats de 2006 pour les Canadiens de 15*

ans. Ottawa: Ressources humaines et Développement social Canada, Conseil des ministres de l'Éducation (Canada) et Statistique Canada.

- Capéraà, P., & Van Cutsem, B. (1988). *Méthodes et modèles en statistique non paramétrique : exposé fondamental*. Québec: Presses de l'Université Laval.
- Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 20(4), 19-27.
- Conseil des ministres de l'éducation du Canada. (2003). *Cadre commun de résultats d'apprentissage en sciences de la nature*. Toronto: Gouvernement du Canada.
- Crahay, M., & Detheux, M. (2005). L'évaluation des compétences, une entreprise impossible? (Résolution de problèmes complexes et maîtrise de procédures mathématiques). *Mesure et évaluation en éducation*, 28(1), 57-78.
- Dionne, E. (1999). État de l'application du constructivisme en sciences physiques 416-436 dans une commission scolaire québécoise, *Mémoire de maîtrise inédit*. Montréal: Université de Montréal.
- Dionne, E. (2004). L'évaluation des apprentissages dans le cadre de la réforme : un défi de taille qu'il faudra bien relever... *Spectre*, 33(3), 22-23.
- Dionne, E. (2005). Enjeux en évaluation des apprentissages dans le cadre de programmes d'études formulés par compétences : l'exemple de science et technologie. *Mesure et évaluation en éducation*, 28(2), 49-66.
- Dionne, E., & Blais, J.-G. (2004). L'évaluation dans le cadre des nouveaux programmes de science et technologie au secondaire : constats et pistes de réflexion, *Communication présentée dans le cadre de la 26ième session d'étude de l'Association pour le Développement de la Mesure et de l'Évaluation en Éducation (A.D.M.É.É.)*. Laval.
- Dionne, E., & Potvin, P. (2005). Étude des difficultés et enjeux liés à l'implantation d'un programme par compétence en science et technologie dans une école secondaire

québécoise, *Communication présentée dans le cadre du 73ième congrès de l'Association francophone pour le savoir - Acfas*. Chicoutimi (Québec).

Doran, R. L., & Hejaily, N. (1992). Hands-on evaluation: a how to guide. *Science Scope*, 15(6), 9-11.

Dumas-Carré, A., Goffard, M., & Gil, D. (1992). Difficultés des élèves liées aux différentes activités cognitives de résolution de problèmes. *Aster*, 14, 53-75.

Duschl, R. A. (2003). Assessment of inquiry. Dans J. M. Atkin & J. E. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 41-59). Arlington, VA: NSTApress.

Éducation Formation professionnelle et jeunesse Manitoba. (2001). *Sciences de la nature Secondaire 1 Programme d'études : Document de mise en oeuvre*. Winnipeg, Manitoba: Non publié.

Excel. (2003). Excel (Version 11.5612.5606). Redmond, Washington: Microsoft Corporation.

Fabre, M. (1999). *Situations-problèmes et savoir scolaire*. Paris: Presses universitaires de France.

Fogarty, R. (1991). Ten ways to integrate curriculum. *Educational Leadership*, 49(2), 61-65.

Fourez, G. (1994). *Alphabétisation scientifique et technique : Essai sur la finalité de l'enseignement des sciences*. Bruxelles: De Boeck Université.

Gao, X. (1992). *Generalizability of a state-wide science performance assessment*. University of California, Santa Barbara.

Gitomer, D. H. (1993). Performance assessment and educational measurement. Dans R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement : issues in constructed response, performance testing, and portfolio*

*assessment* (pp. 241-263). Hillsdale, New-Jersey: Lawrence Erlbaum Associates, inc.

Gordon, S. P., & Reese, M. (1997). High-Stakes Testing : Worth the Price ? *Journal of School Leadership*, 7(4), 345-368.

Gott, R. (1987). The Assessment of Practical Investigations in Science. *School Science Review*, 68(244), 411-421.

Gott, R., & Duggan, S. (2002). Problems with the Assessment of Performance in Practical Science: Which Way Now? *Cambridge Journal of Education*, 32(2), 183-201.

Hadji, C. (1997). *L'évaluation démystifiée*. Paris: ESF.

Haury, D. L., & Rillero, P. (1994). *Perspectives of Hands-On Science Teaching*. Columbus, Ohio: ERIC Clearinghouse for Science, Mathematics and Environmental Education.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for supervision and curriculum development.

Hibbard, M. K. (1996). *A teacher's guide to performance-based learning and assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Howell, D. C. (1998). *Méthodes statistiques en sciences humaines* (M. Rogier, Trans.). Paris: DeBoeck Université.

Huberman, M. A., & Miles, M. B. (1991). *Analyse des données qualitatives : recueil de nouvelles méthodes* (C. De Backer & V. Lamongie, Trans.). Bruxelles: De Boeck Université.

- Hunter, D. M., Jones, R. M., & Randhawa, B. S. (1996). The use of holistic versus analytic scoring for large-scale assessment of writing. *The canadian journal of program evaluation, 11*(2), 61-85.
- International Technology Education Association. (2000). *Standards for technological literacy : content for the study of technology*. Virginia (USA): International technology education association.
- Jackson, V. (2001). The multidimensional assessment of student performance in middle school science. Dans D. P. Shepardson (Ed.), *Assessment in science : A guide to professional development and classroom practice* (pp. 181-196). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Jovanovic, J., Solano-Flores, G., & Shavelson, R. J. (1994). Performance-based assessments: will gender differences in science achievement be eliminated? *Education and urban society, 26*(4), 352-366.
- Klein, S. P., Jovanovic, J., Stecher, B. M., McCaffrey, D., Shavelson, R. J., Haertel, E., et al. (1997). Gender and racial/ethnic differences on performance assessments in science. *Educational Evaluation and Policy Analysis, 19*(2), 83-97.
- Klein, S. P., Stecher, B. M., Shavelson, R. J., McCaffrey, D., Ormseth, T., Bell, R. M., et al. (1998). Analytic versus Holistic Scoring of Science Performance Tasks. *Applied Measurement in Education, 11*(2), 121-137.
- Kober, N. (1993). *What We Know about Science Teaching and Learning*. (No. RP91002001, RP91002010): Council for Educational Development and Research, Washington, DC.[BBB10294].
- Lasnier, F. (2000). *Réussir la formation par compétences*. Montréal: Guérin.
- Laurier, M. D., Tousignant, R., & Morissette, D. (2005). *Les principes de la mesure et de l'évaluation des apprentissages* (3 ed.). Montréal: Gaëtan Morin éditeur.

- Lawrenz, F., Huffman, D., & Welch, W. (2001). The science achievement of various subgroups on alternative assessment formats. *Science Education*, 85(3), 279-290.
- Legendre, M.-F. (2001). Favoriser l'émergence de changements en matière d'évaluation des apprentissages. *Vie pédagogique*, 120, 15-19.
- Leighton, J. P., Rogers, W. T., & Maguire, T. O. (1999). Assessment of Student Problem-Solving on Ill-Defined Tasks. *Alberta Journal of Educational Research*, 45(4), 409-427.
- Linacre, J. M. (2006). Winsteps Rasch Measurement (Version 3.60.1). Chicago: Winsteps.com.
- Lump, A. T., & Oliver, J. S. (1991). Dimensions of hands-on science. *The american biology teacher*, 53(6), 345-348.
- MacDonald, R. P. (1999). *Test theory a unified treatment*. New-Jersey: Lawrence Erlbaum associates.
- Maine Department of Education. (2001). Bouncing balls. Retrieved 2006-02-16, 2006, from <http://www.maptasks.org/index.php>
- McColskey, W., & O'Sullivan, R. (1993). *How To Assess Student Performance in Science: Going beyond Multiple-Choice Tests. A Resource Manual for Teachers* (Non-Classroom Use No. RP91002010): Southeastern Regional Vision for Education (SERVE), Tallahassee, FL.[BBB29512].
- McMillan, J. H. (2001). *Classroom assessment : Principles and practice for effective instruction* (2 ed.). Needham Heights, MA: Allyn and Bacon.
- McMillan, J. H., & Workman, D. J. (1998). *Classroom Assessment and Grading Practices: A Review of the Literature* (INFORMATION ANALYSES (State-of-the-Art Papers, Research Summaries, Reviews of the Literature on a Topic)): Metropolitan Educational Research Consortium, Richmond, VA.[BBB33103].

- Mehrens, W. A. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice*, 11(1), 3-9.
- Mertler, C. A. (2001). Designing Scoring Rubrics for Your Classroom. *Practical Assessment, Research & Evaluation*, 7(25).
- Millman, J., & Greene, J. (1993). The specification and development of tests of achievement and ability. Dans R. L. Linn (Ed.), *Educational measurement* (pp. 335-366). Phoenix: American Council on Education, series on higher education, Oryx Press.
- Ministère de l'Éducation du Loisir et du Sport. (2005). Statistiques de l'éducation. Retrieved 4 mai, 2007, from [http://www.mels.gouv.qc.ca/stat/stat\\_edu/index.htm](http://www.mels.gouv.qc.ca/stat/stat_edu/index.htm)
- Ministère de l'Éducation du Loisir et du Sport (Writer) (2006a). Échelles de niveaux de compétence, enseignement secondaire, premier cycle. Québec.
- Ministère de l'Éducation du Loisir et du Sport. (2006b). Statistiques détaillées sur les effectifs scolaires et les diplômes décernés. Retrieved 4 mai, 2007, from <http://www.mels.gouv.qc.ca/stat/index.htm>
- Ministère de l'Éducation du Québec. (1997). *Réaffirmer l'école, Rapport du Groupe de travail sur la réforme du curriculum*. Québec.
- Ministère de l'Éducation du Québec. (2001). *Programme de formation de l'école québécoise, éducation préscolaire, enseignement primaire*. Québec.
- Ministère de l'Éducation du Québec. (2003a). *Plan de mise en oeuvre de la politique d'évaluation des apprentissages*. Québec: Gouvernement du Québec.
- Ministère de l'Éducation du Québec. (2003b). *Politique d'évaluation des apprentissages : Formation générale des jeunes, Formation générale des adultes, Formation professionnelle*. Québec.

- Ministère de l'Éducation du Québec. (2003c). *Programme de formation de l'école québécoise, enseignement secondaire, premier cycle*. Québec.
- Ministère de l'Éducation du Québec. (2004). *Résultats obtenus par les élèves québécois aux épreuves de mathématique et de sciences de 2003*. Québec: Gouvernement du Québec.
- Ministère de la communauté française de Belgique. Socle de compétences, éveil initiation scientifique, enseignement fondamental et premier degré de l'enseignement secondaire.
- Mislevy, R. J. (2004). The Case for an Integrated Design Framework for Assessing Science Inquiry. CSE Report 638. *Center for Research on Evaluation Standards and Student Testing CRESST*.
- Morissette, D. (1984). *La mesure et l'évaluation en enseignement*. Sainte-Foy: Les presses de l'Université Laval.
- Moskal, B. M. (2003). Recommendations for developing classroom performance assessments and scoring. *Practical Assessment, Research & Evaluation*, 8(14), Retrieved October 5, 2005 from <http://PAREonline.net/getvn.asp?v=2008&n=2014>.
- Nitko, A. J. (1993). Designing tests that are integrated with instruction. Dans R. L. Linn (Ed.), *Educational measurement* (pp. 447-474). Phoenix, AZ: American Council on Education, series on higher education, ORYX Press.
- O'Neil, J. (1992). Putting performance assessment to the test. *Educational Leadership*, 49(8), 14-19.
- Popham, W. J. (1993). Circumventing the high costs of authentic assessment. *Phi Delta Kappan*, 74(6), 470-473.
- Popham, W. J. (1999a). *Classroom assessment : what teachers need to know* (2 ed.). Needham Heights, MA: Allyn and Bacon.

- Popham, W. J. (1999b). Where Large Scale Educational Assessment Is Heading and Why It Shouldn't. *Educational Measurement: Issues and Practice*, 18(3), 13-17.
- Popham, W. J. (2001a). Teaching to the Test? *Educational Leadership*, 58(6), 16-20.
- Popham, W. J. (2001b). *The truth about testing: an educator's call to action*. Alexandria, Virginia USA: Association for supervision and curriculum development.
- Porter, C. D. (1999). *The Implementation of an Integrated Science Curriculum: A Case of Study*. Inédite Thèse de doctorat présentée à la Faculty of The Graduate College at the University of Nebraska. Major : Interdepartmental area of administration, Curriculum and instruction.), University of Nebraska, Lincoln, NE.
- Potvin, P., & Dionne, E. (2007). Realities and challenges of educational reform in the province of Québec: exploratory research on teaching science and technology. *McGill Journal of Education* Retrieved 2008-01-24, from <http://mje.mcgill.ca/article/view/2394/1850>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagogiske Institut.
- Rey, B., Carette, V., Defrance, A., & Kahn, S. (2003). *Les compétences à l'école, apprentissage et évaluation*. Bruxelles: DeBoeck.
- Roberts, R., & Gott, R. (2006). Assessment of performance in practical science and pupil attributes. *Assessment in Education*, 13(1), 45-67.
- Roegiers, X. (2000). *Une pédagogie de l'intégration : Compétences et intégration des acquis dans l'enseignement*. Bruxelles: DeBoeck Université.
- Ruby, A. (2001). *Hands-on science and student achievement*. Inédite Doctorat, Santa Monica, CA.
- Scallon, G. (1988). *L'évaluation formative des apprentissages. Tome 1 : la réflexion*. Québec: Les Presses de l'Université Laval.

- Scallon, G. (2000). *L'évaluation formative*. Montréal: ERPI.
- Scallon, G. (2001). Pourquoi évaluer ?... Quelle question! *Vie pédagogique*, 120(septembre-octobre 2001), 20-23.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Montréal: Éditions du renouveau pédagogique.
- Shavelson, R. J., & Baxter, G. P. (1992). What We've Learned about Assessing Hands-On Science. *Educational Leadership*, 49(8), 20-25.
- Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of educational measurement*, 30(3), 215-232.
- Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance assessment in science. *Applied measurement in education*, 4(4), 347-362.
- Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Performance assessment: Political rhetoric and measurement reality. *Educational Researcher*, 21(4), 22-27.
- Simon, M., & Forgette-Giroux, R. (2001). *Notation du rendement des élèves à partir de tâches complexes de performance : recension des écrits*: Ministère de l'Éducation de l'Ontario.
- Solano-Flores, G., Jovanovic, J., & Shavelson, R. J. (1994). *Development of an item shell for the generation of performance assessments in physics*. Paper presented at the Annual meeting of the American Educational Research Association, New Orleans, LA.
- Solano-Flores, G., Jovanovic, J., Shavelson, R. J., & Bachman, M. (1999). On the Development and Evaluation of a Shell for Generating Science Performance Assessments. *International Journal of Science Education*, 21(3), 293-315.

- Solano-Flores, G., & Shavelson, R. J. (1997). Development of Performance Assessments in Science: Conceptual, Practical, and Logistical Issues. *Educational Measurement: Issues and Practice*, 16(3), 16-25.
- Solano-Flores, G., Shavelson, R. J., Ruiz-Primo, M. A., Schults, S. E., Wiley, E. W., & Brown, J. H. (1997). *On the Development and Scoring of Classification and Observation Science Performance Assessments*. Paper presented at the Annual conference of the American Educational Research Association.
- SPSS inc. (2006). SPSS 15.0 pour Windows (Version 15.0.1). Chicago, Illinois: SPSS inc.
- Stecher, B. M., & Klein, S. P. (1997). The Cost of Science Performance Assessments in Large-Scale Testing Programs. *Educational Evaluation and Policy Analysis*, 19(1), 1-14.
- Stecher, B. M., Klein, S. P., Solano-Flores, G., McCaffrey, D., Robyn, A., Shavelson, R. J., et al. (2000). The Effects of Content, Format, and Inquiry level on Science Performance Assessment Scores. *Applied Measurement in Education*, 13(2), 139-160.
- Stecher, B. M., Klein, S. P., Solano-Flores, G., McCaffrey, D., Robyn, A., Shavelson, R. J., et al. (1998). *Do content, format, and level of inquiry affect scores on open-ended science tasks?* Paper presented at the Annual meeting of the American Educational Research Association, San Diego, California.
- Stiggins, R. J. (1987). Design and development of performance assessments. *Educational Measurement: Issues and Practice*, 6(3), 33-42.
- Stiggins, R. J. (1994). *Student-centered classroom assessment*. New-York: Macmillan Publishing Company.
- Tardif, J. (1993). L'évaluation dans le paradigme constructiviste. Dans R. Hivon (Ed.), *L'évaluation des apprentissages : Réflexions, nouvelles tendances et formation* (Éditions du CRP ed., pp. 27-56). Sherbrooke.

- Tardif, J. (2006). *L'évaluation des compétences : documenter le parcours de développement*. Montréal: Chenelière éducation.
- Walvoord, B. E., & Anderson, V. J. (1998). *Effective grading : a tool for learning and assessment*. San Francisco, Californie: Jossey-Bass.
- Ward, J. D., & Lee, C. L. (2002). A Review of Problem-Based Learning. *Journal of Family and Consumer Sciences Education*, 20(1), 16-26.
- Watson, R., Goldsworthy, A., & Wood-Robinson, V. (1999). What is not fair with investigations? *School Science Review*, 80(292), 101-106.
- Weiss, J. (1977). *Vers une appréciation par objectifs globaux du travail scolaire des élèves en mathématique*. Neuchâtel: Institut romand de recherches et de documentation pédagogiques.
- Wiggins, G. P. (1993a). *Assessing student performance : Exploring the purpose and limits of testing*. San Francisco: Jossey-Bass Publishers.
- Wiggins, G. P. (1993b). Assessment: Authenticity, Context, and Validity. *Phi Delta Kappan*, 75(3), 200-208,210-214.
- Wilson, M., & Scalise, K. (2003). Reporting progress to parents and others: Beyond grades. Dans J. Myron Atkin & J. E. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 89-108). Arlington, VA: NSTA Press.

## **Annexes**

---

## Appendice A

### Liste des membres du comité d'experts

Les personnes suivantes ont procédé à la validation des situations d'évaluation au cours.

Experts	Fonctions	Institutions
M. Patrice Potvin	Professeur-chercheur en didactique des sciences et rédacteur (1 <sup>er</sup> cycle) du nouveau programme de science et technologie	UQÀM
M. Michel D. Laurier	Professeur titulaire en mesure et évaluation, doyen de la Faculté des sciences de l'éducation	Université de Montréal
M. Jesus Vazquez-Abad	Professeur agrégé en didactique des sciences	Université de Montréal
M <sup>me</sup> Paule Bellehumeur	Conseillère pédagogique et rédactrice du nouveau programme (1 <sup>er</sup> cycle) de science et technologie	Commission scolaire des Portages-de-l'Outaouais (CSPO)
M. Pierre Lachance	Animateur au service national du RÉCIT dans le domaine de la Mathématique, de la Science et de la Technologie, membre du comité élargi de rédaction des programmes de S&T (1 <sup>er</sup> et 2 <sup>ième</sup> cycle)	RÉCIT national MST

## **Appendice B**

### **Outil d'évaluation des situations d'évaluation remis aux membres du comité d'experts**

---

Grille d'évaluation permettant de juger de l'adéquation entre la situation d'évaluation et les fondements du nouveau programme de science et technologie.

#### Consigne pour l'évaluation de la situation d'évaluation

Pour chacun des critères, cochez la case correspondant à votre appréciation en vous basant sur l'échelle présentée. Au besoin, précisez votre appréciation ou émettez vos suggestions dans l'encadré correspondant.

Nom de l'expert

Date

Fonction

Institution

Situation d'évaluation

Critères d'évaluation	Description
1. La mise en situation est significative.	<ul style="list-style-type: none"><li>• Elle a un caractère motivant qui incite les élèves à s'investir dans la tâche.</li><li>• Elle trouve ancrage dans la vraie vie en étant associée à l'un ou l'autre des domaines généraux de formation (DGF). Elle permet également de développer l'un des axes de développement du DGF ciblé.</li></ul>
<b>Cochez le choix correspondant à votre appréciation</b>	
Totalelement en désaccord	<input type="checkbox"/>
En désaccord	<input type="checkbox"/>
D'accord	<input type="checkbox"/>
Totalelement en accord	<input type="checkbox"/>
Commentaires ou suggestions	

Critères d'évaluation	Description
2. La mise en situation présente un défi réaliste et des contraintes.	<ul style="list-style-type: none"> <li>• La mise en situation ne suggère pas une solution évidente à laquelle tous les élèves vont souscrire d'emblée.</li> <li>• Le défi est bien dosé en ce sens qu'il est exigeant mais accessible aux élèves.</li> <li>• Les contraintes rendent le défi réaliste et gérable selon les contraintes (matériel, espace disponible dans la classe, etc.) rencontrées dans le contexte scolaire.</li> </ul>
Cochez le choix correspondant à votre appréciation	
Totalemment en désaccord	<input type="checkbox"/>
En désaccord	<input type="checkbox"/>
D'accord	<input type="checkbox"/>
Totalemment en accord	<input type="checkbox"/>
<p>Commentaires ou suggestions</p>	

Critères d'évaluation	Description
3. La situation fait appel à différentes disciplines.	<ul style="list-style-type: none"><li>• La tâche fait appel à plus d'un univers notionnel en science et technologie.</li></ul>
<b>Cochez le choix correspondant à votre appréciation</b>	
Totaletemnt en désaccord	<input type="checkbox"/>
En désaccord	<input type="checkbox"/>
D'accord	<input type="checkbox"/>
Totaletemnt en accord	<input type="checkbox"/>
Commentaires ou suggestions	

Critères d'évaluation	Description
4. La situation permet à l'élève de faire des choix à différentes étapes de la démarche.	<ul style="list-style-type: none"><li>• Les contraintes permettent aux élèves de faire certains choix à des moments névralgiques et ciblés de leur démarche.</li><li>• Les choix peuvent concerner autant le produit que le processus.</li></ul>
Cochez le choix correspondant à votre appréciation	
Totalemment en désaccord	<input type="checkbox"/>
En désaccord	<input type="checkbox"/>
D'accord	<input type="checkbox"/>
Totalemment en accord	<input type="checkbox"/>
Commentaires ou suggestions	

Critères d'évaluation	Description
5. La situation permet une découverte de la part de l'élève.	<ul style="list-style-type: none"><li>• L'élève est en mesure de faire de nouveaux apprentissages.</li></ul>
<b>Cochez le choix correspondant à votre appréciation</b>	
Totalemment en désaccord	<input type="checkbox"/>
En désaccord	<input type="checkbox"/>
D'accord	<input type="checkbox"/>
Totalemment en accord	<input type="checkbox"/>
<b>Commentaires ou suggestions</b>	

Critères d'évaluation	Description
6. La situation permet le développement d'une ou de plusieurs compétence(s).	<ul style="list-style-type: none"><li>• La situation permet de développer plusieurs compétences disciplinaires ou transversales.</li></ul>
<b>Cochez le choix correspondant à votre appréciation</b>	
Totalemment en désaccord	<input type="checkbox"/>
En désaccord	<input type="checkbox"/>
D'accord	<input type="checkbox"/>
Totalemment en accord	<input type="checkbox"/>
Commentaires ou suggestions	

Critères d'évaluation	Description
7. La situation peut être facilement enrichie afin de favoriser la différenciation des apprentissages.	<ul style="list-style-type: none"> <li>• La tâche est évolutive et peut être bonifiée avec des éléments d'enrichissement.</li> <li>• La tâche peut être adaptée afin d'offrir un défi adéquat autant aux élèves plus lents qu'aux élèves plus rapides.</li> <li>• La tâche peut être utilisée ultérieurement dans un autre contexte et être à nouveau adéquate en permettant d'approfondir les apprentissages.</li> </ul>
<b>Cochez le choix correspondant à votre appréciation</b>	
<p data-bbox="250 716 646 753">Totalelement en désaccord <input type="checkbox"/></p> <p data-bbox="250 753 646 791">En désaccord <input type="checkbox"/></p> <p data-bbox="250 791 646 829">D'accord <input type="checkbox"/></p> <p data-bbox="250 829 646 863">Totalelement en accord <input type="checkbox"/></p>	
<p data-bbox="250 863 646 900"><b>Commentaires ou suggestions</b></p>	

Critères d'évaluation	Description
8. La situation exige que l'élève fasse montre de son autonomie.	<ul style="list-style-type: none"><li>• La tâche offre un défi à la mesure des élèves et leur permet de montrer leur autonomie</li></ul>
<b>Cochez le choix correspondant à votre appréciation</b>	
Totalemt en désaccord	<input type="checkbox"/>
En désaccord	<input type="checkbox"/>
D'accord	<input type="checkbox"/>
Totalemt en accord	<input type="checkbox"/>
<b>Commentaires ou suggestions</b>	

Critères d'évaluation	Description
9. La situation nécessite que l'élève soit intellectuellement actif.	<ul style="list-style-type: none"><li>• L'élève développe ses compétences en mettant de l'avant ses propres démarches;</li><li>• La situation implique que l'élève fasse des choix et les défende.</li></ul>
<b>Cochez le choix correspondant à votre appréciation</b>	
Totalemment en désaccord	<input type="checkbox"/>
En désaccord	<input type="checkbox"/>
D'accord	<input type="checkbox"/>
Totalemment en accord	<input type="checkbox"/>
<b>Commentaires ou suggestions</b>	

Critères d'évaluation	Description
10. La situation est complexe.	<ul style="list-style-type: none"><li>• La situation est complexe (permet le développement de compétences) et non pas compliquée (une suite d'exercices simples).</li><li>• La tâche s'inscrit dans une problématique dans laquelle les élèves doivent obligatoirement faire des liens (au niveau des savoirs essentiels, des démarches, des stratégies, etc.).</li></ul>
<b>Cochez le choix correspondant à votre appréciation</b>	
Totalemment en désaccord	<input type="checkbox"/>
En désaccord	<input type="checkbox"/>
D'accord	<input type="checkbox"/>
Totalemment en accord	<input type="checkbox"/>
Commentaires ou suggestions	

Critères d'évaluation	Description
11. La situation offre des attentes claires.	<ul style="list-style-type: none"><li>• Les critères d'évaluation sont explicites et connus des élèves au moment où la tâche leur est soumise.</li><li>• Les critères d'évaluation font référence à la démarche, au produit ou à ces deux dimensions.</li></ul>
Cochez le choix correspondant à votre appréciation	
Totalemment en désaccord	<input type="checkbox"/>
En désaccord	<input type="checkbox"/>
D'accord	<input type="checkbox"/>
Totalemment en accord	<input type="checkbox"/>
Commentaires ou suggestions	

Critères d'évaluation	Description
12. La situation est bien formulée et intelligible pour un élève de première secondaire	<ul style="list-style-type: none"><li>• Le vocabulaire est adapté pour un élève du premier cycle.</li><li>• Il n'y a pas de biais (sexisme, racisme, etc.) apparent.</li></ul>
<b>Cochez le choix correspondant à votre appréciation</b>	
Totalemment en désaccord	<input type="checkbox"/>
En désaccord	<input type="checkbox"/>
D'accord	<input type="checkbox"/>
Totalemment en accord	<input type="checkbox"/>
<b>Commentaires ou suggestions</b>	

Critères d'évaluation	Description
13. Les critères d'évaluation sont pertinents	<ul style="list-style-type: none"><li>• Les critères permettent de rendre compte des principales dimensions évaluées.</li></ul>
Cochez le choix correspondant à votre appréciation	
Totalemment en désaccord	<input type="checkbox"/>
En désaccord	<input type="checkbox"/>
D'accord	<input type="checkbox"/>
Totalemment en accord	<input type="checkbox"/>
Commentaires ou suggestions	

## Synthèse de l'évaluation de la situation d'évaluation

1. Est-ce que globalement cette situation d'évaluation respecte les fondements du nouveau programme ?

Oui

Non

Commentaires/Suggestions

2. Est-ce que vous recommandez l'utilisation de cette situation d'évaluation pour les fins de cette recherche ?

Oui

Non

Commentaires/Suggestions

## **Appendice C**

### **Lettre de présentation du projet et formulaires de consentement**

---

#### **Document d'information pour les enseignants qui expérimenteront les situations d'évaluation**

##### Type de situation d'évaluation

Les situations d'évaluation qui seront administrées visent toutes à mesurer le niveau d'atteinte de la compétence 1. Il s'agit donc de situations d'évaluation pour lesquelles les élèves doivent concevoir un plan expérimental afin de résoudre un problème. Les attentes de fin de cycle (MEQ, 2003, p. 277)<sup>1</sup> ainsi que les échelles de niveau de compétence élaborées et diffusées par le MELS - sans être des références absolues puisque leur but est de rendre compte du niveau d'atteinte en fin de cycle - peuvent néanmoins donner une idée de la nature des démarches attendues de la part des élèves.

##### Exigences matérielles

Les situations d'évaluation ont été conçues en tenant compte des nombreuses contraintes inhérentes au travail en laboratoire dans une école secondaire. Ainsi, la passation des situations d'évaluation ne nécessite aucun matériel spécial que les écoles ne possèdent pas déjà. Une liste de matériel sera remise aux écoles environ deux semaines avant le début de l'expérimentation.

##### Apport des techniciens de laboratoire

Les situations d'évaluation s'apparentent en forme et en contenu à une situation d'apprentissage. En ce sens, cela n'impose aucune contrainte supplémentaire aux

---

<sup>1</sup> Ministère de l'Éducation du Québec. (2003). *Programme de formation de l'école québécoise, enseignement secondaire, premier cycle*. Québec.

techniciens de laboratoire pour les aspects techniques liés à ces situations. Cependant, la présence du technicien de laboratoire en classe lors de la passation des situations d'évaluation est fortement recommandée voir indispensable.

### Concepts prescrits

Les situations d'évaluation ont été bâties en s'appuyant principalement sur deux univers : l'univers vivant et l'univers matériel. Dans une perspective d'évaluation « bilan », il importe que les contenus de formation (concepts, techniques, stratégies, attitudes) aient été vus et maîtrisés de la part des élèves en cours d'année. Les concepts généraux susceptibles d'être mobilisés dans l'une ou l'autre des situations d'évaluation apparaissent dans le tableau suivant (à compléter).

Contenu admissible et non admissible concernant les épreuves expérimentées.

Univers	Concepts généraux		
	Propriétés	Transformations	Organisation
<i>Matériel</i>			
<i>Vivant</i>	Diversité de la vie	Perpétuation des espèces	Maintien de la vie
<i>Terre et Espace</i>	Caractéristiques générales de la Terre	Phénomènes géologiques et géophysiques	Phénomènes astronomiques
<i>Technologique</i>	Ingénierie	Systèmes technologiques	Forces et mouvement

Légende :

Cellules grisées : Concepts généraux susceptibles d'être mobilisés dans les situations d'évaluation.

Autres cellules : Concepts généraux qui ne devront pas obligatoirement être mobilisés dans les situations d'évaluation.

### Conditions de passation

Les situations d'évaluation ont été conçues pour être administrées en 4 ou 5 périodes de 75 minutes. Selon le rythme de travail de chacun des élèves, cette estimation peut être

revue à la hausse ou à la baisse. Afin de pallier aux imprévues, il faut prévoir environ 3 périodes tampons supplémentaires. Le tableau de la page suivante présente un exemple de gestion du temps pour les situations d'évaluation qui seront administrées.

Avant le début de l'expérimentation, il est à prévoir deux rencontres visant, dans un premier temps, à mieux définir les objectifs de la recherche et les mandats des participants et, dans un deuxième temps, à expliquer les conditions concrètes d'expérimentation compte tenu des contraintes diverses de chacun des enseignants.

### Correction

Les cahiers d'élève doivent être d'abord photocopiés pour ensuite être remis à Eric Dionne. Par la suite, l'enseignant pourra procéder à la correction des situations d'évaluation selon les standards qu'il aura retenu. Des grilles d'évaluation crititériées seront remis aux enseignants qui désireront utiliser ces outils de notation.

Un exemple pour la gestion du temps des situations d'évaluation expérimentées

	Période 1	Période 2	Période 3	Période 4	Période 5	Période 6	Période 7	Période 8
Exemples d'activités réalisées par les élèves	Présentation du problème					Période tampon	Période tampon	Période tampon
	Esquisse des premières solutions	<i>Esquisse de solutions</i>	<i>Esquisse de solutions</i>					
		Ébauche des premières solutions	<i>Ébauche des premières solutions</i>					
			Préparation des aspects techniques	<i>Préparation des aspects techniques</i>				
				Expérimentation		Élaboration et discussion de la solution finale retenue		

Notes aux lecteurs

1. Chaque période prévue est de 75 minutes.
2. Les exemples d'activités réalisées par les élèves ne sont pas exhaustifs. La démarche réalisée par les élèves peut également être très différentes et plus itérative que celle ici illustrées et être tout aussi valable.

Montréal, le 30 janvier 2006

Objet : Recherche d'enseignant(e)s de première secondaire en science et technologie afin de participer à une recherche visant l'expérimentation d'un modèle d'évaluation-bilan.

Chèr(e) collègue,

Depuis plusieurs mois je m'intéresse aux différentes problématiques liées à l'évaluation des compétences en science et technologie. Plus particulièrement, le contexte de l'évaluation-bilan semble poser des défis particulièrement intéressants au secondaire. À ce sujet, je vous invite, enseignant(e)s de première secondaire, à participer à une recherche visant à expérimenter un modèle d'évaluation-bilan comportant des situations d'évaluation susceptibles éventuellement d'être administrées de façon plus formelle vers la fin du premier cycle en science et technologie.

La participation à cette étude ne demande qu'un investissement en temps et en travail relativement faible puisqu'il s'agit essentiellement d'administrer une situation d'évaluation « clé en main » c'est-à-dire une situation déjà conçue qu'il ne vous reste qu'à administrer. Au moment opportun, tous les documents nécessaires vous seront transmis électroniquement.

Si vous choisissez de participer à cette étude, voici ce qui est attendu plus spécifiquement de votre part :

- Administrer, auprès d'un ou deux groupes d'élèves, une situation d'évaluation d'une durée prévue d'environ 6-7 périodes de 75 minutes. Cette situation permettra de juger essentiellement du niveau de la première compétence (C1) dans un contexte strictement scientifique. La situation d'évaluation devrait être administrée, selon votre convenance, entre la mi-mai et la mi-juin 2006.
- Répondre à un court questionnaire ayant pour objet votre pratique professionnelle.
- Rendre disponible une compilation des résultats des élèves aux étapes précédentes (ex. résultats de chaque élève à chacun des trois premiers bulletins).

- Rendre disponibles les cahiers de réponse qui seront photocopiés et remis au chercheur (afin que vous conserviez les originaux).

Votre participation à un tel projet comporte plusieurs avantages. D'abord, la possibilité d'expérimenter, avant l'heure, un modèle d'évaluation qui tient compte des multiples contraintes liées au contexte d'évaluation-bilan afin de mieux vous permettre d'appuyer votre jugement professionnel. Si ce modèle s'avère efficace, il vous sera toujours possible de l'utiliser tel quel ou de l'adapter selon vos besoins spécifiques à la lumière de votre expérience. Ensuite, cette recherche vous offre l'opportunité de mettre à l'essai des situations « clé en main » bâties dans une optique stricte d'évaluation tout en respectant les fondements et les orientations du nouveau programme. Force est de constater que ce type de situation est, pour l'instant, relativement rare. Au terme de l'expérimentation, vous pourrez réutiliser les situations d'évaluation ou les adapter (vous les aurez en format .doc) à votre guise.

Si vous souhaitez participer à cette recherche ou si vous avez des questions, je vous invite à contacter votre conseiller(ère) pédagogique (qui veillera à me transmettre les informations) ou moi-même aux coordonnées qui apparaissent plus bas. Et si vous décidez de participer à cette recherche, de plus amples renseignements vous seront acheminés à différents moments d'ici le moment de l'expérimentation en mai-juin.

Espérant compter sur votre participation,

Eric Dionne

  
Téléphone : 514-528-1337

Enseignant en science et technologie, Commission scolaire des Samares  
Étudiant-chercheur, Faculté des sciences de l'éducation, Université de Montréal

## **Appendice D**

### **Les situations d'évaluation**

---

Première situation (SE1) : La tasse de chocolat chaud

- Canevas de la situation
- Les situations SE1-1, SE1-2, SE1-3
- Outil d'évaluation

## SE1 : La tasse de chocolat chaud

## Canevas

1. Titre de la tâche			
La tasse de chocolat chaud			
2. Intention pédagogique et/ou évaluative			
<i>Pourquoi faire réaliser cette tâche par les élèves?</i>			
1. Évaluer le niveau de l'élève au regard de la première compétence.			
3. Clientèle visée			
	Cycle :	Sec. 1	Moment dans l'année : Mai-Juin
4. Temps consacré à la tâche			
Environ 6 périodes de 75 minutes			
5. Type d'évaluation			
	<input type="checkbox"/> Formative	<input checked="" type="checkbox"/> Certificative (Bilan)	<input type="checkbox"/> Diagnostique
6. Compétence disciplinaire ciblée			
	<input checked="" type="checkbox"/> Compétence 1	<input type="checkbox"/> Compétence 2	<input checked="" type="checkbox"/> Compétence 3
<i>Expliquez brièvement pourquoi la tâche permet de développer la compétence.</i>			
C1 : Les élèves doivent résoudre un problème <i>hands-on</i> pour lequel il existe plusieurs solutions et dont la valeur de la démarche est fonction des explications fournies par les élèves.			
C3 : Les élèves sont invités à faire part de leurs résultats et de leur démarche aux autres élèves dans un langage clair et rigoureux.			
7. Compétence(s) transversale(s) ciblée(s)			
Ordre intellectuel	Ordre de la communication	Ordre personnel et social	Ordre méthodologique
<input type="checkbox"/> Exploiter l'information	<input type="checkbox"/> Communiquer de façon appropriée	<input type="checkbox"/> Actualiser son potentiel	<input type="checkbox"/> Se donner des méth. trav. efficaces
<input checked="" type="checkbox"/> Résoudre des problèmes		<input type="checkbox"/> Coopérer	<input type="checkbox"/> Exploiter les TIC
<input type="checkbox"/> Exercer son jugement critique			
<input checked="" type="checkbox"/> Mettre en œuvre sa pensée créatrice			
<i>Expliquez brièvement pourquoi la tâche permet de développer la compétence ciblée.</i>			
L'élève est invité à imaginer une solution originale au problème qui lui est soumis.			
8. DGF touché			
<input type="checkbox"/> Médias	<input type="checkbox"/> Santé et bien-être	<input type="checkbox"/> Vivre ensemble et citoyenneté	
<input type="checkbox"/> Orientation et entrepreneuriat	<input checked="" type="checkbox"/> Environnement et consommation		
<i>Expliquez en quoi l'activité permet de réaliser des apprentissages en lien avec les axes de développement du DGF touché.</i>			
La situation d'évaluation vise à évaluer le troisième axe de développement de la compétence transversale : Consommation et utilisation responsable de biens et de services (PDFEQ, p. 26)			
9. Univers notionnel(s) visé(s)			
<input checked="" type="checkbox"/> Univers matériel	<input type="checkbox"/> Univers vivant	<input type="checkbox"/> Univers technologique	<input type="checkbox"/> Terre et espace

<b>Contenu obligatoire / Contenu facultatif</b>		
<b>Initiation</b>	<b>Consolidation</b>	<b>Maîtrise</b>
		<b>Propriétés caractéristiques</b>
		<b>Température</b>
		<b>Matériau</b>
		<b>Volume</b>
		<b>Masse</b>
<i>Techniques</i>		
		<b>Utilisation sécuritaire du matériel de labo</b>
		<b>Utilisation d'instruments de mesure</b>
<i>Attitudes</i>		
		<b>Sens de l'initiative</b>
		<b>Goût du risque intellectuel</b>
		<b>Considérations de solutions originales</b>
<i>Stratégies</i>		
		<b>Explorer diverses pistes de solution</b>
		<b>Recourir à des outils permettant de représenter des données sous forme de tableaux et de graphiques ou de tracer des diagrammes</b>
<b>10. Repères culturels</b>		
<i>Indiquez les principaux repères culturels touchés dans cette tâche.</i>		
<ul style="list-style-type: none"> <li>• Organismes voués à la protection de l'environnement (ex. Greenpeace, Recyc-Québec, etc.)</li> </ul>		

Nom de l'élève

Âge de l'élève

Masculin	<input type="checkbox"/>
Féminin	<input type="checkbox"/>

Groupe

Nom de l'enseignant(e)

Nom de l'école

Nom de la commission scolaire

SE1-1 : La tasse de chocolat chaud

**ATTENTION!!!**

**Assure-toi de répondre à toutes les sections de cette page**

Juin 2006

## Consignes générales

- Tu disposes de 4 périodes de 75 minutes pour résoudre ce problème.
- Ce document (cahier de l'élève) doit être remis à ton enseignant(e) à la fin de **CHAQUE** période. Tu n'as pas le droit de l'apporter à la maison.
- À moins que ton enseignant(e) ne te donne des indications contraires, cette activité se déroule individuellement.
- Au besoin, utilise le verso de chaque page de ton cahier de réponse.
- Tu as le droit de consulter des livres de référence pour t'aider à résoudre ce problème.

## Ton mandat

Les responsables de la cafétéria persistent à vendre leur chocolat chaud dans des tasses en polystyrène qui sont non biodégradables. En t'appuyant sur des arguments variés, tente de convaincre les responsables d'offrir leur chocolat chaud dans des contenants qui sont à la fois tout aussi efficaces et plus écologiques.

**Première étape : Je cerne le problème !****ATTENTION !!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

L'élève peut poursuivre son travail : Oui \_\_\_ Non \_\_\_

**Deuxième étape : Je détermine une façon de résoudre le problème**

Maintenant, tu dois te préparer à réaliser l'expérience qui te permettra de résoudre le problème. Indique le matériel dont tu auras besoin pour réaliser ton expérience.



**Troisième étape : Je réalise l'expérience**

Prends ici en note toutes les informations pertinentes au cours de ton expérience.

**Quatrième étape : J'analyse les résultats de mon expérience**



Nom de l'élève

Âge de l'élève

Masculin	<input type="checkbox"/>
Féminin	<input type="checkbox"/>

Groupe

Nom de l'enseignant(e)

Nom de l'école

Nom de la commission scolaire

SE1-2 : La tasse de chocolat chaud

**ATTENTION!!!**

**Assure-toi de répondre à toutes les sections de cette page**

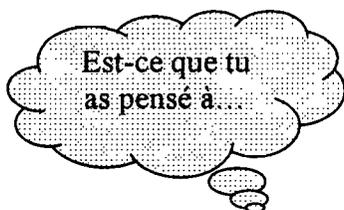
Juin 2006

## Consignes

- Tu disposes de 4 périodes de 75 minutes pour résoudre ce problème.
- Ce document (cahier de l'élève) doit être remis à ton enseignant(e) à la fin de **CHAQUE** période. Tu n'as pas le droit de l'apporter à la maison.
- À moins que ton enseignant(e) ne te donne des indications contraires, cette activité se déroule individuellement.
- Tu as le droit de consulter des livres de référence pour t'aider à résoudre ce problème.

## Ton mandat

Les responsables de la cafétéria persistent à vendre leur chocolat chaud dans des tasses en polystyrène qui sont non biodégradables. En vous appuyant sur des arguments variés, tentez de convaincre les responsables d'offrir leur chocolat chaud dans des contenants qui sont à la fois tout aussi efficaces et plus écologiques.

**Première étape : Je cerne le problème !**

1. reformuler le mandat dans tes propres mots ?
2. trouver la définition des mots que tu ne comprenais pas dans la section « Ton mandat » ? En particulier ceux qui sont soulignés...
3. indiquer quels étaient les arguments que tu allais utiliser pour résoudre le problème et expliquer pourquoi tu utilisais ces arguments ?
4. énoncer le but de l'expérience que tu allais mener et l'hypothèse que tu vérifierais ?
5. identifier quelle variable tu mesurerais ?
6. chercher dans des livres de référence afin de t'aider à mieux comprendre ce problème ?

**ATTENTION!!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

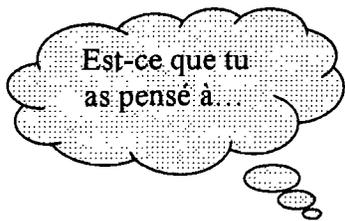
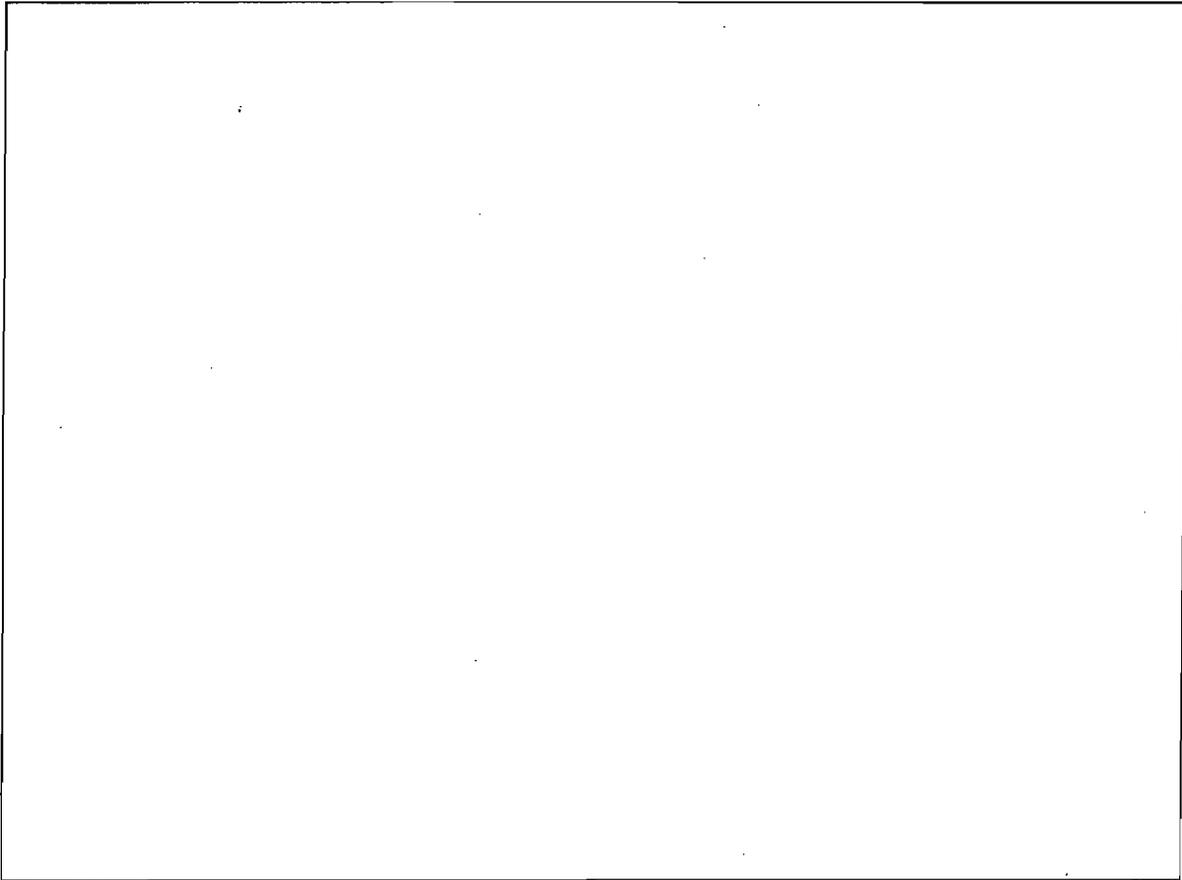
L'élève peut poursuivre son travail : Oui \_\_\_ Non \_\_\_

**Deuxième étape : Je détermine une façon de résoudre le problème**

Maintenant, tu dois te préparer à réaliser l'expérience qui te permettra de résoudre le problème. Voici le matériel que tu peux utiliser. Au besoin, complète cette liste en indiquant le matériel que tu as utilisé.

Thermomètres	Chronomètre
Laine isolante	Carton, plastique (pour couvrir les contenants)
Verres en polystyrène	Feuille de polystyrène
Papier aluminium	Couvercles pour verre en polystyrène
Variétés de tasse (plastique, verre, etc.)	

Indique les manipulations que tu dois effectuer pour faire ton expérience. Au besoin, utilise le verso de cette feuille pour compléter.



1. compléter la liste du matériel ?
2. indiquer tes manipulations dans l'ordre où tu dois les réaliser ?
3. écrire tous les détails des manipulations afin qu'une personne qui ne sait pas ce qu'elle doit faire puisse réaliser exactement ton expérience en suivant ces manipulations ?
4. représenter tes manipulations sous forme de dessin si cela permet d'être plus clair ?

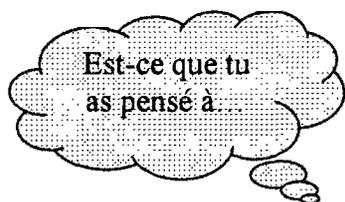
**ATTENTION!!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

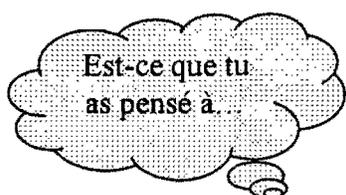
L'élève peut poursuivre son travail : Oui \_\_\_ Non \_\_\_

**Troisième étape : Je réalise l'expérience**

Prends ici en note toutes les informations pertinentes au cours de ton expérience.



1. construire un tableau qui te permettra de consigner tes données ?
2. recourir à un témoin (afin de comparer tes données) ?
3. noter toutes les observations utiles que tu as pu voir durant ton expérience ?

**Quatrième étape : J'analyse les résultats de mon expérience**

1. présenter tes données à l'aide d'un graphique ou d'un diagramme à bandes ?
2. mesurer la température du contenant qui conservait le mieux la chaleur ?
3. identifier des causes d'erreurs expérimentales durant tes manipulations ?



## Auto-évaluation

**Consigne : Noircit la case (Oui, Non, Un peu) qui correspondant le mieux à l'énoncé de la colonne de gauche. Au besoin, apporte les correctifs nécessaires dans ton cahier de réponse.**

1. J'ai bien compris le problème et j'ai été capable de le reformuler dans mes mots.	Oui	Non	Un peu
2. J'ai identifié la variable importante que je voulais investiguer dans mon expérience.	Oui	Non	Un peu
3. J'ai formulé une hypothèse vérifiable.	Oui	Non	Un peu
4. J'ai élaboré des manipulations expérimentales qui permettent de reproduire l'expérience.	Oui	Non	Un peu
5. J'ai identifié les causes d'erreurs expérimentales.	Oui	Non	Un peu
6. La solution que je propose s'appuie sur les données expérimentales que j'ai recueillies.	Oui	Non	Un peu
7. J'ai vérifié et corrigé les fautes d'orthographe.	Oui	Non	Un peu
8. Je me suis appliqué à bien écrire et à présenter soigneusement mes réponses.	Oui	Non	Un peu
9. J'ai utilisé les bons symboles scientifiques pour présenter mes données.	Oui	Non	Un peu

**ATTENTION!!!**

Prends le temps de réviser ton cahier de réponse afin de t'assurer que tu as répondu à toutes les questions. Quand tu as terminé ta révision, tu remets ton cahier à ton enseignant(e).

Nom de l'élève

Âge de l'élève

Masculin

Féminin

<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------

Groupe

Nom de l'enseignant(e)

Nom de l'école

Nom de la commission scolaire

SE1-3 : La tasse de chocolat chaud

**ATTENTION!!!**

**Assure-toi de répondre à toutes les sections de cette page**

Juin 2006

## Consignes

- Tu disposes de 4 périodes de 75 minutes pour résoudre ce problème.
- Ce document (cahier de l'élève) doit être remis à ton enseignant(e) à la fin de **CHAQUE** période. Tu n'as pas le droit de l'apporter à la maison.
- À moins que ton enseignant(e) ne te donne des indications contraires, cette activité se déroule individuellement.
- Tu as le droit de consulter des livres de référence pour t'aider à résoudre ce problème.

## Ton mandat

Les responsables de la cafétéria persistent à vendre leur chocolat chaud dans des tasses en **polystyrène** qui sont non **biodégradables**. En vous appuyant sur des **arguments** variés, tentez de convaincre les responsables d'offrir leur chocolat chaud dans des contenants qui sont à la fois tout aussi **efficaces** et plus **écologiques**.

**Première étape : Je cerne le problème !**

1. En utilisant un dictionnaire ou un livre de référence, donne la définition des mots apparaissant en gras dans l'énoncé du problème et qui apparaissent ci-dessous.

**Polystyrène**

Définition :

---

---

---

---

---

Référence (indique dans quel livre tu as pris la définition) :

---

**Biodégradable**

Définition :

---

---

---

---

---

Référence (indique dans quel livre tu as pris la définition) :

---

**Argument**

Définition :

---

---

---

---

---

Référence (indique dans quel livre tu as pris la définition) :

---

**Efficace**

Définition :

---

---

---

---

---

Référence (indique dans quel livre tu as pris la définition) :

---

**Écologique**

Définition :

---

---

---

---

---

Référence (indique dans quel livre tu as pris la définition) :

---

2. En utilisant tes propres mots (sans copier), indique le problème que tu dois résoudre.

---

---

---

---

---

---

---

---

---

---

3. Afin de résoudre le problème, on indique qu'il faut s'appuyer sur des **arguments variés**. Encerle dans la liste qui suit les trois (3) arguments les plus importants sur lesquels tu te baseras pour résoudre le problème.

a) Coût de chaque contenant	d) Poids du contenant
b) Capacité du contenant à conserver la chaleur	e) Volume du contenant
c) Beauté du contenant	f) Contenant fabriqué de matériaux recyclable

Explique pourquoi tu as choisi ces trois (3) arguments.

---

---

---

---

---

---

---

4. Quel est le but de l'expérience ?

---

---

---

---

---

5. Quelle est l'hypothèse que tu souhaites vérifier ?

---

---

---

---

---

**ATTENTION!!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

L'élève peut poursuivre son travail : Oui \_\_\_ Non \_\_\_

<b>Deuxième étape : Je détermine une façon de résoudre le problème</b>
--

Maintenant, tu dois te préparer à réaliser l'expérience qui te permettra de résoudre le problème.

6. En te basant sur la liste de matériel et les manipulations qui suivent, fais un résumé des manipulations en dessinant sur la page 8 ce que tu as fait à chaque étape.

Voici la liste du matériel disponible pour cette activité :

Thermomètres	Différents liquides (eau, chocolat au lait, etc.)
Cylindre gradué	Carton, plastique (pour couvrir les contenants)
Verres en polystyrène	Feuille de polystyrène
Agitateur	Béchers de différents formats
Laine isolante	Chronomètre
Papier aluminium	Couvercles pour verre en polystyrène
Variétés de tasse (plastique, verre, etc.)	Carton

Voici les manipulations que tu dois exécuter :

1. Fais chauffer l'eau ou le liquide choisi jusqu'à une certaine température (ex. 70 degrés Celsius).
2. Isole le contenant de ton choix avec le(s) matériau(x) isolant(s) que tu as choisi(s).
3. S'il y a lieu, fabrique un couvercle pour fermer ton contenant. Prévois un espace pour y insérer un thermomètre.
4. Verse la même quantité de liquide dans chacun des contenants (le verre en polystyrène et ton contenant à toi).
5. Insère un thermomètre dans chacun des contenants.
6. Ferme chacun des contenants avec un couvercle.
7. Prends en note la température initiale du liquide pour chacun des contenants.
8. Actionne le chronomètre.
9. Prends en note la température à toutes les minutes.
10. Note toutes les observations pertinentes et les erreurs de manipulation que tu as pu faire.
11. Tu arrêtes l'expérience après 15 minutes.

Pour chacune des étapes des manipulations, fais un dessin représentant ce que tu as fait.

1.	2.
3.	4.
5.	6.

7.	8.
9.	10.
11.	

**ATTENTION!!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

L'élève peut poursuivre son travail : Oui \_\_\_\_ Non \_\_\_\_

<b>Troisième étape : Je réalise l'expérience</b>
--

7. Complète le tableau des données qui suit durant ton expérience.

Temps (minute)	Température (°C) du liquide du contenant en polystyrène	Température (°C) du liquide de ton contenant	Observations
0			← Température initiale
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			← Température finale



---

---

---

12. Quel contenant a le mieux conservé la chaleur ? Pourquoi ?

---

---

---

---

---

---

---

---

---

---

13. Si tu pouvais recommencer l'expérience, qu'est-ce que tu modifierais à celle-ci ?

---

---

---

---

---

---

---

---

---

---

**Cinquième étape : Je propose une solution au problème**

14. Dans quel genre de contenant pourrait-on verser le chocolat chaud offert à la cafétéria afin que ce soit plus écologique et tout aussi efficace qu'un verre en polystyrène ? Pourquoi ?

---

---

---

---

---

---

---

---

---

---



**ATTENTION!!!**

Prends le temps de réviser ton cahier de réponse afin de t'assurer que tu as répondu à toutes les questions. Pour t'aider, utilise la grille d'auto-évaluation qui apparaît ci-dessous. Quand tu as terminé ta révision, tu remets ton cahier à ton enseignant(e).

**Auto-évaluation**

**Consigne : Noircit la case (Oui, Non, Un peu) qui correspondant le mieux à l'énoncé de la colonne de gauche. Au besoin, apporte les correctifs nécessaires dans ton cahier de réponse.**

1. J'ai bien compris le problème et j'ai été capable de le reformuler dans mes mots.	Oui	Non	Un peu
2. J'ai identifié la variable importante que je voulais investiguer dans mon expérience.	Oui	Non	Un peu
3. J'ai formulé une hypothèse vérifiable.	Oui	Non	Un peu
4. J'ai élaboré des manipulations expérimentales qui permettent de reproduire l'expérience.	Oui	Non	Un peu
5. J'ai identifié les causes d'erreurs expérimentales.	Oui	Non	Un peu
6. La solution que je propose s'appuie sur les données expérimentales que j'ai recueillies.	Oui	Non	Un peu
7. J'ai vérifié et corrigé les fautes d'orthographe.	Oui	Non	Un peu
8. Je me suis appliqué à bien écrire et à présenter soigneusement mes réponses.	Oui	Non	Un peu
9. J'ai utilisé les bons symboles scientifiques pour présenter mes données.	Oui	Non	Un peu

	Critères d'évaluation généraux (c.f. programme d'études)	Référence dans le cahier de l'élève	Critères d'évaluation spécifiques	Niveau de performance		
				0	1	2
C 1	Représentation adéquate de la situation	Première étape	1. Reformulation du problème	0	1	2
C 1	Représentation adéquate de la situation	Première étape	2. Identification des variables, connaissances et techniques	0	1	2
C 1	Élaboration d'une démarche pertinente pour la situation	Deuxième étape	3. Formulation d'une hypothèse	0	1	2
C 1	Élaboration d'une démarche pertinente pour la situation	Deuxième étape	4. Élaboration du protocole	0	1	2
C 1	Mise en œuvre adéquate de la démarche	Troisième étape	5. Consignation des données expérimentales	0	1	2
C 1	Élaboration de conclusions ou d'explications pertinentes	Quatrième et cinquième étape	6. Identification des erreurs expérimentales ou des limites	0	1	2
C 1	Élaboration de conclusions ou d'explications pertinentes	Quatrième et cinquième étape	7. Interprétation des données	0	1	2
C 1	Élaboration de conclusions ou d'explications pertinentes	Quatrième et cinquième étape	8. Qualité de la solution ou de l'explication	0	1	2
C 3	Respect de la terminologie, des règles et des conventions propres à la science et à la technologie dans la production de messages	Première à la cinquième étape	9. Qualité de la langue	0	1	2
C 3	Respect de la terminologie, des règles et des conventions propres à la science et à la technologie dans la production de messages	Première à la cinquième étape	10. Qualité de la présentation	0	1	2
C 3	Respect de la terminologie, des règles et des conventions propres à la science et à la technologie dans la production de messages	Première à la cinquième étape	11. Utilisation des termes et symboles scientifiques	0	1	2

**Deuxième situation (SE2) : La balle de tennis**

- Canevas de la situation
- Les situations SE2-1, SE2-2, SE2-3
- Outil d'évaluation

## Canevas de la situation d'évaluation SE2 : La balle de tennis

1. Titre de la tâche			
La balle de tennis			
2. Intention pédagogique et/ou évaluative			
<i>Pourquoi faire réaliser cette tâche par les élèves?</i>			
1. Familiariser les élèves à une démarche scientifique plus « ouverte ».			
2. Évaluer la capacité des élèves à résoudre un problème scientifique en usant de créativité.			
3. Clientèle visée			
	Cycle :	Sec. 1	Moment dans l'année : Mai-Juin
4. Temps consacré à la tâche			
Environ 6 périodes de 75 minutes			
5. Type d'évaluation			
	<input type="checkbox"/> Formative	<input checked="" type="checkbox"/> Certificative (Bilan)	<input type="checkbox"/> Diagnostique
6. Compétence disciplinaire ciblée			
	<input checked="" type="checkbox"/> Compétence 1	<input type="checkbox"/> Compétence 2	<input checked="" type="checkbox"/> Compétence 3
<i>Expliquez brièvement pourquoi la tâche permet de développer la compétence.</i>			
C1 : Les élèves doivent résoudre un problème <i>hands-on</i> pour lequel il existe plusieurs solutions et dont la valeur de la démarche est fonction des explications fournies par les élèves.			
C3 : Les élèves sont invités à faire part de leurs résultats et de leur démarche aux autres élèves dans un langage clair et rigoureux.			
7. Compétence(s) transversale(s) ciblée(s)			
Ordre intellectuel	Ordre de la communication	Ordre personnel et social	Ordre méthodologique
<input type="checkbox"/> Exploiter l'information	<input type="checkbox"/> Communiquer de façon appropriée	<input type="checkbox"/> Actualiser son potentiel	<input type="checkbox"/> Se donner des méth. trav. efficaces
<input checked="" type="checkbox"/> Résoudre des problèmes		<input type="checkbox"/> Coopérer	<input type="checkbox"/> Exploiter les TIC
<input type="checkbox"/> Exercer son jugement critique			
<input type="checkbox"/> Mettre en œuvre sa pensée créatrice			
<i>Expliquez brièvement pourquoi la tâche permet de développer la compétence ciblée.</i>			
L'élève est invité à imaginer une solution originale au problème qui lui est soumis.			
8. DGF touché			
<input type="checkbox"/> Médias	<input type="checkbox"/> Santé et bien-être	<input type="checkbox"/> Vivre ensemble et citoyenneté	
<input type="checkbox"/> Orientation et entrepreneuriat	<input type="checkbox"/> Environnement et consommation		
<i>Expliquez en quoi l'activité permet de réaliser des apprentissages en lien avec les axes de développement du DGF touché.</i>			
La situation est surtout à caractère scientifique. Aucun axe de développement formel n'est spécifiquement développé.			
9. Univers notionnel(s) visé(s)			

X Univers matériel	<input type="checkbox"/> Univers vivant	<input type="checkbox"/> Univers technologique	<input type="checkbox"/> Terre et espace
<b>Contenu obligatoire / Contenu facultatif</b>			
Initiation	Consolidation	Maîtrise	
		<b>Propriétés caractéristiques</b>	
		<b>Changements physiques</b>	
		<b>Effets d'une force</b>	
		<b>Matériau</b>	
		Masse	
		Température	
		Volume	
<i>Techniques</i>			
		<b>Utilisation sécuritaire du matériel de labo</b>	
		<b>Utilisation d'instruments de mesure</b>	
<i>Attitudes</i>			
		<b>Sens de l'initiative</b>	
		<b>Goût du risque intellectuel</b>	
		<b>Considérations de solutions originales</b>	
<i>Stratégies</i>			
		Explorer diverses pistes de solution	
		Recourir à des outils permettant de représenter des données sous forme de tableaux et de graphiques ou de tracer des diagrammes	
<b>10. Repères culturels</b>			
<i>Indiquez les principaux repères culturels touchés dans cette tâche.</i>			
<ul style="list-style-type: none"> <li>• Vedettes de tennis d'hier et d'aujourd'hui (Lendl, Navratilova, Hingis, Borg, McEnroe, Sampras, Agassi, Rodick, Williams, Henin-Ardenne, etc.)</li> <li>• Tournois du grand chelem : Roland Garros, Wimbledon, US Open, Open d'Australie.</li> <li>• Raquette d'hier et d'aujourd'hui (bois, aluminium, carbone, etc.)</li> <li>• Balles d'hier et d'aujourd'hui</li> <li>• L'étiquette au tennis</li> <li>• Le tennis et les classes sociales</li> <li>• La recherche et le développement dans les équipements sportifs</li> </ul>			

Nom de l'élève

Âge de l'élève

Masculin	<input type="checkbox"/>
Féminin	<input type="checkbox"/>

Groupe

Nom de l'enseignant(e)

Nom de l'école

Nom de la commission scolaire

**ATTENTION!!!**

Assure-toi de répondre à toutes les sections de cette page

## Consignes générales

- Tu disposes de 4 périodes de 75 minutes pour résoudre ce problème.
- Ce document (cahier de l'élève) doit être remis à ton enseignant(e) à la fin de **CHAQUE** période. Tu n'as pas le droit de l'apporter à la maison.
- À moins que ton enseignant(e) ne te donne des indications contraires, cette activité se déroule individuellement.
- Au besoin, utilise le verso de chaque page de ton cahier de réponse.
- Tu as le droit de consulter des livres de référence pour t'aider à résoudre ce problème.

## Ton mandat

Les balles jouent un rôle fort important au tennis. Elles doivent être bien conçues afin de résister aux chocs violents de la raquette en plus de bien rebondir. Déterminez une variable qui peut influencer le rebond de la balle et démontrez comment ce facteur influence le rebond.

**Première étape : Je cerne le problème !**

**ATTENTION !!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

L'élève peut poursuivre son travail : Oui \_\_\_ Non \_\_\_

**Deuxième étape : Je détermine une façon de résoudre le problème**

Maintenant, tu dois te préparer à réaliser l'expérience qui te permettra de résoudre le problème. Indique le matériel dont tu auras besoin pour réaliser ton expérience.



**Troisième étape : Je réalise l'expérience**

Prends ici en note toutes les informations pertinentes au cours de ton expérience.

**Quatrième étape : J'analyse les résultats de mon expérience**



Nom de l'élève

Âge de l'élève

Masculin	<input type="checkbox"/>
Féminin	<input type="checkbox"/>

Groupe

Nom de l'enseignant(e)

Nom de l'école

Nom de la commission scolaire

**ATTENTION!!!**

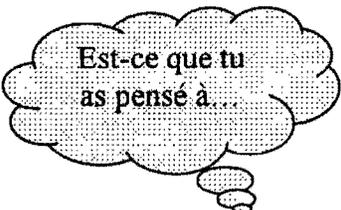
**Assure-toi de répondre à toutes les sections de cette page**

## Consignes générales

- Tu disposes de 4 périodes de 75 minutes pour résoudre ce problème.
- Ce document (cahier de l'élève) doit être remis à ton enseignant(e) à la fin de **CHAQUE** période. Tu n'as pas le droit de l'apporter à la maison.
- À moins que ton enseignant(e) ne te donne des indications contraires, cette activité se déroule individuellement.
- Au besoin, utilise le verso de chaque page de ton cahier de réponse.
- Tu as le droit de consulter des livres de référence pour t'aider à résoudre ce problème.

## Ton mandat

Les balles jouent un rôle fort important au tennis. Elles doivent être bien conçues afin de résister aux chocs violents de la raquette en plus de bien rebondir. Déterminez une variable qui peut influencer le rebond de la balle et démontrez comment cette variable peut influencer le rebond.

**Première étape : Je cerne le problème !**

Est-ce que tu  
as pensé à...

1. reformuler le mandat dans tes propres mots ?
2. trouver la définition des mots que tu ne comprenais pas dans la section « Ton mandat » ? En particulier ceux qui sont soulignés...
3. énoncer le but de l'expérience que tu allais mener et l'hypothèse que tu vérifierais ?
4. identifier quelle variable tu mesurerais ?
5. chercher dans des livres de référence afin de t'aider à mieux comprendre ce problème ?

**ATTENTION !!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

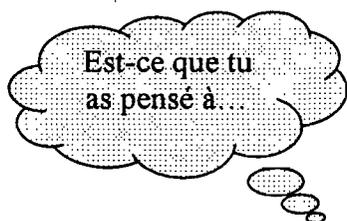
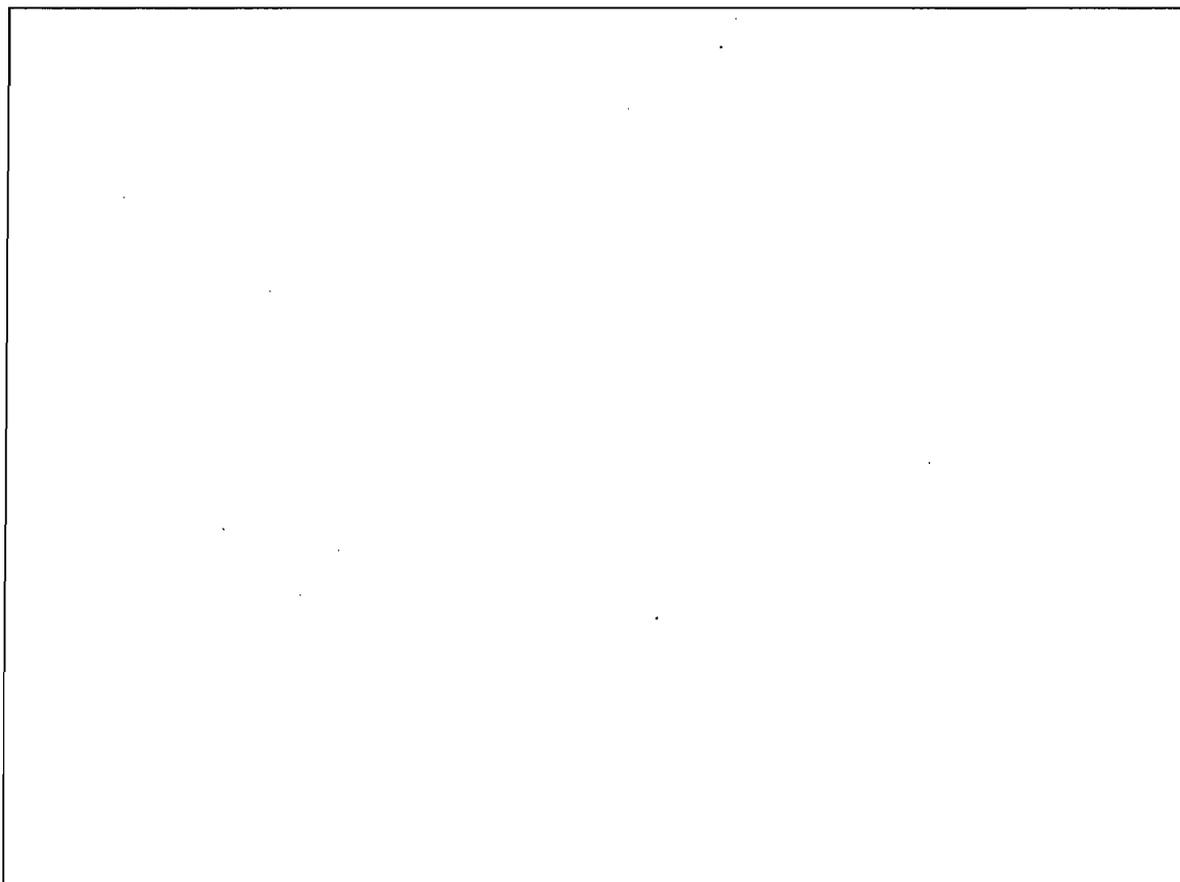
L'élève peut poursuivre son travail : Oui \_\_\_ Non \_\_\_

**Deuxième étape : Je détermine une façon de résoudre le problème**

Maintenant, tu dois te préparer à réaliser l'expérience qui te permettra de résoudre le problème. Voici le matériel que tu peux utiliser. Au besoin, complète cette liste en indiquant le matériel que tu as utilisé.

Papier sablé	Pieds à coulisse
Mètres à mesurer	Supports universels
Diverses balles de tennis	Diverses masses
Béchers	Couteau utilitaire

Indique les manipulations que tu dois effectuer pour faire ton expérience. Au besoin, utilise le verso de cette feuille pour compléter.



1. compléter la liste du matériel ?
2. indiquer tes manipulations dans l'ordre où tu dois les réaliser ?
3. écrire tous les détails des manipulations afin qu'une personne qui ne sait pas ce qu'elle doit faire puisse réaliser exactement ton expérience en suivant ces manipulations ?
4. représenter tes manipulations sous forme de dessin si cela permet d'être plus clair ?

**ATTENTION!!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

L'élève peut poursuivre son travail : Oui \_\_\_ Non \_\_\_

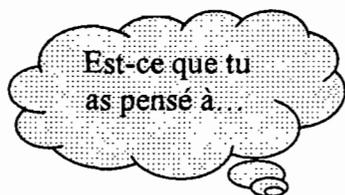
### Troisième étape : Je réalise l'expérience

Prends ici en note toutes les informations pertinentes au cours de ton expérience.



1. construire un tableau qui te permettra de consigner tes données ?
2. recourir à un témoin (afin de comparer tes données) ?
3. noter toutes les observations utiles que tu as pu voir durant ton expérience ?

**Quatrième étape : J'analyse les résultats de mon expérience**



1. présenter tes données à l'aide d'un graphique ou d'un diagramme à bandes ?
2. mesurer la variable que tu voulais vraiment investiguer ?
3. identifier des causes d'erreurs expérimentales durant tes manipulations ?



## Auto-évaluation

**Consigne : Noircit la case (Oui, Non, Un peu) qui correspondant le mieux à l'énoncé de la colonne de gauche. Au besoin, apporte les correctifs nécessaires dans ton cahier de réponse.**

1. J'ai bien compris le problème et j'ai été capable de le reformuler dans mes mots.	Oui	Non	Un peu
2. J'ai identifié la variable importante que je voulais investiguer dans mon expérience.	Oui	Non	Un peu
3. J'ai formulé une hypothèse vérifiable.	Oui	Non	Un peu
4. J'ai élaboré des manipulations expérimentales qui permettent de reproduire l'expérience.	Oui	Non	Un peu
5. J'ai identifié les causes d'erreurs expérimentales.	Oui	Non	Un peu
6. La solution que je propose s'appuie sur les données expérimentales que j'ai recueillies.	Oui	Non	Un peu
7. J'ai vérifié et corrigé les fautes d'orthographe.	Oui	Non	Un peu
8. Je me suis appliqué à bien écrire et à présenter soigneusement mes réponses.	Oui	Non	Un peu
9. J'ai utilisé les bons symboles scientifiques pour présenter mes données.	Oui	Non	Un peu

**ATTENTION!!!**

Prends le temps de réviser ton cahier de réponse afin de t'assurer que tu as répondu à toutes les questions. Quand tu as terminé ta révision, tu remets ton cahier à ton enseignant(e).

Nom de l'élève

Âge de l'élève

Masculin

Féminin

Groupe

Nom de l'enseignant(e)

Nom de l'école

Nom de la commission scolaire

**ATTENTION!!!**

**Assure-toi de répondre à toutes les sections de cette page**

## Consignes

- Tu disposes de 4 périodes de 75 minutes pour résoudre ce problème.
- Ce document (cahier de l'élève) doit être remis à ton enseignant(e) à la fin de **CHAQUE** période. Tu n'as pas le droit de l'apporter à la maison.
- À moins que ton enseignant(e) ne te donne des indications contraires, cette activité se déroule individuellement.
- Tu as le droit de consulter des livres de référence pour t'aider à résoudre ce problème.

## Ton mandat

Les balles jouent un rôle fort important au tennis. Elles doivent être bien conçues afin de résister aux chocs violents de la raquette en plus de bien **rebondir**. Déterminez une **variable** qui peut influencer le rebond de la balle et démontrez comment ce facteur influence le rebond.

**Première étape : Je cerne le problème !**

1. En utilisant un dictionnaire ou un livre de référence, donne la définition des mots apparaissant en gras dans l'énoncé du problème et qui apparaissent ci-dessous.

**Rebondir**

Définition :

---

---

---

---

---

Référence (indique dans quel livre tu as pris la définition) :

---

**Variable**

Définition :

---

---

---

---

---

Référence (indique dans quel livre tu as pris la définition) :

---

2. En utilisant tes propres mots (sans copier), indique le problème que tu dois résoudre.

---

---

---

---

---

---

---

---

---

---

---

3. Afin de résoudre le problème, on indique qu'il faut déterminer une **variable** qui peut influencer le rebond de la balle. Dans la liste qui suit, choisis une variable que tu souhaites investiguer parmi les trois qui te sont présentées.

a) Température de la balle
b) Taux d'humidité de la balle
c) La résistance du caoutchouc de la balle à se déformer

Explique pourquoi tu as choisi cette variable.

---

---

---

---

---

---

---

4. Quel est le but de l'expérience ?

---

---

---

---

---

5. Quelle est l'hypothèse que tu souhaites vérifier ?

---

---

---

---

---

**ATTENTION!!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

L'élève peut poursuivre son travail : Oui \_\_\_ Non \_\_\_

**Deuxième étape : Je détermine une façon de résoudre le problème**

Maintenant, tu dois te préparer à réaliser l'expérience qui te permettra de résoudre le problème.

6. En te basant sur la liste de matériel et les manipulations qui suivent, réalise chacune des étapes des protocoles qui suivent.

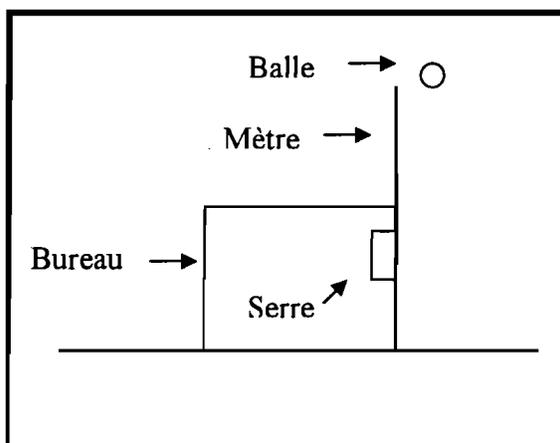
Voici la liste du matériel disponible pour cette activité :

Thermomètres	Lime
Mètres à mesurer	Supports universels
Diverses balles de tennis	Diverses masses
Béchers	Couteau utilitaire
Eau	
Pieds à coulisse	
Papier sablé	

Selon la variable que tu as choisie, réalise l'un des trois protocoles qui suit :

**Protocole #1 : Pour mesurer l'effet de la température sur le rebond de la balle...**

1. Réalise le montage qui suit.



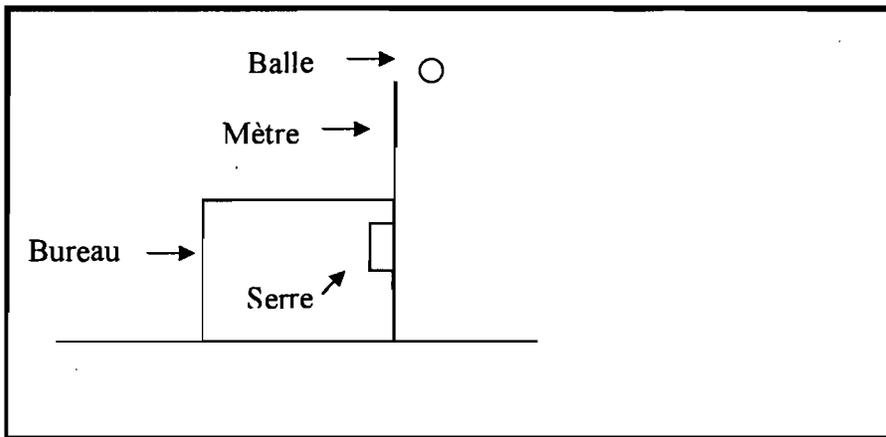
2. Choisis deux balles identiques.

3. Place l'une des balles (balle #1) au congélateur pendant environ 20 minutes.

4. En attendant, utilise l'autre balle (balle #2) et mesure la hauteur à laquelle elle rebondit en prenant soin de toujours la laisser tomber à partir du même point (ex. extrémité du mètre à mesurer). Laisse tomber la balle 5 fois et note, à chaque fois, la hauteur à laquelle elle rebondit.
5. Réalise exactement les mêmes manipulations avec la balle #1 lorsque cette dernière a passé 20 minutes au congélateur. Prends en note les valeurs obtenues à chaque essai.

Protocole #2 : Pour mesurer l'effet de l'humidité sur le rebond de la balle...

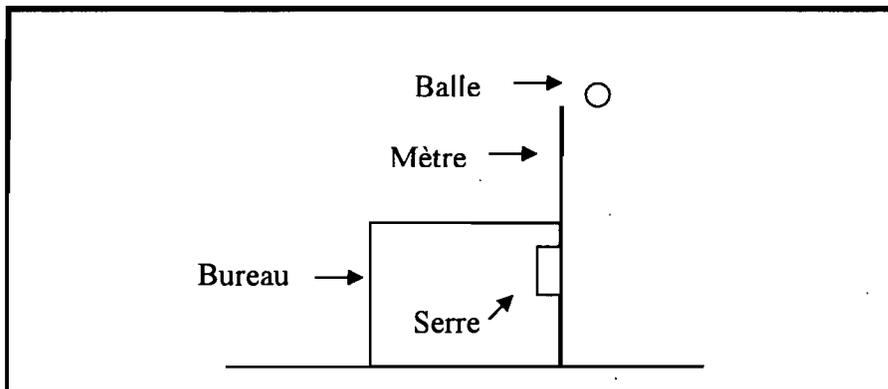
1. Réalise le montage qui suit.



2. Choisis deux balles identiques.
3. Verse une certaine quantité d'eau (c'est toi qui choisit la quantité d'eau) dans un bécher de 500 mL. Plonge l'une des balles (balle #1) dans le bécher en prenant soin de l'humecter d'eau sur toute sa surface.
4. Utilise l'autre balle (balle #2) et mesure la hauteur à laquelle elle rebondit en prenant soin de toujours la laisser tomber à partir du même point (ex. extrémité du mètre à mesurer). Laisse tomber la balle 5 fois et note, à chaque fois, la hauteur à laquelle elle rebondit.
5. Réalise exactement les mêmes manipulations avec la balle #1. Prends en note les valeurs obtenues à chaque essai.

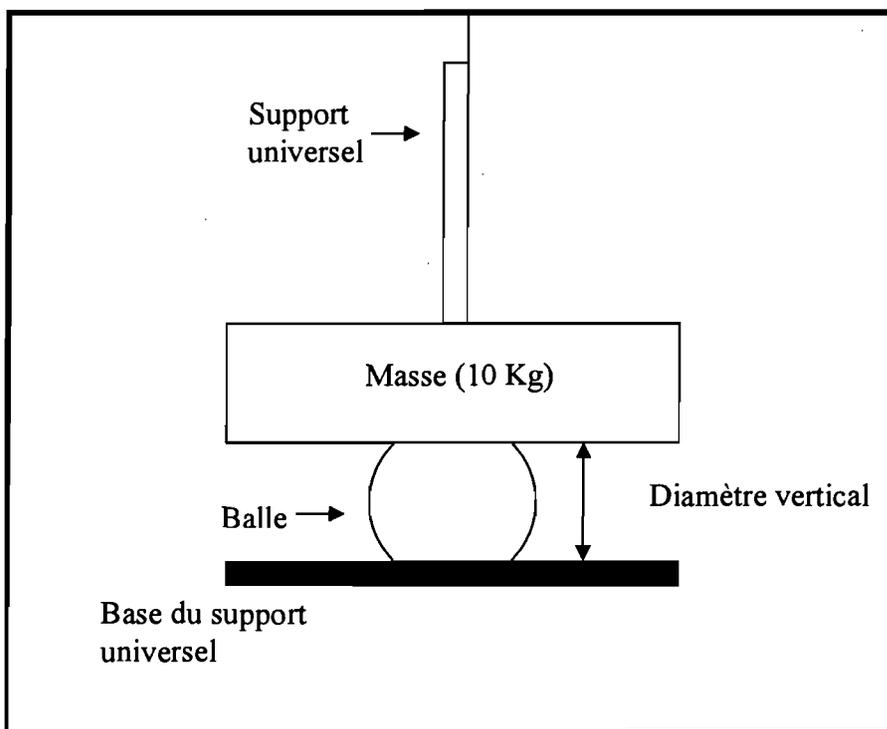
Protocole #3 : Pour mesurer l'effet de la résistance du caoutchouc sur le rebond de la balle...

1. Réalise le montage qui suit.



2. Choisis deux balles différentes.

3. Applique une masse d'environ 10 Kg sur chacune des balles et note le diamètre vertical de la balle à ce moment.



4. Prends ensuite chacune des balles et mesure la hauteur à laquelle elles rebondissent en prenant soin de toujours la laisser tomber à partir du même point (ex. extrémité du mètre à mesurer). Laisse tomber chacune des balles 5 fois et note, à chaque fois, la hauteur à laquelle elles rebondissent.

**ATTENTION!!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

L'élève peut poursuivre son travail : Oui \_\_\_ Non \_\_\_

**Troisième étape : Je réalise l'expérience**

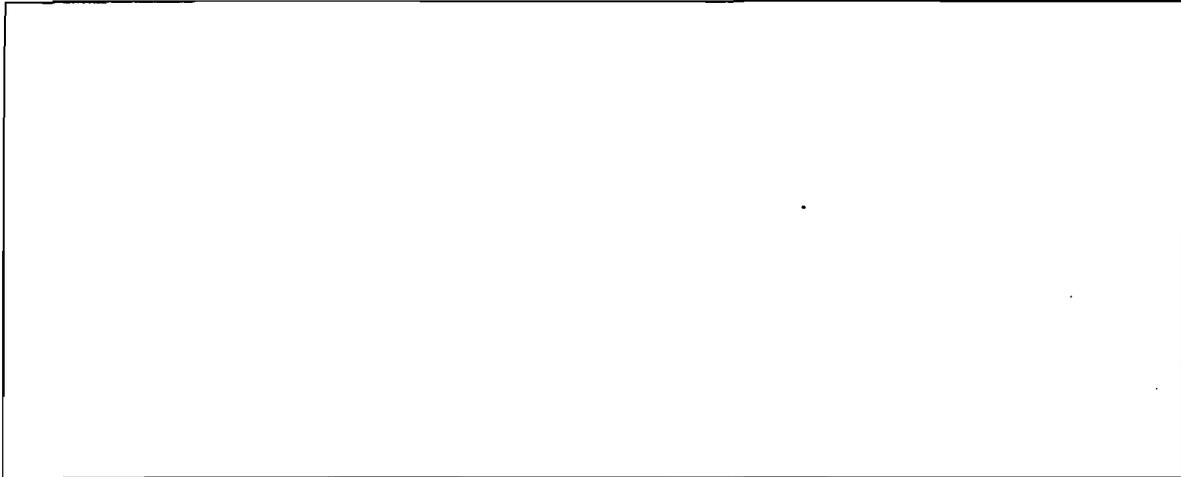
7. Complète le tableau des données qui suit durant ton expérience.

Variable investiguée : \_\_\_\_\_

	Hauteur du rebond de la balle #1	Hauteur du rebond de la balle #2	Observations
Essai 1			
Essai 2			
Essai 3			
Essai 4			
Essai 5			

**Quatrième étape : J'analyse les résultats de mon expérience**

8. Calcule la moyenne de la hauteur obtenue pour la balle #1 et ensuite pour la balle #2 au cours des 5 essais.



9. Est-ce qu'il y a une différence entre les moyennes calculées au numéro précédent ? Si oui, est-ce que cette différence est importante ?

---

---

---

---

---

---

---

---

---

---

10. Explique pourquoi on t'a demandé de recommencer l'expérience à trois reprises.

---

---

---

---

---

---

---

---

---

---

11. Quelle balle a le mieux rebondit ? Explique pourquoi cette balle a de meilleurs rebonds.

---

---

---

---

---

---

---

---

---

---

12. Que peux-tu conclure à propos de l'expérience que tu viens de mener ?

---

---

---

---

---

---

---

---

---

---

13. Si tu pouvais recommencer l'expérience, qu'est-ce que tu modifierais à celle-ci afin d'obtenir de meilleurs résultats ?

---

---

---

---

---

---

---

---

---

---



**ATTENTION!!!**

Prends le temps de réviser ton cahier de réponse afin de t'assurer que tu as répondu à toutes les questions. Quand tu as terminé ta révision, tu remets ton cahier à ton enseignant(e).

## Auto-évaluation

**Consigne :** Noircit la case (Oui, Non, Un peu) qui correspondant le mieux à l'énoncé de la colonne de gauche. Au besoin, apporte les correctifs nécessaires dans ton cahier de réponse.

1. J'ai bien compris le problème et j'ai été capable de le reformuler dans mes mots.	Oui	Non	Un peu
2. J'ai identifié la variable importante que je voulais investiguer dans mon expérience.	Oui	Non	Un peu
3. J'ai formulé une hypothèse vérifiable.	Oui	Non	Un peu
4. J'ai élaboré des manipulations expérimentales qui permettent de reproduire l'expérience.	Oui	Non	Un peu
5. J'ai identifié les causes d'erreurs expérimentales.	Oui	Non	Un peu
6. La solution que je propose s'appuie sur les données expérimentales que j'ai recueillies.	Oui	Non	Un peu
7. J'ai vérifié et corrigé les fautes d'orthographe.	Oui	Non	Un peu
8. Je me suis appliqué à bien écrire et à présenter soigneusement mes réponses.	Oui	Non	Un peu
9. J'ai utilisé les bons symboles scientifiques pour présenter mes données.	Oui	Non	Un peu

	Critères d'évaluation généraux (c.f. programme d'études)	Référence dans le cahier de l'élève	Critères d'évaluation spécifiques	Niveau de performance		
				0	1	2
C 1	Représentation adéquate de la situation	Première étape	1. Reformulation du problème	0	1	2
C 1	Représentation adéquate de la situation	Première étape	2. Identification des variables, connaissances et techniques	0	1	2
C 1	Élaboration d'une démarche pertinente pour la situation	Deuxième étape	3. Formulation d'une hypothèse	0	1	2
C 1	Élaboration d'une démarche pertinente pour la situation	Deuxième étape	4. Élaboration du protocole	0	1	2
C 1	Mise en œuvre adéquate de la démarche	Troisième étape	5. Consignation des données expérimentales	0	1	2
C 1	Élaboration de conclusions ou d'explications pertinentes	Quatrième et cinquième étape	6. Identification des erreurs expérimentales ou des limites	0	1	2
C 1	Élaboration de conclusions ou d'explications pertinentes	Quatrième et cinquième étape	7. Interprétation des données	0	1	2
C 1	Élaboration de conclusions ou d'explications pertinentes	Quatrième et cinquième étape	8. Qualité de la solution ou de l'explication	0	1	2
C 3	Respect de la terminologie, des règles et des conventions propres à la science et à la technologie dans la production de messages	Première à la cinquième étape	9. Qualité de la langue	0	1	2
C 3	Respect de la terminologie, des règles et des conventions propres à la science et à la technologie dans la production de messages	Première à la cinquième étape	10. Qualité de la présentation	0	1	2
C 3	Respect de la terminologie, des règles et des conventions propres à la science et à la technologie dans la production de messages	Première à la cinquième étape	11. Utilisation des termes et symboles scientifiques	0	1	2

**Troisième situation (SE3) : Les lombrics**

- Canevas de la situation
- Les situations SE3-1, SE3-2, SE3-3
- Outil d'évaluation

## Canevas de la situation d'évaluation SE3 : Les lombrics

1. Titre de la tâche			
Les lombrics			
2. Intention pédagogique et/ou évaluative			
<i>Pourquoi faire réaliser cette tâche par les élèves ?</i>			
1. Familiariser les élèves à une démarche scientifique plus « ouverte ».			
2. Évaluer la capacité des élèves à résoudre un problème scientifique en usant de créativité.			
3. Clientèle visée			
	Cycle :	Sec. 1	Moment dans l'année : Mai-Juin
4. Temps consacré à la tâche			
Environ 6 périodes de 75 minutes			
5. Type d'évaluation			
	<input type="checkbox"/> Formative	<input checked="" type="checkbox"/> Certificative (Bilan)	<input type="checkbox"/> Diagnostique
6. Compétence disciplinaire ciblée			
	<input checked="" type="checkbox"/> Compétence 1	<input type="checkbox"/> Compétence 2	<input checked="" type="checkbox"/> Compétence 3
<i>Expliquez brièvement pourquoi la tâche permet de développer la compétence.</i>			
C1 : Les élèves doivent résoudre un problème <i>hands-on</i> pour lequel il existe plusieurs solutions et dont la valeur de la démarche est fonction des explications fournies par les élèves.			
C3 : Les élèves sont invités à faire part de leurs résultats et de leur démarche aux autres élèves dans un langage clair et rigoureux.			
7. Compétence(s) transversale(s) ciblée(s)			
Ordre intellectuel	Ordre de la communication	Ordre personnel et social	Ordre méthodologique
<input type="checkbox"/> Exploiter l'information	<input type="checkbox"/> Communiquer de façon appropriée	<input type="checkbox"/> Actualiser son potentiel	<input type="checkbox"/> Se donner des méth. trav. efficaces
<input checked="" type="checkbox"/> Résoudre des problèmes		<input type="checkbox"/> Coopérer	<input type="checkbox"/> Exploiter les TIC
<input type="checkbox"/> Exercer son jugement critique			
<input type="checkbox"/> Mettre en œuvre sa pensée créatrice			
<i>Expliquez brièvement pourquoi la tâche permet de développer la compétence ciblée.</i>			
L'élève est invité à imaginer une solution originale au problème qui lui est soumis.			
8. DGF touché			
<input type="checkbox"/> Médias	<input type="checkbox"/> Santé et bien-être	<input type="checkbox"/> Vivre ensemble et citoyenneté	
<input type="checkbox"/> Orientation et entrepreneuriat	<input type="checkbox"/> Environnement et consommation		
<i>Expliquez en quoi l'activité permet de réaliser des apprentissages en lien avec les axes de développement du DGF touché.</i>			
La situation est surtout à caractère scientifique. Aucun axe de développement formel n'est spécifiquement développé.			
9. Univers notionnel(s) visé(s)			

<input type="checkbox"/> Univers matériel	<input type="checkbox"/> Univers vivant	<input type="checkbox"/> Univers technologique	<input type="checkbox"/> Terre et espace
<b>Contenu obligatoire / Contenu facultatif</b>			
<b>Initiation</b>	<b>Consolidation</b>	<b>Maîtrise</b>	
		<b>Adaptations physiques et comportementales</b>	
		<b>Niche écologique</b>	
		<b>Habitat</b>	
		Caractéristiques du vivant	
		Acidité/basicité	
<i>Techniques</i>			
		<b>Utilisation sécuritaire du matériel de labo</b>	
		<b>Utilisation d'instruments de mesure</b>	
<i>Attitudes</i>			
		<b>Sens de l'initiative</b>	
		<b>Goût du risque intellectuel</b>	
		<b>Considérations de solutions originales</b>	
<i>Stratégies</i>			
		Explorer diverses pistes de solution	
		Recourir à des outils permettant de représenter des données sous forme de tableaux et de graphiques ou de tracer des diagrammes	
<b>10. Repères culturels</b>			
<i>Indiquez les principaux repères culturels touchés dans cette tâche.</i>			
<ul style="list-style-type: none"> <li>• La pêche</li> <li>• Les décomposeurs et leur rôle dans la chaîne alimentaire</li> <li>• Etc.</li> </ul>			

Nom de l'élève

Âge de l'élève

Masculin

Féminin

<input type="checkbox"/>
<input type="checkbox"/>

Groupe

Nom de l'enseignant(e)

Nom de l'école

Nom de la commission scolaire

**ATTENTION!!!**

**Assure-toi de répondre à toutes les sections de cette page**

## Consignes générales

- Tu disposes de 4 périodes de 75 minutes pour résoudre ce problème.
- Ce document (cahier de l'élève) doit être remis à ton enseignant(e) à la fin de **CHAQUE** période. Tu n'as pas le droit de l'apporter à la maison.
- À moins que ton enseignant(e) ne te donne des indications contraires, cette activité se déroule individuellement.
- Au besoin, utilise le verso de chaque page de ton cahier de réponse.
- Tu as le droit de consulter des livres de référence pour t'aider à résoudre ce problème.

## Ton mandat

L'été approche et la saison de la pêche va reprendre de plus belle. Afin de te faire un peu d'argent de poche, tu décides de cultiver des lombrics pour les vendre aux pêcheurs. Tu voudrais bien leur offrir les meilleures conditions de vie possible afin d'obtenir de beaux gros lombrics bien en santé qui attireront les poissons ! Parmi tous les facteurs qui peuvent influencer l'abondance et la santé des lombrics, choisis-en un et démontre comment ce facteur agit sur les lombrics.

**Première étape : Je cerne le problème !**

**ATTENTION !!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

L'élève peut poursuivre son travail : Oui \_\_\_ Non \_\_\_

**Deuxième étape : Je détermine une façon de résoudre le problème**

Maintenant, tu dois te préparer à réaliser l'expérience qui te permettra de résoudre le problème. Indique le matériel dont tu auras besoin pour réaliser ton expérience.



**Troisième étape : Je réalise l'expérience**

Prends ici en note toutes les informations pertinentes au cours de ton expérience.

**ATTENTION**

**Les lombrics sont des êtres vivants qui doivent être manipulés avec soin.**

**Quatrième étape : J'analyse les résultats de mon expérience**

**Cinquième étape : Je propose une solution au problème**

Rédige une publicité (1/2 à 1 page) s'adressant à tes clients afin de les informer et les convaincre de l'abondance et de la santé de tes lombrics.

**ATTENTION !!!**

Prends le temps de réviser ton cahier de réponse afin de t'assurer que tu as répondu à toutes les questions. Quand tu as terminé ta révision, tu remets ton cahier à ton enseignant(e).

Nom de l'élève

Âge de l'élève

Masculin	
Féminin	

Groupe

Nom de l'enseignant(e)

Nom de l'école

Nom de la commission scolaire

**ATTENTION!!!**

**Assure-toi de répondre à toutes les sections de cette page**

## Consignes générales

- Tu disposes de 4 périodes de 75 minutes pour résoudre ce problème.
- Ce document (cahier de l'élève) doit être remis à ton enseignant(e) à la fin de **CHAQUE** période. Tu n'as pas le droit de l'apporter à la maison.
- À moins que ton enseignant(e) ne te donne des indications contraires, cette activité se déroule individuellement.
- Au besoin, utilise le verso de chaque page de ton cahier de réponse.
- Tu as le droit de consulter des livres de référence pour t'aider à résoudre ce problème.

## Ton mandat

L'été approche et la saison de la pêche va reprendre de plus belle. Afin de te faire un peu d'argent de poche, tu décides de cultiver des lombrics pour les vendre aux pêcheurs. Tu voudrais bien leur offrir les meilleures conditions de vie possible afin d'obtenir de beaux gros lombrics bien en santé qui attireront les poissons ! Parmi tous les variables qui peuvent influencer l'abondance et la santé des lombrics, choisis-en une et démontre comment cette variable agit sur les lombrics.

**Première étape : Je cerne le problème !**

1. reformuler le mandat dans tes propres mots ?
2. trouver la définition des mots que tu ne comprenais pas dans la section « Ton mandat » ? En particulier ceux qui sont soulignés...
3. chercher dans des livres de référence afin de t'aider à mieux comprendre ce problème ?
4. énoncer le but de l'expérience que tu allais mener et l'hypothèse que tu vérifierais ?
5. identifier quelle variable tu mesurerais ?

**ATTENTION !!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

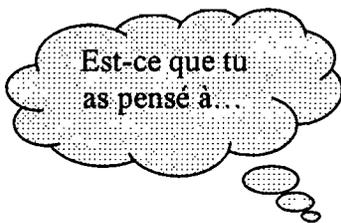
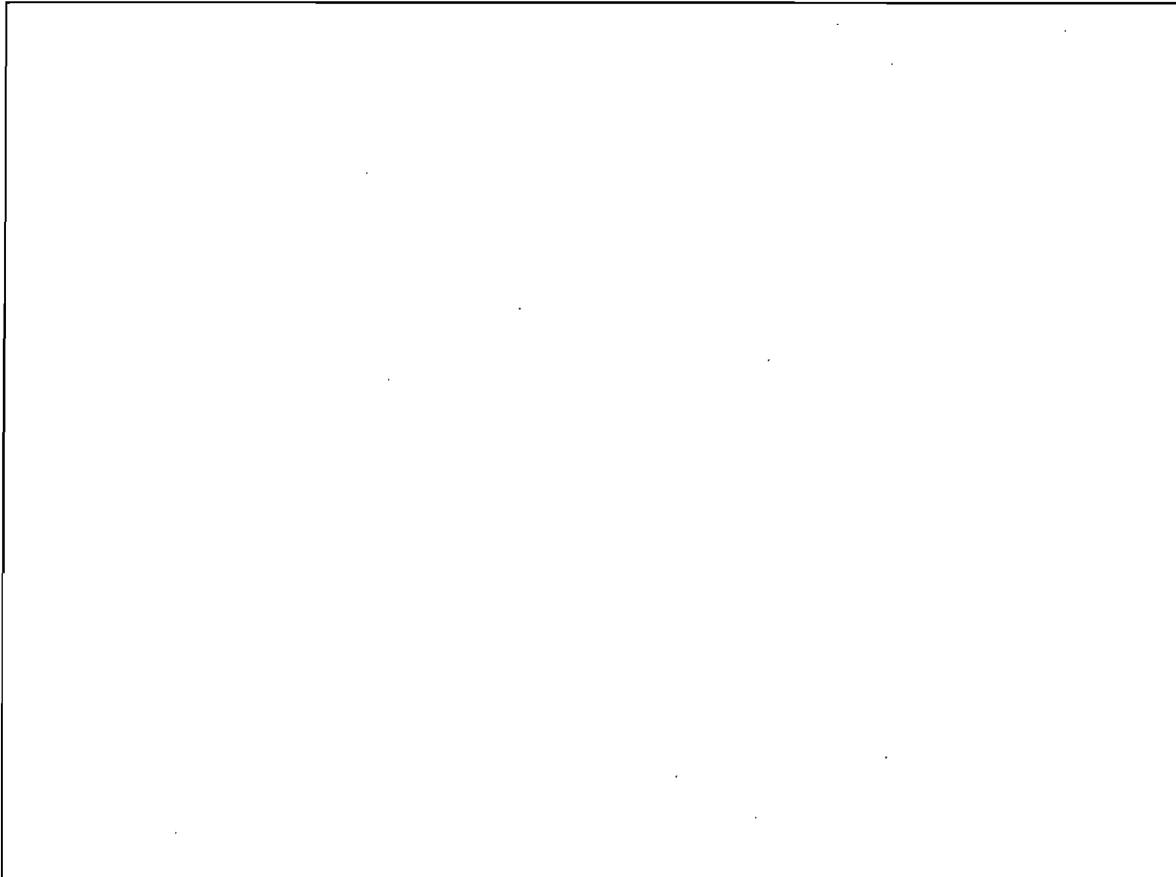
L'élève peut poursuivre son travail : Oui  Non

**Deuxième étape : Je détermine une façon de résoudre le problème**

Maintenant, tu dois te préparer à réaliser l'expérience qui te permettra de résoudre le problème. Voici le matériel que tu peux utiliser. Au besoin, complète cette liste en indiquant le matériel que tu as utilisé.

Lombrics	Glace
Boîte de carton (ex. boîte à chaussure)	Terre
Source lumineuse	Séparateurs (carton, plastique, etc.)
Gants protecteurs	

Indique les manipulations que tu dois effectuer pour faire ton expérience. Au besoin, utilise le verso de cette feuille pour compléter.



1. compléter la liste du matériel ?
2. indiquer tes manipulations dans l'ordre où tu dois les réaliser ?
3. écrire tous les détails des manipulations afin qu'une personne qui ne sait pas ce qu'elle doit faire puisse réaliser exactement ton expérience en suivant ces manipulations ?
4. représenter tes manipulations sous forme de dessin si cela permet d'être plus clair ?

**ATTENTION !!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

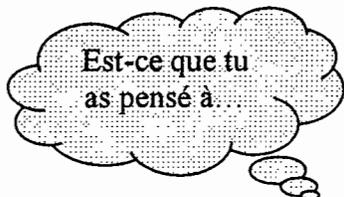
L'élève peut poursuivre son travail : Oui \_\_\_ Non \_\_\_

**Troisième étape : Je réalise l'expérience**

Prends ici en note toutes les informations pertinentes au cours de ton expérience.

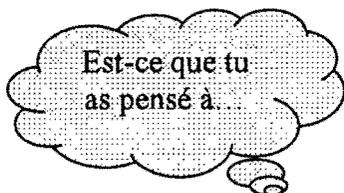
**ATTENTION**

**Les lombrics sont des êtres vivants qui doivent être manipulés avec soin.**



1. construire un tableau qui te permettra de consigner tes données ?
2. recourir à un témoin (afin de comparer tes données) ?
3. noter toutes les observations utiles que tu as pu voir durant ton expérience ?

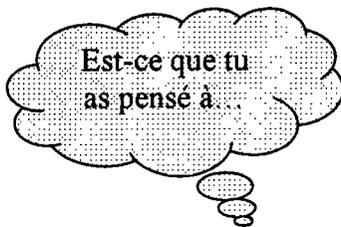
**Quatrième étape : J'analyse les résultats de mon expérience**



1. présenter tes données clairement en t'aidant par exemple de schémas ?
2. observer la variable que tu souhaitais investiguer ?
3. identifier des causes d'erreurs expérimentales durant tes manipulations ?

**Cinquième étape : Je propose une solution au problème**

Rédige une publicité (1/2 à 1 page) s'adressant à tes clients afin de les informer et les convaincre de l'abondance et de la santé de tes lombrics.



1. relire ton mandat afin de t'assurer de bien répondre au problème ?
2. utiliser les données de ton expérience pour convaincre les responsables de la production des balles de tennis ?

## Auto-évaluation

**Consigne : Noircit la case (Oui, Non, Un peu) qui correspondant le mieux à l'énoncé de la colonne de gauche. Au besoin, apporte les correctifs nécessaires dans ton cahier de réponse.**

1. J'ai bien compris le problème et j'ai été capable de le reformuler dans mes mots.	Oui	Non	Un peu
2. J'ai identifié la variable importante que je voulais investiguer dans mon expérience.	Oui	Non	Un peu
3. J'ai formulé une hypothèse vérifiable.	Oui	Non	Un peu
4. J'ai élaboré des manipulations expérimentales qui permettent de reproduire l'expérience.	Oui	Non	Un peu
5. J'ai identifié les causes d'erreurs expérimentales.	Oui	Non	Un peu
6. La solution que je propose s'appuie sur les données expérimentales que j'ai recueillies.	Oui	Non	Un peu
7. J'ai vérifié et corrigé les fautes d'orthographe.	Oui	Non	Un peu
8. Je me suis appliqué à bien écrire et à présenter soigneusement mes réponses.	Oui	Non	Un peu
9. J'ai utilisé les bons symboles scientifiques pour présenter mes données.	Oui	Non	Un peu

**ATTENTION!!!**

Prends le temps de réviser ton cahier de réponse afin de t'assurer que tu as répondu à toutes les questions. Quand tu as terminé ta révision, tu remets ton cahier à ton enseignant(e).

Nom de l'élève

Âge de l'élève

Masculin	<input type="checkbox"/>
Féminin	<input type="checkbox"/>

Groupe

Nom de l'enseignant(e)

Nom de l'école

Nom de la commission scolaire

**ATTENTION!!!**

**Assure-toi de répondre à toutes les sections de cette page**

## Consignes

- Tu disposes de 4 périodes de 75 minutes pour résoudre ce problème.
- Ce document (cahier de l'élève) doit être remis à ton enseignant(e) à la fin de **CHAQUE** période. Tu n'as pas le droit de l'apporter à la maison.
- À moins que ton enseignant(e) ne te donne des indications contraires, cette activité se déroule individuellement.
- Tu as le droit de consulter des livres de référence pour t'aider à résoudre ce problème.

## Ton mandat

L'été approche et la saison de la pêche va reprendre de plus belle. Afin de te faire un peu d'argent de poche, tu décides de cultiver des **lombrics** pour les vendre aux pêcheurs. Tu voudrais bien leur offrir les meilleures conditions de vie possible afin d'obtenir de beaux gros lombrics bien en santé qui attireront les poissons ! Parmi toutes les **variables** qui peuvent influencer l'abondance et la santé des lombrics, choisis-en un et démontre comment ce facteur agit sur les lombrics.

**Première étape : Je cerne le problème !**

1. En utilisant un dictionnaire ou un livre de référence, donne la définition des mots apparaissant en gras dans l'énoncé du problème et qui apparaissent ci-dessous.

**Lombric**

Définition :

---

---

---

---

---

Référence (indique dans quel livre tu as pris la définition) :

---

**Variable**

Définition :

---

---

---

---

---

Référence (indique dans quel livre tu as pris la définition) :

---



4. Quel est le but de l'expérience ?

---

---

---

---

---

5. Quelle est l'hypothèse que tu souhaites vérifier ?

---

---

---

---

---

**ATTENTION!!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

L'élève peut poursuivre son travail : Oui \_\_\_ Non \_\_\_

**Deuxième étape : Je détermine une façon de résoudre le problème**

Maintenant, tu dois te préparer à réaliser l'expérience qui te permettra de résoudre le problème.

6. En te basant sur la liste de matériel et les manipulations qui suivent, réalise chacune des étapes des protocoles qui suivent.

**ATTENTION**

**Les lombrics sont des êtres vivants qui doivent être manipulés avec soin.**

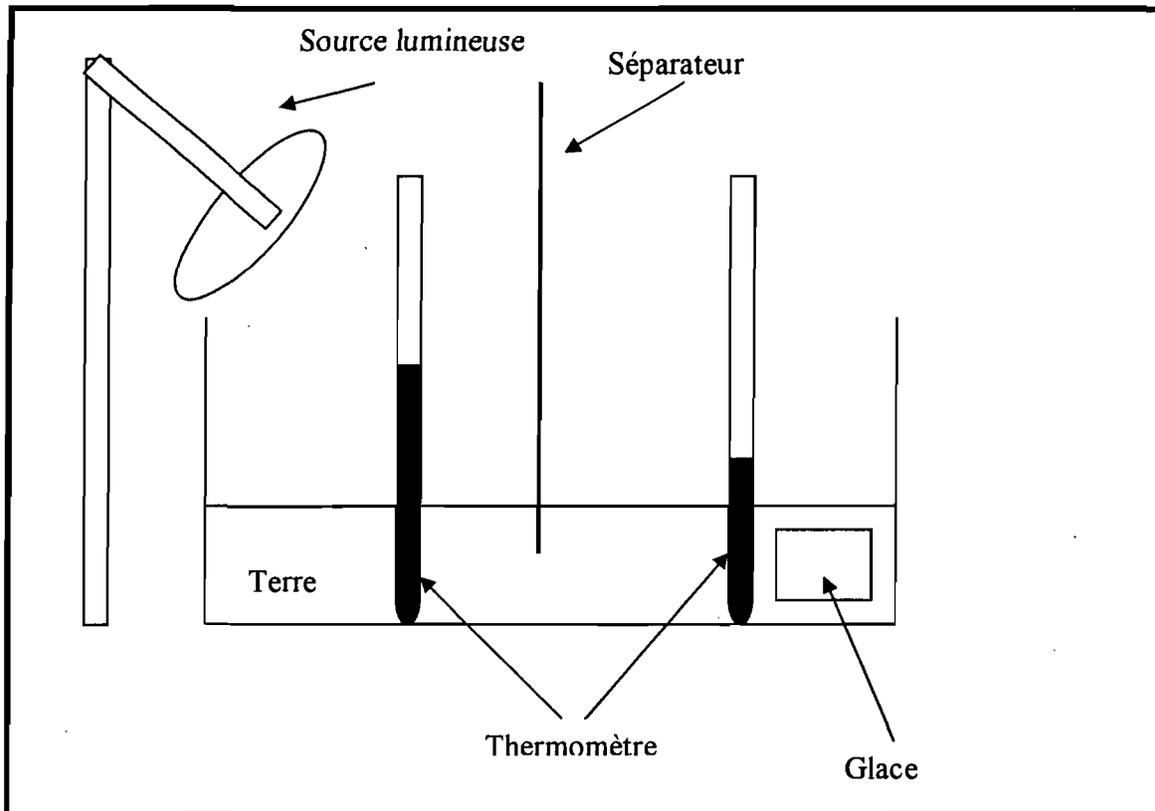
Voici la liste du matériel disponible pour cette activité :

Lombrics	Glace
Eau	Thermomètre
Béchers	Terre
Boîte de carton (ex. boîte à chaussure)	Séparateurs (carton, plastique, etc.)
Source lumineuse	Gants protecteurs
Loupe	

Selon la variable que tu as choisie, réalise l'un des trois protocoles qui suit :

Protocole #1 : Pour mesurer l'effet de la chaleur sur l'abondance des lombrics...

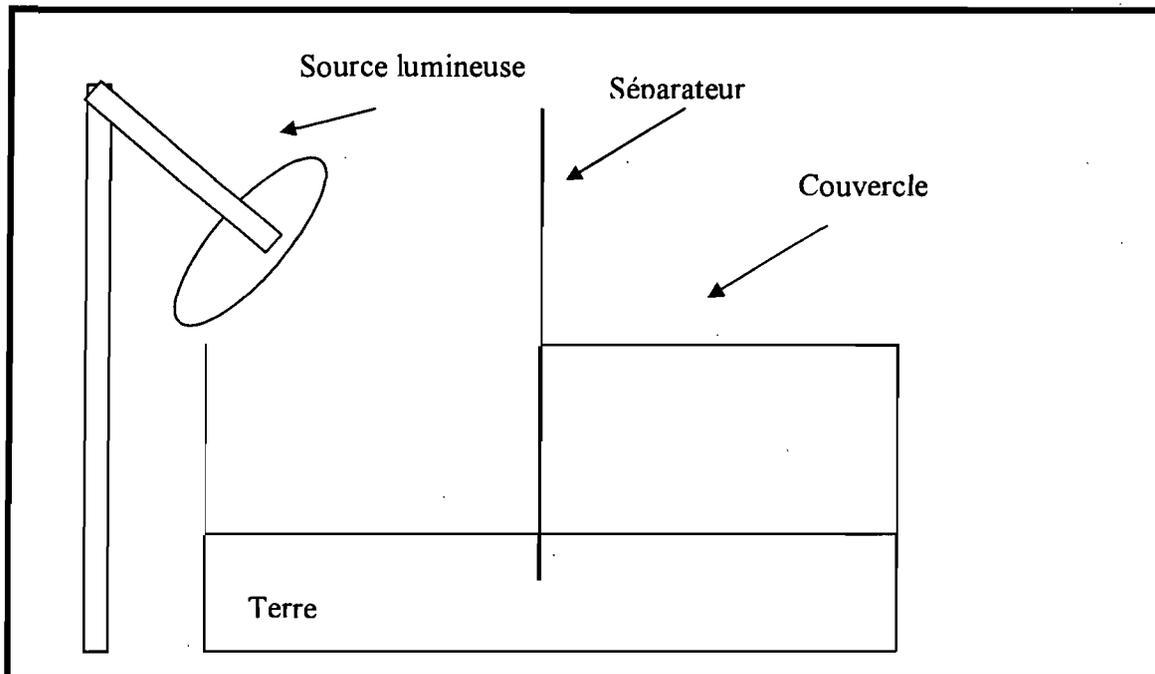
1. Réalise le montage qui suit.



2. Lors que tu as complété le montage ci-dessus, prends un lombric et pose-le délicatement au centre de ta boîte vis-à-vis le séparateur.
3. Observe le lombric et note tes observations.
4. Fais de même avec deux autres lombrics.

**Protocole #2 : Pour mesurer l'effet de la lumière sur l'abondance des lombrics...**

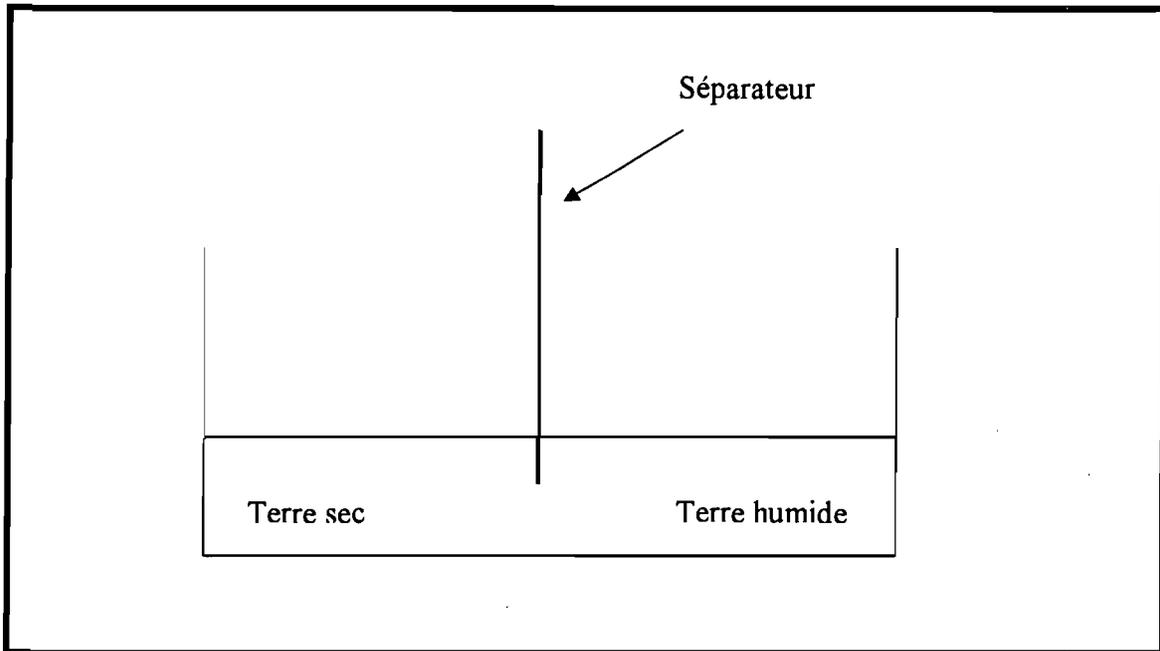
1. Réalise le montage qui suit.



2. Lors que tu as complété le montage ci-dessus, prends un lombric et pose-le délicatement au centre de ta boîte vis-à-vis le séparateur.
3. Observe le lombric et note tes observations.
4. Fais de même avec deux autres lombrics.

Protocole #3 : Pour mesurer l'effet de l'humidité sur l'abondance des lombrics...

1. Réalise le montage qui suit.



2. Lors que tu as complété le montage ci-dessus, prends un lombric et pose-le délicatement au centre de ta boîte vis-à-vis le séparateur.
3. Observe le lombric et note tes observations.
4. Fais de même avec deux autres lombrics.

**ATTENTION!!!**

Avant de poursuivre, demande à ton enseignant de vérifier ton travail jusqu'à maintenant.

L'élève peut poursuivre son travail : Oui \_\_\_ Non \_\_\_

**Troisième étape : Je réalise l'expérience**

7. Complète le tableau des données qui suit durant ton expérience.

Variable investiguée : \_\_\_\_\_

	Observations
Essai avec le lombric #1	
Essai avec le lombric #2	
Essai avec le lombric #3	

**Quatrième étape : J'analyse les résultats de mon expérience**

8. Comment se sont comportés les lombrics durant ton expérience ?

---

---

---

---

---

---

---

---

---

---

9. Pourquoi as-tu refait l'expérience avec trois lombrics différents ?

---

---

---

---

---

---

---

---

---

---

10. Que peux-tu conclure à propos de l'expérience que tu viens de mener ?

---

---

---

---

---

---

---

---

---

13. Si tu pouvais recommencer l'expérience, qu'est-ce que tu modifierais à celle-ci ?

---

---

---

---

---

---

---

---

---

---

**Cinquième étape : Je propose une solution au problème**

Rédige une publicité (1/2 à 1 page) s'adressant à tes clients afin de les informer et les convaincre de l'abondance et de la santé de tes lombrics.

**ATTENTION!!!**

Prends le temps de réviser ton cahier de réponse afin de t'assurer que tu as répondu à toutes les questions. Pour t'aider, utilise la grille d'auto-évaluation qui apparaît ci-dessous. Quand tu as terminé ta révision, tu remets ton cahier à ton enseignant(e).

## Auto-évaluation

**Consigne : Noircit la case (Oui, Non, Un peu) qui correspondant le mieux à l'énoncé de la colonne de gauche. Au besoin, apporte les correctifs nécessaires dans ton cahier de réponse.**

1. J'ai bien compris le problème et j'ai été capable de le reformuler dans mes mots.	Oui	Non	Un peu
2. J'ai identifié la variable importante que je voulais investiguer dans mon expérience.	Oui	Non	Un peu
3. J'ai formulé une hypothèse vérifiable.	Oui	Non	Un peu
4. J'ai élaboré des manipulations expérimentales qui permettent de reproduire l'expérience.	Oui	Non	Un peu
5. J'ai identifié les causes d'erreurs expérimentales.	Oui	Non	Un peu
6. La solution que je propose s'appuie sur les données expérimentales que j'ai recueillies.	Oui	Non	Un peu
7. J'ai vérifié et corrigé les fautes d'orthographe.	Oui	Non	Un peu
8. Je me suis appliqué à bien écrire et à présenter soigneusement mes réponses.	Oui	Non	Un peu
9. J'ai utilisé les bons symboles scientifiques pour présenter mes données.	Oui	Non	Un peu

	Critères d'évaluation généraux (c.f. programme d'études)	Référence dans le cahier de l'élève	Critères d'évaluation spécifiques	Niveau de performance		
				0	1	2
C 1	Représentation adéquate de la situation	Première étape	1. Reformulation du problème	0	1	2
C 1	Représentation adéquate de la situation	Première étape	2. Identification des variables, connaissances et techniques	0	1	2
C 1	Élaboration d'une démarche pertinente pour la situation	Deuxième étape	3. Formulation d'une hypothèse	0	1	2
C 1	Élaboration d'une démarche pertinente pour la situation	Deuxième étape	4. Élaboration du protocole	0	1	2
C 1	Mise en œuvre adéquate de la démarche	Troisième étape	5. Consignation des données expérimentales	0	1	2
C 1	Élaboration de conclusions ou d'explications pertinentes	Quatrième et cinquième étape	6. Identification des erreurs expérimentales ou des limites	0	1	2
C 1	Élaboration de conclusions ou d'explications pertinentes	Quatrième et cinquième étape	7. Interprétation des données	0	1	2
C 1	Élaboration de conclusions ou d'explications pertinentes	Quatrième et cinquième étape	8. Qualité de la solution ou de l'explication	0	1	2
C 3	Respect de la terminologie, des règles et des conventions propres à la science et à la technologie dans la production de messages	Première à la cinquième étape	9. Qualité de la langue	0	1	2
C 3	Respect de la terminologie, des règles et des conventions propres à la science et à la technologie dans la production de messages	Première à la cinquième étape	10. Qualité de la présentation	0	1	2
C 3	Respect de la terminologie, des règles et des conventions propres à la science et à la technologie dans la production de messages	Première à la cinquième étape	11. Utilisation des termes et symboles scientifiques	0	1	2

**Appendice F**  
**Le questionnaire administré aux enseignants**

---

**QUESTIONNAIRE ADMINISTRÉ AUX ENSEIGNANTS**

**PREMIÈRE PARTIE : IDENTIFICATION DE L'ENSEIGNANT**

Nom

Prénom :

École :

Commission scolaire :

**DEUXIÈME PARTIE : PROFIL DE L'ÉCOLE**

1. Est-ce que l'école possède une vocation particulière (ex. voie scientifique, voie artistique, etc.) ? Si oui, laquelle ou lesquelles ?

2. Est-ce que l'école réalise une opération de sélection des élèves à l'entrée des élèves en première secondaire ? Si oui, sur quels critères se base-t-on pour réaliser la sélection ?

3. Est-ce que l'école est située en milieu urbain, rural ou semi-rural ?

**TROISIÈME PARTIE : PROFIL PROFESSIONNEL DE L'ENSEIGNANT**

4. Combien d'année d'expérience comptez-vous en enseignement au secondaire ?

5. Combien d'année d'expérience comptez-vous en enseignement des sciences au secondaire ?

6. a) Combien d'années de scolarité avez-vous ?

b) Quel(s) est(sont) le(s) diplôme(s) universitaire(s) que vous détenez ?

a)

b)

7. a) À combien de journées de formation (en évaluation des apprentissages, disciplinaire : science et technologie, approches pédagogiques, congrès, etc.) avez-vous assisté au cours des deux dernières années ?

b) Quelles sont les activités de formation (en évaluation des apprentissages, disciplinaire : science et technologie, approches pédagogiques, congrès, etc.) que vous avez suivi au cours des deux dernières années ? Décrivez brièvement la teneur de chacune de ces activités (objectifs de la formation, ce que vous avez fait ou appris, etc.).

a)

b)

**QUATRIÈME PARTIE : STYLE PÉDAGOGIQUE DE L'ENSEIGNANT**

8. Durant la présente année scolaire, est-ce que vos élèves ont réalisé des situations d'apprentissage qui ressemblaient à celles proposées dans le cadre de cette recherche ? Si oui, décrivez brièvement l'une de ces situations d'apprentissage.

## CINQUIÈME PARTIE : LE RENDEMENT DES ÉLÈVES

9. En utilisant une liste de classe contenant le numéro d'identification<sup>1</sup> des élèves participant à cette recherche, classez chacun des élèves dans l'une ou l'autre des trois catégories qui suivent :

	Catégorie d'élève	Commentaire
Catégorie 1	Élève cheminant avec difficulté en science et technologie.	Élève qui, à votre avis, pourrait rencontrer de profonde difficulté à atteindre le niveau de compétence minimal au terme du cycle (élève potentiellement en échec).
Catégorie 2	Élève cheminant normalement en science et technologie.	Élève qui, à votre avis, devrait normalement atteindre le niveau de compétence minimal au terme du cycle (élève potentiellement en réussite).
Catégorie 3	Élève cheminant remarquablement bien en science et technologie.	Élève qui, à votre avis, ne rencontreront aucune difficulté à atteindre le niveau de compétence minimal au terme du cycle.

Ex.

Liste de classe de l'enseignant Hubert Reeves

Élève	Catégorie
1.	1
2.	3
3.	2
4.	2
5.	3
6.	1
...	...
32	1

<sup>1</sup> Comment assigner le numéro d'identification à un élève ? Simplement en remplaçant le nom de l'élève par sa position dans la liste de classe. Le premier élève figurant sur la liste de classe (ex. Aaron, James) se voit assigner le numéro d'identification 1, le deuxième élève (Baron, Sylvie) se voit administrer le numéro d'identification 2, etc. Simplement vous assurer que – **pour chacun des élèves - le numéro d'identification figurant sur la liste de classe soit le même que celui apparaissant sur le cahier de l'élève.**