

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600



Université d'Ottawa • University of Ottawa

**LA JUSTESSE DES RÉSULTATS,
CRITÈRE DE QUALITÉ DE L'ÉVALUATION
DES APPRENTISSAGES EN SALLE DE CLASSE**

par : Micheline Bercier-Larivière

Thèse présentée à l'École des études supérieures et de la recherche
de l'Université d'Ottawa en vue de
l'obtention du Ph. D. en éducation

Ottawa, Canada, 1998

© Micheline Bercier-Larivière, Ottawa, Canada, 1998



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-36765-7

Canada

RECONNAISSANCES

Un sujet de recherche passionnant et de bonnes conditions de travail sont deux atouts précieux pour s'engager dans une thèse de doctorat. Mais il faut aussi pouvoir compter sur un comité à la fois critique, formatif et positif. Je remercie sincèrement les professeurs Jean-Paul Dionne, Dany Laveault et Marielle Simon de l'avoir été. Merci aussi aux cinq enseignants qui ont bien voulu m'aider à réaliser la partie empirique de cette recherche; sans eux je n'aurais pas pu réaliser cette étude.

L'accompagnement de ma directrice de thèse, Renée Forgette-Giroux, a cependant été l'élément déterminant dans la réalisation de ce projet. Aucun mot ne saurait dire combien son professionnalisme, sa disponibilité, sa générosité, sa confiance et son respect m'ont été précieux. C'est en très grande partie à elle que je dois cet accomplissement. Merci beaucoup Renée!

Mille mercis aussi à mes proches, Clémence et Sylvain. Vous m'avez permis de me consacrer, corps et âme, à la réalisation de ce projet, sans aucune animosité. Quel témoignage d'amour et d'affection. Merci encore une fois!

Micheline

RÉSUMÉ

Le désir de comprendre l'origine de la disparité des pratiques d'évaluation des apprentissages en salle de classe et des incohérences qui souvent les caractérisent est le point de départ de cette recherche. Force fut de constater que l'absence d'un modèle de qualité des résultats d'évaluation cohérent avec la nature et le contexte des apprentissages est en grande partie responsable du laxisme observé.

Qu'ils tirent leur origine du paradigme traditionnel de la psychométrie ou de celui plus moderne de l'édu-métrie, les éléments concourant à des résultats justes ont été regroupés en quatre composantes : la pertinence de la tâche d'évaluation par rapport aux habiletés visées, sa cohérence avec l'activité pédagogique, sa transparence à l'égard de l'élève et l'absence de désavantage circonstanciel. Tous ces éléments, retenus pour leur apport à l'obtention de résultats justes, c'est-à-dire pour situer correctement un élève dans son développement, forment le concept de justesse des résultats. Ce dernier a été confronté à la réalité de la salle de classe, en collaboration avec cinq enseignants d'écoles élémentaires de l'Ontario.

De l'avis des participants, les quatre composantes du concept de justesse sont toutes pertinentes à l'obtention de résultats justes. Cependant, cette conception de l'évaluation fait ressortir comme étant problématiques certaines situations d'évaluation en salle de classe, comme l'évaluation de tâches effectuées en équipe et celle d'habiletés multiples enseignées de manière intégrée. La discussion tente d'établir si ces faits sont dus au manque de réalisme du concept de justesse ou s'ils font plutôt ressortir des situations d'évaluation en salle de classe menant à des résultats d'évaluation de piètre qualité.

L'expérience menée ici ne suffit pas à établir la viabilité du critère de justesse comme modèle de qualité des résultats d'évaluation des apprentissages en salle de classe. Toutefois, cette expérience suggère fortement que cette organisation simplifiée de la réalité est compatible avec la nature contemporaine de l'apprentissage et rend bien compte de la complexité de son évaluation. De ce point de vue, elle s'avère supérieure aux critères de validité et de fidélité présentement utilisés pour guider les enseignants dans leurs interventions d'évaluation en salle de classe.

**LA JUSTESSE DES RÉSULTATS,
CRITÈRE DE QUALITÉ DE L'ÉVALUATION
DES APPRENTISSAGES EN SALLE DE CLASSE**

TABLE DES MATIÈRES

	page
INTRODUCTION	1
Chapitre 1 : CADRE CONCEPTUEL	6
A - NOUVEAU PARADIGME ÉDUCATIONNEL	8
1. Grands buts	8
2. Conception de l'apprentissage	9
3. Liens entre apprentissage et évaluation	11
4. Stratégies complémentaires d'évaluation	13
4.1 Évaluation de la performance	15
4.2 Évaluation authentique	17
4.3 Autoévaluation	18
B - NOTIONS DE QUALITÉ DES RÉSULTATS	19
1. Fondements du modèle psychométrique	21
1.1 Critères de qualité psychométriques	22
2. Éduométrie	24
2.1 Définition de l'éduométrie	24
2.2 Dernières tentatives pour conserver le modèle traditionnel	26
2.3 Questions de terminologie	28
C - CONCEPT DE JUSTESSE DES RÉSULTATS	29
1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées.	31
2. Cohérence de l'évaluation avec l'activité pédagogique	34
3. Transparence de l'activité évaluative	35
4. Absence de désavantage circonstanciel	36

TABLE DES MATIÈRES

iii

page

D - REVUE DE LA DOCUMENTATION	40
1. Démarches d'évaluation	41
2. Lien avec l'activité pédagogique	42
3. Participation des élèves au processus d'évaluation	43
4. Approche méthodologique et crédibilité des données	44
E - RÉSUMÉ DE LA PROBLÉMATIQUE ET QUESTIONS DE RECHERCHE	45
 Chapitre 2 : MÉTHODOLOGIE	50
A - CONTEXTE SOCIAL ET POLITIQUE DE MISE À L'ESSAI	52
B - PARTICIPANTS	56
1. Recrutement	56
2. Description des participants	57
2.1 Christine	57
2.2 Kevin	58
2.3 Marie	58
2.4 Matilde	60
2.5 Marie-Chantal	60
3. Description de l'auteure de la recherche	62
C - COLLECTE DES DONNÉES	63
1. Première question de recherche	65
2. Deuxième question de recherche	65
D - PLAN D'ANALYSE DES DONNÉES	69
E - CRÉDIBILITÉ DES RÉSULTATS	70
1. Orientation des observations	70
2. Respect des participants	71
3. Transparence	73
4. Triangulation et saturation	74

TABLE DES MATIÈRES

iv

page

Chapitre 3 : PREMIÈRE QUESTION DE RECHERCHE	77
A - PRÉSENTATION DES RÉSULTATS	79
1. Résumés des cinq situations d'évaluation documentées	80
1.1 Situation d'évaluation 1	81
1.2 Situation d'évaluation 2	83
1.3 Situation d'évaluation 3	85
1.4 Situation d'évaluation 4	86
1.5 Situation d'évaluation 5	88
2. Éléments marquants des pratiques évaluatives	90
2.1 Attribution de notes chiffrées	91
2.2 Préparation des élèves aux épreuves	93
3. Justesse des résultats	96
3.1 Objet de l'évaluation	97
3.2 Stratégies utilisées	99
3.3 Traces accumulées	101
3.4 Circonstances entourant l'évaluation	102
3.5 Compétence des enseignants	103
B - INTERPRÉTATION DES RÉSULTATS	105
1. Attribution continue de notes chiffrées	105
2. Critères et schèmes de notation	107
3. Conservation de traces	109
4. Évaluation, reflet du type d'apprentissage valorisé	109
5. Évaluation, reflet de la compétence enseignante	110
C - SYNTHÈSE DES RÉSULTATS	113
Chapitre 4 : DEUXIÈME QUESTION DE RECHERCHE	120
A - PRÉSENTATION DES RÉSULTATS	123
1. Entrevue finale avec Christine	123
2. Entrevue finale avec Kevin	125

TABLE DES MATIÈRES

v

	page
3. Entrevue finale avec Marie	127
4. Entrevue finale avec Matilde	129
5. Entrevue finale avec Marie-Chantal	131
B - INTERPRÉTATION DES RÉSULTATS	133
1. Discussion sur la pertinence du concept de justesse des résultats	133
1.1 Utilité	134
1.2 Réalisme	137
1.3 Intégration	140
1.4 Améliorations	142
2. Discussion sur la viabilité du concept de justesse des résultats	144
2.1 Incompatibilités rencontrées	145
2.1.1 Travail en équipe	145
2.1.2 Matières intégrées	147
2.2 Intégrabilité	147
C - SYNTHÈSE DES RÉSULTATS	155
CONCLUSION	159
A - CONCEPT DE JUSTESSE DES RÉSULTATS	164
1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées	164
2. Cohérence avec l'activité pédagogique	165
3. Transparence de l'activité évaluative	165
4. Absence de désavantage circonstanciel	165
5. Viabilité du concept de justesse des résultats	167
B - LIMITES ET PISTES DE RECHERCHES	170
RÉFÉRENCES	174

ANNEXES

page

A - Assentiment du Comité de déontologie de l'Université d'Ottawa	189
B - Guide de la première entrevue	191
C - Synopsis de la première entrevue	193
D - Clé de codification des protocoles verbaux, première entrevue	203
E - Bilan de codification des protocoles verbaux, première entrevue	208
F - Guide de formation des participants	210
G - Plans d'évaluation des situations 6 à 10	268
H - Analyse de la justesse des résultats d'évaluation, situations 6 à 10	277
I - Guide de l'entrevue finale	283

LISTE DES TABLEAUX

page

1. Première composante du concept de justesse des résultats révisée suite à la mise à l'essai	151
2. Deuxième composante du concept de justesse des résultats révisée suite à la mise à l'essai	152
3. Troisième composante du concept de justesse des résultats révisée suite à la mise à l'essai	153
4. Quatrième composante du concept de justesse des résultats révisée suite à la mise à l'essai	154
5. Clé de codification des protocoles verbaux, première entrevue (annexe D)	204-207
6. Bilan de codification des protocoles verbaux, première entrevue (annexe E)	209

LISTE DES FIGURES

page

1. Schématisation des composantes du concept de justesse des résultats.	39 et 122
2. Matériel documentaire accumulé en vue de répondre aux deux questions de recherche	68
3. Répartition des diverses activités de la collecte des données et leur durée respective . . .	78
4.1 Utilité et possibilité de respecter les composantes du concept de justesse des résultats selon Christine	124
4.2 Utilité et possibilité de respecter les composantes du concept de justesse des résultats selon Kevin	126
4.3 Utilité et possibilité de respecter les composantes du concept de justesse des résultats selon Marie	128
4.4 Utilité et possibilité de respecter les composantes du concept de justesse des résultats selon Matilde	130
4.5 Utilité et possibilité de respecter les composantes du concept de justesse des résultats selon Marie-Chantal	132
5. Analyse de la justesse des résultats d'évaluation, situation 6	278
6. Analyse de la justesse des résultats d'évaluation, situation 7	279
7. Analyse de la justesse des résultats d'évaluation, situation 8	280
8. Analyse de la justesse des résultats d'évaluation, situation 9	281
9. Analyse de la justesse des résultats d'évaluation, situation 10	282

LA JUSTESSE DES RÉSULTATS,
CRITÈRE DE QUALITÉ DE L'ÉVALUATION
DES APPRENTISSAGES EN SALLE DE CLASSE

INTRODUCTION

L'enseignement traditionnel, caractérisé par la transmission de connaissances et la sélection des meilleurs élèves* est graduellement disparu. Il a cédé sa place à l'apprentissage dynamique visant la maîtrise d'habiletés complexes transdisciplinaires. Celles-ci sont susceptibles d'amener tous les élèves à «acquérir les connaissances et les habiletés qui assureront leur réussite dans le contexte de la mondialisation de l'économie, tout en leur permettant de mener avec intégrité une vie civique et personnelle satisfaisante» (Ministère de l'Éducation et de la Formation de l'Ontario, 1997b, p. 3). Conséquemment, l'évaluation des apprentissages a évolué, tentant de s'adapter aux nouvelles philosophies du système d'éducation, aux théories d'apprentissage qu'il adopte et aux missions qu'il se donne.

La qualité de l'évaluation menée par l'enseignant dans sa salle de classe a suscité peu d'intérêt jusqu'à maintenant, comparativement aux épreuves à grandes échelles. Ces dernières, préparées par des experts et administrées à grands frais, génèrent des résultats largement diffusés et comparés, plus susceptibles d'affecter la réputation de grandes nations. Malgré ce fait, l'évaluation au quotidien a été suffisamment explorée en

* L'emploi exclusif de la forme masculine dans ce document ne vise qu'à alléger le texte et n'exclut aucunement les femmes.

Amérique du Nord pour qu'on puisse affirmer qu'elle affecte plus le vécu des élèves du fait qu'elle sert de fondement à de nombreuses décisions. Il apparaît aussi que la plupart des enseignants ne reçoivent, au mieux, qu'une formation minimale en évaluation des apprentissages, que celle-ci s'avère la plupart du temps inadéquate pour leurs besoins, et qu'ils ne disposent pratiquement d'aucune ressource pouvant les aider à s'améliorer. Pas étonnant que le *Rapport de la Commission royale d'enquête sur l'éducation* (1995) ait recommandé la mise en application de multiples moyens pour favoriser une évaluation des apprentissages plus juste et plus équitable.

Un modèle théorique est une représentation simplifiée d'une réalité complexe; il sert à communiquer une vision spécifique, organisée selon une logique particulière (Keeves, 1988). Or, aussi paradoxal que cela puisse paraître, l'éducation n'a pas de modèle théorique adéquat de la qualité des résultats d'évaluation des apprentissages en salle de classe. Elle n'a pas de schème auquel se référer pour donner de la rigueur et de la cohérence à la démarche, de manière à arriver à des résultats justes, représentant bien le développement de chaque élève.

Un tel schème faciliterait pourtant la communication entre intervenants et rendrait la formation plus simple et plus efficace. Il permettrait d'uniformiser les exigences et les contextes d'évaluation selon les normes à atteindre; ceci favoriserait donc un traitement plus équitable de la clientèle étudiante. La recherche sur l'évaluation en salle de classe basée sur un modèle de qualité des résultats serait-elle aussi plus efficace si elle s'appuyait sur une vue d'ensemble cohérente d'éléments à étudier et à analyser. Pour le moment, elle est souvent fragmentée et superficielle. Toutes ces raisons soutiennent

l'importance de rechercher un modèle théorique approprié. Jusqu'à maintenant, la documentation ne fait état d'aucune tentative empirique en ce sens.

Des chercheurs tentent d'adapter au contexte de la salle de classe les critères de validité et de fidélité élaborés d'abord en psychologie, pour les «recycler» en principes éducatifs. D'autres favorisent plutôt la définition de concepts nouveaux et de principes propres au domaine, ancrés dans le contexte de l'enseignement-apprentissage en salle de classe. D'après ces derniers, les notions de validité et de fidélité ont été galvaudées dans le contexte de l'éducation, à la faveur de l'apparence de précision, de scientificité et d'objectivité que leur conféraient les calculs statistiques. La confrontation de ces arguments fait ressortir les préoccupations fondamentales concernant la production de résultats d'évaluation de qualité et contribue à l'émergence du concept de justesse des résultats d'évaluation des apprentissages. Ce concept est essentiellement axé sur la nécessité d'arriver à des résultats justes, tant en situation formative que sommative, à des jugements fondés et à des décisions éclairées. Sans prétendre pouvoir éclipser tout autre critère pour juger de la qualité de l'évaluation, le concept de justesse des résultats englobe toutes les étapes de la démarche d'évaluation et la situe par rapport au contexte d'apprentissage. Il souligne la nécessité de décrire les capacités de l'apprenant en termes de forces et de faiblesses, tout en se préoccupant de cohérence entre les expériences d'apprentissage vécues en salle de classe et les situations d'évaluation proposées.

La présente thèse met en évidence les composantes de ce critère de qualité de l'évaluation des apprentissages et leur interrelation, dans le but d'orienter les pratiques des enseignants. Il serait possible d'apprécier la justesse d'un résultat de deux manières

différentes. Soit *a priori*, en considérant toutes les composantes du critère de justesse avant de procéder à une évaluation, s'assurant ainsi de produire les résultats les plus justes possibles, soit *a posteriori*, en utilisant les composantes du critère pour estimer la justesse d'un résultat après coup, en s'interrogeant sur les conditions dans lesquelles l'évaluation a eu lieu. Bien que cette conception de la qualité d'un résultat d'évaluation des apprentissages en salle de classe paraisse appropriée à n'importe quel niveau scolaire, elle est d'abord éprouvée aux niveaux scolaires de 3^e à 8^e année.

Le premier chapitre du présent document décrit d'abord l'apprentissage dans le nouveau paradigme éducationnel et y situe l'évaluation, telle qu'elle devrait être pour soutenir l'intervention pédagogique. Les critères de qualité élaborés en psychométrie – validité et fidélité – sont confrontés aux besoins de la salle de classe. Même si souvent les auteurs prennent position en faveur du recyclage des critères psychométriques ou de la définition de nouveaux critères propres à l'éduométrie, cette discussion fait ressortir les aspects importants de l'évaluation des apprentissages de qualité dans le nouveau paradigme. C'est ainsi qu'émergent les quatre composantes du concept de justesse des résultats. La revue de la documentation démontre ensuite le besoin d'un modèle pour assurer la qualité des résultats en salle de classe mais aussi le besoin de leadership et de formation des enseignants pour son application. Ce chapitre se termine par une synthèse de la problématique discutée et l'énoncé des questions de recherche.

Le chapitre deux expose la méthodologie servant à éprouver la viabilité et la pertinence du critère proposé. Avec la collaboration de cinq enseignants d'écoles élémentaires de l'Ontario, les quatre composantes du concept de justesse de l'évaluation ont été

confrontées au quotidien de la salle de classe. Les données à ce sujet sont colligées dans un **recueil de dix situations d'évaluation témoins**, complétées d'un **journal de bord** et d'entrevues individuelles réalisées à deux reprises avec chacun des participants. Une section de ce chapitre situe le contexte social et politique de cette recherche et de l'année scolaire 1996-1997; elle décrit aussi le profil professionnel des participants, y compris celui de la chercheure.

La présentation et la discussion des résultats fait l'objet de deux chapitres. La première question de recherche occupe le troisième chapitre et la deuxième question de recherche, le quatrième. De multiples extraits tirés des protocoles verbaux et du journal de bord complètent le texte, en vue de donner de la transparence à l'analyse des faits observés et de répondre le plus honnêtement possible aux questions de recherche. Chacun de ces chapitres se termine par une synthèse des données présentées et discutées.

La conclusion de cette recherche résume l'essentiel de la problématique explorée ici et des résultats obtenus. Une seule mise à l'essai n'est pas, bien entendu, suffisante pour démontrer la viabilité du concept; elle constitue tout de même un premier pas vers une conception de la qualité des résultats d'évaluation des apprentissages adaptée à sa nature et aux besoins de la salle de classe. Ce chapitre expose aussi plusieurs pistes de recherches susceptibles de faire progresser l'évaluation en général et plus spécifiquement, celle en salle de classe.

CADRE CONCEPTUEL

La difficulté à prédire ce dont les jeunes d'aujourd'hui ont besoin pour devenir des adultes fonctionnels et intégrés à leur société de demain a entraîné des changements importants dans la mission de l'école (Romberg, 1992). Celle-ci mise désormais sur le développement d'habiletés multidisciplinaires, la résolution de problèmes, la pensée critique, le transfert de ces habiletés à des situations nouvelles, etc. De cette façon, les élèves deviendront des apprenants perpétuels, souples et adaptables aux besoins de la future société (Leder, 1992). Ce nouveau paradigme éducationnel exacerbe certaines incohérences dans le système, notamment au niveau de la qualité des résultats d'évaluation des apprentissages (Archbald, 1991; Chase-Furman, 1994; Gipps, 1994a; Hipps, 1993; Taylor, 1994; Wolf, Bixby, Glenn & Gardner, 1991). Il fait ressortir le besoin de balises pour guider les pratiques, de manière à produire des résultats d'évaluation fiables, utiles, valables et signifiants, à partir d'une variété de stratégies (Dassa, 1988; Traub & Maraun, 1994). Malheureusement, il n'existe pas présentement de modèle conceptuel de ce qui contribue à la qualité des résultats d'évaluation en salle de classe qui soit cohérent avec la nature contemporaine des apprentissages scolaires (Wilson & Rees, 1990). Ces nouveaux besoins d'évaluation font ressortir à quel point les critères élaborés en psychométrie pour juger de la qualité des résultats d'évaluation ne conviennent pas du tout à la réalité scolaire.

Les enseignants sont un pilier de cette nouvelle vision; ils ont pour tâche d'amener l'élève

à s'approprier son apprentissage en le faisant participer activement à son évaluation (Stiggins, 1994). Or, jusqu'à maintenant, la formation des enseignants à l'évaluation des apprentissages n'a pas été très valorisée, ni même vue comme une expertise à développer (Biggs, 1995; Stiggins, 1991a). Même s'il est nécessaire de reconnaître les acquis à des moments précis du développement (évaluation sommative), c'est l'évaluation en vue de faciliter l'apprentissage (évaluation formative) qui doit dorénavant primer dans les écoles primaires et secondaires (Crooks, 1988; Simmons, 1994; Smith, 1992). De plus, les tests papier-crayon ne sont plus suffisants pour dresser un portrait complet d'apprentissages multidimensionnels; il faut d'autres stratégies, variées et complémentaires, adaptées à la nature des apprentissages visés.

Il apparaît fondamental de se familiariser d'abord avec le nouveau paradigme d'éducation et ses besoins en matière d'évaluation; il constitue donc le sujet de la première partie du présent chapitre. La deuxième partie est consacrée à définir les critères de qualité des résultats d'évaluation en psychométrie et ce qui serait cohérent avec le nouveau paradigme éducationnel. L'opposition entre les tenants de l'adaptation des principes traditionnels de mesure au nouveau paradigme de l'éducation et les promoteurs de la définition de concepts nouveaux, propres au domaine, fait ressortir les préoccupations fondamentales liées à la qualité de l'évaluation des apprentissages. Ces préoccupations esquissent les grandes lignes de référence susceptibles de guider l'évaluation sommative ou formative des habiletés complexes, objet de la troisième partie. Le concept de justesse, en quatrième partie, fait ressortir la pertinence d'une évaluation des apprentissages basée sur l'établissement claire des habiletés visées, du choix d'une tâche appropriée, de sa cohérence avec l'activité pédagogique, de sa transparence à l'égard de

l'élève et de l'absence de désavantage circonstanciel. Les deux dernières parties de ce chapitre (D et E) sont consacrées à la revue de la documentation pertinente et à un résumé de la problématique de recherche, suivis de l'énoncé des questions de recherche.

A - NOUVEAU PARADIGME ÉDUCATIONNEL

Un paradigme éducationnel est un ensemble de croyances, de valeurs et de conceptions présentant une vision globale de l'éducation, guidant l'élaboration de théories et suggérant des pratiques appropriées (Sauvé, 1992). C'est une façon de voir le monde qui «oriente l'action des praticiens sans qu'ils doivent recourir à de longues considérations existentielles ou épistémologiques» (DeLandsheere, 1992, p. 220). L'éducation vit présentement une mutation profonde dans ses pensées, ses perceptions et ses valeurs; elle change de paradigme (Brown & Campione, 1994; Mislevy, 1993). Cette remise en question n'est pas soudaine, elle a germé à l'époque post-industrielle. Ce sont toutefois les événements des deux dernières décennies qui ont donné le coup d'envoi à cette ère nouvelle, en particulier la récession économique, la redevabilité du système d'enseignement, la diversité ethnique de la clientèle étudiante et l'expansion des droits individuels et collectifs (Stiggins, 1994).

1. Grands buts

La mauvaise performance du système d'éducation nord-américain, ces dernières années, soulignée par les examens internationaux et le manque de compétitivité de la nouvelle

génération dans l'économie mondiale (Champagne & Newell, 1992; Garcia & Pearson, 1994), a forcé plusieurs constatations. L'enseignement en général ne va pas en s'améliorant (Caouette, 1992) et les habiletés élémentaires de la pensée, comme la mémorisation et les automatismes, principal contenu des programmes scolaires, ne sont d'aucune utilité s'ils ne sont pas appliqués à bon escient. L'abondance de l'information disponible, son évolution rapide et sa facilité d'accès mettent au premier plan la capacité de débrouillardise et la connaissance de sources variées de renseignements pertinents. L'énergie autrefois consacrée à la mémorisation de faits et d'algorithmes doit dorénavant être investie dans le développement d'habiletés de pensée critique, de savoirs transdisciplinaires et de stratégies métacognitives (Redding, 1992). L'apprentissage d'un niveau décent de savoir doit être intégré, extérieurement manifestable et non pas théorique, morcelé et décontextualisé.

2. Conception de l'apprentissage

L'apprentissage est «un processus d'effet plus ou moins durable par lequel des comportements nouveaux sont acquis ou des comportements déjà présents sont modifiés en interaction avec le milieu ou l'environnement» (DeLandsheere, 1992, p. 20). Dans sa conception moderne, l'apprentissage est un concept multidimensionnel complexe (Gipps, 1994a), en constante évolution, pleinement signifiant uniquement dans sa globalité et appréciable selon un idéal à atteindre (Gipps & Murphay, 1994; Hipps, 1993).

Selon les théories récentes, l'apprentissage n'a pas besoin d'être linéaire (comme la hiérarchie établie par Bloom); il est possible de raisonner et de résoudre des problèmes

à n'importe quel âge ou n'importe quelle étape de l'apprentissage. L'existence de différents styles d'apprentissage, de multiples formes d'intelligence et des différences culturelles sont reconnus pour intervenir dans l'apprentissage (Bruer, 1994; Darling-Hammond, 1994; Gardner, 1992). L'équité en éducation passe par l'adaptation de l'enseignement aux caractéristiques personnelles de chacun (Harmon, 1991). La performance est à son meilleur lorsque les buts sont connus, les exigences précisées et les attentes explicites. De plus, la motivation, l'effort et l'estime de soi jouent un rôle déterminant dans l'apprentissage (Tardif, 1994).

Dans la vraie vie, les problèmes sont souvent résolus en groupe, avec la collaboration de personnes expertes. La discussion et l'argumentation deviennent des stratégies d'apprentissage; elles forcent la réflexion et mettent à l'épreuve les schèmes de pensée. Le travail en coopération oblige la réflexion logique et les discussions constructives (Herman, 1992). Cette manière d'aborder l'apprentissage devrait se refléter à la fois sur l'enseignement et sur l'évaluation (Harnish, 1994; Lussier & Turner, 1995; Tardif, 1993) et donner un rôle actif à l'apprenant.

Pour bien jouer son rôle, l'enseignant doit savoir comment se construit la connaissance (Leder, 1992; Tourneur, Noël & Depover, 1989); il lui faut aussi comprendre le raisonnement de l'élève, sa conception des choses (Rieben, 1988) et la source de ses erreurs pour être en mesure de rejoindre sa pensée (Gipps, 1994b; Glaser, Lesgold & Lajoie, 1984; Harlen, 1994). La formation adéquate des enseignants à cet effet influence directement leur capacité à fournir à chacun des élèves l'occasion réelle d'apprendre (Herman & Klein, 1996).

3. Liens entre apprentissage et évaluation

L'évaluation des apprentissages consiste à mener une démarche stimulant l'apparition d'indices observables dans le but d'éclairer un jugement au sujet de son développement (Nitko, 1989; Wolf et al., 1991). Puisque l'apprentissage n'est pas observable directement, il est nécessaire de provoquer, au moyen d'un stimulus quelconque, l'apparition de comportements qui révèlent son existence, afin d'en apprécier la qualité ou l'étendue (Bloom, Madaus & Hasting, 1981; Tousignant & Morissette, 1990). L'apprentissage ne peut être qu'estimé, toujours de manière imprécise et toujours selon le contexte dans lequel il est observé (Gipps & Murphay, 1994). Puisque l'élève apprend constamment, les résultats d'une évaluation donnée deviennent vite périmés (Shepard, 1992a, 1992b).

L'intervention éducative efficace, selon le nouveau paradigme d'éducation, a besoin de l'évaluation pour adapter l'aide, la communication et le renforcement (Abrecht, 1991; Bertrand & Tremblay-Desrochers, 1985; Brossard, 1986; Cardinet, 1985; Dassa & Gaudreau, 1989; Ouellette, 1989; Rowntree, 1987). L'évaluation est étroitement liée au processus pédagogique et rend ce dernier plus dynamique en s'intéressant davantage au processus de pensée et au cheminement qu'aux résultats (Herman, Aschbacher & Winters, 1992; Kulm, 1994; Ministère de l'Éducation du Québec, 1985a; Scallon, 1988a). Sa forme doit être constructive, formative et intégrée à l'apprentissage (Birzée, 1982; Davaud & Cardinet, 1992; Noizet & Caverni, 1978). Elle sert à assurer le succès des apprentissages de tous et chacun (Darling-Hammond, 1994; Dempster, 1992; Gagné & Thouin, 1991; Stiggins, 1991a). D'une part, elle permet de mieux cibler l'enseignement, par exemple d'adapter des tâches d'apprentissage aux intérêts personnels (Harnisch &

Mabry, 1993; Leder, 1992; Lussier & Turner, 1995). D'autre part, elle donne régulièrement le pouls de l'apprenant, indique si ce dernier progresse normalement et si la stratégie est bonne (Baron, 1991a). L'implication active des élèves au processus d'évaluation vise à les responsabiliser dans leur apprentissage et à développer leurs stratégies métacognitives (Cardinet & Laveault, 1996). L'autoévaluation contribue à favoriser la prise en charge personnelle et l'intégration des exigences.

Aujourd'hui, l'école doit s'ajuster à de nouvelles raisons d'être; l'évaluation sommative est nécessaire pour établir des bilans et éclairer des décisions (Auger & Dassa, 1992; Bloom et al., 1981; Harlen, 1994; Shepard, 1993), à des fins de sanction ou de promotion. Mais, il s'agit d'une évaluation ponctuelle consommant beaucoup moins de temps que l'évaluation formative (Lussier & Turner, 1995). Comme le précise Thouin, «il n'y a souvent pas de démarcation nette entre les stratégies pédagogiques et les stratégies d'évaluation ... les examens standardisés et l'évaluation sommative, même s'ils peuvent être utiles de temps à autre, ne constituent pas l'approche privilégiée» (1993, p. 51-52).

Il n'est plus strictement question de savoirs théoriques, de mémorisation et d'application d'algorithmes et l'évaluation doit s'adapter à son nouveau mandat (Gipps, 1994a; Kulm, 1994). Les questionnaires papier-crayons, en particulier ceux à choix de réponses, ne peuvent à eux seuls révéler comment un élève pense, réfléchit et s'adapte aux nouvelles situations (Kulm & Stuessy, 1991; Mislevy, 1993). Apparaît donc le besoin de nouvelles stratégies d'évaluation dites complémentaires, comme les questionnaires, les grilles d'observations, les fiches d'appréciation, le questionnement oral, l'entrevue, les dossiers d'apprentissage, la démonstration ou le journal personnel (Peddie, 1992). Des stratégies

complémentaires aux tests traditionnels sont particulièrement utiles, voire même essentielles; elles facilitent le jugement de l'apprentissage d'habiletés complexes et d'habiletés supérieures de la pensée (Herman et al., 1992; Horowitz, 1995; Linn, Baker & Dunbar, 1991; O'Neil, 1992; Szetela & Nicol, 1992). L'autoévaluation conscientise l'élève à ses forces et à ses faiblesses dans l'apprentissage des habiletés ciblées. Les stratégies complémentaires d'évaluation devraient être plus susceptibles d'aider à juger de la capacité de penser des élèves (Baron, 1991b; Herman, 1992; Wolf, 1993) et à établir un portrait plus complet des apprentissages réalisés. Et «même le rôle réservé à l'évaluateur s'élargit; il ne relève pas seulement de l'enseignant, les apprenants ont eux aussi un rôle à y jouer par le biais de l'autoévaluation et de l'évaluation par les pairs» (Lussier & Turner, 1995, p. 143).

4. Stratégies complémentaires d'évaluation

Un grand nombre d'ouvrages spécialisés qualifient de complémentaire ou d'alternative toute forme d'évaluation qui n'a pas recours à une épreuve papier-crayon avec réponses choisies. Tous ces autres moyens de générer des informations sur l'apprentissage nécessitent une appréciation personnelle des traces laissées par l'apprenant, afin de qualifier ou de quantifier sa maîtrise des habiletés évaluées. Aux États-Unis, le terme «*performance assessment*» désigne toute forme d'évaluation ayant recours à l'exécution d'une tâche plutôt qu'à des questionnaires à choix de réponses (Dunbar, Koretz & Hoover, 1991; Gipps, 1994a). Cette forme d'évaluation est née du besoin de court-circuiter l'enseignement systématique de notions fragmentaires souvent visées par les épreuves nationales à questions objectives (Baron, 1991a). Même si les épreuves à choix de

réponse possèdent certaines qualités (Biggs, 1995), comme l'expliquent Fitzpatrick-Reardon, Scott et Verre (1994), cette nouvelle formule est apparue comme étant réellement capable d'améliorer l'éducation parce que moins facilement corrompible, tout en attirant l'attention sur les vraies préoccupations de l'école.

Les stratégies complémentaires d'évaluation (questions ouvertes, tâches concrètes, enquêtes, expériences, journal de bord, dossier d'apprentissage) ont pour but d'apporter des informations variées, en vue d'améliorer l'enseignement et l'apprentissage (Kulm, 1994). Le recours à une performance ne rend pas nécessairement l'évaluation meilleure (Hambleton & Murphy, 1992; Herman et al., 1992; Wiggins, 1992) mais mieux on connaît l'attitude, les intérêts et l'expérience d'un élève, plus on est en mesure de l'aider. Ces stratégies offrent aussi la possibilité d'observer plus facilement divers niveaux de savoir, peuvent être plus stimulantes, comporter une variété de réponses possibles à partir de diverses stratégies, faire appel en même temps à des notions théoriques et pratiques.

Parmi les nouvelles tendances en évaluation, trois types retiennent l'attention, tant pour leur complexité que leur grande utilité dans l'évaluation des éléments valorisés dans le nouveau paradigme éducationnel. Il s'agit de l'évaluation de la performance, de l'évaluation authentique et de l'autoévaluation. Ces trois appellations comportent aussi, comme il arrive souvent aux termes nouveaux et pas encore standardisés, une part d'ambiguïté. Pour les besoins de la présente recherche, afin d'éliminer la confusion et de faire ressortir aussi leur complexité, il semble approprié de définir sommairement le sens généralement attribué à ces trois stratégies complémentaires d'évaluation.

4.1 Évaluation de la performance

Une performance est la démonstration d'habiletés acquises grâce à l'assimilation de connaissances et d'expérience pour circonscrire et résoudre des problèmes spécifiques dans leur globalité. Elle sert à démontrer une compétence (le savoir-faire en situation, en utilisant des connaissances acquises et des méthodes) par une activité concrète facilement observable (Winfield, 1995). Elle fait appel à des comportements de l'ordre du savoir, du savoir-faire et du savoir-être (MÉQ, 1985b), sans égard à une discipline particulière (Jaeger, 1989). Certains auteurs recourent au terme «performance complexe» pour désigner la démonstration d'une compétence exigeant de multiples savoirs interreliés et intégrés (Gipps, 1994a; Mislevy, 1993). «En ressemblant davantage aux situations de vie réelle, ces situations sont souvent plus motivantes et plus enrichissantes pour les élèves» (Herman et al., 1992, p. 9).

En optant pour la performance comme indicateur des compétences développées lors des évaluations à grande échelle, il est ainsi devenu difficile pour l'élève de régurgiter un savoir peu intégré pour l'adapter à une situation réaliste, nouvelle et complexe. L'évaluation externe s'est ainsi rapprochée du curriculum et de l'apprentissage (Haertel, 1992; Kirst, 1991; Resnick, 1994) et est devenue promoteur d'un enseignement davantage axé sur l'atteinte des buts de l'éducation (Baron, 1991a; Dunbar et al., 1991). Même si le phénomène du «*Measurement Driven Instruction*» est moins important au Canada, le mouvement vers des méthodes d'évaluation plus authentiques gagne en importance, et ce, partout en Amérique du Nord.

Cette manière d'évaluer nécessite un degré minimal d'inférence, augmentant la qualité

des résultats qui s'en suivent (Baron 1991b; Meyer, 1992; Resnick, 1994; Taylor, 1994; Wiggins, 1989, 1992), mais comme l'observation qu'elle nécessite consomme généralement beaucoup de temps (Mehrens, 1991; Romberg, 1992; Hardy, 1996), elle réduit le nombre d'activités d'évaluation possibles (Herman et al., 1992; Moss, 1994; Satterly, 1994). Il faut donc la réserver pour ce qui ne peut être évalué autrement (Biggs, 1995; Peddie, 1992). De plus, sa pertinence doit être justifiée par la spécification des habiletés qu'elle cible et la manière dont chacune d'elles est susceptible de se manifester, sans quoi la qualité du résultat pourrait grandement en souffrir (McComas, 1989; Wiley & Haertel, 1996).

Les résultats exprimés sous forme de note chiffrée sont loin d'être adéquats dans ce genre d'évaluation parce qu'ils sont difficilement réinvestissables. L'approche qualitative, plus descriptive, produit des résultats plus utiles pour améliorer les apprentissages (Biggs, 1995; Scallon, 1988b). Ainsi, si une note symbole réfère à une description, elle communique à l'élève l'appréciation de sa performance et la manière de l'améliorer. Kulm (1994) explique :

L'utilisation de notes chiffrées peut à elle seule anéantir le bénéfice d'un schème de notation ou d'une situation d'évaluation complémentaire. Un score global capte l'attention au détriment des critères ou des normes établis et détourne l'attention des processus et des habiletés à développer. Par exemple, un profil comportant trois notes – l'une attribuée à la compréhension du problème, la deuxième à la solution proposée et la troisième à l'explication du problème – est beaucoup plus signifiant que

n'importe quelle autre format de notation. L'élève voit exactement et rapidement sur quel aspect de la résolution de problème il doit mettre davantage d'effort la prochaine fois. (p. 99, traduction libre)

La difficulté de cette forme d'évaluation semble résider dans la détermination de tâches équivalentes (Baxter & Shavelson, 1994; Shavelson, Baxter & Gao, 1993; Shavelson, Baxter & Pine, 1991) et dans la définition des critères et des indicateurs devant guider les observations, afin d'apprécier les habiletés dans leur globalité, mais aussi dans leur interaction (Linn, 1989; Taylor, 1994). Contrairement à l'évaluation d'une habileté à partir de plusieurs items homogènes de type «réponse choisie», l'évaluation de la performance peut faire appel à une seule tâche pour évaluer de multiples habiletés (Wiley & Haertel, 1996).

4.2 Évaluation authentique

Le terme «authentique» ou «écologique» est encore plus ambigu et galvaudé que le terme «performance». Il s'agirait, en fait, d'une manière spécifique d'évaluer la performance (Biggs, 1995; Elliot, 1991; Gipps, 1994a). L'évaluation authentique exigerait la démonstration publique de la maîtrise d'habiletés, à partir d'une tâche et de critères tirés du domaine d'expertise en question (Wiggins, 1989). Ces tâches de la «vraie vie» favoriseraient l'engagement intégral de l'élève et permettraient d'apprécier ses capacités réelles (Archbald, 1991; Baron, 1991a; Bateson, 1994; Moss, 1992); on éviterait ainsi de se tromper entre la non-maîtrise et le manque d'intérêt (Linn, 1992; Stone-Wiske, 1994; Traub, 1990). Pour certains experts, ce serait le seul type d'évaluation acceptable (Wiggins, 1993); pour d'autres il servirait surtout à certifier la maîtrise de compétences

minimales dans l'apprentissage d'un métier (Jaeger, 1989).

L'authenticité remet en cause la limite des ressources permises durant une tâche d'évaluation, notamment l'aide de l'enseignant; dans la «vraie vie», il est tout aussi important de savoir trouver les personnes capables d'aider à résoudre un problème que d'exposer le problème en question (Archbald, 1991; Wiggins, 1992). La pertinence des limites de temps et l'individualité des travaux sont aussi remis en question (Wiggins, 1992; Thouin, 1993), tout comme la rétroaction formative et la disponibilité d'outils d'autoévaluation (Taylor, 1994; Lussier & Turner, 1995).

4.3 Autoévaluation

Même si elle recourt souvent à des outils de type papier-crayon avec réponses choisies, «l'autoévaluation fait partie intégrante de l'acte d'apprentissage» (Lussier & Turner, 1995, p. 68). En forçant l'analyse critique d'une tâche au moyen d'une procédure structurée, l'autoévaluation amène l'élève à réfléchir sur sa manière de penser et d'apprendre. Il peut ainsi reconnaître ses faiblesses et ses forces et modifier ou améliorer sa démarche en conséquence. L'autoévaluation est aussi un moyen d'amener l'élève à se responsabiliser; en étant habilité à s'évaluer lui-même, il devient plus autonome et détient plus de contrôle sur ce qui lui arrive. Les rares expériences structurées en ce sens se sont avérées très fructueuses (Arter, Spandel, Culham & Pollard, 1994).

Pour arriver à s'autoévaluer, l'élève doit connaître et avoir intégré les critères d'évaluation, les indicateurs à observer et les normes établies (Mitchell, 1992; Zahorik, 1994); il peut ensuite s'en servir pour développer précisément les habiletés visées (Arter et al., 1994).

L'évaluation devient ainsi plus transparente et réduit considérablement la tension qui accompagne généralement les épreuves.

Il est incontestable que ces trois stratégies d'évaluation peuvent compléter d'autres types d'évaluation, mais elles ne sauraient suffire à la tâche. Toutefois, qu'il s'agisse de l'une ou l'autre de ces stratégies complémentaires d'évaluation, la préoccupation de générer des indices de qualité demeure vive.

B - NOTIONS DE QUALITÉ DES RÉSULTATS

La qualité d'un résultat d'évaluation tient essentiellement à la confiance qu'il inspire en vue d'une fonction précise (Harlen, 1994). Dans l'ancien paradigme éducationnel axé sur la transmission de connaissances factuelles ou déclaratives, l'évaluation servait surtout à la sélection de candidats et à la certification. La qualité des résultats d'évaluation, alors établie selon leur validité et leur fidélité, établissait la confiance qu'ils pouvaient inspirer pour ces deux fonctions précises. Dans le nouveau paradigme d'éducation, la qualité d'un résultat doit être définie selon la nature et l'usage réservé à l'évaluation, en plus d'être cohérent avec la mission actuelle de l'école et sa manière de percevoir l'apprentissage. Ainsi, la définition de la qualité doit convenir aussi à l'apprentissage d'habiletés multidimensionnelles et complexes (Gipps, 1994a; Hipps, 1993). Rappelons que ces savoirs ne peuvent qu'être estimés, toujours avec une certaine marge d'erreur et toujours

selon le contexte dans lequel ils sont observés (Gipps & Murphay, 1994; Wolf et al., 1991). La quantification des apprentissages n'est pas réaliste ni souhaitable. L'important est de pouvoir distinguer ce qui est maîtrisé de ce qui ne l'est pas et de déterminer si le niveau de maîtrise atteint les seuils fixés lorsque ceux-ci existent.

Qu'il s'agisse d'évaluation formative ou sommative, un résultat d'évaluation de qualité doit être sensé, fiable, utile et signifiant. Il doit soutenir les efforts quotidiens en reconnaissant les acquis, mais surtout, en repérant rapidement les erreurs de parcours et en aidant à poser un regard critique sur les réalisations de chacun. Il doit s'exprimer en termes de forces et de faiblesses, selon un résultat idéal à atteindre, toujours dans le but de réinvestir ces informations dans le processus d'apprentissage et de le rendre encore plus efficace. C'est ce que doit soutenir et promouvoir un modèle d'évaluation des apprentissages cohérent avec le nouveau paradigme de l'éducation. Un résultat de qualité contribue à assurer le succès de l'élève en renseignant sur sa manière d'apprendre et en fournissant des informations précises sur ses forces et ses faiblesses. Une évaluation fournissant des résultats de qualité mène à des jugements éclairés et à des décisions adéquates quant au cheminement préférable de chaque élève.

Malheureusement, l'école n'a pas et n'a jamais eu son propre modèle conceptuel de la qualité des résultats d'évaluation en salle de classe (Auger & Dassa, 1992; Biggs, 1995). Jusqu'à maintenant elle s'est accommodé tant bien que mal du modèle d'évaluation développé en psychologie pour apprécier la qualité des résultats d'évaluation. Cependant, l'arrivée des programmes cadres modernes a fait ressortir, plus que jamais, toute l'incongruité du modèle psychométrique pour assurer des critères de qualités

édumétriques (Chase-Furman, 1994). Pour mieux le comprendre, voici un bref aperçu des développements de la psychométrie, de ses objectifs et de ses préoccupations.

1. Fondements du modèle psychométrique

Les réalisations de Binet-Simon et Terman pour mesurer l'intelligence humaine sont à la base de la psychométrie (Trahan & Dassa, 1978). Parce que l'éducation compose avec des attributs non observables, que l'apprentissage est associé de près à l'intelligence et que la préoccupation de l'école traditionnelle était principalement de transmettre des connaissances (Redding, 1992), l'évaluation s'est accommodée assez bien du modèle psychométrique. Il a permis, en comparant les individus entre eux, de repérer les candidats les plus habiles à mémoriser les contenus et à appliquer les algorithmes. L'enseignement personnalisé n'étant pas encore à ce moment-là en vogue, les habiletés personnelles des élèves se reflétaient dans leur capacité à assimiler plus ou moins bien la matière. Ce modèle d'évaluation, bien qu'établi pour des traits ou des attributs personnels, était relativement cohérent avec les anciennes préoccupations de l'école. Le concept traditionnel de «connaissances» était unidimensionnel parce que lié de très près à la capacité de mémoriser et que cette capacité, plutôt stable chez un même individu, était répartie de manière assez variée dans la population en général. L'évaluation des connaissances selon ce modèle respectait assez bien les deux postulats de la mesure des attributs psychologiques : le trait ou la qualité mesuré est unidimensionnel et stable et est réparti dans la population en suivant une courbe de Gauss.

Pour mesurer un trait psychologique et supposer sa distribution normale dans une

population donnée, la mesure doit se concentrer sur ce qui distingue les individus entre eux, dans le but de les comparer et d'en faire ressortir les différences. En agissant comme un microscope sur les légères variations de l'attribut mesuré, il devient possible de le quantifier et de répartir cette qualité selon une courbe de distribution normale. Les individus mesurés peuvent alors être classés selon la quantité qu'ils possèdent de la qualité d'intérêt (Berlak, 1992). Pour repérer les candidats qui se démarquent des autres par rapport à un attribut donné, cette approche convient très bien, mais pour évaluer les apprentissages de manière à produire des résultats cohérents avec le paradigme éducationnel actuel, il en est tout autrement (Auger & Dassa, 1992; Biggs, 1995; Bloom, Hastings & Madaus, 1971; Lane & Glaser, 1995; Rheinberg, 1983).

1.1 Critères de qualité psychométriques

En psychologie, la capacité d'un instrument à bien quantifier un trait mesuré s'exprime par sa validité et sa fidélité. De ces deux indicateurs de qualité, la validité est l'indice de la capacité d'un outil à mesurer réellement ce qu'il prétend mesurer, selon l'utilisation que l'on veut en faire (Popham, 1990); c'est une qualité rattachée aux résultats fournis par une évaluation (Cronbach, 1969; Messick, 1989b; Moss, 1992), selon le lien théorique établi entre le sens accordé aux résultats et l'intention qui a motivé leur collecte (Moss, 1995). La fidélité, pour sa part, est la constance avec laquelle une personne obtiendrait le même résultat dans des conditions semblables.

Il existe des principes d'élaboration d'un instrument de mesure susceptibles de favoriser des résultats fidèles et valides (méthode *a priori*); cependant, la vérification se fait toujours à partir d'une application de l'outil en question (méthode *a posteriori*). La validité et la

fidélité sont en fait des tests d'hypothèse (Satterly, 1994) pour lesquels on applique un modèle mathématique précis à des données recueillies auprès d'un groupe de sujets. La confirmation de la validité et de la fidélité des résultats recueillis avec un instrument de mesure ne s'obtient donc qu'à partir d'une application réelle de l'instrument en question. Pour qu'il puisse maintenir son niveau de qualité, il est essentiel que l'outil en question soit toujours administré de manière uniforme et à des sujets en tous points identiques à la population considérée au départ (Messick, 1989a).

Le modèle psychométrique nécessite une variance maximale des scores, et un assez grand nombre d'items indépendants les uns des autres (Lussier & Turner, 1995). Même la théorie de réponse aux items, y compris le modèle de Rasch, relève de l'analyse factorielle, laquelle a pour postulat l'unidimensionnalité de l'élément mesuré (Gipps, 1994a). Selon Taylor, «cette notion de qualité tient essentiellement aux différences individuelles et à notre habileté à bien les mesurer. Tout se joue sur la longueur de l'épreuve, le type d'items utilisés, la distribution des scores, les conditions d'administration standardisées et les modèles mathématiques établis pour confirmer leur validité et leur fidélité.» (1994, p. 239, traduction libre) Si tous les sujets répondent de manière semblable, les items sont dits sans valeur, autant du point de vue de leur fidélité que de leur validité. De plus, l'évaluation des apprentissages, en particulier les apprentissages complexes et l'évaluation de performances, a recours à un très petit nombre d'items; elle ne peut donc pas escompter une distribution variée des résultats, comme le modèle psychométrique l'exige. Tous ces arguments suggèrent d'aller plutôt vers des approches plus qualitatives afin d'établir la qualité des résultats (Gipps, 1994a, p.151).

2. Éduométrie

En 1963, Glaser a été le premier à soutenir publiquement que la répartition normale des caractéristiques humaines et la nécessité d'amplifier les différences entre les individus sont tout à fait inutiles pour reconnaître les acquis, dépister les difficultés et faciliter l'apprentissage (Rheinberg, 1983; Lussier & Turner, 1995; Trahan et Dassa, 1978). L'éducation, même si elle a beaucoup de choses communes avec la psychologie, s'intéresse à mesurer des éléments différents; elle ne peut donc pas s'accommoder de sa manière d'établir la qualité d'un résultat d'évaluation.

2.1 Définition de l'éduométrie

Pour évaluer les apprentissages à partir de critères absolus, comme le propose Glaser, il faut examiner chaque apprenant par rapport à un résultat visé, indépendamment des autres de son groupe-classe. Ceci fait ressortir les changements ou les progrès à un moment précis du cheminement scolaire (Gipps, 1994a). Ainsi, l'évaluation génère des informations utiles à la régulation des apprentissages, fait ressortir les forces et les faiblesses, les éléments à retravailler et reconnaît les acquis. Le schème d'interprétation critérielle constitue donc le fondement de l'évaluation des apprentissages en salle de classe. Bien que l'approche soit différente de celle développée en psychologie, il demeure toutefois important que les résultats d'évaluation soient fiables, vrais et utiles, pour qu'ils mènent à des jugements éclairés et des décisions appropriées.

Un document marquant dans cette affirmation de l'éduométrie est sans conteste un article de Carver (1974). Il y explique les différences fondamentales entre l'éduométrie et la

psychométrie, laquelle est née à l'époque darwinienne, période particulièrement axée sur les différences individuelles. L'évaluation selon ce modèle visait principalement à rehausser les différences entre les individus; les items qui discriminent bien présentent une variance des résultats élevée, généralement associée à un bon degré de validité et de fidélité. L'édu-métrie s'intéresse plutôt aux changements chez un même individu, par rapport à un critère ou un objectif visé; les bons items sont ceux qui font le mieux ressortir l'apprentissage réalisé suite à l'enseignement; ils ne dépendent nullement de la variance des réponses. Voici des extraits de cet article :

Peu de gens saisissent qu'une épreuve peut à la fois être de très grande qualité, édu-métriquement parlant, tout en étant très mauvaise, psychométriquement (p. 153). . . . Une épreuve est psychométriquement fidèle si elle fait ressortir les mêmes différences entre les individus en étant appliquée à deux reprises, alors qu'elle est édu-métriquement fidèle si elle fait ressortir les mêmes différences chez un même individu (p. 514). ... Parce que l'apprentissage implique nécessairement un changement, il apparaît évident qu'une évaluation basée sur le changement chez une même personne est beaucoup plus utile que celle basée sur les différences individuelles, comme le fait la psychométrie (p. 517).

À remarquer qu'en 1974, il n'était pas encore question d'incompatibilité entre l'unidimensionnalité de la psychométrie, de la complexité des apprentissages et des savoirs transdisciplinaires. À cette époque, aux États-Unis, c'était aussi l'apogée des épreuves à réponses choisies menées à grande échelle, pour lesquelles les questions

de validité et de fidélité des résultats revêtaient une extrême importance; d'énormes ressources humaines et financières y étaient investies et les comparaisons qui s'en suivaient alimentaient la compétitivité des institutions. Il était donc primordial de pouvoir généraliser les résultats, ce qui n'aurait pu se faire sans validité ni fidélité des résultats.

2.2 Dernières tentatives pour conserver le modèle traditionnel

Malgré la logique des arguments de Scriven et de Carver, des chercheurs se sont particulièrement dévoués à adapter les critères psychométriques à l'évaluation des apprentissages; Messick est l'un de ceux-ci. Lorsqu'en 1985, l'*American Psychological Association* a revu et corrigé ses critères de qualité comme elle le fait à intervalles réguliers, Messick a tenté à nouveau de soutenir leur applicabilité au monde de l'éducation. Dans ses dernières publications, il défend l'unicité des concepts de validité de contenu, de construit et de critère. Il y discute du type d'inférence possible, selon chacun de ces types de validité et les évidences à rechercher pour les appuyer. Malgré son enthousiasme, sa fidélité aux concepts traditionnels de psychométrie ne peut faire oublier toutes les incongruités qui demeurent. Il parle encore de «trait» ou de «construit» mesuré (Messick, 1992) et son utilisation du singulier témoigne clairement de l'unidimensionnalité de l'objet mesuré.

Messick ne fait aucune allusion au contexte scolaire : il ne se préoccupe pas de l'évolution constante de l'apprentissage, ni de la congruence nécessaire entre les expériences d'apprentissage et leur évaluation, ni de la connaissance des critères d'évaluation par les évalués, ni de l'utilité des résultats générés pour assurer la réussite des élèves. Il ne manifeste aucune considération à l'égard des arguments de Scriven ni de ceux de Carver,

ni de ceux soulevés dans les nombreux débats qui ont suivi, ni pour les efforts plus récents de chercheurs pourtant respectés dans le domaine de l'éducation, comme Linn, Baker, Dunbar, Frederiksen et Collins.

D'autres chercheurs, comme Hipps, entrevoient plus d'avenir dans une conception purement éduométrique de l'évaluation : «L'intérêt actuel pour une évaluation plus authentique pourrait résulter en de meilleurs moyens de nourrir le processus pédagogique. Les stratégies complémentaires d'évaluation suscitent une préoccupation justifiée de la qualité des résultats. Cependant, ce n'est pas en s'appuyant sur les critères traditionnels de validité, fidélité, généralisabilité et objectivité qu'on y arrivera; il faut reconnaître qu'ils s'appuient sur des postulats complètement différents.» (1993, p. 15, traduction libre)

Avec le temps, les débats suscités par Scriven et Carver ont porté fruits. L'évaluation critérielle s'est enrichie de normes de rendement et de seuils de performance minimale. La tendance s'oriente maintenant vers la description qualitative de la performance ou du rendement plutôt que la quantification. Dans cette autre conception de l'évaluation, la généralisabilité des résultats doit céder sa place à la pertinence de jugements énoncés selon des critères appropriés et des indicateurs précis (Frederiksen & Collins, 1989; Gipps, 1994a; Hipps, 1993; Moss, 1996; Sugrue, 1995; Traub, 1990). Il faut désormais tenir compte du contexte d'exécution d'une tâche, parce qu'il affecte de manière importante la performance de l'élève (Bateson, 1994; Fraser, 1986; Gordon, Engelhard, Gabrielson & Bernknopf, 1996; Moss, 1995; Shavelson et al., 1991; Tardif, 1994). De plus, il est difficile de standardiser le contexte d'exécution d'une tâche (Linn & Burton, 1994); les situations d'évaluation, limitées en nombre mais générant des informations

complémentaires, réduisent l'assurance des interprétations qui s'en dégagent (Airasian, 1991a; Moss, 1994, 1996; Taylor, 1994).

2.3 Questions de terminologie

La nature de l'apprentissage, le milieu collectif de la salle de classe, la complexité des savoirs et leur évolution constante ainsi que les décisions qui sont à l'origine des démarches d'évaluation imposent de nombreuses et sérieuses modifications aux critères traditionnels de qualité pour les transformer en critères éducatifs (Auger & Dassa, 1992; Gipps & Murphay, 1994; Herman et al., 1992). Ceci peut entraîner beaucoup d'ambiguïté (Peddie, 1992). L'essentiel étant de générer des résultats représentant bien l'apprentissage réalisé, des résultats significatifs, fiables, vrais et pertinents (Gipps, 1994b; Harlen, 1994; Herman, Aschbacher & Winters, 1992; Moss, 1995; Peddie, 1992; Shepard, 1992a), il importe peu que la façon d'y arriver s'appelle un principe de validité, de fidélité ou d'équité. L'important c'est de respecter et de promouvoir l'apprentissage comme étant un concept multidimensionnel complexe, en constante évolution, ne pouvant qu'être qu'estimé, pleinement signifiant uniquement dans sa globalité et selon le contexte dans lequel il est observé, en comparaison avec un idéal à atteindre (Gipps & Murphay, 1994; Hipps, 1993).

Certains auteurs croient qu'il ne convient pas de conserver une terminologie psychométrique pour véhiculer des principes tout à fait différents. La validité et la fidélité appartiennent à un paradigme différent de l'éducation et sont basées sur des postulats différents (Hipps, 1993). Utiliser des termes déjà bien établis, tout en en modifiant le sens ne pourrait que créer de la confusion (Moss, 1995; Shepard, 1993). Puisque les notions

de validité et de fidélité appartiennent à un autre domaine et sont encore d'usage, il est préférable d'en choisir d'autres (Gipps, 1994a). Ces derniers arguments motivent le choix d'un terme nouveau, à l'image de l'évaluation qu'il promeut. La prochaine section en discute.

C - CONCEPT DE JUSTESSE DES RÉSULTATS

Malgré le débat entourant la terminologie employée, l'essence même de la qualité des résultats d'évaluation demeure. Il s'agit de la justesse avec laquelle un résultat d'évaluation situe un apprenant dans son cheminement, détermine ses forces et ses faiblesses, ce qu'il maîtrise ou ne maîtrise pas, peu importe qu'il s'agisse d'une habileté complexe ou pas, peu importe la stratégie d'évaluation utilisée. C'est cette essence que doit traduire un cadre conceptuel appuyant la qualité des résultats d'évaluation en salle de classe et son but doit être avant tout de réguler les apprentissages ou de reconnaître les acquis (Frederiksen & Collins, 1989; Hipps, 1993). Il doit appuyer la démarche d'évaluation de manière à maximiser le degré de confiance dans les résultats d'évaluation en les rendant utiles, sensés, représentatifs, fiables, signifiants et entachés du moins d'erreurs possible (Auger & Dassa, 1992; Dunbar et al., 1991; Gipps & Murphay, 1994; Herman, et al., 1992; Hipps, 1993; Moss, 1995; Peddie, 1992; Shepard, 1993). Opter pour un terme nouveau, simple, signifiant pour les enseignants et approprié à la réalité de la salle de classe pourrait faciliter sa mise en application et marquer la transition vers des

principes propres à l'éducation. Le terme «justesse» des résultats pourrait répondre à ces critères.

Le terme *justesse* a un fondement philosophique. Par opposition à la notion de *justice*, laquelle fait abstraction de toutes différences individuelles et réclame le traitement impartial de tous, la *justesse* ou l'équité fait appel au jugement moral pour s'ajuster aux particularités des situations. Dans le domaine de l'évaluation, la notion de justice pourrait être associée aux concepts de validité et de fidélité, critères de qualité des résultats obtenus par mesures standardisées, appliquées uniformément. La justesse des résultats, quant à elle, découlerait plutôt du respect des particularités, des conditions, du contexte et des besoins personnels des évalués pour arriver à des résultats situant correctement chacun dans son développement.

Il existe plusieurs points critiques pour produire des résultats d'évaluation justes. Ils sont souvent interdépendants, rendant difficile d'intervenir sur un seul d'entre eux à la fois (Lane, 1993; Leder, 1992; Stiggins, 1991b). Mais aux fins de discussion, ils sont ici regroupés en quatre composantes :

- pertinence de la tâche d'évaluation par rapport aux habiletés visées
- cohérence de celle-ci avec l'activité pédagogique
- transparence de l'activité évaluative
- absence de désavantage circonstanciel

Ces composantes sont retenues pour leur contribution à l'assurance de résultats

d'évaluation justes en salle de classe, peu importe leur domaine d'origine. Voici une brève description de chacun d'eux.

1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées

L'évaluation, tout comme l'enseignement, nécessite d'établir les habiletés visées par les activités pédagogiques. Que ces habiletés soient simples, comme celle de posséder des connaissances déclaratives ou théoriques, ou plus complexes comme de raisonner, de produire une oeuvre ou de réaliser une tâche à partir de plusieurs habiletés interreliées, il est essentiel de les reconnaître et d'en établir les signes distinctifs. Ceci permet de ne pas les confondre avec d'autres habiletés semblables et d'en établir les dimensions importantes ou les critères d'évaluation (Elliot, 1991; Glaser & Silver, 1994; Taylor, 1994). Cet aspect de la première composante du concept de justesse des résultats s'apparente à l'étape «intention» de la démarche d'évaluation, laquelle devrait donner tout son sens à la procédure d'évaluation engagée (Lussier, 1992).

La tâche d'évaluation est le stimulus susceptible de faire émerger des comportements révélant l'état des apprentissages ciblés (Lane, 1993; Linn et al., 1991; Messik, 1989b; Moss, 1995; Popham, 1994). Sa longueur et son degré de difficulté sont ajustés aux caractéristiques des élèves. Qu'il s'agisse de choisir des réponses, d'en composer, d'effectuer une performance ou d'échanger verbalement, la tâche choisie doit mettre les habiletés ciblées en valeur, permettre de les apprécier, tout en faisant ressortir leur interaction s'il y a lieu (Frederiksen & Collins, 1989; Meyer, 1992; Mislevy, 1993). La stratégie adoptée doit être familière à l'élève mais la tâche suffisamment nouvelle pour

informer de manière fiable au sujet des habiletés ciblées. De plus, des résultats d'évaluation justes sont tirés d'une tâche signifiante pour l'évalué, compatible avec ses intérêts, justifiant pour lui d'investir le temps et l'énergie qu'elle réclame et définie clairement dans un langage qui lui est compréhensible (Gipps, 1994a). Enfin, il va sans dire que la formulation de cette tâche ne comporte pas d'erreur technique.

Le choix de la tâche permet ensuite d'établir les indicateurs ou les signes perceptibles permettant de situer les apprentissages par rapport à chacun des critères établis précédemment (Linn, 1992; Shepard, 1992b). Alors que les critères sont communs à plusieurs tâches pouvant servir à l'évaluation des mêmes habiletés, les indicateurs sont les signes perceptibles propres à la tâche assignée; ils varient donc d'une tâche à l'autre. L'agencement optimal des habiletés visées et le type de tâche choisi pour les faire ressortir s'avère crucial (Stiggins, 1994; Taylor, 1994); c'est le fondement de l'obtention de résultats d'évaluation justes. Des lacunes à ce niveau-ci de la démarche d'évaluation mettraient en péril la justesse des résultats, peu importe la rigueur appliquée aux autres composantes du concept.

La détermination des critères et des indicateurs est aussi essentielle à la mise au point d'activités d'apprentissage qu'à l'établissement des forces et des faiblesses lors de l'évaluation (Frederiksen & Collins, 1989; Moss, 1995; Popham, 1994; Sugrue, 1995). Des critères bien choisis et bien définis aident à juger une performance de manière fidèle, valide et juste et communiquent explicitement ce qui est attendu (Herman et al., 1992). Ils sont propres à l'activité en milieu réel et applicables à une variété de tâches semblables (Quellmalz, 1991). Ils peuvent s'ajuster au niveau de développement des

évalués et à leur cheminement scolaire et leur nombre varie selon la complexité de la performance (Herman et al., 1992). Même si une habileté complexe est momentanément fractionnée en habiletés plus simples pour faciliter son évaluation, celles-ci demeurent liées entre elles. Prises isolément, elles n'ont aucune valeur pour juger de l'atteinte du résultat visé. Il n'est toutefois pas obligatoire que l'évaluation tienne compte à chaque fois de tous ces critères, en particulier en tout début d'apprentissage.

Dans une situation d'évaluation formative, les critères choisis pour juger la performance sont spécifiques à divers aspects des apprentissages visés; ils servent à repérer les erreurs de parcours, à mieux saisir la manière de penser de l'élève, puisque le but poursuivi est d'encourager ses efforts, de mieux cibler les activités d'apprentissage subséquentes ou de consolider les acquis. Dans le cas d'une situation d'évaluation sommative, les critères sont plus englobants, afin de poser un regard critique sur les acquis, en comparaison avec les résultats visés et les normes établies.

Le recours à des situations dites authentiques réduit au minimum le degré d'inférence au sujet des capacités mentales évaluées; elles peuvent donc favoriser des résultats de meilleure qualité (Baron 1991b; Taylor, 1994; Wiggins, 1989). L'observation directe d'un élève au travail peut, dans certains cas, fournir des renseignements inégalables mais, par ailleurs, les situations authentiques sont difficiles à obtenir en salle de classe et, aussi, plus une tâche d'évaluation consomme de temps, plus le nombre de situations différentes en sera limité (Herman, et al., 1992; Moss, 1994).

2. Cohérence de l'évaluation avec l'activité pédagogique

Pour être intégrée à l'apprentissage, l'activité d'évaluation proposée à l'apprenant doit s'insérer convenablement dans l'activité pédagogique (Lane, 1993; Satterly, 1994) et arriver au moment opportun, c'est à dire après avoir eu l'occasion de développer suffisamment les apprentissages en question (Zahorik, 1994). Même si toute situation d'évaluation doit présenter un aspect nouveau, il est essentiel que les apprenants soient familiers avec le genre de situation d'évaluation proposée (Gardner, 1992; Jonnaert & Laveault, 1994), bénéficient d'une mise en situation convenable et puissent demander des précisions s'ils ne comprennent pas bien ce qu'ils doivent faire. Non seulement doivent-ils connaître les critères qui guident l'évaluation, mais avoir déjà eu l'occasion de les appliquer (Lussier & Turner, 1995; Shepard, 1989). De plus, l'évaluation étant habituellement menée dans le but de servir l'apprentissage ou de reconnaître les acquis, les résultats communiquent à l'apprenant une appréciation de sa performance. Les résultats réfèrent donc à des forces et à des faiblesses, avec un degré de spécificité ajusté au type de réinvestissement envisagé : plus détaillés et plus spécifiques pour un réinvestissement immédiat, plus globaux et plus généraux pour un réinvestissement à plus long terme. Le droit à une ré-évaluation permet à un élève de démontrer un progrès éventuel suite à un résultat qu'il juge insatisfaisant.

Seules les activités d'apprentissage vécues précédemment par les élèves permettent de déterminer avec certitude à quelles habiletés la tâche d'évaluation fait réellement référence (Herman et al., 1992). Comme l'expliquent Linn, Baker et Dunbar (1991), une tâche peut aussi bien mettre en valeur des habiletés de mémorisation que des habiletés

complexes, dépendant du degré de nouveauté de la tâche et des expériences vécues antérieurement. Il est primordial d'établir clairement les habiletés impliquées dans la tâche assignée et les indices qui révéleront leur manifestation. Une épreuve serait systématiquement non valide si elle pouvait être réussie avec des stratégies d'apprentissage élémentaires, comme des automatismes, sans en comprendre les principes (Frederiksen & Collins, 1989; Lane, 1993). C'est «le danger de s'appuyer uniquement sur des objectifs définis au moyen de comportements observables» (Thouin, 1993, p. 50).

3. Transparence de l'activité évaluative

La transparence d'une démarche d'évaluation se manifeste par la participation, la discussion et l'appropriation par l'élève des divers aspects de l'évaluation de ses apprentissages, dans le but de s'assurer qu'il sait ce qui est attendu de lui (Frederikson & Collins, 1989). Selon son niveau de maturité, l'élève peut aider à établir les critères d'évaluation, les indicateurs de performance et expérimenter leur application à ses activités d'apprentissage (Quellmalz, 1991; Smith, 1990); l'autoévaluation et l'évaluation par les pairs à des fins formatives sont particulièrement propices à cette prise en charge (Abrecht, 1991; Doyon & Juneau, 1991). L'apprentissage graduel d'une plus grande autonomie le responsabilise et agit indirectement sur sa motivation. Présenter aux élèves et commenter des exemples concrets de travaux de diverses qualités peuvent aussi aider à communiquer ces attentes et favoriser le transfert de ces critères à d'autres activités d'apprentissage (Taylor, 1994).

Les mesures de transparence favorisent une communication enseignant-élève plus

ouverte (Linn et al., 1991), contribuant ainsi à voir venir, réduisant la méfiance et l'anxiété face à une situation d'évaluation (Wiggins, 1993), augmentant ainsi la justesse des résultats, tout en produisant un rendement optimal. La description des critères, éléments cruciaux de cette transparence, appuie aussi le jugement porté sur les réalisations des élèves, favorisant leur traitement objectif et constant de la part de l'enseignant (Frederiksen & Collins, 1989; Glaser & Silver, 1994). Le résultat, formulé en termes de progrès, de difficultés, de maîtrise ou de réussite, selon la visée de l'évaluation, se rattache à la performance fournie et aux critères qui ont servi à la juger. Le retour en classe sur l'activité d'évaluation donne l'occasion à l'élève d'exprimer son accord ou son point de vue sur la performance fournie et le résultat obtenu.

Dans le cas d'habiletés complexes, la référence la plus utile pour apprécier le travail d'un élève semble être le schème descriptif ou qualitatif (Biggs, 1995; Herman et al., 1992; Scallon, 1988a). En décrivant en détail des performances de divers niveaux de qualité, ce schème est le plus susceptible de fournir des renseignements utiles à l'apprentissage, comme la détermination des forces et les faiblesses, ce que ne peut pas faire une note symbole (Gipps & Murphay, 1994); il fait aussi ressortir l'incongruité d'utiliser des résultats produits dans un but précis pour alimenter n'importe quelle décision. Ce schème peut facilement s'enrichir d'un niveau de performance minimal, ou de normes de rendement (Archbald, 1991; Lussier & Turner, 1995; Taylor, 1994).

4. Absence de désavantage circonstanciel

Malgré les nombreuses mises en garde décrites jusqu'à maintenant, il existe encore

d'autres menaces à la justesse des résultats. Même s'il n'est pas possible de tout contrôler, il faut au moins être conscient de ce qui peut fausser les résultats obtenus. Par exemple, à cause de l'absence de limites de temps imposées, les devoirs faits à la maison peuvent d'une certaine manière témoigner des capacités d'un élève avec plus de justesse; par contre, puisqu'il n'est pas possible d'affirmer hors de tout doute que l'élève a fait le travail lui-même, cette stratégie peut aussi réduire la justesse des résultats (Leder, 1992). La situation d'évaluation doit aussi rassembler des conditions favorables à l'élève (Pelligrino, 1992); le bruit excessif, les limites de temps, la fatigue, des événements excitants, peuvent nuire à une réalisation optimale et contribuer à l'instabilité de la performance observée (Dunbar, et al., 1991).

Une stratégie choisie peut aussi être plus favorable à certains élèves qu'à d'autres, dépendant de leurs caractéristiques personnelles (Lussier & Tumer, 1995). Par exemple, l'anxiété ressentie par les élèves timides nuit à leur performance lors d'évaluations orales, comparativement à des élèves sociables et sûrs d'eux; de manière semblable, un élève écrivant avec difficulté est plus à l'aise face à un questionnaire à réponses choisies plutôt qu'à réponses élaborées. De plus, une tâche qui ne comporte pas de biais culturel ni sexiste produirait des résultats d'évaluation justes (Herman et al., 1992; Peddie, 1992). Cette qualité optimale d'une situation d'évaluation étant difficile à atteindre, tout résultat d'évaluation insatisfaisant devrait mener à conclure en l'incapacité de démontrer la maîtrise plutôt qu'à la non-maîtrise des habiletés évaluées, parce que l'apprentissage n'est pas nécessairement en cause (Traub, 1990).

Les quatre composantes de la justesse des résultats d'évaluation sont donc la pertinence

de la tâche par rapport aux habiletés visées, la cohérence avec l'activité pédagogique, la transparence de l'activité évaluative et l'absence de désavantage circonstanciel. La figure 1, à la page suivante, tente de schématiser ces quatre composantes. À noter que cette organisation spatiale n'est pas l'objet de la présente recherche; elle ne sert qu'à faciliter les références qui y sont faites ultérieurement dans le texte, lors des discussions avec les participants et lors de l'analyse des informations recueillies. À la base d'une démarche saine, donc en première place dans cette figure, se trouve l'établissement des bonnes habiletés à évaluer accompagné du choix d'une stratégie susceptible de bien les mettre en valeur. Les trois autres composantes contribuent à mettre en lumière les apprentissages réalisés. La cohérence avec l'activité pédagogique (deuxième composante) et la transparence de l'activité évaluative (troisième composante) sont placées côte à côte parce qu'elles réfèrent à des éléments du même ordre, se déroulant souvent simultanément. En dernier lieu, l'absence de désavantage circonstanciel regroupe un ensemble de circonstances à éviter lors de l'apprentissage ou de son évaluation, afin de ne pas réduire la qualité des résultats. Si l'une de ces quatre composantes est de moins bonne qualité, l'image finale de l'apprentissage évalué en sera moins juste.

Par ailleurs, les limites entre ces quatre composantes ne sont qu'artificielles. Ainsi, les critères d'évaluation sont reliés de près à l'établissement des habiletés ciblées, mais leur communication aux élèves relève de la transparence et leur application, de l'activité pédagogique. Il en est de même de l'expression des forces et des faiblesses. Elle doit être pressentie lors du choix de la tâche, permet un réinvestissement dans l'activité pédagogique et donne de la transparence en communiquant à l'élève une appréciation de ses réalisations en termes concrets.

POUR UN RÉSULTAT D'ÉVALUATION JUSTE	
1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées - Tâche et critère(s) d'évaluation congruents avec les diverses habiletés visées - Tâche signifiante, adaptée, claire, techniquement correcte, nouvelle mais familière	
2. Cohérence avec l'activité pédagogique - Activités d'apprentissage préalables avec mise en valeur des mêmes critères d'évaluation - Mise en situation appropriée et aide disponible - Schème approprié au type d'évaluation et permettant le réinvestissement des résultats - Possibilité de reprise	3. Transparence de l'activité évaluative - Participation de l'élève à toutes les étapes de la démarche d'évaluation - Schème d'évaluation familier aux élèves - Résultats communiqués en termes de forces et de faiblesses - Retour sur l'activité d'évaluation en classe
4. Absence de désavantage circonstanciel - conditions propices à la concentration - collaboration contrôlée - procédure équitable et habituelle - temps amplement suffisant - présence pendant l'apprentissage préalable	

Figure 1. Schématisation des composantes du concept de justesse des résultats.

La justesse comme critère de qualité des résultats semble répondre à deux préoccupations actuelles en évaluation des apprentissages : d'abord, il est cohérent avec la nature même des apprentissages; de plus, il semble convenir à différents types de stratégies d'évaluation. En mettant en valeur des principes fondamentaux d'évaluation des apprentissages, en faisant ressortir les points stratégiques sur lesquels il faut se concentrer et les failles à éviter, il pourrait favoriser des pratiques d'évaluation appropriées et intégrées au processus pédagogique. En donnant plus de rigueur aux démarches d'évaluation entreprises en salle de classe, le concept de justesse pourrait contribuer à produire des résultats d'évaluation justes, corrigeant ainsi une lacune généralisée, comme en témoignent les recherches empiriques à ce sujet.

D - REVUE DE LA DOCUMENTATION

Dans la documentation consultée, on remarque l'existence de nombreux modèles connexes à l'évaluation des apprentissages. Malgré la parution récente d'une taxonomie des modèles d'évaluation (Bonniol & Vial, 1997), aucun d'eux ne peut être comparé au modèle psychométrique pour ce qui est d'assurer la qualité des résultats d'évaluation. De nombreux chercheurs déplorent cette absence (Darling-Hammond, 1994; Gipps, 1994a; Gipps & Murphay, 1994; Mislevy, 1992, 1993; Rogers, 1991; Taylor, 1994). De ce fait, la revue de la documentation se limite à décrire les pratiques évaluatives rapportées dans les recherches empiriques selon les aspects touchant les composantes du concept de justesse des résultats d'évaluation des apprentissages en salle de classe. Elle aborde

d'abord la qualité des épreuves élaborées par les enseignants, ce qu'elles valorisent et la manière dont les résultats sont interprétés. Cet aspect de l'évaluation nous renseigne indirectement sur la capacité des enseignants à déterminer correctement les habiletés à évaluer, comment les évaluer et comment les noter. Cette section est suivie des liens existant entre l'évaluation et l'activité pédagogique, de même que la transparence du processus à l'égard des élèves. Finalement, les observations réalisées lors de recherches de type qualitatif portent à remettre en question toute manière indirecte de collecter les données à ce sujet.

1. Démarches d'évaluation

L'analyse approfondie d'épreuves préparées par les enseignants conclut à des failles importantes (Carter, 1986; McMorris & Boothroyd, 1993; Oescher & Kirby, 1990); elles se limitent souvent aux habiletés élémentaires de la pensée, comme la mémorisation de faits, en particulier à l'école primaire (DePascale, 1990; McLean, 1990; Rogers, 1991; Stiggins, Miller-Griswold & Reed-Wikelund, 1989; Wiggins, 1992). Les enseignants demandent rarement d'appliquer un savoir (Fleming & Chambers, 1983) et ont de la difficulté à cerner les habiletés intellectuelles visées par des questions nécessitant des habiletés supérieures de la pensée (Carter, 1986; Herman & Dorr-Bremme, 1984; McMorris & Boothroyd, 1993; Ross, 1994).

Les enseignants sont préoccupés par la nécessité d'accumuler des notes pour le bulletin et de confirmer les jugements déjà établis, bien plus que par l'équité envers tous les élèves dans les tâches d'évaluation proposées (Airasian, Kellaghan & Madaus, 1977;

Forgette-Giroux, Simon & Bercier-Larivière, 1996; Herman & Klein, 1996; Nicholson & Anderson, 1993; Wilson, 1990). Les valeurs et les croyances personnelles établies depuis longtemps et ancrées dans les pratiques de chacun continuent de déterminer leur manière de faire (Pryor, 1992; Shepard, Flexer, Hiebert, Marion, Mayfield & Weston, 1996). Quand ils disent évaluer, ils se limitent en fait à attribuer une note (Desrosiers, Godbout & Marzouk, 1992; Oescher & Kirby, 1990) et ils le font sans se référer à un schème d'interprétation particulier ni à des critères précis (Bachor & Anderson, 1994; Barnes, 1985; Macdonald, 1993; Pryor, 1992; Wilson, 1993).

2. Lien avec l'activité pédagogique

Les enseignants éprouvent des difficultés à intégrer l'évaluation à l'enseignement, font peu de rétroaction et réinvestissent rarement les résultats dans l'intervention pédagogique (Anderson, 1989; McMorris & Boothroyd, 1993; Rowley, 1988; Stiggins & Conklin, 1992; Van Nieuwenhoven & Jonnaert, 1992; Wilson, 1990). Lors de l'évaluation, ils tiennent parfois compte d'éléments qui ont été ignorés lors de l'apprentissage (Davinroy, 1994) ou qui n'ont pas à voir avec l'habileté visée, comme le comportement en salle de classe, l'effort ou le progrès (Anderson, 1989; Barnes, 1985; Bateson, 1990; Cox, 1995; Rope, 1994). Par exemple, des élèves de 3e année ont remarqué que la compréhension était le critère important lors de l'apprentissage de la lecture mais que lors de son évaluation, c'était plutôt la qualité de réponses écrites à des questions et la qualité de l'élocution lors de la lecture à haute voix (Davinroy, 1994). Pour les bilans, les notes proviennent habituellement de résultats d'évaluation formative combinés à ceux de fin d'étape (Conseil supérieur de l'éducation, 1992; Bachor & Anderson, 1994; Forgette-Giroux, Simon &

Bercier-Larivière, 1996). La plupart du temps, les élèves reçoivent leurs copies corrigées sans pouvoir savoir précisément à quelles forces et faiblesses correspond le score obtenu (Brookhart, 1993; Schwager & Carlson, 1994).

L'évaluation des apprentissages est fréquemment reliée à l'expression de l'autorité. Selon plusieurs recherches empiriques récentes, les enseignants sont conscients des enjeux qui entourent les résultats d'évaluation et du pouvoir que leur confère cette activité (Barnes, 1995; Kusch, 1995; McMorris & Boothroyd, 1993; Stiggins & Conklin, 1992; Tobin, Butler-Kahle & Fraser, 1990), comme en témoigne l'extrait suivant : «les parents, les enseignants et la plupart des élèves accordent beaucoup d'importance aux notes; c'est pourquoi on les utilise pour les contraindre à travailler en classe (Pilcher, 1994, p. 87).

3. Participation des élèves au processus d'évaluation

Les tentatives pour impliquer activement les apprenants au processus d'évaluation ont fait ressortir non seulement des avantages évidents mais aussi la difficulté à donner le maximum de transparence au processus d'évaluation (Arter et al., 1994; Desrosiers et al., 1992; Doyon & Juneau, 1991; Ferrara, Goldberg & Mctighe, 1995; Gagné & Thouin, 1991; Hould, 1992; Oescher & Kirby, 1990). Au mieux, les enseignants s'en tiennent à informer les élèves des objectifs à atteindre ou des critères établis, plutôt que de solliciter leur participation à leur établissement (Desrosiers et al., 1992; Ferrare, et al., 1995). La communication maître-élève ouverte et explicite, en particulier au sujet de l'évaluation, semble être très peu répandue en Ontario (Bercier-Larivière, 1994; Bercier-Larivière & Forgette-Giroux, 1995).

4. Approche méthodologique et crédibilité des données

L'objectivité des enseignants à décrire leur savoir-faire a été mis à l'épreuve lorsque les chercheurs ont opté pour des approches de type plus qualitatives pour documenter les pratiques évaluatives dans leur contexte. Force fut de constater que les propos recueillis lors d'entrevues de groupe, d'entrevues individuelles ou de toute autre méthode indirecte sont vulnérables au désir de bien paraître (Forgette-Giroux et al., 1996; Gravel, 1991; Higgins & Rice, 1991; Stiggins, 1990; Van Nieuwenhoven, & Jonnaert, 1992). Certains chercheurs optent maintenant pour l'observation participante ou l'ethnographie. Stiggins, Bridgeford, Conklin et Brody concluent une de leurs recherches comme suit : «Il nous a été impossible de distinguer ce que les enseignants disent faire de ce qu'ils font vraiment. La seule stratégie viable pour décrire les pratiques d'évaluation en salle de classe reste l'observation directe.» (1985. p.1, traduction libre) Pour leur part, Tobin, Butler-Kahle et Fraser n'ont désormais plus confiance qu'en l'observation prolongée pour reconnaître l'efficacité des enseignants. «L'observation de courte durée, autrefois acceptée, paraît maintenant douteuse. Cette constatation nous amènera peut-être un jour à améliorer les pratiques d'évaluation futures.» (1990, p. 240)

Le coût des méthodes plus qualitatives oblige à se limiter à de plus petits échantillons mais permet par contre d'apprécier une image plus détaillée de la réalité (Higgins & Rice, 1991). Ainsi, il a été démontré que dans une école secondaire où on prétendait mettre l'accent sur les habiletés supérieures de la pensée et le développement de l'esprit critique dans ses cours de sciences, on se limitait en fait à des habiletés élémentaires et accordait davantage d'importance à la couverture des contenus qu'à leur compréhension (Fraser

& Tobin, 1991; Tobin et al., 1990).

E - RÉSUMÉ DE LA PROBLÉMATIQUE ET QUESTIONS DE RECHERCHE

L'évaluation des apprentissages doit obligatoirement s'adapter au nouveau paradigme éducationnel et devenir plus formative, afin de faciliter l'apprentissage. Elle doit permettre la compréhension de la manière de penser des élèves et la correction de leurs raisonnements erronés. En s'adaptant à la variété d'habiletés à développer et en produisant des résultats pouvant être réinvestis dans la démarche pédagogique, l'évaluation des apprentissages devient un outil pédagogique. Elle peut responsabiliser les apprenants en les impliquant activement dans l'évaluation de leurs apprentissages et ceux de leurs pairs. L'équité envers les élèves passe maintenant par leur connaissance des critères servant à juger leur compétence et par un nombre suffisant d'occasions de démontrer leur savoir-faire.

Ces changements majeurs dans le paradigme éducationnel font ressortir que les critères de qualité des résultats d'évaluation psychométrique ne conviennent pas à la nature des apprentissages scolaires. La qualité de la mesure, rattachée aux notions de fidélité et de validité des résultats, est basée sur l'unidimensionnalité de l'objet mesuré et de sa distribution normale. Or, les apprentissages ne peuvent qu'être estimés, toujours de manière imprécise, selon le contexte dans lequel ils sont observés et selon un idéal à atteindre. Bien que les experts en discutent depuis déjà une trentaine d'années, les

enseignants n'ont toujours pas de modèle conceptuel de qualité des résultats sur lequel s'appuyer pour guider leur intervention.

La prescription de nouvelles stratégies d'évaluation ne suffit pas à habilitier les enseignants aux exigences accrues de cette tâche. Déjà, en 1976, le ministère de l'Éducation de l'Ontario recommandait des changements majeurs dans les pratiques évaluatives, notamment au niveau des schèmes d'interprétation et de la fonction formative de l'évaluation. À notre connaissance, aucune recherche n'a démontré la mise en oeuvre de ces recommandations, ni l'habileté des enseignants à le faire, même s'ils sont entièrement responsables de l'évaluation de leurs élèves; les pratiques recensées jusqu'à maintenant permettent même d'en douter. Actuellement, l'évaluation constitue une activité détachée de l'enseignement; les résultats ne sont pas réinvestis dans l'activité pédagogique et les enseignants se servent de l'évaluation, entre autres, pour influencer sur les agissements de leurs élèves. Avec l'avènement de l'imputabilité et de l'égalité des chances en éducation, l'absence d'un modèle conceptuel de qualité des résultats d'évaluation des apprentissages en salle de classe met en jeu la justesse des résultats d'évaluation.

En effet, sans modèle de qualité des résultats d'évaluation des apprentissages en salle de classe, il n'y a pas de base pour inculquer aux enseignants les principes fondamentaux qui doivent inspirer leur intervention en déterminant les points critiques sur lesquels il doivent se concentrer et les failles à éviter. C'est toute la formation des intervenants qui en a souffert, tant initiale qu'en cours d'emploi. Sans modèle disponible, les protocoles de recherche visant à documenter cet aspect de l'activité pédagogique sont plus difficiles à

élaborer; c'est pourquoi la recherche sur l'évaluation des apprentissages en salle de classe a souvent fragmenté le processus, perdant de vue son aspect global et le coupant du contexte pédagogique où il se situe. Par ricochet, l'absence de recherche a nuit au développement de cet aspect essentiel de l'enseignement et à l'adaptation des pratiques.

Le concept de justesse des résultats d'évaluation, tel que décrit précédemment, pourrait peut-être constituer un pas dans le redressement de cette situation. Les éléments stratégiques établis pour assurer des résultats d'évaluation plus justes semblent répondre à deux préoccupations actuelles en évaluation des apprentissages : d'abord, ils sont cohérents avec la nature même des apprentissages, puis, ils semblent convenir à différents types de stratégies d'évaluation. En mettant en valeur des principes fondamentaux d'évaluation des apprentissages, en énonçant les points sur lesquels il faut se concentrer et les failles à éviter, tous ces éléments critiques peuvent promouvoir des pratiques d'évaluation appropriées et intégrées au processus pédagogique.

Si l'on tient compte des habiletés à développer chez les élèves, du choix d'une tâche propice à leur manifestation, des expériences d'apprentissage vécues, de la transparence du processus et de l'absence de désavantage circonstanciel, une situation d'évaluation devrait assurer des résultats d'évaluation justes. Le concept de justesse des résultats pourrait promouvoir l'évaluation équitable d'habiletés complexes selon divers niveaux de performances, dans leur contexte. Il semble pouvoir s'appliquer à n'importe quelle situation d'évaluation, traditionnelle ou complémentaire, assurer des résultats plus sensés, plus utiles, conformes à la nature même des apprentissages visés par les nouveaux programmes.

Étant donné l'importance d'un modèle fiable pour générer des résultats d'évaluation de qualité, avant tout pour le mieux-être des élèves, mais aussi pour une meilleure formation des enseignants et pour l'évolution de la recherche en ce domaine, il apparaît opportun de mettre à l'épreuve la viabilité des composantes présentées précédemment. La présente recherche se propose donc de vérifier empiriquement si elles peuvent assurer la qualité ou la justesse des résultats d'évaluation des apprentissages en salle de classe. Même si ce concept tire son origine d'une analyse exhaustive des écrits, tant théoriques qu'empiriques, cette mise à l'essai représente une contribution importante dans l'avancement en ce domaine, puisqu'aucun modèle de ce genre n'a jamais été, à notre connaissance, expérimenté précédemment. Par conséquent, la partie empirique de cette recherche tentera de déterminer si le critère de justesse des résultats d'évaluation des apprentissages, tel que décrit précédemment, est applicable à l'évaluation en salle de classe.

Jusqu'à maintenant, les pratiques d'évaluation en salle de classe ont été documentées presque exclusivement au moyen d'approches indirectes. Or, ces méthodes de recherche apparaissent aujourd'hui peu fiables. Puisque la collaboration prolongée nécessaire à cette recherche s'y prête bien, il semble approprié de documenter effectivement les pratiques actuelles des enseignants avant d'aborder le concept de justesse des résultats; la description détaillée et contextualisée de quelques situations d'évaluation et la perception des éléments pouvant influencer la capacité des résultats à situer correctement les apprenants dans leur développement, s'avèrent de bons prétextes à l'exploration de la conception actuelle de l'évaluation en salle de classe. De plus, tout en renseignant sur la manière la plus convenable d'aborder le concept de justesse des résultats, la différence

observée entre cette conception d'une évaluation juste et celle des participants permet de mieux comprendre leurs réactions et leurs opinions. Tenant compte de ces faits, la présente recherche tente de répondre aux questions suivantes :

1 a) Quels sont les éléments marquants des pratiques d'évaluation des apprentissages en salle de classe menées par les enseignants?

b) Quels aspects de ces démarches les enseignants considèrent-ils importants ou critiques pour obtenir des résultats justes?

c) Comment leur conception des éléments pouvant intervenir sur la justesse des résultats se compare-t-elle à celle du cadre conceptuel de la présente recherche?

2 a) La pertinence de la tâche d'évaluation par rapport aux habiletés visées, la cohérence de celle-ci avec l'activité pédagogique, la transparence de l'activité évaluative et l'absence de désavantage circonstanciel rendent-elles compte de l'ensemble des éléments pouvant intervenir sur la justesse des résultats d'évaluation des apprentissages en salle de classe?

b) La notion de justesse des résultats telle que présentée dans le cadre conceptuel peut-elle s'intégrer à l'activité scolaire?

MÉTHODOLOGIE

La notion de l'évaluation des apprentissages basée sur la justesse des résultats, développée et exposée au chapitre précédent, résulte d'une analyse exhaustive de la documentation à ce sujet. Il s'agit maintenant d'examiner si elle est fonctionnelle, si elle tient compte de tous les éléments en cause et si elle est cohérente et claire. Cet exercice permettra aussi de raffiner le concept, afin de mieux tenir compte de la réalité de la salle de classe.

Cette façon de concevoir l'évaluation des apprentissages en salle de classe étant à la fois inédite, innovatrice et originale, il est nécessaire d'ancrer son étude dans la conception que s'en font actuellement les participants, afin de nuancer leur prise de position. La première question de recherche sert donc d'entrée en matière en situant la manière de faire habituelle des enseignants-participants, ce qui permet de mieux comprendre par la suite leur évaluation d'une nouvelle façon de faire, en vue de répondre à la deuxième question de recherche.

La méthodologie adoptée pour cette recherche s'inspire en partie des recommandations de recherches empiriques portant sur l'évaluation des apprentissages, mais aussi de la nature de la problématique étudiée et des questions de recherche énoncées. La complexité du milieu étudié et la conjoncture du moment amènent d'abord à exposer le contexte social et politique de cette mise à l'essai. En jetant un éclairage sur le vécu des participants, la deuxième section du présent chapitre permet de contextualiser les

informations recueillies. La troisième partie décrit la démarche prévue pour répondre aux deux questions de recherche et se termine par la procédure de recrutement. Les analyses anticipées sont brièvement présentées en quatrième partie. La dernière partie aborde la rigueur de la méthode utilisée afin d'assurer des données crédibles, fiables, utiles et objectives et des résultats significatifs, permettant une compréhension approfondie de la réalité étudiée.

L'approche naturaliste telle que décrite par Lincoln et Guba (1985) semble la plus appropriée pour documenter les questions de recherche énoncées. Elle peut contrer l'effet de désirabilité résultant de toute méthode indirecte pour collecter des données à ce sujet (Gravel, 1991; Higgins & Rice, 1991; Van Nieuwenhoven & Jonnaert, 1992). En effet, malgré ses exigences, certains chercheurs n'ont dorénavant plus confiance qu'en l'observation participante et l'ethnographie pour fournir une image détaillée et objective de la réalité : « Dans une perspective d'avenir, je crois que la clé du succès dans la compréhension de la nature et de la qualité de l'évaluation en salle de classe repose sur des méthodes de recherche axées sur la profondeur des données recueillies plutôt que sur la généralisabilité des résultats. Ce qui correspond à l'examen de faibles échantillons au moyen de microscopes puissants. » (Stiggins, 1990, p. 94, traduction libre). De plus, la souplesse de cette approche permet de s'ajuster au contexte (Poupart et al., 1997) et de respecter la complexité de l'activité pédagogique (Bercier-Larivière, 1994; Bercier-Larivière & Forgette-Giroux, 1995; Rogers, 1991; Tobin et al., 1990).

Cette recherche a eu recours à la collaboration étroite d'enseignants volontaires pendant une période d'environ six mois. Ce partenariat entre quelques praticiens et une personne ressource visait à opérationnaliser le concept de justesse tout en restant le plus près possible de la réalité, et à l'adapter si nécessaire. Le besoin d'outils d'évaluation exprimé par les participants quant à leur tâche quotidienne ont servi de matière première à cette mise à l'essai (Erickson, 1986). Ce pont entre le concept développé et la réalité de l'univers empirique (Massé, 1992) a permis de vérifier sa pertinence et son utilité.

Les personnes impliquées dans la présente mise à l'essai sont uniques et leurs caractéristiques ont influencé indirectement les résultats à venir. Chez les participants, le désir de s'engager à fond dans cette collaboration a été confronté à leurs autres responsabilités professionnelles et au temps dont ils disposaient. Cette approche faisait aussi de la chercheuse le pivot central de la collecte des données et les enseignants-participants y ont joué un rôle majeur. Les deux prochaines sections présentent donc ces influences importantes de la collecte des données : le contexte social et politique de la collecte des données, suivi d'un portrait professionnel de chacun des cinq participants, de même que celui de l'auteure de la recherche.

A - CONTEXTE SOCIAL ET POLITIQUE DE MISE À L'ESSAI

L'approche qualitative est sensible aux circonstances entourant la recherche, à la nature des informations recueillies et aussi au vécu des participants. Elle est donc particulièrement appropriée pour documenter la problématique étudiée ici. Durant la

période qui nous intéresse, l'atmosphère de l'école et le moral du personnel enseignant sont surtout affectés par des changements politiques. L'arrivée d'un gouvernement conservateur au pouvoir, en juin 1995, a pour le moins chambardé les procédures engagées par le précédent gouvernement, d'allégeance néo-démocrate, dans le renouvellement du système scolaire. Les années 90 ont été difficiles pour le personnel enseignant : augmentation de la charge de travail et réduction ou gel des salaires pendant plusieurs années, augmentation du ratio maître-élèves, coupures de ressources pédagogiques, insécurité d'emploi, réduction des budgets pour matériel pédagogique; même les crédits de suppléance sont diminués. Cette période de difficultés économiques est aussi le lot des parents des écoliers; cette pression supplémentaire sur la cellule familiale se répercute sur la tâche de l'enseignant.

En décembre 1995, le gouvernement a annoncé la création de l'Ordre des enseignantes et des enseignants de l'Ontario et de l'Office de la qualité et de la responsabilité en éducation (OQRÉ). Selon le ministre de l'Éducation et de Formation de l'Ontario de l'époque, monsieur John Snobelen, ces deux organismes assureront une plus grande surveillance du système d'éducation, pour qu'il assume davantage ses responsabilités (Ministère de l'Éducation et de Formation de l'Ontario, 1995a). L'OQRÉ effectuera des enquêtes, publiera des indicateurs de rendement et présentera un rapport annuel au public. De plus, il administrera annuellement à tous les élèves de troisième année et à un échantillon d'élèves de sixième ou de neuvième années de la province des tests de lecture, d'écriture et de mathématiques. En octobre 1996, 900 écoles de la province se sont engagés dans la mise à l'essai d'un bulletin normalisé, en vue de son entrée en

vigueur dès septembre 1997 (MÉFO, 1996). En janvier 1997, au moment de commencer la collecte des données, le Ministre entamait également la restructuration de l'éducation, «réduisant les doubles emplois et le gaspillage, et rationalisant l'administration et la bureaucratie» (MÉFO, 1997c, p. 2) afin, disait-il, d'orienter les ressources vers la salle de classe.

Le *Programme d'études commun* (PÉC) est entré officiellement en vigueur à l'automne 1996, soit quatre mois avant le début de la collecte des données. Ce document politique regroupe les disciplines par champs d'étude et le contenu enseigné est exprimé sous forme de résultats à atteindre en troisième, sixième et neuvième années; les mathématiques, les sciences et la technologie bénéficient de normes de rendement en quatre niveaux et les langues, de normes en six niveaux. Des documents similaires, spécifiques aux deux autres domaines (Arts et Formation personnelle et études sociales) sont à paraître. Des groupes de travail d'enseignants, parrainés par divers conseils scolaires, ont élaboré ces documents politiques en établissant des «mini-résultats» à atteindre et des contenus à enseigner en première, deuxième, quatrième, cinquième, septième et huitième années. À l'automne 1996, le ministère de l'Éducation et de la Formation a entrepris de modifier les normes en mathématiques et en langue, pour en faire des *Cadres d'apprentissage*; ces nouveautés devraient subir une mise à l'essai au printemps 1997 et devenir ensuite officielles en septembre suivant.

En attendant les nouveaux documents, des enseignants ont conservé leur ancien programme et attendu que les choses se placent; d'autres ont chevauché les deux documents, comme il leur convenait. Pour se préparer à la transition, des enseignants ont

accepté de contribuer à la mise à l'essai des futurs Cadres d'apprentissage dans leur classe et de bénéficier ainsi d'aide pour élaborer des activités répondant à ces nouvelles normes. De leur côté, les enseignants de troisième année se sont sentis plus obligés au *Programme d'études commun* puisque leurs élèves subissaient les premiers examens provinciaux en mai 1997.

Au cours du mois de mars 1997, le ministère de l'Éducation a annoncé la disparition du PÉC comme document politique; la sortie et la mise à l'essai des nouveaux *Cadres d'apprentissage* ont été mis en veilleuse. Le suspense s'est terminé à la fin du mois de juin 1997 avec la mise en circulation de deux nouveaux documents : *Le curriculum de l'Ontario, de la 1^{re} à la 8^e année, français* et *Le curriculum de l'Ontario, de la 1^{re} à la 8^e année, mathématiques*. Ces documents sont devenus les programmes cadres officiels, éclipçant du même coup tous les documents précédents.

Durant l'année scolaire 1996-1997 au cours de laquelle a eu lieu la partie expérimentale de la présente recherche, l'évaluation de tous les élèves de troisième année en lecture, écriture et mathématiques a été précédée d'une mise à l'essai chez environ 2000 élèves francophones de quatrième année et l'évaluation d'élèves de sixième année en mathématiques a fait suite à une mise à l'essai chez des élèves de septième année. Une mise à l'essai du bulletin normalisé provincial se déroulait simultanément. Afin d'assurer l'uniformité des procédures d'administration des tests provinciaux, tous les enseignants concernés ont reçu une formation obligatoire; de plus, des enseignants volontaires de la province ont effectué la notation de toutes ces épreuves, ce qui constituait au yeux du gouvernement, un élément de formation professionnelle. Le remplacement de tous ces

enseignants en salle de classe a dérangé la routine de travail.

B - PARTICIPANTS

Les enseignants de la première à la neuvième année inclusivement, oeuvrant dans une école ontarienne de langue française, constituent la population cible. À ce moment-là, ils enseignaient tous à partir du *Programme d'études commun* et des *Normes de performance* imposées par le ministère de l'Éducation et de la Formation de l'Ontario (1995c).

1. Recrutement

Le processus de recrutement des participants a commencé en décembre 1996, dès l'approbation déontologique du projet de recherche (Annexe A). Les directions d'école ont été rejointes par l'entremise d'un surintendant du conseil scolaire mis au courant du projet; l'identité de celui-ci est tenue confidentielle afin de respecter les droits d'anonymat des participants. La chercheure s'est présentée comme étant une étudiante diplômée en éducation, qui n'était pas employée d'un conseil scolaire et qui cherchait des praticiens pour collaborer à une recherche sur l'évaluation en salle de classe. Les directions d'école où des enseignants étaient intéressés par ce projet ont contacté ensuite la chercheure afin d'établir un moment propice pour les rencontrer. À cette occasion, les candidats présents ont été informés du projet de recherche et des exigences de la participation. Des

formulaire de consentement ont été distribués et les personnes qui, après mûre réflexion, étaient intéressées à y participer, ont retourné ce formulaire par la poste.

2. Description des participants

La généreuse participation de cinq enseignants volontaires a permis de poursuivre la présente recherche. Afin de protéger leur anonymat, un nom fictif leur a été attribué. Ce pseudonyme leur sert de référence tout au cours de la collecte des données et du présent document. Ces cinq personnes provenaient de deux différentes écoles d'un conseil scolaire ontarien de langue française, à une heure de distance l'une de l'autre; Christine et Kevin de la première école, Marie, Matilde et Marie-Chantal de la deuxième école. Voici maintenant un bref profil de chacun d'eux.

2.1 Christine

Christine en était à sa sixième année d'expérience, dont cinq dans ce même conseil. Son groupe comportait vingt élèves de cinquième année, dont trois en grande difficulté d'apprentissage. Sa formation initiale est en français aux cycles primaire et moyen. Elle a aussi suivi un cours d'enseignant-associé et un autre sur l'utilisation des ordinateurs en salle de classe. Au moment de la collecte des données, elle était représentante syndicale pour son école, accueillait régulièrement des enseignants en formation et contribuait aux comités social et environnement-santé de son école.

Pendant l'année scolaire 1996-1997, Christine a participé à la notation des tests de troisième année. Elle se considère choyée étant donné son groupe d'élèves peu

nombreux et d'un même niveau scolaire. Elle a accepté de participer à cette recherche pour rendre service.

2.2 Kevin

Diplômé de la Formation à l'enseignement en juin 1996, Kevin est l'un des rares étudiants de sa promotion à s'être trouvé un poste à temps plein dès sa première année d'exercice. Il a été embauché le 25 septembre, alors que la classe de Christine est devenue trop nombreuse pour accommoder à la fois les élèves de cinquième et de sixième années. Christine a gardé les élèves de cinquième année et Kevin a pris en charge les 26 élèves de sixième. Il est spécialisé en musique instrumentale et français langue seconde, aux cycles intermédiaire et supérieur. Il enseigne toutes les matières, sauf éducation physique et art visuel. En plus de sa charge régulière de travail, il s'occupe des sports et de la foire mathématique. Ses élèves ont participé au premier test provincial de mathématiques. Lors du premier contact établi avec son école, le directeur a pensé que Kevin pourrait bénéficier de participer à cette recherche, vu son manque d'expérience pratique dans le domaine de l'évaluation. Kevin s'est engagé dans cette recherche, heureux de suivre les conseils de son directeur.

2.3 Marie

En vue de réintégrer le marché du travail après que ses enfants eurent commencé l'école, Marie s'est dirigée vers la formation à l'enseignement. Son domaine d'études universitaires est la sociologie et sa formation à l'enseignement, aux cycles primaire et moyen. Au moment de la présente recherche, elle en était à sa cinquième année, toujours

dans ce même conseil scolaire, dans cette même école, toujours au cycle primaire. Elle enseignait à nouveau en troisième année et son groupe-classe comportait 25 élèves.

Cette enseignante cherche à adapter son enseignement aux besoins spécifiques des élèves. Les classes de plus en plus hétérogènes et nombreuses, les difficultés particulières, comme l'élève qui n'a pas réussi sa deuxième année mais qui a quand même suivi les élèves de son âge en troisième année, ne peuvent s'accommoder d'une pédagogie adressée à l'élève moyen. Elle croit que l'avenir est dans l'individualisation des services; l'évaluation lui est récemment apparue comme un outil lui permettant de mieux répondre aux besoins de chacun.

En plus de sa tâche régulière, elle a collaboré à la préparation de l'album souvenir de l'école, à la semaine de l'environnement, à la journée de lecture et au spectacle de fin d'année. Elle a aussi contribué à la mise à l'essai du nouveau bulletin provincial. Ses élèves subissaient le premier test provincial en lecture, écriture et mathématiques et elle a participé à la notation des copies des élèves. Pour ces tâches, elle a bénéficié de quelques journées de formation offertes par l'Office de la qualité et de la responsabilité en éducation et les a grandement appréciées.

Selon elle, sa participation à la présente recherche lui permettait de verbaliser ses prises de conscience, d'en discuter avec les autres participants et de se rassurer sur la «normalité» d'une telle remise en question.

2.4 Matilde

Après avoir travaillé dans trois différents conseils scolaires ontariens comme enseignante suppléante et comme enseignante itinérante en musique, de la maternelle à la douzième année, Matilde a obtenu un poste à temps plein en classe régulière. Au moment de la recherche, pour sa sixième année de carrière enseignante, elle avait une classe de vingt-quatre élèves de huitième année. Elle était aussi responsable de la chorale de l'école, du conseil étudiant et du spectacle de fin d'année, et partageait la tâche de l'album souvenir avec Marie. Puisque ses élèves quittaient cette école à la fin de l'année, elle était aussi responsable de leur bal des finissants et menait de front plusieurs activités de levée de fonds pour le voyage d'adieu traditionnel à Toronto.

Sa formation initiale est en français, aux cycles primaire, moyen et intermédiaire. Elle a complété sa formation par divers ateliers portant sur la gestion de classe «participative», l'animation culturelle et l'enseignement de la musique. Sa préoccupation la plus importante est la discipline en salle de classe. Son désir de réduire le volume de correction l'a incité à prendre de plus amples informations au sujet du présent projet de recherche.

2.5 Marie-Chantal

Cette enseignante cherche par tous les moyens à améliorer son enseignement, mais exprime rarement sa satisfaction du travail réalisé. Pour sa cinquième année d'enseignement, son groupe classe était formé de 22 élèves de cinquième année et de quatre élèves de quatrième année. En plus de sa tâche d'enseignement, elle s'occupait du conseil étudiant, du club de plein air, de hockey-salon en cinquième et sixième années,

du ballon-volant en septième et huitième années, de la semaine d'activités d'hiver et du comité social. Marie-Chantal était aussi responsable du service informatique de son école, pour lequel elle a reçu une formation particulière. À l'automne, elle a accueilli des enseignants en formation, elle a aussi participé à la mise à l'essai du bulletin provincial normalisé. Sa grande créativité l'amène à souhaiter un jour pouvoir se consacrer entièrement à l'élaboration de matériel pédagogique. Son intérêt pour la présente recherche résidait dans son désir constant de se perfectionner.

Tous ces participants étaient prêts à investir beaucoup de leurs énergies pour se tailler une place dans la profession. La compétition est forte et se joue sur plusieurs fronts : le rendement des élèves, la discipline en salle de classe, les activités parascolaires, les comités d'écoles et du conseil scolaire et la visibilité de l'école dans la communauté en général. Au dire des enseignants, l'activité en salle de classe tient souvent de la survie, à cause de l'intensité et la diversité des besoins des élèves et la rareté des ressources spécialisées. L'atmosphère du milieu de travail, en grande partie créée par le contexte politique de l'éducation, ne favorise pas nécessairement l'épanouissement professionnel et personnel des enseignants.

La formation à l'enseignement de ces cinq participants est relativement récente; elle date des années 90. On pourrait croire que leur conception de la pédagogie est en harmonie avec les théories récentes d'apprentissage et qu'ils sont familiers avec des notions élémentaires modernes d'évaluation, comme l'évaluation formative et l'évaluation de la

performance. Exception faite de Kevin, à ses premières armes dans la profession, elles sont toutes assez expérimentées pour avoir développé une conception personnelle de l'évaluation des apprentissages en salle de classe.

3. Description de l'auteure de la recherche

Son intérêt pour baliser la démarche d'évaluation des apprentissages en salle de classe est né de diverses expériences. D'abord, une formation à l'enseignement ayant à peine fait allusion à l'évaluation des élèves, comme s'il s'agissait d'un savoir inné, suivie d'une totale absence d'encadrement en la matière une fois en fonction. Ensuite, quelques années plus tard, le point de vue de parent d'un enfant évalué par des enseignants tout aussi mal préparés pour le faire : calculs mathématiques montrés comme preuve de pratiques rigoureuses, consignes confuses, insistance sur la forme davantage que sur le fond, sur la mémorisation plutôt que sur la compréhension et la réflexion. Une évaluation réduite à une collecte de points pour donner du sérieux aux activités en salle de classe, très souvent pour éviter les écarts de discipline. Enfin, l'ampleur du décrochage scolaire chez les adolescents, suite à de nombreuses années à carburger aux notes, au point d'étouffer le plaisir d'apprendre qu'ils avaient eu plus jeunes. Tout cela a engendré le choix du sujet de thèse de maîtrise en mesure et évaluation en éducation : *La motivation des élèves de 6^e, 7^e et 8^e année et leur perception des pratiques d'évaluation des apprentissages.*

L'exercice lui a fait prendre conscience de l'anarchie entourant l'évaluation en milieu scolaire et pire encore, de l'ignorance de ce fait par les dirigeants. La vue d'enseignants

profondément ébranlés dans leurs convictions suite à des laboratoires axés sur l'évaluation en salle de classe a ajouté à l'impression de découvrir un maillon faible de l'activité pédagogique. En parallèle à ce cheminement intérieur, la collaboration à des recherches menées auprès d'enseignants d'écoles élémentaires et secondaires a nourri sa réflexion. C'est ainsi qu'il lui fut donné de constater leur intérêt pour des ateliers portant sur des notions élémentaires d'évaluation, leur utilisation discutable du vocabulaire pertinent, leur difficulté à décrire objectivement leurs pratiques et leur manque de vision harmonisée du processus d'évaluation (Forgette-Giroux et al., 1996; Simon, Forgette-Giroux & Bercier-Larivière, 1993).

Par ailleurs, la crédibilité et le sérieux des enfants lorsqu'ils parlent des notes, de ce qu'ils trouvent juste et de ce qu'ils ne croient pas l'être et leur vulnérabilité face au pouvoir de l'enseignant ont laissé entrevoir une utilisation de l'évaluation qui s'éloigne de l'objectif pédagogique qui devrait être recherché. À force de chercher l'origine de cette faiblesse dans l'activité pédagogique et une manière d'y remédier, l'absence d'un modèle conceptuel de la qualité des résultats d'évaluation des apprentissages en salle de classe est apparue évidente. Le présent projet de recherche était alors né.

C - COLLECTE DES DONNÉES

Le développement de la confiance mutuelle entre les parties impliquées s'avère crucial pour la crédibilité des données recueillies (Stake, 1994). C'est à la chercheuse de démontrer aux candidats qu'il ne s'agit pas d'une évaluation de leurs compétences à

évaluer, ni d'un parasitage de leurs activités, mais bien d'un partenariat en vue d'une compréhension approfondie du phénomène étudié (Lincoln & Guba, 1985; Pourtois & Desmet, 1988). La chercheuse s'est présentée comme une ressource supplémentaire leur fournissant un contexte propice à l'épanouissement de leurs compétences à évaluer les apprentissages en salle de classe, domaine pour lequel ils ne disposent habituellement pas de soutien spécialisé. En échange, elle bénéficiait de leur expertise en milieu de travail en vue de faire avancer les connaissances dans le domaine.

Cette mise à l'essai a été documentée au jour le jour au moyen d'observations notées dans un journal de bord. Les indices extérieurs comme les réactions, les commentaires, les discussions et les échanges lors des ateliers de travail et des observations dans la salle de classe y ont été inscrits, de même que la réaction des élèves. Le déroulement des activités, les difficultés rencontrées et les solutions adoptées, l'atmosphère de travail et les péripéties pertinentes au projet y ont aussi été recueillies (Bogdan & Biklen, 1992; Miles & Huberman, 1994).

Un guide de formation a fourni un minimum d'appui aux enseignants-participants; son contenu a alimenté les échanges des premières rencontres et servi à donner du vocabulaire aux participants, tout en assurant l'utilisation d'une terminologie correcte. Il a facilité aussi la communication du concept de justesse abordé par la suite et illustré au moyen de quelques situations de salle de classe, en vue toujours de les habiliter à se prononcer plus tard sur sa pertinence. Ce document, tout en étant au coeur de la collecte des données, n'était pas destiné à faire la lumière sur les questions de recherche mais constituait plutôt une ressource appuyant la collecte des données.

1. Première question de recherche

Une situation d'évaluation vécue durant l'année scolaire en cours par chacun des cinq enseignants-participants a documenté la première question de recherche portant sur les pratiques d'évaluation des apprentissages avant la mise à l'essai. Ces cinq situations d'évaluation (1 à 5) ont été rassemblées dans un cahier technique. Les aspects de ces situations d'évaluation considérés importants ou critiques par les participants pour obtenir des résultats représentant bien ce que les élèves ont acquis ou développé étaient l'objet du premier entretien individuel. Même si un guide d'entrevue a orienté l'entretien, chacun était centré sur la situation d'évaluation personnelle documentée à l'étape précédente.

2. Deuxième question de recherche

La vulgarisation du concept de justesse des résultats et sa mise à l'essai ont commencé par des ateliers de travail. Ces rencontres étaient informelles, individuelles ou en équipe, selon la disponibilité, le désir et les besoins de chacun, mais toujours en dehors des heures de classe. Les participants assistaient d'abord à de brèves présentations exposant les composantes du concept, puis ils tentaient ensuite, avec l'aide de la chercheure, d'en imaginer l'application à des situations d'évaluation fictives. Les discussions entourant chaque exemple mettaient en valeur les composantes du concept et leur harmonisation et avaient pour but d'amener graduellement les participants à voir sa mise en pratique. L'activité s'étendait ensuite à quelques situations d'évaluation fictives ou réelles pertinentes à l'activité pédagogique quotidienne de chacun.

Par la suite, les participants ont choisi une activité pédagogique réelle à laquelle ils voulaient appliquer le concept de justesse des résultats pour préparer une situation d'évaluation et ensuite l'appliquer dans leur classe. L'opérationnalisation du concept, ou sa confrontation à diverses situations réelles, selon les besoins et le vécu de la salle de classe, a généré des informations permettant d'étudier sa pertinence, sa cohérence et sa clarté et de le réajuster par la suite. Un plan et une description de ces cinq situations, accompagnés de pièces témoins, appuyés parfois de photographies prises en salle de classe et de copies d'élèves, ont été insérés au cahier technique (situations d'évaluation 6 à 10).

Des entretiens individuels menés à la fin du projet et enregistrés sur bande sonore invitaient les participants à exprimer leur opinion sur le concept étudié. Ces entretiens ont lieu vers la fin de l'année scolaire, à un moment choisi par le participant. Les questions, placées dans un ordre logique pour la discussion, portaient sur la pertinence et l'utilité du concept, les aspects les plus difficiles à appliquer selon le vécu de chacun, les répercussions tant positives que négatives de cet essai sur l'activité pédagogique de l'enseignant et les contraintes ressenties lors de sa mise en oeuvre.

Le matériel documentaire accumulé tout au cours de cette collecte de données porte donc sur dix situations d'évaluation des apprentissages en salle de classe et est rassemblé dans un cahier technique. Les cinq premières situations témoignent des pratiques d'évaluation des participants avant la mise à l'essai et les cinq autres illustrent l'application du concept de justesse à l'évaluation des apprentissages en salle de classe. À ce recueil se sont ajoutés deux entretiens individuels enregistrés; le premier portait sur les pratiques

habituelles des participants avant la mise à l'essai et s'intéressait particulièrement aux aspects de ces démarches considérées importants ou critiques pour obtenir des résultats justes. Le deuxième entretien s'est déroulé à la toute fin de la période de mise à l'essai et visait à connaître l'opinion des enseignants-participants sur l'utilité, le réalisme, l'amélioration possible et l'intégrabilité de cette nouvelle conception de l'évaluation des apprentissages des élèves. Un journal de bord a complété le tout. Une entrée à chaque rencontre (visite ou atelier individuel ou collectif, observation en salle de classe) rapporte des indices extérieurs comme les réactions, les commentaires, les sujets de discussions et la teneur des échanges. La figure 2 résume les différents aspects de la collecte des données et les diverses pièces-témoins se rapportant à chacune des questions de recherche.

Première question	Pratiques d'évaluation avant la mise à l'essai	Journal de bord
	Situation d'évaluation 1 : Christine, Mathématiques, 5 ^e année (Multiplications)	
	Situation d'évaluation 2 : Kevin, Musique, 6 ^e année (Théorie musicale)	
	Situation d'évaluation 3 : Marie, Lecture, 3 ^e année (Directeur d'école et directeur de cinéma)	
	Situation d'évaluation 4 : Matilde, Présentation orale (anglais), 8 ^e année (La musique)	
	Situation d'évaluation 5 : Marie-Chantal, Présentation orale (français), 5 ^e année (En fin de semaine dernière)	
	Entrevues préliminaires individuelles : Aspects des situations 1 à 5 estimés importants pour l'obtention de résultats situant bien les élèves dans leur développement	
Deuxième question	Application du concept de justesse des résultats	
	Situation d'évaluation 6 : Christine, Sciences, 5 ^e année (Foire scientifique)	
	Situation d'évaluation 7 : Kevin, Lecture, 6 ^e année (En toute sécurité)	
	Situation d'évaluation 8 : Marie, Mathématiques, 3 ^e année (La multiplication)	
	Situation d'évaluation 9 : Matilde, Robotique, 8 ^e année (Construire une machine)	
	Situation d'évaluation 10 : Marie-Chantal, Projet technologique, 5 ^e année (L'hélicoptère)	
	Entrevues finales individuelles : Opinion des participants sur l'utilité, le réalisme, l'amélioration possible et l'intégrabilité du concept de justesse des résultats	

Figure 2. Matériel documentaire accumulé en vue de répondre aux deux questions de recherche

D - PLAN D'ANALYSE DES DONNÉES

L'analyse des données a débuté par l'examen de la documentation au sujet des cinq situations d'évaluation d'avant la mise à l'essai. Des synopsis de chacune ont été préparés et présentés aux participants pour vérification, puis ajustés si nécessaire. Chaque situation a ensuite été étudiée en détail, afin de déterminer et d'analyser ses particularités, ses points saillants et ses similitudes avec les autres. Les points d'intérêts sont présentés et discutés au prochain chapitre.

Le premier entretien portait sur les aspects de ces situations d'évaluation considérés importants ou critiques pour l'obtention de résultats représentant bien ce que les élèves ont acquis ou développé; il a été transcrit intégralement. À nouveau, des synopsis ont été préparés et soumis aux participants pour approbation. Le traitement des transcriptions a débuté par leur lecture à maintes reprises et la recherche d'unités de sens, jusqu'à ce que des catégories et sous-catégories émergent du matériel documentaire, en harmonie avec le cadre conceptuel établi. Cette structure dépouillée a ensuite été analysée et comparée au contenu du cadre conceptuel présenté ici, en vue de déterminer les éléments communs, les éléments manquants et les éléments étrangers. Des extraits du journal de bord pouvant enrichir cette question ont aussi été retenus. Toutes ces informations, présentées et discutées au chapitre trois, complètent la réponse à la première question de recherche.

Chacune des situations d'évaluation élaborées avec les participants selon les composantes du concept de justesse a été analysée. Ce traitement préliminaire a servi

de base au deuxième entretien portant sur la viabilité du concept de justesse des résultats. Celui-ci a également été transcrit intégralement, suivi de la préparation de synopsis et de leur soumission aux participants. Les protocoles verbaux ont d'abord été fragmentés selon qu'il portaient sur l'utilité, le réalisme, l'amélioration possible ou l'intégrabilité de cette autre conception de l'évaluation. Enfin, le chapitre quatre présente et discute les données en réponse à la deuxième question de recherche, de même que des extraits du journal de bord jugés pertinents.

E - CRÉDIBILITÉ DES RÉSULTATS

L'approche qualitative adoptée pour documenter le problème présenté ici fait de la chercheure le pivot central de la collecte des données (Lincoln & Guba, 1985). Ayant elle-même élaboré le concept, elle est la seule personne à le connaître aussi bien, la mieux placée pour en faire la mise à l'essai, mais aussi la plus intéressée à réussir la mise à l'essai. Pour contrer toutes ces sources possibles de biais, la présente section fait état des mesures adoptées pour collecter et traiter le matériel documentaire avec rigueur, afin que les résultats de la présente recherche bénéficient de toute la crédibilité qu'il leur revient.

1. Orientation des observations

D'abord, le concept de justesse des résultats a été défini explicitement dans le cadre conceptuel de la présente recherche. Selon Hammersley (1993) et Van der Maren (1995),

ceci a permis de cibler précisément et efficacement les observations susceptibles de faire la lumière sur les questions de recherche. De plus, le critère de justesse étant supposé convenir à n'importe quel champ d'études, domaine ou niveau scolaire, à l'évaluation d'habiletés simples ou complexes, à des fins formative ou sommative, il était souhaitable de l'éprouver dans des situations variées. Bien qu'ici le nombre de situations différentes mettant le concept à l'épreuve soit modeste, ce sont toutes des situations réelles, choisies par les participants, en réponse aux besoins ressentis dans leurs activités pédagogiques du moment. Parce qu'elles étaient imprévisibles et authentiques, elles informaient sur la polyvalence du concept, son utilité, sa flexibilité, son intégrabilité et sa pertinence (Keeves, 1988; Mucchielli, 1996).

La documentation approfondie de quelques situations d'évaluation et l'observation directe en salle de classe présentées dans la présente étude, y compris les tâches assignées, leur notation et le contexte dans lequel elles s'insèrent, en font un portrait plus global, plus concret et plus fiable que l'utilisation de questionnaires (Gravel, 1991; Higgins & Rice, 1991; Stiggins et al., 1985; Tobin et al., 1990). De plus, les longues discussions, avec les participants et entre eux, ont fait émerger beaucoup d'informations complémentaires aux situations décrites.

2. Respect des participants

Afin d'obtenir le maximum d'information crédible sur la présente problématique de recherche, la chercheure, très consciente de l'effet de halo qu'aurait pu initialement avoir le statut d'experte en évaluation sur les participants, s'en est tenue à celui d'étudiante

diplômée en éducation. Ce n'est qu'après plusieurs semaines de collaboration que les compétences en matière d'évaluation des apprentissages en salle de classe ont pu être appréciées. À ce moment-là, les bons rapports et la grande importance accordée à leur opinion sincère et à leur expérience pratique de l'activité pédagogique n'auraient pu être altérés par l'évidence de toutes les possibilités d'avancement que présentait pour eux la présente recherche.

La collecte des données a duré environ six mois. Étant donné les nombreuses rencontres et le travail assidu, cette période a suffi à développer la confiance mutuelle, vérifier les ambiguïtés et éclaircir d'éventuelles incongruités (Van der Maren, 1995). La communication avec les participants ne laissait aucun doute sur la valeur capitale d'une opinion honnête et franche, plutôt que des propos favorables, polis et réservés (Altheide & Johnson, 1994; Lessard-Hébert, Goyette & Boutin, 1996). Ils se sentaient respectés dans leur intégrité et l'intervieweuse manifestait son objectivité en réprimant toute forme de jugement sur les propos échangés (Fontana & Frey, 1994; Pourtois & Desmet, 1988).

La formation des enseignants en matière d'évaluation des apprentissages étant limitée, le vocabulaire utilisé lors des discussions et dans les entrevues est simple, clair et accessible (Erickson, 1986; Van der Maren, 1995). C'est pour cette raison que les échanges lors des premières rencontres ont porté sur des notions fondamentales d'évaluation des apprentissages, afin de s'entendre sur la terminologie utilisée pour désigner des faits, des pratiques et des concepts spécifiques. Tout au cours de la démarche, la chercheuse est demeurée consciente des risques de méprises et de leurs conséquences sur les résultats de la présente recherche; elle n'a pas hésité à clarifier tout

doute quant au sens des termes utilisés, à tout moment au cours de cette collaboration. La mise à l'essai était accompagnée d'une formation sur mesure des participants. Le but ultime était de les rendre à l'aise avec la notion de qualité des résultats et du concept de justesse pour que leur opinion soit crédible à la fin du projet.

3. Transparence

Comme mentionné précédemment, le chercheur constitue l'outil majeur de la collecte des données. Afin de renseigner les lecteurs et les assurer de la crédibilité de ses propos, le chercheur décrit avec le plus de détails pertinents la fenêtre par laquelle lui parvient l'information qu'il cumule. Une section de ce chapitre s'attarde à décrire l'individualité de chaque participant et insiste sur le contexte politique de la période de collecte des données. La présente recherche s'intéressant à comprendre et à générer une description étoffée d'un phénomène, le profil professionnel des cinq participants présenté dans ce chapitre convainc du caractère plutôt « ordinaire » des participants; ils sont ni plus, ni moins habilités pour cet aspect exigeant de la tâche d'enseignement qu'est l'évaluation des apprentissages. Par ailleurs, leur formation relativement récente, c'est-à-dire durant les années 90, témoigne de manière générale, de la compétence des enseignants en exercice.

Le traitement des données n'est pas une fin en soi mais un moyen de faire avancer la théorisation (Coffey & Atkinson, 1996). Toutefois, les procédures d'analyses ont été documentées avec beaucoup de détails, afin de démontrer leur traitement rigoureux et systématique (Lincoln & Guba, 1985). Comme le suggèrent Erickson (1986) et Muchielli

(1996), les synopsis et les résumés des événements d'intérêts, annexées à ce document, permettent au lecteur d'assister à l'analyse des documents, d'en suivre la progression, de juger de la qualité de l'argumentation et d'en tirer ses propres conclusions. La clé de codification des protocoles verbaux de la première entrevue, le bilan de sa codification, les analyses préliminaires des situations d'évaluation six à dix selon les composantes du concept de justesse, de même que les guides d'entrevue, ont aussi tous été présentés au lecteur, soit intégrés au texte, soit en annexes.

4. Triangulation et saturation

«Puisque aucune procédure en soi n'est suffisante pour rendre compte de la réalité» (Massé, 1992, p. 137), diverses stratégies de triangulation sont appliquées dans cette recherche, en vue de générer des informations riches, utiles et pertinentes, tout en réduisant au minimum les risques de biais (Denzin, 1988). D'abord, toutes les dix situations d'évaluation (avant la mise à l'essai et durant l'application du concept de justesse) sont documentées au moyen de plusieurs pièces-témoins insérées au cahier technique; elles sont également discutées, accompagnées de plusieurs extraits tirés du journal de bord. La prise de position des participants au sujet de la pertinence du concept de justesse des résultats est elle aussi examinée de plusieurs point de vue. Des extraits du journal de bord à ce sujet et les grilles d'analyse des situations six à dix ont servi à alimenter le deuxième entretien et à nuancer les propos des participants. Ces situations d'évaluation étant aussi abordées lors des entretiens, elles ont été documentées de différentes manières. Par ailleurs, le concept de justesse des résultats a été mis à l'épreuve dans cinq situations d'évaluation différentes; il y a donc triangulation des

données. De plus, chaque étape de la collecte des données a été suivie de la préparation de synopsis ou de résumé. Cette analyse préliminaire des informations recueillies a permis de retourner aux acteurs pour vérifier le sens attribué à leurs propos et aux situations observées (Massé, 1992; Pourtois & Desmet, 1988; Van der Maren, 1995), ce que Mucchielli appelle la triangulation indéfinie (Mucchielli, 1996). Enfin, mentionnons simplement que le concept de justesse des résultats est le cadre conceptuel appuyant l'analyse du matériel documentaire puisqu'il est l'objet de la présente recherche; la triangulation théorique n'a donc pas été appliquée à la présente recherche.

La saturation des données est souvent un principe privilégié pour assurer la crédibilité des résultats d'une recherche de type qualitatif; elle indique au chercheur le moment d'arrêter la collecte des données. Puisqu'il a assez d'information pour interpréter les données (Poisson, 1990), il évite ainsi le gaspillage des ressources (Poupart et al., 1997). Dans le cas présent, cinq situations d'évaluation différentes ont été estimées suffisantes pour une première mise à l'essai du concept de justesse des résultats. Cette limite établie avant la mise à l'essai se voulait une manière de respecter les participants et de ne pas abuser de leur générosité. De plus, cinq situations apparaissaient suffisantes pour inspirer les meilleures stratégies à adopter pour faire évoluer le concept de justesse des résultats en évaluation des apprentissages en salle de classe, à plus long terme.

La méthode adoptée pour documenter les deux questions de recherche énoncées au chapitre précédent se voulait un dosage équilibré de rigueur et de flexibilité, afin d'accumuler des données utiles et riches, sans profiter abusivement de l'amabilité des participants, ni dénaturer leur contexte habituel de travail. Les procédures choisies ont permis de documenter quelques pratiques habituelles d'évaluation des apprentissages en salle de classe dans leur globalité, sans les fragmenter et sans les sortir de leur contexte, ce que peu de recherches ont fait jusqu'à maintenant. Ces informations ont servi ensuite à mieux rejoindre les participants dans leur propre conception de l'évaluation, pour amorcer la vulgarisation du concept de justesse des résultats. De plus, l'écart existant entre ces deux conceptions de l'évaluation des apprentissages en salle de classe a permis de nuancer l'opinion des participants au sujet de la viabilité du concept de justesse des résultats.

L'application concrète du concept de justesse des résultats à l'activité évaluative en salle de classe mettait sa polyvalence et sa cohérence à l'épreuve. L'opinion des participants à la fin de la période de mise à l'essai risquait fort d'apporter des éléments enrichissants aux composantes du concept de justesse et d'orienter la recherche future.

Chapitre 3

PREMIÈRE QUESTION DE RECHERCHE

La partie expérimentale de la présente recherche s'est échelonnée sur une période de six mois, en collaboration avec cinq enseignants-participants oeuvrant au sein d'un conseil scolaire francophone de l'Ontario. En plus du nombre incalculable d'heures de préparation, ce projet s'est concrétisé par 45 heures de formation et d'échange individuel ou en groupe, plus de 15 heures d'observation directe en salles de classe, dix entrevues individuelles enregistrées, pour un total de 50 visites dans les deux écoles participantes. La figure 3, à la page suivante, illustre la répartition de ces activités et leur durée respective.

Ce troisième chapitre traite de la question de recherche suivante :

- 1 a) Quels sont les éléments marquants des pratiques d'évaluation des apprentissages en salle de classe menées par les enseignants?
- b) Quels aspects de ces démarches d'évaluation les enseignants considèrent-ils importants ou critiques pour obtenir des résultats justes?
- c) Comment leur conception des éléments pouvant intervenir sur la justesse des résultats se compare-t-elle avec celle du cadre conceptuel de la présente recherche?

École 1		École 2		
Christine	Kevin	Marie	Matilde	Marie-Chantal
rencontre indiv. 30 minutes	rencontre indiv. 30 minutes	rencontre en équipe, 60 min.		
rencontre en équipe, 60 min.		rencontre en équipe, 60 minutes		
rencontre en équipe, 70 min.		rencontre en équipe, 75 minutes		
rencontre en équipe, 70 min.		rencontre en équipe, 75 minutes		
rencontre en équipe, 60 min.		rencontre en équipe, 75 minutes		
rencontre indiv. 30 minutes	rencontre indiv. 30 minutes	rencontre en équipe, 75 minutes		
entrevue 1 40 minutes	entrevue 1 25 minutes	rencontre indiv. 60 minutes	rencontre indiv. 30 minutes	rencontre indiv. 30 minutes
rencontre indiv. 40 minutes		entrevue 1 45 minutes	entrevue 1 45 minutes	entrevue 1 60 minutes
rencontre en équipe, 70 min.		rencontre en équipe, 90 minutes		
rencontre en équipe, 70 min.		rencontre indiv. 60 minutes	rencontre indiv. 60 minutes	rencontre indiv. 90 minutes
observation en classe 3 heures	observation en classe 2 ½ heures	rencontre indiv. 90 minutes	rencontre indiv. 45 minutes	rencontre indiv. 40 minutes
observation en classe 3 ½ heures		rencontre indiv. 90 minutes	rencontre indiv. 30 minutes	rencontre indiv. 30 minutes
observation en classe 90 minutes		observation en classe 3 ½ heures	rencontre indiv. 30 minutes	rencontre indiv. 60 minutes
rencontre en équipe, 75 min.		rencontre indiv. 20 minutes	rencontre en équipe, 90 minutes	
entrevue 2 30 minutes	entrevue 2 30 minutes	rencontre indiv. 60 minutes		observation en classe 40 minutes
		entrevue 2 105 minutes	entrevue 2 45 minutes	entrevue 2 50 minutes

Figure 3. Répartition des diverses activités de la collecte des données et leur durée respective

La première partie du présent chapitre expose d'abord le résumé des situations d'évaluation documentées avant la mise à l'essai, puis en élabore les aspects les plus marquants. La troisième section aborde ensuite les facteurs qui peuvent, selon les participants, contribuer ou nuire à la justesse des résultats. La deuxième partie du chapitre consiste en une discussion des éléments de justesse abordés jusqu'à maintenant et des écarts observés entre la conception des participants et celle du cadre conceptuel de la présente thèse. Ce chapitre se termine par une synthèse des éléments de réponse à cette première question de recherche.

A - PRÉSENTATION DES RÉSULTATS

À prime abord, la première partie de cette question de recherche peut sembler détachée du concept de justesse des résultats, mais sa teneur permet en premier lieu de situer les participants dans leur manière de faire habituelle. Par la suite, en saisissant mieux l'écart existant entre leur conception de l'évaluation avant l'expérimentation et celle qui est exprimée dans le concept de justesse, il est possible de mieux comprendre leurs réactions et opinions à l'égard du concept de justesse abordé en second lieu.

Le projet a commencé par une période de familiarisation avec le contexte de travail de chacun et la documentation détaillée d'une situation d'évaluation de leur choix, vécue au cours de l'année scolaire courante (situations d'évaluation 1 à 5). Cette étape de la collecte des données a grandement contribué à développer un rapport de confiance avec

la chercheure. Même si en principe l'activité aurait pu prendre peu de temps, le déroulement de ces rencontres s'ajustait, autant que possible, au désir et à l'humeur des participants. Certaines rencontres ont été écourtées pour permettre de respecter d'autres engagements ou obligations, d'autres se sont prolongées, devant l'intérêt des participants à parler de leur vécu en salle de classe.

L'intégration de la chercheure au milieu scolaire durant une période prolongée a permis d'observer de nombreux détails au sujet de la tâche des enseignants; ceux-ci sont estimés importants parce qu'ils contribuent à contextualiser l'expérience et à rattacher les informations recueillies à un milieu de vie réel et dynamique. Les enseignants ajustent les activités de la salle de classe avec une telle spontanéité que l'observation de certaines étapes des démarches d'évaluation a parfois été difficile. Les préparations sont généralement faites à la dernière minute et le fait d'avoir à inclure une tierce personne dans ce processus, en l'occurrence la chercheure, dérangeait leurs habitudes.

Les résumés des cinq situations d'évaluation documentées en vue de décrire les pratiques évaluatives des participants sont maintenant présentés, puis les éléments les plus marquants sont soulignés. La notion de justesse des résultats avant la mise à l'essai a ensuite été explorée chez les participants.

1. Résumés des cinq situations d'évaluation documentées

Une situation d'évaluation élaborée par chacun des participants avant le début de la mise à l'essai a permis d'esquisser un profil de leurs pratiques évaluatives. Un résumé de chacune d'elles est présenté ici. Ils ont été soumis aux participants pour vérification et

modifiés lorsque nécessaire en vue d'une représentation fidèle de la situation vécue.

1.1 Situation d'évaluation 1 :

Multiplication de nombres entiers, 5^e année, classe de Christine

Pour l'enseignement des mathématiques, Christine se réfère au manuel Découvertes mathématiques 5 (B. Kelly et al., 1988). L'unité 3 porte sur la multiplication de nombres entiers et fait référence à l'estimation d'un produit après avoir arrondi les nombres. L'épreuve discutée ici est tirée du Guide du maître; Christine y a fait quelques réajustement pour qu'il soit plus court et a réécrit un problème (copie témoin). Pour le réussir, il est essentiel de connaître les tables de multiplications par coeur. Il faut aussi connaître l'algorithme permettant de garder les chiffres à leur place lorsqu'il s'agit de facteurs à deux ou trois chiffres. Il faut aussi savoir arrondir les nombres. Seules les réponses comptent.

Cet unité d'apprentissage a duré environ six semaines de classe, en comptant l'épreuve et le retour sur l'épreuve après. À chaque jour, les élèves avaient un exercice tiré du livre à faire. S'ils ne le finissaient pas en classe, ils devaient le terminer en devoirs. Le lendemain, l'exercice était revu en classe. Des élèves allaient au tableau montrer comment ils avaient fait un certain numéro de l'exercice. Chacun pouvait corriger son devoir dans son cahier ou échanger son cahier avec celui d'un voisin. À cette occasion, les élèves qui ne comprenaient pas pouvaient poser des questions. Lorsque la correction était terminée, l'enseignante faisait un survol rapide des résultats en demandant, par exemple, d'indiquer en levant la main, ceux qui n'avaient pas eu de faute, ceux qui avaient

eu une faute, deux fautes, etc. L'enseignante offrait de l'aide individuelle à ceux qui avaient eu une proportion importante d'erreurs, durant le temps alloué en classe pour commencer le prochain exercice.

Tous les élèves ont une calculatrice personnelle mais Christine ne leur permet pas de l'utiliser pour les mini-tests. Elle les met en garde aussi contre le danger de l'utiliser pour faire les devoirs, plutôt que d'apprendre les tables de multiplications par coeur. Ils sont avertis qu'il sera interdit d'utiliser la calculatrice pour le test à la fin de cette unité et qu'ils s'exposent à d'autres difficultés lorsqu'on abordera l'unité sur les divisions, plus tard dans l'année.

Presqu'à chaque jour, la période de mathématiques commence par un mini-test, dans le but d'encourager les élèves à apprendre leurs tables de multiplication par coeur. Les résultats de chacun à ce mini-test sont consignés dans le registre de l'enseignante (copie témoin). Une fois que l'élève a démontré qu'il sait toutes ses tables de multiplications, il n'a plus à se plier à cet exercice quotidien. Il continue toutefois à faire les exercices assignés en devoirs.

Christine ne se réfère pas aux notes accumulées lors des 14 mini-tests (total de 185 points) pour attribuer une note globale sur cette unité. Elle utilise ces notes pour stimuler les élèves et s'y réfère à l'occasion, lorsqu'elle a à décider si un élève mérite un 1 ou un + dans son bulletin. Les élèves n'ont jamais été mis explicitement au courant de cette pratique. Ils savent cependant que le fait d'être dispensés des mini-tests ne les pénalise pas.

L'épreuve est sur 55, avec note attribuée à chaque numéro indiquée dans la marge. Ce résultat est ensuite transformé en pourcentage puis traduit dans le barème du bulletin (95 à 100 = +; 80 à 94 = 1; 60 à 79 = 2; inférieur à 60 = 3). Cette dernière note est encadrée; l'élève s'y réfère pour savoir s'il dépasse l'objectif visé (+), atteint l'objectif visé (1), est en voie d'atteindre l'objectif visé (2) ou a besoin de soutien et progresse à son propre rythme (3). Cette note est ensuite combinée à d'autres pour être rapportée dans le bulletin sous la rubrique Comprend les concepts d'arithmétique.

Après que Christine a eu noté le test, elle l'a remis aux élèves et l'a revu en entier en classe. Ils ont pu poser des questions et relever des erreurs dans la notation, s'il y en avait. Des 19 élèves dans ce groupe-classe, 2 n'ont pas subi cette épreuve parce qu'ils ont des plans d'intervention personnalisée (PIP). Il y a eu un seul échec au test (36%); 2 élèves ont obtenu la note +, 9 élèves ont obtenu la note 1, 4 élèves ont obtenu la note 2 et 2 élèves ont obtenu la note 3. Christine était contente de ces résultats.

1.2 Situation d'évaluation 2 :

Notions élémentaires de théorie en musique, 6^e année, classe de Kevin

Kevin est un enseignant spécialisé en musique. Alors qu'il n'existe pas de programme précis dans cette discipline et que les enseignants sont laissés à eux-mêmes pour établir la manière de traiter cette matière, Kevin en profite pour enseigner la flûte à bec. En plus de la période hebdomadaire où tous les élèves jouent ensemble, il trouve l'occasion

d'écouter chacun de ses élèves jouer afin de bien l'observer, corriger sa technique et lui prodiguer des recommandations.

Certains de ses élèves apprennent à jouer d'un instrument de musique par le biais de cours privés, certains ont été initiés à diverses formes de musique et finalement, d'autres ne manifestent de l'intérêt que pour la musique populaire du jour. Bref, le groupe est très hétérogène au point de vue connaissances musicales. Après avoir constaté que bon nombre d'élèves avaient de la difficulté à jouer de la musique en classe parce qu'ils ne maîtrisaient pas les notions élémentaires de théorie musicale, Kevin a décidé d'y consacrer du temps. Il a préparé deux exercices (copies témoins), lesquels ont été assignés aux élèves après des explications et directives données par l'enseignant (valeurs des notes, rythmes, mesures, termes musicaux). Les travaux ont été corrigés en classe, accompagnés d'explications pour ceux qui avaient besoins de précisions. Kevin a noté le résultat de chacun sur 15 points et a ensuite traduit ce résultat dans le barème du bulletin. Par exemple, 15 points = +; 12 à 14 points = 1; 10 ou 11 points = 2; moins de 10 points = 3.

Suite à ces deux exercices, Kevin a administré une épreuve (copie témoin). Celle-ci porte sur les mêmes notions que les deux exercices; son contenu vaut 61 points. Les copies sont notées par Kevin et les résultats sont inscrits dans le registre selon le barème du bulletin. Des 26 élèves, il y a eu quelques scores parfait et un seul échec. Le test a été revu en classe.

L'ensemble de ces résultats ne sont pas rapportés dans le bulletin, la seule rubrique pertinente à la musique étant la suivante : Participe et apprécie les activités musicales.

Pour accorder une note à ce sujet, Kevin a évalué à quelques reprises l'effort et la coopération durant la classe de musique. Pour ce faire, il a à 2 ou 3 reprises, noté l'intervention des élèves, comme de lever la main pour répondre aux questions, poser des questions et tenter d'appliquer ce qui leur était demandé; il a aussi tenu compte de l'effort manifesté à jouer correctement, selon le temps personnel mis à s'exercer, et le progrès démontré lors des performances individuelles et en groupe.

1.3 Situation d'évaluation 3 :

Lecture, 3^e année, classe de Marie

Cette activité d'évaluation s'est déroulée en janvier 1997. L'exercice est tiré du manuel MEMO 3 (1989). Chaque élève a reçu un exemplaire des pages 26 à 28 (copie témoin). Marie a préparé des questions auxquelles les élèves devaient répondre afin de démontrer leur compréhension du texte (copie témoin). Ces questions nécessitaient de dégager l'idée principale du texte, de repérer des informations et de faire des inférences.

Marie a prévenu les élèves de l'importance de faire de leur mieux parce qu'il s'agissait d'une évaluation rapportée dans le prochain bulletin. Comme elle fait habituellement pour un exercice de lecture, elle a fait une mise en situation et commencé la lecture avec eux. Elle a écrit des mots au tableau, avec des synonymes, elle a attiré leur attention sur l'organisation du texte (sous-titres) et sur les illustrations accompagnant le texte. Elle a abordé brièvement la profession de réalisateur et établi le parallèle avec celle de directeur d'école. Elle a insisté pour que les élèves répondent aux questions par écrit, ce

qui est habituellement fait oralement, au niveau de la classe quand il ne s'agit pas d'une évaluation. Elle a lu les questions avec les élèves avant qu'ils ne se mettent au travail.

Dès le début de la période de travail individuel, Marie a vu plusieurs mains levées. Elle a constaté que plusieurs élèves avaient de la difficulté à comprendre le sens des questions; d'autres semblaient démunis devant la tâche de trouver la partie du texte qui leur permettrait de répondre aux questions.

Lors de la notation des copies, Marie a constaté que la moyenne des résultats était de 65,8%. Des 25 élèves de sa classe, 3 étaient absents. Il y a eu 13 échecs; la meilleure note était de 12/13, obtenue par deux élèves.

1.4 Situation d'évaluation 4 :

Expression orale en langue seconde, 7^e et 8^e années, classe de Matilde

Pour favoriser l'apprentissage de l'expression orale en langue seconde, Matilde assigne à deux ou trois reprises dans une session, des thèmes pour lesquels ses élèves doivent préparer un exposé oral et le présenter devant la classe. Cette fois-ci, le thème choisi est celui de la musique. Chaque élève peut aborder de la manière qu'il veut (compositeurs, instruments, préférences musicales, etc.). Les élèves disposent de quatre périodes de préparation en classe durant lesquelles ils préparent un plan et le montre à l'enseignante; elle le regarde et fait des suggestions. Les élèves peuvent aussi pratiquer leur présentation en travaillant en dyade. Les élèves reçoivent cette tâche environ deux

semaines avant la première présentation, laquelle devrait durer entre deux et trois minutes, ne comptant pas la présentation d'extraits sonores, illustrations et questions. La plupart des élèves agrémentent leur présentation d'illustrations, d'enregistrement de pièces musicales ou d'objets reliés au thème traité, ce qui augmente considérablement le temps consacré à chacune des présentations. Durant la présentation d'un élève, les autres élèves notent des commentaires ou des questions à poser suite à cette présentation. Elles ont lieu à un rythme qui varie de 10 à 12 présentations par semaines, par groupe-classe. Cette fois-ci, les présentations se sont échelonnées sur une période de deux semaines, (fin janvier - début février 97).

Lors d'une tâche semblable assignée précédemment, Matilde s'était référée à une grille d'évaluation proposée par une autre enseignante (copie témoin). La trouvant trop détaillée, elle a préféré en préparer une nouvelle, plus brève (copie témoin). Elle s'est inspirée de la grille d'évaluation des discours oratoires parrainés annuellement par les Club Optimiste de la province (copie témoin). Cette nouvelle grille a été présentée aux élèves, juste avant la première présentation. Matilde remplit cette grille et y ajoute des détails précis sur la performance de chacun, immédiatement après chaque présentation, mais ne les remets pas tout de suite. Elle se réserve le temps de revenir sur son jugement, la première impression passée. Elle retourne ce formulaire à tous les élèves du groupe au même moment, une fois l'activité terminée.

Les évaluations de présentation orale en langue seconde sont rapportées dans le bulletin sous les rubriques suivantes :

- choisir et organiser les idées en fonction de son intention,

- formuler convenablement les idées et
- présenter efficacement les idées.

Matilde fait une moyenne des notes obtenues au cours du trimestre, pour chacune des rubriques, selon les grilles utilisées lors des présentations. Elle traduit ces notes dans le barème du bulletin, selon l'échelle suivante :

80 % et plus = 1 (L'élève le fait très facilement);

79 % à 70% = 2 (L'élève le fait facilement);

69 % à 60% = 3 (L'élève le fait partiellement);

59 % à 50% = 4 (L'élève le fait difficilement);

moins de 50 % = 5 (L'élève ne le fait pas).

1.5 Situation d'évaluation 5 :

Expression orale, 4^e et 5^e années, classe de Marie-Chantal

Marie-Chantal aborde cette activité dès le début de l'année scolaire et la présente comme une occasion de se pratiquer à parler devant un groupe, comme ils ont à le faire à diverses occasions, soit pour exposer le contenu d'un travail de classe, exprimer une idée dans un rassemblement d'élèves, contribuer aux annonces du matin à l'interphone ou participer à un concours oratoire. Grâce à une rétroaction immédiate, les participants

connaissent tout de suite leurs points forts et leurs faiblesses, et reçoivent des suggestions pour s'améliorer.

Voici comment se passe cette activité hebdomadaire du lundi matin :

L'enseignante demande cinq élèves volontaires pour raconter devant la classe des éléments intéressants de leur fin de semaine (vers la fin d'un trimestre, elle donne priorité aux élèves qui ne se sont pas encore exercés). Les élèves ne sont pas limités dans le temps. Normalement l'exposé dure entre une et trois minutes. Lorsque l'élève fige, l'enseignante l'aide en lui posant une question.

Pendant qu'un élève s'adresse à la classe, les autres l'écoutent et le regardent afin de pouvoir commenter sa présentation par la suite. Le contenu est examiné selon les principes d'expression orale ou écrite mis en valeur en tout temps, comme d'éviter les anglicismes, le langage populaire, faire des phrases complètes, donner des idées complètes, utiliser correctement les verbes, etc. Les éléments d'intérêts sont les suivants :

Gestuelle : regard, posture, gesticulation

Voix : Ton, volume, début

Contenu : Structure des phrases, vocabulaire, ordre des idées

Pendant la présentation, Marie-Chantal note de manière abrégée tout ce qu'elle remarque de positif et de négatif pertinent à ces critères (copie témoin). Quand la présentation est terminée, les élèves dans l'auditoire sont invités à commenter la présentation. Marie-Chantal associe alors à chaque détail qu'elle a noté, le nom de l'élève qui a relevé

cette information; ainsi, elle tient note de tous ceux qui ont participé activement, soit en présentant, soit en écoutant pour identifier des caractéristiques et suggérer des améliorations. Quand les élèves ont terminé de commenter, l'enseignante fait part des éléments qu'elle a notés et qui n'ont pas été soulignés par les élèves.

Marie-Chantal accumule sur une même page ses notes personnelles au sujet de toutes les présentations d'une même journée. Après la présentation, cette feuille est insérée dans un registre, avec celles des présentations précédentes. Elle se réfère à ces notes et à toutes autres performances en expression orale en classe pour accorder une note au bulletin. C'est la même chose pour l'écoute.

2. Éléments marquants des pratiques évaluatives

Tout au cours de la collecte des données, des détails se sont ajoutés au sujet des pratiques évaluatives des participants. Des informations notées dans le journal de bord relatent des détails glanés lors des rencontres et complètent le tableau des pratiques habituelles précédant la familiarisation au concept de justesse des résultats. Ces éléments complémentaires sont abordés sommairement dans la section qui suit. Il s'agit du sens accordé à l'évaluation en général mais de manière spécifique à la notation, puis des périodes de révision avant les épreuves.

À noter que les enseignants ne voient pas l'évaluation des apprentissages en salle de classe avec les schèmes de référence des spécialistes, ni leur jargon habituel. À cette

étape-ci de la recherche, avoir tenté de traduire ou d'adapter leurs propos au vocabulaire standard de la mesure et de l'évaluation aurait été une trop grande menace aux rapports de confiance en voie de s'établir.

L'activité d'évaluation en salle de classe est réglée sur le bulletin. Dans le conseil scolaire où s'est déroulé la présente étude, les enseignants doivent produire trois bulletins durant l'année scolaire (fin novembre, début mars, mi-juin). Il y a donc une période intensive d'évaluation dans les deux semaines précédant la fin du terme, suivie d'une période d'environ quatre à six semaines plus calme à ce point de vue. Les évaluations recommencent ensuite lentement pour culminer vers la période précédant le bulletin suivant.

En documentant les pratiques évaluatives des enseignants-participants, deux thèmes retiennent l'attention : l'attribution de notes chiffrées et la préparation des élèves aux épreuves. Ce sont maintenant les sujets abordés.

2.1 Attribution de notes chiffrées

Les participants à cette recherche utilisent de préférence des épreuves déjà préparées, quitte à les modifier si nécessaire. C'est ce que Matilde a tenté de faire, pour finalement se résoudre à en préparer une elle-même. Elle trouve cependant difficile de concilier les aspects essentiels d'une présentation orale en langue seconde et les énoncés du bulletin. Après plus de la moitié de l'année scolaire d'écoulée, Matilde réajuste encore la grille d'évaluation des présentations orales en langue seconde. « J'ai rien qui évalue le geste, j'ai rien qui évalue le regard. ... souvent les élèves vont arriver en avant avec ...un texte,

puis ils vont le lire. Donc si j'évalue le regard, si ils regardent leur public, ils vont le dire plus dans leurs propres mots plutôt que le lire. (Matilde, entrevue 1, ligne 20)

En troisième année, Marie remarque que les élèves n'ont pas encore acquis le sens des notes chiffrées. Cependant, Christine et Kevin, en cinquième et sixième années, accumulent beaucoup de points mais ils ne comptabilisent pas. Est-ce parce «les élèves veulent toujours que ça compte»? (Matilde, entrevue 1, ligne 112). Cette dernière poursuit en disant «je leur dis que ça compte, puis ça compte pas vraiment, pour qu'ils le fassent, ou je le compte pour le vrai, puis j'attribue un petit pourcentage par rapport au test ou à l'évaluation vraiment formative». Christine dit avoir été contrainte d'agir ainsi au cours de sa première année d'enseignement par le directeur d'école, mais qu'elle ne fait plus cela. Cependant, ses initiatives démontrent encore une tendance à attribuer des points aux élèves pour faire les activités d'apprentissage.

Le besoin se faisant sentir, il a fallu préciser la distinction entre «faire des activités d'apprentissage» et «évaluer l'apprentissage réalisé». L'extrait suivant est tiré du journal de bord de la chercheure :

Je dis qu'en apprentissage, l'enseignant aide les élèves à évoluer et à progresser, à aller plus loin; on ne peut pas contribuer au projet de l'élève avec des suggestions et des corrections d'une part et d'autre part, accorder des notes au travail effectué - lequel comporte une contribution parfois importante de la part de l'enseignant. ... C'est la stupéfaction! ... Elles [les trois participantes présentes] ne peuvent pas concevoir de faire travailler l'élève sans lui attribuer des points - ces points qui constituent en fait, dans bien des cas, la note qui est sensée

représenter où l'élève se situe dans son développement. (journal de bord, p. 30-31)

Marie-Chantal se distingue des autres participants en utilisant réellement l'évaluation comme un outil d'apprentissage. Lors de l'activité documentée, tous les élèves sont invités à commenter de manière positive la performance d'un des leurs, et non pas lui accorder une note. L'activité a lieu tous les lundis matins, moment le plus approprié pour raconter un événement de la fin de semaine; elle choisit des élèves volontaires, jugeant important qu'ils soient motivés à la tâche. Les traces, conservées dans un registre, indiquent la participation de chacun des élèves et permettent de documenter un progrès éventuel. C'est la seule occasion qu'il a été donnée d'observer l'application de l'évaluation formative, au vrai sens du terme.

2.2 Préparation des élèves aux épreuves

Lors de notre première rencontre, le directeur de l'école où Kevin et Christine enseignent a précisé sa vision de l'évaluation : «vaut mieux prévenir que guérir». Il s'est expliqué en disant recommander à ses enseignants de prendre les mesures nécessaires pour que leurs élèves réussissent. La stratégie encouragée est de donner ce qu'il appelle des pré-tests. Voici comment se reflète cette approche chez les deux enseignants-participants de cette école.

Il y a toujours une période de révision avant le test, qui est habituellement une semaine avant l'évaluation, parce que j'annonce mes tests une semaine à l'avance. ... Un moment donné si je l'annonce 2-3 jours avant, ils me disent «bien

Madame, c'est pas une semaine». Alors je leur dis : «non, c'est vrai, cette fois-ci c'était comme 3 jours, mais je vais passer plus de temps avec vous pour vous montrer comment étudier, quoi étudier». Alors la révision se fait toujours : «Voici, ça c'est important, ça c'est moins important, ça lisez-le rapidement ...» (Christine, ligne 63)

Dans ce test-là je peux te dire que si ils ont oublié, qu'ils savent pas comment arrondir, c'est pas par ma faute parce qu'au tableau j'avais fait quatre exemples. Je me souviens, avant, je leur ai dit : «Vous vous souvenez, vous devez arrondir, c'est quoi arrondir? On en a fait quatre exemples au tableau» avant de leur remettre le test entre les mains. *(Et c'est resté au tableau le temps du test?)* Non, c'est pas resté mais je leur ai dit : «Regardez ça, puis souvenez-vous en». Puis quand je l'ai remis [le test], je leur ai dit «le numéro 15, 16, 17 c'est ce qu'on a fait au tableau». (Christine, entrevue 1, ligne 151)

Et voici comment Kevin dit préparer ses élèves aux épreuves :

Je fais des exemples de questions que je pourrais mettre sur l'épreuve, des exercices, qu'ils peuvent faire le transfert, l'appliquer dans d'autres choses ... j'essaie d'avoir des questions pour les préparer, ... la journée avant l'épreuve ... je donne des questions exactement comme je vais faire à l'épreuve. Pour qu'il y ait aucune surprise. Des fois la question est difficile à comprendre; ils connaissent la réponse mais ils sont pas toujours certains comment répondre. Donc, aussi la journée de l'épreuve, je m'assure, je distribue les feuilles puis je m'assure que chaque numéro est clair. Je lis les questions avec eux puis j'essaye de faire un

exemple de chaque numéro, afin de les guider, pour faciliter leur tâche (Kevin entrevue 1, ligne 47)

Malgré cela, il arrive à Kevin de trouver que les résultats ne rendent pas justice à un élève, ce qui l'amène à devoir réajuster la note:

Je prends quand même sa copie mais j'essaye d'aller chercher d'autres notes dans son cahier, sur des travaux, ou différents exercices qu'on a fait dans la salle de classe, qui pourraient mieux me guider parce que je sais que la personne comprend bien la matière. Souvent c'est juste la façon, qu'il se sent pas bien, ou qu'il a manqué des jours, et qu'il va se reprendre. Pour que ce soit juste. (*Tu compenses avec d'autres informations?*) ... d'autres informations pour cet élève-là. Parce que je sais que vraiment il est capable puis qu'il comprend vraiment la matière, c'est juste qu'il a de la difficulté cette journée-là, puis ça arrive à tout le monde. Tout le monde peut avoir une journée qu'il se sent pas bien puis qui va pas bien réussir (Kevin, entrevue 1, ligne 79).

Kevin a tendance à se faire assez rapidement une opinion des capacités de ses élèves et à rechercher ensuite des comportements qui appuient les opinions déjà établies. Il a confiance dans la concordance des résultats obtenus à diverses occasions, comme il l'explique en parlant de l'évaluation de la théorie musicale : «ceux qui maîtrisaient bien l'instrument, là, y'avaient pas de problème, c'est ceux qui ont très bien réussi [l'épreuve théorique]» (Kevin, entrevue 1, ligne 12)

3. Justesse des résultats

Après avoir documenté les pratiques d'évaluation des enseignants-participants, ceux-ci ont été invités à discuter des aspects de leurs démarches d'évaluation qu'ils considèrent importants ou critiques pour obtenir des résultats représentant bien ce que l'élève a acquis ou développé. L'entretien individuel a d'abord été précédé de la lecture du résumé de la situation d'évaluation documentée chez le participant; ceci lui a permis de se rafraîchir la mémoire et de se recontextualiser dans la situation d'évaluation décrite en première partie du présent chapitre. Une copie des questions d'entrevue a ensuite été présentée aux participants et à leur demande, quelques minutes leur ont été accordées pour noter des éléments de réponse. Comme on peut voir dans le guide d'entrevue utilisé (annexe B), ces questions abordent de diverses manières l'élément d'intérêt; il y est à la fois question des facteurs favorables et des facteurs nuisibles à la justesse des résultats spécifiques à la situation d'évaluation documentée, mais aussi de manière générale.

Une fois terminés, ces entretiens ont été transcrits intégralement. Après plusieurs séances d'écoute, des synopsis ont été préparés et soumis à chaque participant pour fin de vérification. Ceux-ci résument l'essentiel de la discussion tout en conservant le dynamisme de l'échange (annexe C); ils permettent au lecteur de s'approcher le plus près possible des situations réelles et, par le fait même, de mieux suivre la discussion qui suit. Les protocoles verbaux ont ensuite été étudiés; une clé de codification des unités de sens a été élaborée (annexe D) et le codage a été revu plusieurs fois pour s'assurer de sa convenance, évitant le glissement de la procédure tout au cours de cette démarche.

Comme on peut le voir, les composantes du concept de justesse, telles que décrites dans le cadre conceptuel, constituent l'essentiel de la clé de codification. Elles font ressortir les éléments déjà énoncés et attirent l'attention sur ceux qui y sont étrangers. Un bilan des thèmes abordés au cours de chaque entretien illustre le sujet des propos tenus à cette occasion (annexe E). Dans la discussion qui suit, il est question des circonstances entourant l'évaluation, de l'objet de l'évaluation, de la variation des questions utilisées, de traces des présentations orales et de la compétence des enseignants.

3.1 Objet de l'évaluation

Au cours des échanges, les participants ne se réfèrent pas souvent aux habiletés ciblées, tant par l'activité pédagogique que par l'évaluation. Ceci explique qu'une seule catégorie ait été attribuée pour cette composante du concept de justesse dans la clé de codification des protocoles d'entrevue; les propos des enseignants à ce sujet ne sont pas assez élaborés pour être divisés en sous-catégories, comme le confirme la constatation suivante notée dans le journal de bord de la chercheure : «J'aborde alors les composantes du concept de justesse et attire leur attention sur la nécessité d'établir d'abord les habiletés visées par ce projet. ... C'est comme si je parlais une autre langue; elles me parlent de choses que les élèves doivent faire. Je dis qu'elles doivent bien viser à développer certaines habiletés bien précises quand elles enseignent? Quelles sont donc ces habiletés?» (journal de bord, p. 30). La même difficulté apparaît lorsque Marie-Chantal travaille à la préparation du projet technologique, «tout ce qu'elle peut préciser, c'est qu'elle veut contraindre les élèves à utiliser un moteur électrique plutôt que la force humaine pour actionner le mécanisme» (journal de bord, p. 34).

La notion de critère est liée de près à l'objet d'évaluation. À ce sujet, Matilde insiste sur l'importance d'évaluer selon des critères et d'appliquer les mêmes pour tous les élèves. Comme les autres participants, elle n'a pas l'habitude de préciser les critères d'évaluation au moment d'assigner une tâche d'évaluation, ni de les définir clairement, ni de les communiquer aux élèves. Il n'est donc pas question de s'assurer qu'ils les ont intégrés. Certains participants, comme Matilde, se préoccupent des critères au moment de la notation des travaux; d'autres se basent sur la concordance entre la réponse de l'élève et la réponse attendue pour décider du nombre de points à attribuer à la réponse. En général, les enseignants-participants préfèrent noter au moyen de «points» et ne pas communiquer ou discuter des critères d'évaluation avec les élèves.

Plusieurs participants ont exprimé leur préoccupation pour des tâches d'évaluation motivantes, sans erreur technique, de longueur et de difficulté appropriée, nouvelle mais familière. Cependant, quel sens accordent-ils à ces qualificatifs? Les mettent-ils en pratique? Mais surtout, se préoccupent-ils de les agencer aux habiletés visées par l'activité d'évaluation? Matilde et Marie semblent conscientes, intuitivement, de l'importance de bien cibler les habiletés évaluées. Matilde, par exemple, dit être consciente que de pénaliser un élève en lui enlevant des «points» s'il n'a pas respecté une échéance réduit sa moyenne, laquelle ne le situe plus correctement dans son développement. Mais elle n'a pas pour le moment d'autres moyens de contraindre ses élèves à respecter les échéances. De son côté, Marie parle de l'évaluation provinciale de la lecture, elle croit qu'on peut embrouiller les résultats d'évaluation en faisant appel indirectement à d'autres habiletés pas encore maîtrisées. À son avis, le niveau d'autonomie d'un élève de troisième année peut considérablement biaiser les résultats

d'évaluation de la lecture, comme de toutes autres habiletés.

L'évaluation provinciale de la lecture, de l'écriture et des mathématiques en troisième année a créé des remous. Après la journée de préparation offerte par l'*Office de la qualité et de la responsabilité en éducation* en vue de l'administration des activités évaluatives, Marie a constaté que ses élèves n'étaient pas aussi autonomes que l'exigeait cet examen. Elle a donc tenté de développer leur autonomie dans le contexte de l'évaluation de la lecture, puisque de toute façon elle devait évaluer ses élèves pour le prochain bulletin. Elle a modifié une épreuve de lecture fournie dans le livre du maître (*Mémo*, 3e année). L'année précédente, ses élèves avaient eu de la difficulté avec le vocabulaire des questions et le feuillet de réponse lui avait paru compliqué. Malheureusement, « ... il y avait quand même des difficultés. Les élèves n'ont pas l'habitude tellement d'écrire leurs réponses. On fait beaucoup de travail à l'oral. On lit en groupe et après on discute en groupe ... au fond peut-être qu'ils ont besoin de plus de cette activité-là individuelle pour aller chercher et écrire leurs réponses eux-mêmes. » (Marie, entrevue 1, ligne 28)

3.2 Stratégies utilisées

Parmi les stratégies utilisées pour s'assurer d'arriver à des résultats situant bien les élèves dans leur développement, Christine rappelle une recommandation fréquente des guides techniques de rédaction des épreuves : varier le type de question. Voici un extrait de son témoignage :

Mes évaluations ou mes travaux ou les devoirs, ou peu importe ... ce qui mérite d'être évalué, c'est ça, varie beaucoup d'une fois à l'autre, et varie

souvent dans le même travail. Par exemple, je vais donner un exemple concret, mes évaluations, habituellement, ont une partie où ils doivent donner une définition ou expliquer du vocabulaire. Que ce soit en sciences, que ce soit en sciences sociales, que ce soit en français, ... peu importe. Alors, il y a une partie descriptive, description, définition. Ensuite j'ai toujours une partie vrai ou faux, une partie choix multiples, une partie remplir les tirets, une partie relier la colonne de droite à celle de gauche, mini mots-croisés, jeux de mots, pour que l'enfant qui peut s'exprimer, s'exprimer facilement, puisse répondre très bien aux définitions, mais l'enfant qui ne peut pas s'exprimer, mais qui connaît la définition, mais il n'a pas les mots, il va pouvoir s'exprimer quand-même avec un autre numéro. (Christine, entrevue 1, ligne 45)

Cet extrait de l'entretien ne portait pas spécifiquement sur l'épreuve de mathématiques documentée (situation d'évaluation 1); les propos de Christine laissent donc croire qu'elle applique ce principe à tout type d'évaluation. Cependant, les pièces témoins de la situation d'évaluation documentée chez cette enseignante étaient entièrement composées de multiplications décontextualisées où les élèves devaient donner la bonne réponse. Pourquoi n'applique-t-elle pas ce principe dans toutes les disciplines, comme elle dit le faire?

L'influence du bulletin se fait sentir dans les propos des participants. Même s'ils connaissent l'importance d'évaluer au moment opportun, c'est-à-dire après une période d'apprentissage convenable, certains reconnaissent que de vouloir ou de devoir produire

des notes pour un bulletin peut nuire à l'obtention de résultats d'évaluation justes. «J'arrive des fois puis il me manque des notions, il me manque des notes pour le bulletin, ... donc là je me dis bon bien faut que j'enseigne quelques petites choses et puis que j'évalue, juste pour que je me sente rassurée là dedans.» (entrevue 1, Christine, ligne 123).

Paradoxalement, même si elles acceptent que l'apprentissage évolue sans cesse, rendant les résultats vite périmés, Matilde et Marie-Chantal auraient davantage confiance en plusieurs évaluations pour arriver à des résultats justes, parce qu'elles pourraient faire des moyennes. Elles disent cependant avoir rarement le temps de ré-évaluer les mêmes compétences et seule Marie-Chantal permet à ses élèves de reprendre une évaluation s'ils le désirent.

3.3 Traces accumulées

Dans le cas des présentations orales, l'évaluation s'est échelonnée sur deux semaines; Matilde voudrait pouvoir revenir sur la performance de chaque élève pour s'assurer de la stabilité des exigences. «Ça m'est arrivé d'enregistrer des présentations orales et puis de cette façon là je peux ré-écouter dans un milieu qui est peut-être plus calme, parce qu'en classe, souvent, mon attention est portée vers les élèves qui n'écoutent pas puis je manque des bouts et puis en ré-écoutant, ça me permet de mieux voir ... entendre ce que l'élève a dit.» (Matilde, entrevue 1, ligne 55). Elle revient tout de même sur les notes attribuées après quelques copies ou performances, pour essayer d'appliquer le même degré de sévérité à chacune. Au sujet de l'évaluation des présentations orales, elle dit : «au niveau des présentations orales, ils me remettent souvent, s'ils ont un texte à écrire, je regarde le texte, je le relis pour vérifier ... le vocabulaire» (Matilde, entrevue 1, ligne

61). Elle souligne à nouveau l'importance de garder des traces des performances des élèves.

La stabilité des exigences préoccupe aussi Marie-Chantal. Selon elle, l'humeur de l'enseignant est un facteur pouvant intervenir sur la justesse des résultats d'évaluation. Elle-même, si elle se sent fatiguée, corrige plus sévèrement. «Peut-être pas autant dans des domaines comme les mathématiques, où est-ce que souvent c'est une réponse précise que tu trouves ... mais plutôt, peut-être, au niveau de l'écriture, puis des projets, quelque chose comme ça.» (Marie-Chantal, entrevue 1, ligne 146)

3.4 Circonstances entourant l'évaluation

Dans l'ensemble, les participants connaissent bien les circonstances pouvant altérer la justesse des résultats. Même s'ils ne peuvent intervenir sur la disposition mentale de chacun, sur les répercussions d'absences répétées et d'éléments distrayants dans l'environnement, ils disent se soucier du moment de la journée et de la semaine où ont lieu les activités d'évaluation; ils ont instauré une routine pour sécuriser leurs élèves et ils s'assurent d'allouer assez de temps, même aux élèves les plus lents, pour qu'ils puissent compléter les épreuves. Ils aiment être disponibles pendant que les élèves s'exécutent, soit pour apporter des précisions ou simplement rassurer des élèves inquiets en répondant à leurs questions. Seule Marie a mentionné l'importance d'une mise en situation motivante. Dans la classe de Kevin, les élèves ne bénéficieront pas de mise en situation avant une activité d'évaluation de la lecture; les textes assemblés pour cette activité se passeront même de page titre, laquelle faisait le lien entre des extraits portant sur la sécurité.

3.5 Compétence des enseignants

Dans ses écrits, Mislevy (1993) souligne l'importance de connaître la manière dont se construit le savoir, pour l'évaluer correctement. Or, des enseignants-participants admettent bien modestement qu'ils n'ont pas les connaissances nécessaires pour bien enseigner tous les contenus des programmes prescrits. Par exemple, Marie dit se sentir incompétente à enseigner la danse ou le dessin et par le fait même, de l'évaluer. Il arrive à Marie-Chantal de douter du choix des activités d'apprentissage si ses élèves ne réussissent pas une évaluation aussi bien qu'elle s'y attendait. Christine, elle ne peut admettre qu'un projet de recherche comme celui qui est élaboré pour la foire scientifique puisse comporter aussi bien une question de recherche qu'une d'hypothèse. Elle dit aussi que généralement, un élève qui ne sait pas ses tables de multiplications n'apprend pas l'algorithme de la multiplication; elle ne voit donc pas la nécessité de distinguer ces deux habiletés lors de l'évaluation.

Kevin ne maîtrise pas l'enseignement de la lecture comme il maîtrise l'enseignement de la musique. Plus tard, lors d'une situation d'évaluation de la lecture, il spécifiera à ses élèves, en début d'activité, que les réponses doivent être textuelles. Pourtant, il avait dit bien connaître l'ouvrage de référence pour l'enseignement et l'évaluation de la lecture (Giasson, 1995) pour l'avoir étudié l'année précédente à la formation à l'enseignement.

À eux cinq, les enseignants-participants ont énuméré de nombreux éléments pouvant affecter la justesse avec laquelle un résultat d'évaluation situe un élève dans son développement. Certains de ceux-ci étaient attendus, d'autres sont nouveaux. Selon le bilan de la codification des protocoles verbaux de l'entretien (annexe E), Marie-Chantal

est de loin la participante la plus avertie en ce qui peut altérer la justesse des résultats d'évaluation; on se rappellera aussi qu'elle est la seule participante à avoir démontré une application de l'évaluation formative au vrai sens du terme. Marie, enseignante de troisième année, semble centrée sur l'importance de bien cibler les habiletés à évaluer et elle se préoccupe beaucoup de la tâche choisie pour le faire. Serait-ce par hasard parce qu'elle enseigne à un niveau scolaire où la maîtrise de la lecture et le degré d'autonomie de l'élève peut sérieusement altérer la justesse des résultats de toutes autres habiletés?

Les éléments attendus mais mentionnés par un seul participant (la plupart du temps par Marie-Chantal) sont de s'assurer que les élèves ont intégré les critères servant à juger la qualité de leur performance, de formuler les résultats en termes de forces et de faiblesses, identifiant du même coup ce qui est maîtrisé et ce qui ne l'est pas et de faire participer l'élève à l'évaluation de ses apprentissages. Les mises en situations préalablement à une évaluation sont rares, tout comme les droits de reprise et l'identification claire des contenus concernés. Comme ce bilan le démontre, la composante transparence est la moins appliquée par les participants, suivie du lien de l'évaluation avec l'apprentissage.

Les éléments de justesse présentés dans le chapitre 1 mais omis par tous les participants renseignent tout autant sur ce qui contribue, à leurs yeux, à produire des résultats d'évaluation justes que les nouveaux éléments qu'ils ont apportés. Ces derniers peuvent soit enrichir le concept de justesse, soit simplement témoigner de conceptions erronées

de ce qui peut influencer la justesse des résultats. La pertinence de tous les éléments nouveaux est discuté dans cette section.

B - INTERPRÉTATION DES RÉSULTATS

La deuxième partie de ce chapitre aborde maintenant la conception générale de l'évaluation selon les participants et de ce qui constitue, à leurs yeux, des facteurs pouvant affecter la justesse des résultats d'évaluation.

Les thèmes qui ressortent de cette discussion sont l'attribution continue de points, l'importance des critères et des schèmes de notation, la conservation de traces et l'évaluation, reflet du type d'apprentissage valorisé et de la compétence enseignante.

1. Attribution continue de notes chiffrées

Les constatations faites lors de la collecte des données et de leur analyse ne diffèrent guère de ce que la recension des écrits avait laissé présager : le sens restreint accordé au mot évaluation (Forgette-Giroux et al., 1996), son processus parfois bâclé et, malgré tout, la confiance des enseignants dans leur manière de faire. Contrairement à Desrosiers, Godbout et Marzouk (1992), rien ne porte même à croire que d'autres participants que Marie-Chantal comprennent les principes fondamentaux d'évaluation formative. Les situations exemplaires sont très rares.

Pour les participants, évaluer c'est attribuer des «points» pour l'exécution d'une tâche reliée à un contenu enseigné dans le but d'encourager les élèves à travailler; ceux-ci, habitués à accumuler des points pour toute forme de travail, incluant les activités d'apprentissage, voient mal la nécessité de travailler quand «ça ne compte pas». Et l'engrenage est engagé. Comme le dit Brookhart : «En milieu scolaire, c'est comme si les notes étaient traitées comme des jetons que l'on peut mériter en contrepartie du travail effectué» (1993, p. 139, traduction libre). Les enseignants se plaignent de cette dépendance aux notes, mais attribuer des «points» aux activités d'apprentissage correspond, pour eux, à de l'évaluation continue ou de l'évaluation formative (Bachor & Anderson, 1994), utilisée plus tard à une fin sommative. Ils ne semblent s'être jamais interrogés sur la convenance de telles pratiques.

L'importance de résultats justes semble survenir trois fois durant l'année scolaire : quand il faut préparer un bulletin. Afin d'arriver à des résultats justes, les enseignants participants aiment bien avoir accumulé plusieurs résultats dans chaque discipline pour en faire une moyenne; celle-ci leur semble alors détachée de toute subjectivité. Pourtant, situer un élève dans son développement à partir d'une moyenne va à l'encontre du fait que l'apprentissage est en continuelle évolution; le résultat obtenu lors des premières évaluations ne reflète qu'un niveau de maîtrise provisoire. Comme il y a apprentissage même lors d'une activité d'évaluation, il est important d'arriver, à partir d'une seule séance d'évaluation, à un résultat juste. Cependant, appliquer des stratégies variées pour évaluer les mêmes contenus est contraire au concept de justesse selon lequel il y aurait pour chaque habileté, un type de tâche d'évaluation particulièrement approprié. Cette préoccupation mentionnée par Christine ne semble d'ailleurs pas bien comprise puisqu'il

s'agit en fait du même genre de question : questions à réponses courtes et questions objectives, types correspondant au rappel de connaissances.

La nécessité d'accorder une valeur moindre aux résultats d'évaluations menées avant que les apprentissages soient consolidés, comme Matilde dit le faire, est d'une certaine manière rassurante. Une telle démarche reflète le besoin intuitif de neutraliser, en partie, les résultats qui ne sont pas justes. Mais, d'une autre manière, peu importe le calcul effectué, il ne rendra jamais justice à l'élève. De plus, du point de vue de la transparence de la démarche, comment justifier une telle pratique à ses élèves? Les notes utilisées comme des «jetons» scolaires n'ont pas leur place dans le concept de justesse des résultats; en plus d'apporter une surcharge de correction et de gestion, cette pratique nuit à la confiance à avoir dans les résultats pour situer un apprenant dans son développement.

2. Critères et schèmes de notation

Que des enseignants soulignent l'application uniforme des critères d'évaluation pour convaincre de l'objectivité de leurs pratiques n'est pas sans rappeler la proximité de l'interprétation normative du rendement des élèves. Cependant, la présence de critères ne garantit pas nécessairement l'objectivité des pratiques; c'est pourquoi le concept de justesse des résultats insiste sur la clarté de leur définition et encourage la description de performances en termes qualitatifs. Toutefois, que ces critères soient réalistes ajoute à leur qualité; cette caractéristique appuie le choix d'une tâche motivante et complète la définition de critères typiques des habiletés en milieu authentique. Voilà donc un élément

enrichissant de la composante «Pertinence de la tâche d'évaluation par rapport aux habiletés visées» du concept de justesse.

L'humeur de l'enseignant au moment de la notation et la comparaison des performances fournies entre elles avant d'arrêter son jugement démontrent la nécessité de recourir à des schèmes de notation détaillés, de préférence de type descriptif, plutôt qu'à des échelles numériques. Un tel schème n'est pas nécessairement à toute épreuve mais offre tout de même un appui ferme à l'établissement d'un jugement approprié. D'ailleurs, l'enseignant qui tient à ce que les élèves respectent les échéances devrait en faire un critère de notation, à condition bien sûr de promouvoir une habileté présentée dans le programme scolaire.

Vérifier la concordance d'un résultat d'évaluation avec ceux qui sont généralement obtenus par un même élève n'est pas une pratique nouvelle. Airasian, Kellaghan et Madaus (1977) en faisaient état dans une recherche menée il y a plus de 20 ans. Comme l'explique Wilson (1990), une telle pratique neutralise tous les efforts déployés pour arriver à des résultats justes. Modifier une note en fonction d'une intuition personnelle profonde ouvre la porte à un traitement inéquitable des élèves (Brookhart, 1993). Le concept de justesse ne justifie en aucun temps la modification d'un résultat, même en cas de désavantages circonstanciels; ce concept ne sert qu'à doser le degré de confiance dans un résultat et à encourager à reprendre une évaluation si nécessaire. De plus, modifier une note sans s'appuyer sur des preuves tangibles est une façon efficace de convaincre les élèves de la subjectivité des notes et de leur absence d'influence sur leur destinée. L'appui d'un schème de notation descriptif pourrait prévenir de tels comportements. La

description de la qualité d'exécution d'une tâche selon des critères précis plutôt que l'utilisation d'une note chiffrée ferait ressortir l'incongruence de modifier un résultat sans preuves suffisantes.

3. Conservation de traces

Matilde a mentionné l'importance d'accumuler des traces de la performance évaluée afin de pouvoir y revenir, si nécessaire, après une période de recul. Elle fait ressortir ce besoin après avoir elle-même évalué des présentations orales en classe échelonnées sur plusieurs jours. Cette stratégie n'est pas mentionnée explicitement dans le concept de justesse et la revue de la documentation ne l'avait pas fait ressortir non plus. L'utilisation d'un magnétophone ou d'un magnétoscope pourrait cependant ajouter au stress de la situation d'évaluation, en plus d'en faire une situation fort différente des situations d'apprentissage vécues auparavant. Cette préoccupation, bien réelle, est tout à fait particulière à l'évaluation en salle de classe et constitue un ajout considérable à la composante de la «Pertinence de la tâche d'évaluation par rapport aux habiletés visées», sans toutefois régler la difficulté présentée par la collecte de ces traces de manière discrète.

4. Évaluation, reflet du type d'apprentissage valorisé

Même si le concept de justesse souligne l'importance de prévenir les élèves d'une évaluation éventuelle, l'insistance de Christine à mettre une semaine complète entre le moment d'annoncer une évaluation et le moment où elle a lieu n'apparaît pas justifiée. Ce temps consacré à préparer ses élèves à l'épreuve, le réchauffement tout juste avant de

distribuer les copies et l'utilisation de questions en tous points semblables à l'exercice de revue effectuée la veille, comme Kevin dit le faire, témoignent plutôt de deux faits. D'abord, les épreuves font appel à des habiletés de pensée de bas niveau, comme le rappel de connaissances et l'application d'algorithmes, et non pas de réflexion, de jugement, d'analyse ou de synthèse. Deuxièmement, ces habiletés sont peu intégrées. Leur apprentissage superficiel exige la mise en forme des élèves juste avant de les évaluer; l'utilisation des questions de revue en témoigne aussi. Autre indice de leur peu d'habitude aux questions exigeant la réflexion : l'évaluation suite à la foire scientifique exigeait de compléter un remue-méninge déjà commencé et d'en tirer un problème, une question ou une hypothèse à explorer. Kevin et Christine ont trouvé cette épreuve beaucoup plus difficile que celle qu'ils auraient préparée eux-mêmes. Pour celle-ci, les élèves avaient beau étudier à l'avance, ils ne pouvaient pas mémoriser la réponse; ils devaient démontrer leur compréhension après avoir été exposés à plusieurs projets différents, tous nés d'un remue-méninge. Annoncer une épreuve longtemps à l'avance et préparer formellement les élèves en vue de cet événement, comme Christine et Kevin le croient nécessaire, contribuent fort probablement à biaiser les résultats obtenus. Cette procédure pourtant si habituelle pour eux témoigne plutôt du peu de complexité des apprentissages valorisés dans ces salles de classe.

5. Évaluation, reflet de la compétence enseignante

Kevin, enseignant de sixième année et spécialisé en musique, n'est pas affecté par l'absence de programme rigoureux d'enseignement de la musique. Alors que beaucoup de ses confrères se limitent à faire écouter de la musique variée afin d'élargir les

connaissances musicales de leurs élèves, lui, enseigne la flûte à bec et la guitare, intègre les notions théoriques pertinentes et trouve même le temps d'évaluer individuellement la technique instrumentale. Il choisit spontanément des stratégies efficaces pour amener ses élèves à développer leur talent musical. La connaissance approfondie du développement de la performance musicale amène Kevin à aborder cette discipline de manière efficace et originale tout en s'appuyant sur l'évaluation pour aider ses élèves à améliorer leur technique. Cependant, il se sent obligé d'évaluer les connaissances théoriques de manière traditionnelle, sans s'interroger sur la pertinence de recourir à une épreuve papier-crayon et de comptabiliser les bonnes réponses et les erreurs. Le temps consacré à gérer l'aspect évaluation de cette situation-témoin aurait pu être utilisé à meilleur escient. Comme l'ont conclu Fraser & Tobin (1991), les limites des connaissances pédagogiques de l'enseignant déterminent la qualité de son enseignement et par ricochet, son évaluation.

Des participants soulignent l'importance de la compétence dans le domaine visé, comme s'ils reconnaissaient devoir bien connaître l'apprentissage et son développement spécifique pour être en mesure d'accompagner l'apprenant correctement (Tobin et al., 1990). Or, au cours de la présente recherche, une méprise souvent observée est l'évaluation de la compréhension de textes précédée par la lecture à haute voix des morceaux choisis, soit par l'enseignant, soit par les élèves, à tour de rôle. Aussi, la directive de Kevin à ses élèves spécifiant que les réponses aux questions de compréhension de texte doivent être textuelles, force la constatation suivante : aucun modèle d'évaluation des apprentissages ne peut compenser une lacune au niveau de la

formation des enseignants. De plus, un enseignant qui ne connaît pas ses faiblesses est peu susceptible de chercher à s'améliorer.

Le concept de justesse a pour avantage d'attirer l'attention sur la nécessité de déterminer correctement les habiletés visées; il faut cependant que les enseignants en comprennent tout le sens. Mais n'est-il pas réaliste de penser qu'un enseignant qui ne se sent pas à l'aise avec l'enseignement et l'évaluation de certaines habiletés, comme le dessin, la danse ou même la lecture, engage une démarche personnelle pour combler ses déficiences didactiques? Les programmes cadres et les manuels d'enseignement de ces disciplines devraient faciliter l'établissement des critères d'évaluation des performances des élèves. Mais le problème demeure pour les enseignants inconscients de leurs carences professionnelles et qui ne cherchent pas, par conséquent, à s'améliorer.

Dans la même veine d'idées, l'enseignant optant pour une démarche d'enseignement inefficace devrait s'apercevoir, s'il est aux aguets, que les habiletés ne se développent pas proprement, bien avant d'en arriver à une évaluation formelle. Et même si ce n'était pas le cas, l'absence généralisée de progrès lors de cette évaluation devrait mettre la puce à l'oreille et, par conséquent, indiquer que ces résultats d'évaluation ne sont pas justes.

C - SYNTHÈSE DES RÉSULTATS

Comme discuté au deuxième chapitre, les politiques du ministère de l'Éducation et de la Formation de l'Ontario ont considérablement changé au cours des quatre dernières années. Les revirements soudains ont miné la motivation des enseignants à se familiariser avec les nouveaux programmes et les mettre en oeuvre le plus vite possible. Par ailleurs, *Le curriculum de l'Ontario* (MÉFO, 1997a, 1997b) présente un recul par rapport aux approches didactiques privilégiées dans *Le programme d'études commun* (1995b) et revient à la prescription de contenus spécifiques plus morcelés. Enfin, l'implantation du bulletin provincial avec une seule note par discipline, en pourcentage à partir de la septième année, n'encourage pas la distinction des habiletés évaluées, comme le concept de justesse le valorise. Sans chercher à promouvoir spécifiquement ce concept, de telles conditions présentent un frein à l'évolution des pratiques actuelles d'évaluation, à leur rôle, à leur déroulement et à l'implication active des élèves.

L'activité d'évaluation en salle de classe est réglée sur le bulletin. Dans le conseil scolaire où s'est déroulé la présente étude, les enseignants doivent produire trois bulletins durant l'année scolaire (fin novembre, début mars, mi-juin). Il y a donc une période intensive d'évaluation dans les deux semaines précédant la fin du terme, suivie d'une période d'environ quatre à six semaines plus calme à ce point de vue. Les évaluations recommencent ensuite lentement pour culminer vers la période précédant le bulletin suivant.

Les enseignants participant à la présente recherche semblent ignorer que l'évaluation en

salle de classe renseigne l'élève sur ses propres capacités et sert de base à de nombreuses décisions déterminantes pour son avenir (Stiggins & Conklin, 1992). Ils évaluent surtout par convention, pour faire rapport aux parents au sujet d'un contenu enseigné, beaucoup plus que par conviction ou faire état des forces et des faiblesses des élèves au sujet d'habiletés à développer.

Pour les épreuves, les enseignants-participants recourent si possible à des épreuves toutes faites, quitte à les modifier. Cependant, une grande part de l'évaluation faite en salle de classe consiste en fait à noter les activités d'apprentissage, pour contraindre les élèves à travailler. Cette forme d'évaluation continue — ou continue — est dommageable surtout parce qu'elle comporte un jugement chiffré plutôt qu'un jugement descriptif ou qualitatif. Une note chiffrée est automatiquement coupée de son contexte. De plus, les «points» accumulés tout au long de la période d'apprentissage influencent la note finale, laquelle devrait normalement situer un élève dans son développement après une période d'apprentissage raisonnable. Pour les enseignants-participants, situer un élève dans son développement semble n'être nécessaire qu'au moment de préparer un bulletin et de le présenter au parent. Les calculs mathématiques pour faire les moyennes leur apparaissent dépourvus de toute subjectivité.

Les enseignants-participants sont conscients de l'importance du choix de la tâche pour arriver à des résultats justes, du moment opportun pour l'évaluation et de leur présence sur les lieux à ce moment-là. Ils ont appris qu'il faut utiliser des stratégies variées et évaluer souvent, mais n'en conçoivent pas le bien-fondé. Ils ne ciblent pas toujours bien

l'objet de l'évaluation; ils s'en tiennent souvent à la mémorisation de faits et à l'application d'algorithmes.

Ils se soucient peu de donner de la transparence à l'activité évaluative. Même si des expériences ont démontré la faisabilité et l'effet bénéfique d'intégrer l'élève à l'évaluation de ses apprentissages (Arter et al., 1994; Doyon, & Juneau, 1991), ici, l'élève ne paraît pas être un partenaire responsabilisé et respecté à ce point de vue. Même pour l'évaluation considérée formative, les participants se préoccupent peu de formuler les résultats en termes de forces et de faiblesses, tant pour leur réinvestissement dans l'activité pédagogique que pour le sens qu'ils pourraient prendre pour l'élève. L'habitude d'accorder des points pour contraindre les élèves à travailler et le besoin pressenti de protéger ses jugements n'y est pas propice. Les attentes sont souvent implicites et le schème de notation, absent. L'évaluation par les pairs et par le groupe classe est confrontée au paradigme valorisant les notes chiffrées, additionnées et divisées pour faire des moyennes, ce qui prendrait beaucoup de temps. De plus, l'attribution de points par les pairs briserait le monopole détenu jusqu'à présent par l'enseignant.

L'importance d'un schème de notation pour réduire le plus possible la subjectivité lors de la notation transparaît des propos analysés. Que ce soit pour évaluer une rédaction, un problème de mathématiques ou une présentation orale, le schème de notation devrait être clair et précis, afin d'assurer la stabilité de la notation d'une journée à l'autre. Cela étant dit, si le contexte des présentations orales ne permet pas à l'enseignant de s'engager entièrement dans l'appréciation de la performance de l'élève qui présente, il serait pertinent de recourir à un moyen d'y revenir dans un contexte plus favorable à un

traitement correct. Il ne faut cependant pas oublier que d'enregistrer ou de filmer la performance d'un élève peut augmenter le stress de l'évalué et par conséquent nuire autrement à la justesse des résultats.

Chez les élèves ne maîtrisant pas encore complètement la lecture et l'écriture, l'évaluation des apprentissages présente un défi différent. Même si la recension des écrits n'a pas fait ressortir la spécificité de cette situation, les élèves de troisième année ont paru être au seuil de l'autonomie nécessaire pour effectuer seuls des tâches d'évaluation de type papier-crayon. Leur capacité de concentration sur une tâche et le temps qu'elle peut durer interviennent sur l'utilisation optimale de l'évaluation. Leur degré de maturité moindre exige des stratégies d'évaluation particulièrement bien ciblées; c'est aussi le temps de les accompagner vers plus de responsabilités et une culture différente de l'évaluation en salle de classe.

La possibilité pour les élèves d'apprendre dépend de la compétence des enseignants, mais plusieurs de ceux-ci ne connaissent pas suffisamment comment se construit le savoir pour pouvoir aider efficacement l'élève dans son apprentissage (Herman & Klein, 1996). Deux éléments sont vitaux pour l'enseignement d'habiletés supérieures de pensée : la connaissance de la pédagogie en général et celle spécifique à la discipline (Tobin et al., 1990). Le manque de l'une ou de l'autre se manifeste par des activités limitées aux niveaux élémentaires de pensée (Fraser & Tobin, 1991), ce que nous avons pu observer dans des classes participantes, notamment celles de Kevin et de Christine. L'établissement des habiletés visées apparaît être une force du concept de justesse parce

qu'il incite à distinguer le rappel de connaissances et l'application d'algorithmes du développement d'habiletés plus complexes.

Selon Kulm (1993), si les enseignants changent leur manière d'évaluer, ils changent aussi leur manière d'enseigner, mais un changement profond comme l'exige l'application du concept de justesse ne se fait pas sans soutien professionnel majeur (Schwager & Carlson, 1994). Il est essentiel de s'attaquer d'abord à leur conception personnelle; avec une vision globale de l'évaluation intégrée à l'apprentissage, moins administrative, et des principes fondamentaux, les enseignants pourraient réduire considérablement le nombre de situations d'évaluation nécessaires parce qu'elles seraient plus justes (Macdonald, 1993). Ils réduiraient aussi le temps consacré à l'évaluation, ce qui semble actuellement alourdir considérablement leur tâche. Avec une formation plus complète, ils auraient moins tendance à réajuster les notes (Brookhart, 1993) construiraient des épreuves de meilleure qualité (McMorris & Boothroyd, 1993) et bénéficieraient de l'évaluation formative comme d'un véritable outil pédagogique.

De nouvelles pratiques s'instaurent mieux en début d'année scolaire (Kulm, 1994); généralement les élèves reconnaissent la liberté des enseignants à choisir leurs pratiques, particulièrement s'ils sont convaincus eux-mêmes de leur bien-fondé et prennent le temps d'expliquer ce choix à leurs élèves. C'est probablement le premier pas dans la voie de la transparence de l'activité évaluative, mais tout doit débiter par le désir de l'enseignant de changer ses pratiques et de connaître comment faire mieux, de manière pratique.

Les préoccupations et les stratégies mentionnées par les participants pour arriver à des résultats d'évaluation justes peuvent être regroupées en trois genres. Il y a d'abord celles

déjà abordées lors de la présentation des composantes du concept de justesse (figure 1, page 39), d'autres ensuite apparaissent comme fautives, bien que tout à fait attendues compte tenu de la recension des écrits. Enfin, deux préoccupations nouvelles viennent enrichir le concept de justesse des résultats :

- Accumuler des traces de la performance évaluée afin de pouvoir y revenir si nécessaire pour appuyer son jugement après une période de recul,
- Recourir à des critères réalistes pour évaluer.

Les éléments contenus dans le concept de justesse mais ignorés des enseignants-participants font ressortir le peu de transparence de leurs pratiques d'évaluation quotidiennes. Leur conception de l'apprentissage et de son évaluation, comparée à celle véhiculée par le concept de justesse des résultats, présente un décalage énorme. Qu'il soit attribuable à une formation déficitaire ou à un manque de leadership et de cohérence de la part du Ministère, l'expérience n'a pas été assez longue pour combler cet écart, malgré la confiance établie entre les participants et la chercheure. C'est avec cette optique que la viabilité du concept de justesse doit être abordée.

Autre fait important : l'analyse des protocoles verbaux d'entrevue portant sur la notion de justesse des résultats avant la mise à l'essai a confirmé l'avantage de sa confrontation avec la réalité de la salle de classe. D'abord, elle a forcé le raffinement des éléments impliqués dans la justesse des résultats. Par ricochet, cet exercice a fait ressortir l'interrelation pressentie des quatre composantes. Par exemple, les critères d'évaluation :

bien qu'ils doivent être établis au moment même d'assigner la tâche (élément de la tâche) pour que les élèves les connaissent bien (éléments de transparence), ils doivent également être appliqués lors d'activités d'apprentissage préalables (cohérence avec l'activité pédagogique). La classification de ces éléments selon les quatre composantes du critère de justesse est donc difficile, certains éléments ayant des affinités avec plus d'une composante. L'exercice a fait ressortir aussi le grand nombre de détails pouvant se rattacher à une situation d'évaluation, rendant sa description contextualisée difficile.

DEUXIÈME QUESTION DE RECHERCHE

Le présent chapitre aborde les données se rapportant directement au concept de justesse des résultats. Les participants ont été initiés à ce concept par des discussions, illustrées de nombreux exemples concrets. Le guide de formation distribué à ce moment-là de la collecte des données (annexe F) inclut les notions abordées et les thèmes de discussion choisis par les participants durant la période de familiarisation. Le tout est agencé de manière à faire ressortir le concept de justesse et ses composantes. Les enseignants participants ont ainsi pu saisir comment cette conception de l'évaluation des apprentissages en salle de classe pouvait s'intégrer à leurs pratiques pédagogiques et combien elle pouvait réellement soutenir leur enseignement. Chaque participant a ensuite choisi une situation d'évaluation à laquelle il voulait appliquer, avec l'aide de la chercheure, le concept de justesse. L'expérience personnelle de chacun et l'échange avec d'autres participants ont habilité ces collaborateurs à se prononcer sur la pertinence et la viabilité du concept de justesse, lors d'un deuxième entretien individuel.

Les plans d'évaluations élaborées en vue de l'opérationnalisation du concept de justesse sont présentés à l'annexe G; ils correspondent aux situations d'évaluation 6 à 10. L'analyse de ces situations d'évaluation selon les quatre composantes du concept de justesse (annexe H), sert de base au deuxième entretien (guide de l'entrevue finale, annexe I). Pour en faciliter la lecture, le schéma du concept de justesse et de ses

composantes présenté à la page 39 est repris à la page suivante, juste avant le résumé des entrevues finales.

L'analyse des situations d'évaluation après leur déroulement, selon les quatre composantes établies, est en fait un bilan de toutes les conditions entourant chaque activité d'évaluation; ces conditions auraient pu intervenir de manière favorable ou défavorable sur la justesse des résultats obtenus par les élèves. Après en avoir discuté avec chaque participant, chacun a été invité à exposer sa perception de la viabilité de ce concept dans le quotidien de la salle de classe et à estimer l'utilité et le réalisme de chacune des composantes en lui attribuant une note, selon le barème de son choix. Le contenu de ces deuxième entrevues individuelles permet de répondre à la deuxième question de recherche :

2 a) La pertinence de la tâche d'évaluation par rapport aux habiletés visées, la cohérence de celle-ci avec l'activité pédagogique, la transparence de l'activité évaluative et l'absence de désavantage circonstanciel rendent-elles compte de l'ensemble des éléments pouvant intervenir sur la justesse des résultats d'évaluation des apprentissages en salle de classe?

b) La notion de justesse des résultats telle que présentée dans le cadre conceptuel peut-elle s'intégrer à l'activité scolaire?

POUR UN RÉSULTAT D'ÉVALUATION JUSTE	
1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées - Tâche et critère(s) d'évaluation congruents avec les diverses habiletés visées - Tâche signifiante, adaptée, claire, techniquement correcte, nouvelle mais familière	
2. Cohérence avec l'activité pédagogique - Activités d'apprentissage préalables avec mise en valeur des mêmes critères d'évaluation - Mise en situation appropriée et aide disponible - Schème approprié au type d'évaluation et permettant le réinvestissement des résultats - Possibilité de reprise	3. Transparence de l'activité évaluative - Participation de l'élève à toutes les étapes de la démarche d'évaluation - Schème d'évaluation familier aux élèves - Résultats communiqués en termes de forces et de faiblesses - Retour sur l'activité d'évaluation en classe
4. Absence de désavantage circonstanciel - conditions propices à la concentration - collaboration contrôlée - procédure équitable et habituelle - temps amplement suffisant - présence pendant l'apprentissage préalable	

Figure 1 (reprise). Schématisation des composantes du concept de justesse des résultats.

Étant donnée que cette question est l'élément central de la présente étude, les réactions des participants au concept de justesse des résultats sont d'abord présentées. Cette première partie du chapitre est complétée par l'essentiel de ces opinions, accompagné d'extraits des protocoles d'entrevue à ce sujet, dans le but d'en faire ressortir les traits communs. La deuxième partie du chapitre expose ensuite une discussion en vue d'établir la viabilité du concept de justesse des résultats et de ses composantes. La troisième partie est une synthèse de la réponse à cette deuxième question de recherche.

A - PRÉSENTATION DES RÉSULTATS

1. Entrevue finale avec Christine

Christine espérait que la participation à ce projet lui soit bénéfique mais avait de gros doutes à ce sujet. Contre toute attente, l'exercice a suscité une prise de conscience, une remise en question sur ce qu'elle faisait depuis longtemps; elle devait dire pourquoi elle le faisait et c'était stimulant. Aujourd'hui, elle dit que l'expérience va marquer définitivement sa façon de faire l'évaluation.

Voici ce qu'elle pense de l'utilité des composantes du concept de justesse pour arriver à des résultats d'évaluation justes et si elle croit possible de les respecter dans la pratique pédagogique :

Utilité pour arriver à un résultat juste		Est-ce réalisable?	
1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées : Très bien		1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées : Très bien	
2. Cohérence avec l'activité pédagogique : Superbe. On ne peut pas passer à côté	3. Transparence de l'activité évaluative : Très bien	2. Cohérence avec l'activité pédagogique : Très bien	3. Transparence de l'activité évaluative : Très bien
4. Absence de désavantage circonstanciel : Assez bien		4. Absence de désavantage circonstanciel : Peu réalisable, trop de contraintes imposées par la structure	

Figure 4.1 Utilité et possibilité de respecter les composantes du concept de justesse des résultats selon Christine

D'après Christine, non seulement les trois premières composantes sont-elles intégrables à la pratique régulière en salle de classe mais on ne devrait pas penser pouvoir faire autrement. Par contre, la quatrième composante est très difficile à intégrer à cause des contraintes de temps imposées par la vie scolaire; elle n'a pas assez de contrôle sur ces variables pour pouvoir en tenir compte quand elle évalue. Par contre, elle trouve cette façon de voir l'évaluation très sensée et n'a pas de modifications à suggérer.

Quand on l'interroge sur le type d'enseignant qui serait le plus susceptible de

bénéficier d'une telle expérience, elle répond que selon elle, toute personne pourrait en bénéficier. Qu'on en soit à sa première année d'enseignement ou à sa vingtième année, tous pourraient en tirer profit. Pour sa part, elle aimerait bien poursuivre cette activité l'an prochain. Elle croit que l'exercice serait encore plus profitable s'il était échelonné sur deux années; une fois engagé dans l'activité, ça devient beaucoup plus concret et profitable pour le participant.

2. Entrevue finale avec Kevin

Kevin n'avait pas d'attente en participant à ce projet, il s'inquiétait plutôt de ce qu'il pourrait y apporter étant donné son peu d'expérience dans l'enseignement. Une fois l'activité terminée, il dit que ça l'a fait réfléchir et que ça a amélioré sa façon d'évaluer.

Voici ce qu'il pense de l'utilité des composantes du concept de justesse pour arriver à des résultats d'évaluation justes et s'il croit possible de les respecter dans la pratique pédagogique :

Utilité pour arriver à un résultat juste		Est-ce réalisable?	
1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées : Très important (10/10)		1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées : Oui, très bien	
2. Cohérence avec l'activité pédagogique : Important (9/10)	3. Transparence de l'activité évaluative : Important (8/10)	2. Cohérence avec l'activité pédagogique : Oui, très bien	3. Transparence de l'activité évaluative : Oui, très bien
4. Absence de désavantage circonstanciel : C'est utile d'en être conscient mais en même temps difficile à respecter (7/10)		4. Absence de désavantage circonstanciel : On peut intervenir sur certaines conditions mais pas toutes. C'est la composante la plus difficile à respecter.	

Figure 4.2 Utilité et possibilité de respecter les composantes du concept de justesse des résultats selon Kevin

Kevin n'a pas de modification à suggérer pour améliorer cette façon de voir l'évaluation, si ce n'est d'examiner la possibilité d'intégrer la quatrième composante aux trois autres. Dans les années futures, il a l'intention de se référer à cette conception de l'évaluation pour élaborer des activités d'évaluation pour ses élèves.

À son avis, cette expérience pourrait bénéficier à n'importe quel enseignant. De s'arrêter et de réfléchir sur sa façon d'évaluer peut être profitable à n'importe quel professionnel, tant aux premiers niveaux scolaires qu'aux niveaux scolaires supérieurs d'enseignement.

3. Entrevue finale avec Marie

Marie s'est intéressée à ce projet de recherche parce qu'elle pensait peut-être y trouver un moyen de mieux cibler son intervention en salle de classe et elle y a découvert que l'évaluation est un outil d'enseignement. «C'est vraiment la première année que je m'attarde autant à l'évaluation et je réalise maintenant comment cela peut m'aider dans mon enseignement». Elle a aussi réalisé l'importance de reconnaître les habiletés visées, autant par l'enseignement que l'évaluation, avant d'intervenir. Cette habitude-là sera un atout précieux dans sa pratique l'an prochain, étant donné la particularité du contexte dans lequel elle travaillera (classe pour enfants en grande difficulté d'apprentissage). Voici ce qu'elle pense de l'utilité des composantes du concept de justesse pour arriver à des résultats d'évaluation justes et combien elle croit qu'il est possible de les respecter dans sa pratique pédagogique :

Utilité pour arriver à un résultat juste		Est-ce réalisable?	
1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées : Excellent, même essentiel, autant à l'enseignement qu'à l'évaluation, autant à l'enseignant qu'à l'élève et au parent		1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées : Oui, il <u>faut</u> le faire	
2. Cohérence avec l'activité pédagogique : Même chose que (1)	3. Transparence de l'activité évaluative : Bien	2. Cohérence avec l'activité pédagogique : Oui, il <u>faut</u> le faire	3. Transparence de l'activité évaluative : Oui, même si c'est difficile à faire, avec le temps ça peut s'apprendre
4. Absence de désavantage circonstanciel : Bien		4. Absence de désavantage circonstanciel : Non, je n'ai pas assez de contrôle sur ce qui se passe dans ma classe pour penser pouvoir en tenir compte pour tous les élèves	

Figure 4.3 Utilité et possibilité de respecter les composantes du concept de justesse des résultats selon Marie

Marie considère qu'une amélioration à apporter à cette conception de l'évaluation c'est de l'intégrer à la démarche pédagogique, parce que l'évaluation ça fait dorénavant partie de l'enseignement; elle aimerait que comme elle, tous les enseignants le réalisent enfin. Pour sa part, la prise de conscience au sujet des habiletés visées lui sera d'un grand secours dans son enseignement futur. Elle devra probablement compter beaucoup sur l'observation pour évaluer ses élèves et elle croit avoir découvert un atout précieux pour

y arriver.

Selon elle, l'exercice vécu à travers ce projet de recherche pourrait être utile à n'importe quel enseignant mais surtout à un enseignant chevronné. Les avantages à en retirer sont aussi grands que le temps qu'on veut y investir. Pour sa part, cette activité a insufflé beaucoup d'enthousiasme à sa carrière d'enseignante.

4. Entrevue finale avec Matilde

Matilde espérait vérifier à travers ce projet sa manière d'évaluer et trouver des moyens de réduire le volume de correction à faire. L'expérience lui a aidé à mieux comprendre le processus d'évaluation, ce qu'il implique et ça l'a fait beaucoup réfléchir. Avec l'expérience, elle intègre de plus en plus d'éléments du concept de justesse à sa pratique quotidienne; elle connaît maintenant beaucoup mieux l'évaluation.

Voici ce qu'elle pense de l'utilité des composantes du concept de justesse pour arriver à des résultats d'évaluation justes et si elle croit possible de les respecter dans sa pratique pédagogique :

Utilité pour arriver à un résultat juste		Est-ce réalisable?	
1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées : Les composantes 1 et 2 sont très liées. S'il y a cohérence de l'évaluation avec l'activité pédagogique, la tâche d'évaluation sera pertinente par rapport aux habiletés visées 7/10		1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées : Très réaliste	
2. Cohérence avec l'activité pédagogique : 9/10	3. Transparence de l'activité évaluative : 9/10	2. Cohérence avec l'activité pédagogique : On manque souvent de temps	3. Transparence de l'activité évaluative : Très réaliste, ça s'apprend
4. Absence de désavantage circonstanciel : 8/10		4. Absence de désavantage circonstanciel : Difficile à respecter pour tous les élèves	

Figure 4.4 Utilité et possibilité de respecter les composantes du concept de justesse des résultats selon Matilde

Dans cette conception de l'évaluation des apprentissages en salle de classe, Matilde trouve difficile de faire systématiquement des activités d'apprentissage avant toutes les évaluations. Par exemple, le projet développé ensemble était la première activité de robotique de l'année; il fallait donc compter sur une seule situation pour faire à la fois l'apprentissage et son évaluation, à cause du manque de temps. Mais le plus difficile reste encore d'éliminer les désavantages circonstanciels, parce qu'on a peu de

contrôle sur eux.

Le concept de justesse lui apparaît complet avec ces quatre composantes. Pour elle, il comporte l'essentiel et croit que si on arrivait à le respecter, en entier, ce serait déjà très satisfaisant. À son avis, tous les enseignants pourraient bénéficier d'une telle expérience, tant à la maternelle qu'à l'école secondaire, dans tous les champs d'études. Le concept convient à tout le monde qui doit faire de l'évaluation.

5. Entrevue finale avec Marie-Chantal

Marie-Chantal espérait trouver dans ce projet des moyens plus efficaces d'évaluer, consommant moins de temps tout en arrivant à des résultats justes. L'expérience a suscité beaucoup de réflexion chez elle. L'établissement des habiletés visées, tant par l'activité d'apprentissage que par l'activité d'évaluation, a été pour elle le clou de cette expérience. L'utilisation d'indicateurs de performance pour situer un élève dans son développement en a été un autre. Voici ce qu'elle pense de l'utilité des composantes du concept de justesse pour arriver à des résultats d'évaluation justes et si elle croit possible de les respecter dans la pratique pédagogique :

Utilité pour arriver à un résultat juste		Est-ce réalisable?	
1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées : 10/10		1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées : 8/10	
2. Cohérence avec l'activité pédagogique : 10/10	3. Transparence de l'activité évaluative : 10/10	2. Cohérence avec l'activité pédagogique : 9/10	3. Transparence de l'activité évaluative : 8/10
4. Absence de désavantage circonstanciel : 10/10		4. Absence de désavantage circonstanciel : 8/10	

Figure 4.5 Utilité et possibilité de respecter les composantes du concept de justesse des résultats selon Marie-Chantal

Tout en reconnaissant l'importance de toutes ces composantes pour arriver à des résultats d'évaluation justes, Marie-Chantal trouve que le temps l'empêche de pouvoir les respecter autant qu'elle le voudrait. L'expérience vécue au cours de l'année a surtout stimulé la réflexion et apporté de nouvelles idées en ce qui a trait à l'évaluation en salle de classe. L'an prochain elle compte y revenir et en intégrer certaines, prudemment, tout en prenant le temps d'en analyser les répercussions. Ce qu'elle retient aussi de ce projet, c'est la difficulté d'évaluer de manière juste pour tous les élèves les activités effectuées en équipe. Pour le moment elle ne considère pas avoir assez expérimenté le concept pour pouvoir suggérer des changements. Puisqu'elle aime bien intégrer les matières quand elle prépare des activités d'apprentissage, l'importance de reconnaître les habiletés visées par

l'activité pédagogique avant de s'engager dans toute forme d'intervention la laisse pour le moment perplexe. Elle se demande s'il est réaliste de penser pouvoir viser un grand nombre d'habiletés avec une activité intégrée et d'évaluer chacune d'elles si on ne se préoccupe pas de leur apprentissage de manière plus systématique.

D'après elle, la participation à ce projet aurait été bénéfique à n'importe qui parce que le concept de justesse est pertinent à n'importe quel niveau scolaire et à toute forme d'enseignement; toutefois, c'est chez les enseignants d'expérience que la réflexion serait la plus enrichissante.

B - INTERPRÉTATION DES RÉSULTATS

1. Discussion sur la pertinence du concept de justesse des résultats

Comme pour la première entrevue, les protocoles verbaux ont été transcrits et des synopsis ont été soumis pour approbation aux cinq participants. Le contenu des entretiens transcrits intégralement, a été regroupé selon qu'il portait sur l'utilité, le réalisme, l'amélioration ou l'intégrabilité du concept de justesse. Les réponses fournies étant tellement ciblées, il n'a pas paru nécessaire d'instaurer un système de codification pour l'analyse de contenu. La section suivante présente chacun des volets de cette question de recherche, accompagné d'extraits des entretiens à leur sujet.

À cette occasion, les participants ont aussi noté chacune des composantes du concept de justesse. Bien qu'ils aient utilisé des types d'échelle variés, allant du quantitatif au qualitatif, leur appréciation de l'utilité et de l'intégrabilité des composantes est relativement élevée. Seule l'absence de désavantage circonstanciel est perçue peu intégrable mais en même temps, considérée très utile pour expliquer un résultat inattendu. Le texte qui suit apporte plus de détails à ce sujet.

1.1 Utilité

Les participants sont unanimes, le concept de justesse pourrait être utile à «n'importe quel genre d'enseignant, ... n'importe qui qui fait affaire avec de l'évaluation. Que ce soit de l'évaluation au niveau social, au niveau psychomoteur, au niveau ... Tant qu'à moi ... y'a des composantes qui vont changer ou qui vont s'entrelacer mais ça se ressemble.» (Christine, entrevue 2, ligne 189). «Je pense que tout le monde a quelque chose à tirer de ça. Que ce soit au jardin, enseignante de jardin, ou de 8^e année ou de secondaire, je pense que c'est un modèle qui s'applique à tout, tous les types d'évaluation, pour tous les niveaux aussi, pour tous les champs d'études en même temps, c'est général. (Matilde, entrevue 2, ligne 113). «Je pense qu'il y a un besoin maintenant surtout en évaluation. On en parle beaucoup puis on commence à être beaucoup plus conscient de l'importance que ça a.» (Marie, entrevue 2, ligne 198). Le concept «pourrait être appliqué dans les plus bas niveaux, comme au niveau supérieur, qui vise un bon résultat final» (Kevin, entrevue 2, ligne 119).

Bien qu'il était ici question d'utilité pour l'enseignant de se référer au concept de justesse

des résultats pour évaluer les apprentissages, Marie, elle, lui voit une plus grande portée, particulièrement en ce qui à trait à l'établissement des habiletés visées. «C'est important, pas juste pour nous comme enseignant là, mais aussi pour les élèves, je pense. C'est «mieux cerné». Si nous on peut mieux cerner, certain que les élèves aussi vont pouvoir le cerner. Puis après ça, possiblement les parents aussi, pour aider, pour qu'on s'en aille tous dans une même direction. (Marie, entrevue 2, ligne 36).

Des recherches ont démontré qu'il n'est pas facile pour un enseignant de changer ses pratiques (Parker, 1992; Macdonald, 1993) et que la formation à cet effet peut faire toute la différence (Schwager, 1994; Sawyer, 1994; Cox, 1994). Les participants les plus expérimentés croient que l'activité a été davantage bénéfique pour eux : «il faut que tu aies un peu vécu dans de l'évaluation pour en parler». (Christine, entrevue 2, ligne 180). Malgré cela, Kevin, qui en était à sa première année d'enseignement, s'est dit heureux d'avoir vécu l'expérience. «Je trouve que ça a amélioré ma façon d'évaluer ... ça m'a permis de réfléchir sur la façon que j'évalue» (Kevin, entrevue 2, ligne 13). Le témoignage de Marie porte à croire qu'il vaut mieux tôt que tard : «Au début de mon enseignement, l'évaluation, je pensais pas à ça, c'était pas important d'évaluer mes élèves. Ce qui était important c'était d'enseigner. D'être avec eux, tout ça. Et je portais des jugements, parce qu'il faut, on a des bulletins, tout ça là, mais c'était de façon très holistique par rapport à ce que je faisais avec eux en classe» (Marie, entrevue 2, ligne 126). Cette année, «c'est vraiment la première année que je m'attarde autant au niveau de l'évaluation, mais je vois maintenant plus où est-ce que ça peut mener, comment je peux mieux enseigner en salle de classe avec ... en portant une attention à l'évaluation». (Marie, entrevue 2, ligne 9).

Toutefois, sa participation à cette recherche n'est pas seule en cause; l'évaluation provinciale de ses élèves y est aussi pour beaucoup.

Lorsque l'utilité de chacune des composantes est abordée, pour Matilde, la première composante (Pertinence de la tâche par rapport aux habiletés visées) est la moins utile parce que, dit-elle, on la retrouve implicitement dans la cohérence avec l'activité pédagogique. «Je pense que quand on regarde la Cohérence avec l'activité pédagogique, je pense que ça revient là dedans. Si tu regardes qu'est-ce que t'as fait, puis là tu vas évaluer sur ce que t'as fait, puis là dans le fond tu vas les rechercher tes habiletés visées quand tu penses à la Cohérence.» (Matilde, entrevue 2, ligne 25)

Toutefois, il a été donné d'observer un dérapage, justement dans cette composante: Matilde dit se référer au texte écrit pour juger la composante vocabulaire dans les présentations orales, ce qui consiste à se référer à l'écrit pour évaluer l'oral. Aussi, lors de la planification de l'évaluation du projet de robotique, Matilde se serait limitée à attribuer des «points» au rapport d'expérience, ce qui l'amène loin des habiletés précises visées par cette activité. Cependant, pour Marie, il s'agit de la composante la plus importante :

Décortiquer les habiletés, c'est ça je pense, ne m'être jamais arrêtée pour le faire. (*Ça a cliqué cette année?*) Oui c'est ça. Vraiment. Décortiquer ça pour mieux ... Et ça aide à mieux comprendre. Oui les choses sont devenues plus claires. Ce que je voulais faire avec les élèves, puis pourquoi, pourquoi je faisais certaines choses avec les élèves

(*intuitivement?*) oui c'est ça. Mais là ça aide à mieux comprendre et ça aide après ça à mieux cerner les difficultés et les habiletés des élèves.

(Marie, entrevue 2, ligne 168)

Marie-Chantal a elle aussi été marquée par cette composante, mais plutôt parce que celle-ci a ébranlé sa confiance dans des pratiques pédagogiques qu'elle apprécie particulièrement. Il en sera davantage question plus avant dans le texte.

1.2 Réalisme

Même si les enseignants participants sont convaincus de la pertinence de toutes les composantes du concept et qu'elles sont susceptibles d'intervenir sur la justesse d'un résultat d'évaluation, ils croient impossible de toutes les respecter, en tout temps. L'absence de désavantage circonstanciel leur semble le plus difficile à obtenir. Même s'ils essaient de créer une atmosphère propice à la concentration, ils peuvent difficilement considérer les conditions personnelles de chacun des élèves parce que la plupart du temps, ils ne savent pas ce que vivent les élèves en dehors de l'école, sauf, après coup, si la situation dégénère en crise. Et il serait impossible de coordonner les activités d'évaluation à un moment opportun pour chacun sans apporter d'autres sources d'erreurs.

Quant aux autres composantes du concept de justesse, Marie, enseignante de 3^e année, exprime son ambivalence à l'égard du réalisme de la «Transparence de l'activité évaluative». Comme de nombreux autres enseignants, elle s'est peu préoccupée de cet aspect de l'évaluation jusqu'à maintenant. Sa participation à la présente recherche l'ayant

contraint à trouver des moyens de le faire, elle y a vu, malgré la taille du défi, de bonnes raisons de le faire. Elle croit que plus tôt on arrive à impliquer activement l'élève dans son évaluation, plus il sera capable de s'engager dans son apprentissage : «le plus vite qu'ils arrivent à le voir, aussi, c'est après ça qu'ils peuvent s'investir. Parce que c'est ça, quand ils ne voient pas finalement le but, la raison d'être, bien alors ils s'investissent pas. Y'a pas de raisons.» (Marie, entrevue 2, ligne 71) Les défenseurs de la participation des élèves à leur évaluation ne sauraient pas être plus convaincants.

De l'avis des participants, le temps et les ressources disponibles, le travail en équipe et l'enseignement intégré sont d'autres facteurs qui interviennent sur le réalisme du concept de justesse. Selon Marie-Chantal, le peu de matériel concret disponible pour le projet technologique (activité d'évaluation 10) et la nécessité d'utiliser le temps de classe à bon escient l'ont amenée à diriger simultanément deux types d'activités d'apprentissage, ce qui l'a rendue très sollicitée par les élèves et a réduit sa capacité à observer objectivement leur performance. Elle a même, à un certain moment, gardé quelques élèves à la récréation pour pouvoir les évaluer dans un contexte plus calme; dans le bilan de cette situation d'évaluation, ceci a été relevé comme étant une pratique pouvant nuire à l'intérêt des élèves à démontrer leur plein potentiel, et donc nuire à la justesse des résultats.

À cause du peu de matériel concret disponible, c'est à dire des blocs Lego, petits moteurs et autre matériel du genre, Matilde a opté pour le travail en équipe. Des auteurs croient

que faire des activités d'apprentissage en équipe devrait normalement être suivi d'évaluation en groupe (Glaser et al., 1984; Hamish, 1994; Lussier & Turner, 1995). Or, les écrits au sujet de l'évaluation à partir d'une tâche effectuée en équipe reconnaissent la situation comme étant problématique (Blumenfeld, Marx, Soloway & Krajcik, 1996; Herman et al., 1992; Kulm, 1994; Leder, 1992). D'une part, travailler en équipe constitue en soi une gamme d'habiletés à développer : écouter les autres, partager ses idées, respecter les idées des autres, partager la charge de travail et s'entraider; ces habiletés sociales interviennent sur les autres habiletés à développer. D'autre part, le partage des responsabilités peut permettre à des coéquipiers d'être exemptés complètement de certaines tâches et donc, de démontrer leur maîtrise de certains apprentissages. Les enseignants-participants ont reconnu que la notation des travaux d'équipe est délicate et souvent injuste pour certains membres du groupe, mais ils la trouvent très pratique.

Souvent, les enseignants voient les projets comme de bonnes occasions de faire de l'intégration de matières. Comme les activités en salle de classe sont généralement guidés par les contenus à enseigner, ils rassemblent des contenus jugés compatibles pour en faire un thème intéressant. Ils attribuent généralement une note globale pour l'équipe si le produit est réalisé en groupe. L'application du concept de justesse, nécessitant d'établir les habiletés visées par l'activité d'évaluation, dérange Marie-Chantal :

«Moi j'essaie de faire le plus d'intégration possible là dedans, puis je trouve qu'à ce moment-là c'est un petit peu difficile de bien définir la tâche et les habiletés quand il y en a plusieurs, compte tenu de l'intégration. C'est à ce

moment là qu'il faut vraiment que tu comprennes, c'est quoi exactement que je vise, ici? C'est ça que j'ai pas appris à faire encore. Donc c'est bien beau les projets d'intégration, il y a peut-être 40 habiletés mais, lesquelles? Puis c'est quelles, lesquelles que tu vas évaluer, ça c'est difficile alors c'est quelque chose que j'essaie de voir. Je disais bon, l'élève va développer ci, ci, ci, ça, mais on prenait peut-être pas le temps de le développer...» et «quand je les fais travailler en équipe, jusqu'à quel point on peut bien évaluer le travail de chacun?» (Marie-Chantal, entrevue 2, ligne 53 et 66).

Matilde pense que de bien dissocier l'apprentissage de son évaluation rend l'application du concept de justesse irréaliste; elle ne dispose pas d'assez de temps pour distinguer les deux. De plus, même si elle n'en parle pas, son habitude d'attribuer des notes à presque tout ce que font les élèves constitue un obstacle majeur à cette conception de résultats d'évaluation justes. Elle devrait alors trouver un autre moyen de «motiver» ses élèves à travailler, cumuler moins de «points» dans un trimestre, ce qui rendrait ses moyennes moins fiables, dans son optique.

1.3 Intégration

La conception de justesse des résultats d'évaluation a bouleversé certaines habitudes et fait ressortir le caractère problématique des pratiques d'évaluation antérieurement adoptées par les enseignants-participants. Toutefois, ces derniers disent avoir été suffisamment influencés par cette conception de l'évaluation pour vouloir y adhérer dans l'avenir. «Oùï je me suis posée des questions, puis oùï, je pense que d'année en année,

quand je vais faire de l'évaluation, ce modèle va me rester dans la tête» (Christine, entrevue 2, ligne 49). Marie, assignée l'an prochain à une classe d'élèves en grandes difficultés d'apprentissage est convaincue : «je le sais que je pourrai pas le faire en tout temps, mais je me dis, déjà, d'y penser, puis de le faire de façon plus systématique pour certaines choses, puis après ça, si je suis bien là-dedans, je vais bâtir des choses. (Marie, entrevue 2, ligne 98) ... Comme j'ai dit tantôt ... j'ai aucune idée de ce que je vais faire au juste l'an prochain. Mais je sais que je dois évaluer ces élèves là et je dois voir leur progrès, et même qui sait, peut-être je vais m'en servir même plus ... il va y avoir beaucoup ... d'observations. Alors comment observer pour porter un jugement sur ce que j'observe? (Marie, entrevue 2, ligne 148). «Toujours les habiletés visées, moi c'est ça qui m'a le plus plu ... Cette année, qu'est-ce que je veux chez tel enfant, qu'est-ce que je vise? Après ça c'est d'observer dans ce sens-là.» (Marie, entrevue 2, ligne 157)

Leur position face à la possibilité d'intégrer cette conception de l'évaluation à leur pratique future pourrait bien être influencée par l'expérience vécue au cours de la recherche. Par exemple, à prime abord, Matilde n'était pas chaude à l'idée de mettre en pratique l'évaluation par les pairs; elle ne trouvait pas la pratique réaliste, surtout pour l'ampleur du travail qu'elle y voyait. Elle ne l'aurait par conséquent pas mise en pratique. Cependant, la stratégie développée (situation d'évaluation 9) dans le cadre de la présente recherche l'a amenée à voir les choses différemment : «j'avais pensé être obligée de faire des calculs. ... l'évaluation par les pairs je voyais plus comme tu donnes 8/10, toi, puis y'en a 24 qui évaluent, après ça faire la moyenne ... Mais là quand c'est avec des couleurs c'est beaucoup mieux. C'est tout de suite. T'as pas de calculs à faire, puis ça,

l'année prochaine je va m'en servir» (Matilde, entrevue 2, ligne 153). Elle dit aussi avoir déjà commencé à appliquer le concept de justesse en fournissant à ses élèves les critères d'évaluation au même moment qu'elle leur assigne une tâche.

Au sujet de l'intégration du concept de justesse à ses pratiques futures, Marie-Chantal est prudente : «J'espère bien. C'est mon intention. C'est certain, ça va dépendre des projets, des activités, ça va dépendre de mes objectifs, de ce que je vise, mais pour ce qui est des projets, des travaux sous forme de projet, c'est certain que je va me servir du modèle pour ... bien pas juste pour les projets, pour l'évaluation en général, mais je vais y aller petit à petit, parce que je veux pas tout changer du jour au lendemain. (Marie-Chantal, entrevue 2, ligne 45). Toutefois, il y a deux genres de situations qui posent problème à son application, deux genres de situations fréquentes dans sa classe : les travaux effectués en équipe et les projets d'envergure intégrant des matières.

Kevin, quant à lui, dit qu'il va s'y référer à l'avenir, mais sans plus de précision; il semble éprouver de la difficulté à faire le transfert des notions apprises à d'autres situations semblables. L'expérience a quand même concrétisé pour lui une manière de bien faire les choses : «je pense que la foire scientifique, c'est quelque chose qu'on a fait bien» (Kevin, entrevue 2, ligne 140)

1.4 Améliorations

Trois des cinq participants n'avaient pas d'ajustements ou des modifications à proposer au concept de justesse des résultats, tant globalement qu'à certaines de ses

composantes. Pour Christine, les composantes se chevauchent bien et englobent bien l'activité évaluative; Marie-Chantal préfère s'abstenir de commentaires à ce sujet parce qu'elle dit ne pas avoir le recul nécessaire pour se prononcer sur le sujet. Quant à Matilde, elle dit : «ça me semble complet ... juste d'essayer de suivre ça ... si quelqu'un réussit à respecter tout ça, toujours, dans ses évaluations, ça va être ... très très bien. Puis ça prend, je pense, des années d'expériences avant d'arriver à le respecter parce que c'est comme, ... c'est un idéal. Le modèle idéal.» (Matilde, entrevue 2, ligne 67).

Kevin et Marie sont les seuls à bien avoir voulu émettre des suggestions à ce sujet. Kevin croit que cette conception de l'évaluation serait plus pratique si elle était organisée en étapes coïncidant avec le déroulement de l'activité pédagogique. «Donc j'aurais eu plus de facilité, avoir peut-être un petit modèle à côté, un modèle de leçon déjà faite avec le modèle. Par exemple chaque étape, qu'est-ce qu'on a fait. Donc ce serait plus facile de voir chaque étape. Parce que là on avait les étapes, mais des fois je pensais comment l'appliquer? Disons que je trouve que les étapes se suivent très bien. Donc, peut-être intégrer aussi le quatrième, vers la fin. Je sais pas si 1, 2, 3 même la 4 prend aussi avec le 1^{er} par exemple : choix d'équipe, le dérangement. C'est quelque chose qu'on peut voir au début aussi, donc c'est la seule chose que je pourrais voir.» (Kevin, entrevue 2, ligne 123)

Sans émettre de suggestions au niveau du concept de justesse comme tel, Marie trouve qu'il devrait avoir sa place dans l'apprentissage. Peut-être parce que les circonstances de cette année l'ont amenée à cheminer beaucoup en matière d'évaluation, Marie trouve

que : «au niveau des ... activités d'apprentissage, déjà là, c'est là où il faut qu'on voie quelles sont les habiletés visées, Alors ... c'est ça, de travailler l'évaluation en même temps que travailler nos activités d'enseignement comme telles. Alors je sais pas comment ça pourrait être inclus ça dans le modèle, de vraiment le rattacher à l'enseignement. Comme l'évaluation c'est une partie ... ça va de pair avec l'enseignement.» (Marie, entrevue 2, ligne 118).

2. DISCUSSION SUR LA VIABILITÉ DU CONCEPT DE JUSTESSE DES RÉSULTATS

La mise à l'essai, en collaboration avec les cinq enseignants-participants, démontre que le concept de justesse des résultats est applicable à l'évaluation en salle de classe. Les enseignants reconnaissent que toutes les composantes comportent des éléments pouvant influencer le degré de confiance à accorder à un résultat pour situer correctement un élève dans son développement, déterminer ses forces et ses faiblesses et ce qu'il maîtrise ou pas. Malgré cette position, ils ne croient pas qu'il soit possible d'appliquer les composantes en tout temps, et ce, pour plusieurs raisons. Il y a d'abord l'impossibilité de réunir toutes les conditions nécessaires à l'absence de désavantage circonstanciel pour tous les élèves au même moment, mais d'autres incompatibilités se sont aussi manifestées, surtout en ce qui a trait à l'évaluation d'une tâche effectuée en équipe et l'évaluation d'apprentissages intégrés.

2.1 Incompatibilités rencontrées

Les participants ont relevé deux freins importants à l'application intégrale du concept de justesse des résultats à toute forme d'évaluation : les tâches effectuées en équipe et l'intégration de matières. Ces deux pratiques pédagogiques constituent des stratégies d'enseignement très populaires chez les enseignants parce qu'elles créent un contexte stimulant pour les élèves. Le travail en groupe favorise les échanges, force le raisonnement, exige le développement des relations interpersonnelles et la considération des idées des autres (Blumenfeld et al., 1996). L'intégration des matières, de son côté, permet de rendre le contexte d'apprentissage plus signifiant, plus authentique et favorise le développement d'habiletés complexes (Herman, Ashbacher et Winters, 1992). Ces deux stratégies sont reconnues pour leur capacité à améliorer l'attitude face au travail scolaire. Il arrive aussi qu'elles soient jumelées, donnant des activités d'apprentissage intégrées, effectuées en collaboration par un groupe d'élèves.

2.1.1 Travail en équipe

Le bon fonctionnement d'un groupe de travail nécessite des habiletés particulières (Good & Brophy, 1994). L'équipe fonctionne mieux lorsque la tâche comporte plusieurs solutions possibles, que ceux qui peuvent aider les autres veulent bien le faire et que ceux qui ont besoin d'aide veulent bien le dire et savent aussi comment dire ce qu'ils ne comprennent pas (Blumenfeld, et al., 1996). La collaboration de personnes ayant des habiletés très différentes peut créer des frictions et les élèves faibles sont rarement considérés comme des atouts.

Même si des auteurs (Harnish, 1994; Lussier & Turner, 1995; Mislavy, 1992, 1993; Thouin, 1993; Wiggins, 1992) affirment que l'évaluation en équipe est une suite logique à l'apprentissage en équipe, la présente recherche permet de constater que cette situation risque fort de nuire à la justesse des résultats. L'apprentissage d'un certain contenu et l'apprentissage de la collaboration sont deux aspects différents qui peuvent être évalués, mais il est difficile de se pencher sur les deux en même temps; on pourrait donc choisir d'évaluer l'un ou l'autre aspect (Kulm, 1994). Toutefois, si une seule note est attribuée à l'équipe, elle n'est habituellement pas méritée par tous les membres. Du point de vue motivation, une note commune réduit l'impression de contrôle que chacun a de sa performance, réduisant aussi sa motivation intrinsèque (Deci, & Ryan, 1985). Par contre, si chaque membre du groupe reçoit une note personnelle, la collaboration peut sembler une perte de temps. Dans les deux cas, les rapports interpersonnels à l'intérieur du groupe peuvent en souffrir.

Pour que l'évaluation mène à des résultats justes, il faudrait une gestion rigoureuse de la composition des équipes, du nombre de coéquipiers et de la répartition des tâches à accomplir (Blumenfeld, *et al.*, 1996), ce qui peut devenir lourd pour l'enseignant. Par contre, afin de valoriser le travail en équipe sans pénaliser personne, il est possible d'assigner des tâches individuelles connexes au contenu travaillé en équipe, à chacun des participants et de les noter individuellement, forçant ainsi chacun à trouver profit dans la collaboration. C'est une façon de distinguer l'apprentissage de son évaluation, tout en favorisant des résultats plus justes.

Pour évaluer la contribution au groupe, l'autoévaluation et l'évaluation par les pairs semblent appropriées. Si l'équipe note ouvertement chacun de ses membres à partir d'une description de performance à divers niveaux, en comparant le travail de chacun et en obtenant l'assentiment des autres, la notation sera probablement réfléchie et juste.

2.1.2 Matières intégrées

Du côté des matières intégrées, les vraies préoccupations ont émergé : A-t-on vraiment développé toutes les habiletés qu'on pense avoir développé? L'apprentissage est-il suffisant pour justifier une évaluation? A-t-on assez de traces de ces habiletés pour croire que le résultat est juste? L'interaction des habiletés concernées, si elles sont nombreuses et complexes, ne vient-elle pas brouiller les traces permettant de juger de l'apprentissage réalisé? Plusieurs questions émergent, paraissant toutes très importantes. Même si l'enseignement de matières intégrées est installé depuis longtemps dans la routine pédagogique, le problème qu'elle entraîne lors de l'évaluation ne semble pas avoir déjà émergé dans la documentation, aucune allusion n'ayant été relevée à ce sujet lors de la recension des écrits.

2.2 Intégrabilité

Le manque de connaissance des moyens de faire autrement limite la vision qu'ont les enseignants de la pertinence et de l'intégrabilité du concept de justesse. L'implantation d'une stratégie simple permettant de faire participer les élèves à l'évaluation de leurs pairs lors d'une présentation en salle de classe l'a démontré. L'adoption d'une nouvelle

conception de l'évaluation des apprentissages nécessite un cheminement important et un accompagnement adéquat. Il faudrait probablement de nombreux exemples concrets pour les convaincre de s'y mettre.

Ce n'est pas l'apprentissage en groupe d'élèves ou l'apprentissage de manière intégrée qui porte préjudice à l'élève mais bien l'évaluation dans ce contexte. D'accorder une seule et même note à chaque coéquipier pénalise souvent les plus habiles, au profit de ceux qui le sont moins; ceci nuit à l'esprit d'entraide nécessaire pour bien fonctionner. Tout compte fait, il existe deux possibilités :

- 1) l'activité en équipe se limite à l'apprentissage; l'évaluation, elle, est faite à partir d'une tâche effectuée individuellement.
- 2) l'activité d'apprentissage en équipe met l'accent sur les habiletés sociales à développer et l'évaluation se limite à celles-ci et est faite sur une base individuelle.

Par ailleurs, il est techniquement possible de sélectionner parmi les habiletés impliquées dans l'apprentissage intégré, celles qui sont suffisamment développées pour faire l'objet d'une évaluation. De toute manière, qu'il s'agisse d'évaluation formative ou d'évaluation sommative, la tâche d'évaluation devrait être distincte de celle d'apprentissage, peu importe que l'apprentissage ait été intégré ou pas, qu'il ait eu lieu collectivement ou individuellement. Somme toute, c'est l'habitude bien établie d'octroyer des «points» à tout ce que font les élèves, y compris les activités d'apprentissage, qui crée problème et non

pas les contextes d'apprentissage comme tels. Malheureusement, les discussions à ce sujet démontrent peu de disposition à changer. Cette façon de faire, bien ancrée dans la routine, est presque un élément fondamental de la culture scolaire. Seuls le temps, la discussion de la non pertinence de cette façon de faire et surtout, la connaissance de moyens de faire autrement apparaissent être les meilleures avenues pour motiver des changements à cet effet.

Cette constatation fait surgir le questionnement suivant : Le concept de justesse présente-t-il des limites dans son applicabilité? Ou est-ce plutôt son application qui justement fait ressortir une situation d'évaluation fréquente pour laquelle les résultats d'évaluation ne sont pas justes? Face aux deux obstacles qui viennent d'être discutés, la deuxième option semble la plus juste : le concept de justesse fait ressortir deux pratiques d'évaluation, plutôt répandues, qui ne sont pas susceptibles de générer des résultats justes : l'évaluation de tâches effectuées en équipe et l'évaluation d'apprentissages intégrés.

Chez les participants, le concept de justesse des résultats a été perçu comme un idéal. Ils le conçoivent comme un aide-mémoire visant à éliminer toutes les influences pouvant intervenir sur la justesse des résultats. La perfection étant difficile à atteindre, cette conception de l'évaluation bien menée sert à doser la confiance que devrait inspirer un résultat d'évaluation. Les manquements pourraient permettre d'expliquer un résultat inattendu, mais jamais de le réajuster intuitivement.

À la lumière de ce premier exercice d'application du concept de justesse des résultats à

des situations réelles de la salle de classe, les quatre composantes peuvent être précisées davantage. Non seulement est-il possible de les décrire avec plus de détails mais aussi d'y intégrer des éléments complémentaires, particulièrement en ce qui a trait à l'absence de désavantage circonstanciel. Les tableaux qui suivent présentent donc une version plus élaborée des composantes du concept de justesse, actualisée suite à sa mise à l'essai.

Tableau 1

Première composante du concept de justesse des résultats révisée suite à la mise à l'essai.Pertinence de la tâche d'évaluation par rapport aux habiletés visées**Habiletés :**

- Les habiletés complexes sont décortiquées pour mieux cibler les habiletés mises en relation.
- Le recours à des habiletés complexes en voie d'être atteintes n'est pas implicitement nécessaire.

Critères :

- Les critères sont congruents avec les habiletés ciblées, réalistes et typiques des habiletés en milieu authentique.
- Ils sont les mêmes que ceux valorisés lors de l'apprentissage.
- Ils excluent des considérations étrangères aux habiletés visées (qualité du français, calligraphie, propreté).
- Ils sont définis et communiqués en même temps que la tâche

Tâche :

- La tâche d'évaluation est distincte de celle d'apprentissage (pas de note attribuée pour faire les activités d'apprentissage).
- Elle est appropriée aux habiletés visées, définie clairement, succinctement, en particulier pour les jeunes élèves et techniquement bien construite. (La démonstration d'habiletés complexes est impossible avec des automatismes)
- Son niveau de difficulté convient aux élèves évalués et sa durée respecte la capacité de concentration des élèves.
- Elle vaut la peine d'y investir l'énergie et le temps requis pour bien la faire (motivante et signifiante).
- Elle est nouvelle mais familière.
- Son exécution laisse des traces de la performance des évalués

Tableau 2

Deuxième composante du concept de justesse des résultats révisée suite à la mise à l'essai.

Cohérence de l'activité d'évaluation avec l'activité pédagogique

- Les activités d'apprentissage précèdent l'évaluation et mettent en valeur les mêmes critères de performance.
 - L'activité d'évaluation est en soi une situation d'apprentissage.
 - La situation d'évaluation est familière.
 - La mise en situation est appropriée et stimulante.
 - Du matériel concret ou des ressources sont disponibles, comme pour l'apprentissage
 - De l'aide est disponible pour obtenir des précisions, comme pour l'apprentissage.
 - Les résultats sont exprimés de manière à pouvoir être éventuellement réinvestis (profils d'habiletés).
 - Il y a possibilité de reprise.
-

Tableau 3

Troisième composante du concept de justesse des résultats révisée suite à la mise à l'essai.

Transparence de l'activité évaluative

-
- L'activité d'évaluation est prévue d'avance.
 - Les critères d'évaluation et les indicateurs de rendement sont clairs pour l'élève et connus avec la tâche.
 - Des modèles de travaux sont utilisés pour communiquer les attentes.
 - Les résultats correspondent à des forces et des faiblesses (l'élève peut juger s'ils sont fondés) et sont rattachés à une appréciation de la performance (bien, très bien, exceptionnel).
 - L'élève participe à l'évaluation de ses apprentissages, selon sa maturité et son niveau d'autonomie (autoévaluation, définition des critères, évaluation par les pairs. autoévaluation à l'aveugle, etc.).
 - Il y a retour en classe sur l'activité d'évaluation après la notation.
 - Il est possible de réajuster la note en présence d'arguments convaincants.
-

Tableau 4

Quatrième composante du concept de justesse des résultats révisée suite à la mise à l'essai.

Absence de désavantage circonstanciel

- Les absences de l'élève n'ont pas nuit à un apprentissage convenable.

Travail individuel :

- La tâche ne comporte pas de biais.
- Le temps alloué est amplement suffisant.
- La collaboration entre élèves est documentée, si elle est permise.
- La tâche est effectuée en classe, autant que possible (pas à la maison, à la récré ni à l'heure du dîner).
- Les conditions sont propices à la concentration (absence de : dérangement à l'horaire, d'activité spéciale, de chaleur excessive, de bavardage durant les présentations, de préoccupations personnelles sérieuses).

Travail en équipe :

- Le nombre de personnes convient à l'ampleur de la tâche.
 - Le partage des responsabilités ne permet pas à certains membres d'être exemptés complètement certaines habiletés.
 - Les partenaires sont compatibles (niveau d'engagement, motivation).
 - Le regroupement est géré si les tâches sont susceptibles de comporter des stéréotypes, (par exemple, une tâche impliquant de la mécanique ou la préparation d'aliments sera effectuée par groupes unisexes).
-

Bien que les matières intégrées et les tâches effectuées en équipe comportent toutes les deux de nombreux avantages au point de vue de l'apprentissage, leur évaluation est délicate. En exigeant d'abord l'établissement des habiletés visées, l'application du concept de justesse des résultats fait ressortir les dangers encourus lors de l'évaluation, dans ces deux situations précises. Même si une tâche regroupe plusieurs habiletés, celles-ci ne sont pas nécessairement toutes autant développées les unes que les autres. Il est donc important d'établir lesquelles le sont assez pour être évaluées, comment elles le seront, tout en limitant l'intervention de celles qui ne le sont pas assez. Du côté des tâches effectuées en équipe, les habiletés sociales naturelles ne sont généralement pas également développées chez tous les membres de l'équipe et peuvent masquer, comme c'est souvent le cas, les autres habiletés développées en parallèle. Puisqu'il est difficile de tout évaluer à la fois, il serait préférable d'établir des priorités en décidant à l'avance quel groupe d'habiletés sera évalué et de le faire sur une base individuelle. Dans les deux cas, les habiletés visées par l'évaluation doivent être précisées et les notes attribuées individuellement.

C - SYNTHÈSE DES RÉSULTATS

Jusqu'à tout récemment, la tâche des participants avait été centrée sur l'enseignement, non pas sur l'évaluation; leurs intérêts et leurs pratiques témoignaient cependant du besoin de formation en matière d'évaluation des apprentissages. Leur participation à cette

recherche, malgré un emploi du temps chargé et une sursollicitation par bien d'autres choses, les a conscientisés à la complémentarité de l'évaluation et de l'enseignement dans une activité pédagogique efficace. Malgré le déséquilibre qu'elle a pu apporter momentanément, les participants se disent heureux de l'expérience et auraient aimé qu'elle dure plus longtemps.

Les enseignants-participants ont beaucoup cheminé au cours de ces 45 heures de formation et de discussions, en particulier ceux qui avaient le plus d'expérience; leurs propos sont beaucoup mieux articulés qu'au début lorsqu'ils parlent d'évaluation des apprentissages. Sans prétendre qu'ils soient aptes à comparer différents modèles de qualité des résultats, s'il en existait, la formation qu'ils ont reçue et leur connaissance pratique de l'activité en salle de classe donnent du poids à leur opinion. Même si leurs pratiques d'évaluation sont probablement encore fautives à l'occasion et qu'ils ne peuvent pas toujours justifier leurs façons de faire, leur avis sur la pertinence et l'intégrabilité des composantes du concept de justesse des résultats paraît crédible.

Une fois familiers avec le concept de justesse des résultats et après l'avoir appliqué à une situation de salle de classe, les cinq enseignants participants sont unanimes. Le concept de justesse et ses quatre composantes témoignent bien de la réalité; il rassemble des éléments qui peuvent tous intervenir sur la capacité d'un résultat d'évaluation à situer correctement un élève dans son développement. Ils le trouvent complet et fidèle à la réalité, ils n'ont pas d'ajout ou de retrait à proposer, si ce n'est qu'il devrait être intégré à toute approche d'enseignement.

Malgré cela, ils croient que le concept de justesse ne peut pas toujours être entièrement appliqué, en tout temps; l'absence de désavantage circonstanciel est impossible à réaliser pour tous les élèves en même temps. Aussi, le manque de temps et les ressources matérielles limitées sont des obstacles à la recherche de cohérence entre l'activité évaluative et l'activité pédagogique. De plus, le concept de justesse identifie comme problématique l'évaluation au moyen de tâches effectuées en groupe et prêche la prudence suite à l'enseignement intégré.

Une fois imprégnés de cette manière de penser, les participants disent avoir l'intention de la mettre graduellement en pratique dans leur salle de classe; certains ont d'ailleurs déjà commencé à le faire. Cependant, il n'est pas certain qu'ils sauront tous faire le transfert de ce qu'ils ont appris à d'autres situations analogues; la préférence pour un format ressemblant plutôt à une liste d'étapes à suivre selon le déroulement des activités pédagogiques, telle que mentionnée par Kevin, incite à en douter. Mais peu importe le degré d'intégration de ce concept chez des enseignants, le concept de justesse des résultats d'évaluation ne saura jamais combler les lacunes au niveau de la pédagogie et de la didactique..

L'implantation du concept de justesse des résultats d'évaluation en salle de classe est toutefois confronté à un obstacle majeur : l'habitude d'accorder des «points» à tout ce que font les élèves. La comptabilisation de ces notes avec celles de véritables évaluations nuisent à la justesse des résultats obtenus, en plus de saboter le droit des élèves à l'apprentissage préalable à toute forme d'évaluation. Des recherches ont d'ailleurs déjà

étudié le pouvoir des notes et dénoncé leur utilisation abusive pour contraindre les élèves à travailler (Brookhart, 1993; Cox, 1994).

Enfin, cette collaboration étroite a permis de préciser la mise en application du concept de justesse des résultats. Comme l'affirme Keeves (1988), un modèle a très peu de valeur s'il n'a jamais été confronté à la réalité. L'exercice a permis de préciser les composantes du concept de justesse avec plus de détails, tout en en faisant ressortir les éléments les plus rarement appliqués jusqu'à maintenant dans la salle de classe.

CONCLUSION

Un modèle théorique est une représentation simplifiée d'une réalité complexe; il sert à communiquer une vision spécifique, organisée selon une logique particulière. En éducation il y a plusieurs modèles théoriques de l'apprentissage, **mais pas de modèle théorique de la qualité des résultats d'évaluation des apprentissages en salle de classe**. Il y a une trentaine d'années, des experts dénonçaient l'absence d'un schème adéquat pour appuyer l'évaluation des apprentissages scolaires et assurer des résultats justes, représentatifs du développement de chaque élève. L'éducation a toujours utilisé celui de la psychométrie, même si les concepts de validité et de fidélité ne conviennent pas à la nature des apprentissages et ne peuvent pas s'appliquer à la salle de classe.

Depuis, de nombreux auteurs discutent de moyens d'adapter les concepts de validité et de fidélité des résultats à l'éducation, alors que d'autres soutiennent plutôt la définition de nouveaux concepts propres à ce domaine. À partir de l'analyse approfondie de cette documentation, et plus spécifiquement de ce qui y est reconnu comme étant crucial pour l'obtention de résultats justes dans l'évaluation des apprentissages en salle de classe, cette recherche s'est consacrée à combler cette carence. Il en résulte un ensemble d'éléments coordonnés, rassemblés sous quatre composantes, tous retenus pour leur contribution au concept de justesse des résultats d'évaluation. Ce concept unificateur n'a pas encore entièrement prouvé sa capacité à rendre compte, fidèlement, de la réalité de la salle de classe, mais l'expérience en ce sens encourage à poursuivre la démarche.

L'approche utilisée pour éprouver le concept de justesse des résultats, avec la

collaboration de cinq enseignants d'écoles élémentaires, a d'abord permis une description étoffée et contextualisée de quelques situations d'évaluation en salle de classe. Suite à l'opérationnalisation du concept et à une mise à l'essai, les participants ont pu donner leur opinion sur la pertinence, l'amélioration possible et l'intégrabilité du concept de justesse aux pratiques pédagogiques quotidiennes. Les points saillants des résultats obtenus sont maintenant brièvement discutés.

Pour les enseignants, leur participation à la présente recherche a été l'occasion d'échanger au sujet de l'évaluation. Une fois la confiance établie, les participants ont démontré beaucoup de plaisir à partager leurs expériences personnelles avec les autres. Ces discussions structurées leur ont permis d'exprimer leurs préoccupations, de ventiler leurs frustrations à ce sujet, de stimuler leur réflexion et d'échanger des trucs particulièrement efficaces. La présence d'une personne ressource permettait d'analyser les caractéristiques des situations décrites et d'en faire ressortir les forces et les faiblesses.

Les enseignants ont paru débordés par l'ampleur de leur tâche; la gestion des activités d'apprentissage et la discipline prennent une très grande part de leurs énergies. Les procédures d'évaluation en salle de classe n'étant jamais remises en question, les enseignants y réfléchissent peu et consacrent l'énergie qu'il leur reste à d'autres causes. De toute façon, il n'y a actuellement pas de tribune où ils peuvent exposer leurs préoccupations au sujet de l'évaluation et recevoir une rétroaction constructive. Cette expérience a permis de voir qu'ils sont avides d'informations sur le sujet, qu'ils ont plein de bonnes idées qu'ils ne partagent pas et qu'ils ont besoin d'une structure les soutenant

dans cette tâche particulièrement exigeante. À leur avis, l'arrivée des examens provinciaux vient renforcer ce besoin.

L'évaluation en salle de classe, quoique omniprésente, est réduite à une bien simple expression. Pour la plupart des enseignants-participants, bien que sortis récemment des facultés d'éducation, évaluer veut simplement dire **accorder des points** et ils en accordent à presque tout ce que font les élèves, qu'il s'agisse de tâches d'apprentissage ou de tâches d'évaluation. D'abord au quotidien : des points utilisés comme jetons pour récompenser le bon ordre, la docilité et l'effort, puis, ponctuellement : des points-étalons du degré de rétention d'un contenu précis, une fois sa couverture complétée. Ils disent évaluer parce que les bulletins les y obligent à trois occasions durant l'année scolaire. C'est à ces moments précis et face aux parents que les résultats doivent paraître dénués de subjectivité. Pour ce, ils combinent tous les points attribués depuis le début du trimestre; les calculs mathématiques confèrent aux résultats une aura d'objectivité en laquelle ils croient fermement. L'accumulation d'un grand nombre de points dans chaque discipline leur paraît donc être une pratique consciencieuse, même s'ils ne font aucune distinction du type d'évaluation, ni d'utilisations subséquentes de ces résultats. Tous ces points fondus en un score final inspirent la classification du rendement de l'élève dans l'un des quatre niveaux de performance prévus dans le bulletin (le niveau 3 étant le seul à porter un qualificatif : «pleinement satisfaisant» MÉFO 1997a, 1997b, p. 8).

Selon les observations faites, l'évaluation est encore souvent vue comme un moyen d'exprimer la satisfaction, ou à d'autres occasions, la déception face au respect des exigences établies. Les enseignants ne semblent pas conscients de l'ensemble des

étapes impliquées dans une démarche d'évaluation, ni porter des jugements sur la performance des élèves à chaque fois qu'ils évaluent. Ils n'évaluent pas non plus pour prendre des décisions. Cette mentalité, contraire à la rentabilisation pédagogique de l'évaluation, ne peut pas cohabiter avec une activité évaluative transparente, responsabilisant l'élève et l'intégrant activement dans l'évaluation de ses apprentissages. De plus, la perception réductrice de l'évaluation démontrée par les enseignants-participants demeure encore trop implicite pour être communiquée, discutée et partagée.

La documentation approfondie de quelques situations d'évaluation, telle qu'appliquée ici, met à jour de nombreuses ramifications de cet aspect de l'activité pédagogique et convainc de la place centrale occupée par les habiletés visées, tant dans l'apprentissage que dans son évaluation. Elle fait ressortir l'importance d'examiner les pratiques d'évaluation des apprentissages dans leur contexte et sous toutes leurs facettes. C'est la seule façon de donner un portrait juste de ce qui se passe réellement en salle de classe.

Cette recherche a permis de voir à l'oeuvre des enseignants très préoccupés par l'efficacité et la rectitude de leurs gestes quotidiens, mais d'autres enseignants ont fait germer des doutes sur leur compétence et leur professionnalisme. Par exemple, en exerçant intensivement les élèves à appliquer de manière automatique des algorithmes, des enseignants épargnent à leurs élèves la nécessité de réfléchir. Ils abandonnent, par le fait même, le développement d'habiletés de pensée de niveaux supérieurs, ce qui est un sérieux désengagement de leur part. Lors de la collecte des données, il a été donné

de voir aussi la recherche de résultats confirmant ou appuyant des jugements précédents et le réajustement intuitif des notes estimées injustes.

L'application du concept de justesse des résultats a permis de mettre à jour des conceptions erronées de la pédagogie, notamment celle de l'apprentissage de la lecture. L'exercice individuel, à voix basse, est rarement utilisée comme activité d'apprentissage; c'est la lecture à haute voix par l'enseignant ou par les élèves, à tour de rôle, qui précède habituellement les activités d'évaluation à cet effet. Les diverses habiletés impliquées dans la compréhension de textes, comme de dégager l'idée principale et les idées secondaires, de trouver le sens de mots nouveaux, de faire des inférences et d'établir des liens entre le contenu de la lecture et ses expériences personnelles ne semblent pas être considérées, même en troisième année, là où l'apprentissage de la lecture constitue une partie importante du programme. Ceci est fort inquiétant quand l'on sait que la réussite scolaire est très reliée à la capacité de lecture (Brookhart, 1997). Ces failles ont leur conséquences sur les pratiques d'évaluation en salle de classe et sur la justesse des résultats des élèves.

La prochaine section présente une reformulation des composantes du concept de justesse des résultats, à la lumière de l'expérience vécue lors de leur première confrontation à la réalité scolaire. La section suivante aborde ensuite les limites de la présente recherche et propose des pistes à explorer pour faire évoluer le concept de justesse des résultats.

A - CONCEPT DE JUSTESSE DES RÉSULTATS

Le concept de justesse des résultats est cohérent avec la nature des apprentissages dans le nouveau paradigme de l'éducation. L'évaluation est un outil pédagogique intégré; elle est constructive, favorise l'aide, la communication et le renforcement, responsabilise l'apprenant et s'intéresse à son processus de pensée et à son cheminement.

Un résultat d'évaluation est juste s'il situe correctement un apprenant dans son apprentissage, détermine ses forces et ses faiblesses, ce qu'il maîtrise ou ne maîtrise pas. Un résultat juste est fondé, sensé, représentatif, signifiant et vrai, entaché du moins d'erreur possible. C'est un résultat dans lequel il est possible d'avoir confiance pour situer un élève dans son développement. Pour obtenir un résultat d'évaluation juste, il est nécessaire de respecter les quatre composantes suivantes :

1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées

La situation d'évaluation est distincte de l'apprentissage préalable. La tâche choisie met efficacement en valeur les habiletés développées durant l'apprentissage, sans avoir implicitement recours à des habiletés connexes en voie d'être atteintes. Les critères choisis sont réalistes et typiques des habiletés en milieu réel, et excluent des considérations étrangères aux habiletés visées. Cette tâche est aussi stimulante, définie clairement et vaut la peine d'investir temps et énergie; son niveau de difficulté et sa durée sont appropriés aux caractéristiques des élèves.

2. Cohérence avec l'activité pédagogique

La situation d'évaluation arrive au moment opportun de l'apprentissage, propose à l'élève une tâche familière mais nouvelle et bénéficie d'une mise en situation, comme pour l'apprentissage. Les critères d'évaluation sont les mêmes que ceux valorisés lors des activités d'apprentissage, tout comme les ressources disponibles. La démarche d'évaluation génère des résultats pouvant être réinvestis dans l'activité pédagogique.

3. Transparence de l'activité évaluative

L'élève participe à l'évaluation. Les critères servant à juger la qualité des travaux sont définis clairement et fournis aux élèves au moment où la tâche d'évaluation est assignée. Les résultats sont rattachés à une appréciation de la performance. Le retour sur l'activité d'évaluation en salle de classe après la notation permet à l'élève de situer sa performance par rapport aux attentes énoncées.

4. Absence de désavantage circonstanciel

Autant que possible, l'évaluation se déroule dans des conditions permettant à tous les élèves de donner leur plein rendement. Même s'il est presque impossible de créer des conditions idéales pour tous les élèves au même moment, il faut être conscient que certaines circonstances peuvent nuire à la justesse des résultats. Par exemple, les travaux effectués avec un partenaire qui ne collabore pas, la contribution démesurée d'un parent à une tâche effectuée à la maison, des situations familiales perturbantes, des absences répétées, etc., peuvent nuire à la justesse des résultats.

L'agencement optimal des habiletés visées et du type de tâche choisi pour les faire ressortir s'avère crucial; c'est le fondement de l'obtention de résultats d'évaluation justes. Des lacunes à ce niveau-ci de la démarche d'évaluation mettraient en péril la justesse des résultats, peu importe la rigueur appliquée aux trois autres composantes du concept contribuant à mettre en lumière les apprentissages réalisés. Toutefois, si l'une ou l'autre des composantes est moins bien respectée que les autres, quelle qu'elle soit, l'image finale de l'apprentissage évalué en sera moins juste.

Le concept de justesse des résultats permet d'apprécier la qualité des résultats générés par une situation d'évaluation en faisant ressortir les aspects critiques de son déroulement. Du point de vue de l'enseignant, il met en valeur les éléments déterminants d'une situation d'évaluation bien menée, attire son attention sur les habiletés à évaluer et le force à intégrer son évaluation à l'activité pédagogique.

Il serait possible d'apprécier la justesse d'un résultat de deux manières. Soit *a priori*, en considérant toutes les composantes du concept de justesse avant de procéder à une évaluation, s'assurant ainsi de produire les résultats les plus justes possible, soit *a posteriori*, en utilisant les composantes du concept pour estimer la justesse d'un résultat, après coup, en s'interrogeant sur les conditions dans lesquelles l'évaluation a eu lieu. Dans les deux cas, cette analyse sert à doser la confiance à accorder à un résultat d'évaluation pour situer correctement un élève dans son développement. Il n'est pas ici question de mesurer la justesse, ce qui impliquerait nécessairement de la quantifier (DeLandsheere, 1992), mais simplement d'en apprécier la qualité. Les manquements

peuvent contribuer à expliquer un résultat inattendu, mais jamais à le réajuster intuitivement.

5. Viabilité du concept de justesse des résultats

Se familiariser avec une nouvelle conception de l'évaluation demande du temps et bouleverse des pratiques bien établies. Malgré cela, ce sont les enseignants d'expérience qui ont semblé tirer le plus profit de l'exercice. Après quelques années, ils disent se sentir enfin assez à l'aise dans leur enseignement pour pouvoir penser un peu plus longuement à l'évaluation.

Tous les participants ont trouvé la période d'application du concept de justesse particulièrement enrichissante. Aborder l'évaluation des apprentissages de cette manière les a initiés à une conception générale de l'évaluation cohérente et complémentaire de l'enseignement et de l'apprentissage, ce que peu d'enseignants ont, généralement. À leur avis, malgré le temps qu'elle a nécessité, la participation à cette recherche s'est soldée par un apprentissage important et une prise de conscience qu'ils disent salutaire pour l'exercice de leur profession. Après six mois de collaboration assidue, les participants sont unanimes à dire qu'ils ont aimé cette façon d'aborder l'évaluation. Certains ont même manifesté le désir de poursuivre cette collaboration l'année suivante, bien que la partie empirique de la recherche soit complétée.

En réaction aux quatre composantes du concept de justesse des résultats, tous les participants disent trouver les trois premières composantes pertinentes et intégrables à

leur pratique régulière en salle de classe; il s'agit simplement d'apprendre à le faire. Pour ce qui est de l'absence de désavantage circonstanciel, ils croient important que cette composante demeure. Même s'ils la jugent difficile à respecter, sa présence rappelle qu'il faut en tenir compte autant que possible; elle permet parfois de trouver une explication à un résultat décevant, malgré le respect de toutes les autres composantes.

L'énoncé des habiletés visées apparaît être la force majeure du concept de justesse des résultats en ce qu'il incite à distinguer le rappel de connaissances et de l'application d'algorithmes d'autres habiletés plus complexes. Jusqu'à maintenant, les enseignants se préoccupaient peu des habiletés visées par leurs interventions. Tant pour l'enseignement et l'apprentissage que pour l'évaluation, ils se laissaient guider par un contenu à couvrir. La composante «Pertinence de la tâche d'évaluation par rapport aux habiletés visées» a d'abord fait ressortir, pour la première fois, deux situations d'apprentissage pouvant menacer la justesse des résultats lorsqu'elles sont utilisées aussi pour l'évaluation : le travail en équipe et l'intégration de matières. Ces deux situations présentent des avantages reconnus du point de vue de l'apprentissage mais il ressort que, sans l'établissement rigoureux des habiletés ciblées par l'évaluation, elles risquent de léser les élèves en rendant nécessaires des habiletés connexes en voie d'être atteintes, dans le cas des matières intégrées, ou des habiletés tout à fait étrangères nécessaires au bon fonctionnement du groupe, dans le cas de travail en équipe. Dans les deux cas, le manque d'activités pédagogiques pour développer suffisamment toutes les habiletés impliquées dans l'évaluation pourrait nuire à la justesse des résultats obtenus. De plus, cette composante a fait ressortir toute la place occupée par le développement de

l'autonomie chez les jeunes enfants et comment celle-ci peut masquer la capacité à démontrer la maîtrise de nombreuses autres habiletés à développer.

La composante pertinence de la tâche d'évaluation par rapport aux habiletés visées a aussi mis en relief une faiblesse généralisée de l'intervention pédagogique actuelle : les enseignants n'enseignent pas en vue de développer des habiletés précises chez leurs élèves. Bien que *Le programme d'études commun* (MÉFO, 1995b), en circulation depuis plus de deux ans, ait mis l'accent sur les résultats à atteindre plutôt que sur des contenus à couvrir, l'intervention des enseignants-participants est encore et toujours initiée par des contenus. Le retour à un programme plus classique comme *Le programme cadre de L'Ontario* (MÉFO, 1997a, 1997b) n'améliorera pas l'enseignement de ce point de vue, puisqu'il est revenu à la prescription de contenus.

Le bulletin standardisé de l'Ontario est un autre frein à l'intervention éducative motivée par des habiletés ou des compétences à développer. En exigeant une seule note par discipline, elle perpétue l'habitude de faire des moyennes. De plus, que ces notes soient exprimées en pourcentage en septième et huitième années entretient la croyance que l'apprentissage peut être mesuré avec précision. Aussi, en dénuant le résultat d'évaluation de son contexte et de sa valeur descriptive, il laisse la porte grande ouverte à son utilisation pour n'importe quelle décision.

B - LIMITES ET PISTES DE RECHERCHE

La mise à l'essai a fait ressortir la conception actuelle de l'évaluation des apprentissages chez certains praticiens et révélé l'ampleur du fossé qui la sépare de celle véhiculée par le concept de justesse. L'opinion d'enseignants, bien que très appréciée dans le cas présent, apparaît être limitée par leur connaissance du processus d'évaluation et leur manque de recul par rapport à d'autres manières de faire. Malgré cela, la modeste expérience menée ici confirme la viabilité du concept et permet d'orienter la manière de le perfectionner.

En ce qui a trait aux pistes de recherche, outre sa discussion par des experts en évaluation des apprentissages en salle de classe, l'amélioration de ce concept pourrait se faire de diverses manières. D'abord, le jeune âge des élèves semble être une contrainte à la transparence de la tâche évaluative. Cette question doit donc être examinée de plus près en explorant à fond diverses stratégies pour le faire. Les jeunes enfants sont adaptables et capables de prise en charge personnelle s'ils sont accompagnés convenablement vers l'autonomie. L'exercice mené ici a laissé entrevoir des moyens de rejoindre les élèves de troisième année mais ces moyens doivent être différents de ceux utilisés avec les élèves plus âgés. Avant de pouvoir porter un jugement éclairé sur la flexibilité de cette composante du concept de justesse, il serait nécessaire d'y consacrer plus d'efforts et de temps.

Toujours dans le but d'étudier sa flexibilité, il faut aussi mettre à l'épreuve le concept de justesse des résultats chez de jeunes enfants de première et de deuxième années. Il

semble que la responsabilisation de ces élèves soit souhaitable, mais l'ajuster à leur degré de maturité et d'autonomie pose des défis différents, particulièrement si ces jeunes ne maîtrisent pas encore la lecture. Aussi, il serait intéressant d'explorer également sa pertinence au niveau de l'enseignement secondaire et post-secondaire.

Par ailleurs, il serait utile de mettre à l'essai cette conception de l'évaluation durant toute une année scolaire, en commençant dès le premier jour de classe. Ceci exigerait, d'une part, une connaissance approfondie de ce schème de pensée, une très grande confiance dans cette autre façon de faire et le soutien étroit d'un expert qui y croit encore plus fermement. D'autre part, l'abandon de notes chiffrées comme carburant pour l'activité d'apprentissage, pour toute une année scolaire, pourrait être très révélateur.

Dans la démarche d'évaluation présentée dans de nombreux manuels de mesure et d'évaluation en éducation, l'étape «intention» n'a pas toujours été explicitement présentée comme l'étape précédant la mesure, le jugement et la décision dans la démarche d'évaluation. Cependant, cette manière de concevoir le processus ne laisse aucun doute sur une procédure née de la nécessité de prendre une décision éclairée, soit pour réguler les apprentissages, soit pour reconnaître les acquis. Bien qu'il s'agisse d'une conception de la démarche d'évaluation qui soit logique, la documentation ne rapporte pas de données empiriques sur la pertinence de cette représentation de la réalité; les observations faites au cours de la présente recherche en donne une toute autre image, comme d'autres recherches l'on fait auparavant. Non seulement les enseignants n'évaluent pas du tout pour prendre des décisions, comme Wilson (1990) l'a déjà rapporté, mais ils ne sont pas conscients, non plus, qu'évaluer comporte un jugement de

valeur. Ce n'était pas le mandat de la présente recherche, mais il faudrait éclaircir cette question; elle est déterminante pour l'élaboration d'interventions susceptibles d'améliorer les pratiques d'évaluation en salle de classe.

Enfin, avec l'implantation récente d'examens provinciaux annuels en Ontario, il serait intéressant d'appliquer le concept de justesse aux résultats obtenus dans divers groupes-classes. En examinant les habiletés ciblées par les activités d'évaluation imposées, leur cohérence avec les activités pédagogiques vécues auparavant, leur transparence à l'égard de l'élève et la possibilité de désavantages circonstanciels, il serait possible de doser la confiance à accorder aux résultats de ces examens pour situer correctement les élèves dans leur développement. Le résultat de ces analyses pourraient tempérer les comparaisons possibles d'un groupe à l'autre et expliquer certains écarts de performance.

Pour conclure, disons que cette première opérationnalisation du concept de justesse des résultats d'évaluation s'avère très positive. Le concept est jugé utile, réaliste et intégrable aux activités pédagogiques de la salle de classe, malgré l'écart qu'il présente dans la manière de concevoir l'activité évaluative. Malgré le caractère modeste de son application, limite importante de cette recherche, la mise à l'essai a fait ressortir pour la première fois l'aspect problématique des situations d'évaluation regroupant des élèves et celles d'apprentissages intégrés. Elle a permis de pointer du doigt des pratiques pédagogiques fautives, comme des stratégies d'enseignement de la lecture et la préparation abusive des élèves à l'évaluation en salle de classe. Elle a aussi fait ressortir deux freins importants

à l'instauration de pratiques évaluatives plus justes : l'attribution de points à tout ce que font les élèves et l'intervention pédagogique axée sur des contenus à couvrir plutôt que sur des habiletés à développer.

L'évaluation en salle de classe est la plus déterminante pour le vécu scolaire des élèves. Avec l'avènement de l'imputabilité et de l'égalité des chances en éducation, il apparaît primordial de poursuivre la recherche sur le concept de justesse des résultats parce qu'il pourrait, éventuellement, devenir un schème de référence édumétrique convenable. Il pourrait servir à inculquer aux enseignants les principes fondamentaux d'évaluation des apprentissages, faciliter la recherche en ce domaine et uniformiser davantage les pratiques.

Le projet est ambitieux . . . mais à l'aube d'un nouveau millénaire, n'est-il pas permis de l'être?

RÉFÉRENCES

- Abrecht, R. (1991). *L'évaluation formative, une analyse critique*. Bruxelles : De Boeck.
- Airasian, P. W., Kellaghan, T., & Madaus, G. F. (1977). The stability of teachers' perceptions of pupil characteristics. *The Irish Journal of Education*, 11(1-2), 68-84.
- Anderson, J. (1989). Evaluation of student achievement : Teacher practices and educational measurement. *The Alberta Journal of Educational Research*, 35(2), 123-133.
- Archbald, D. (1991). Authentic assessment : Principles, practices, and issues. *School Psychology Quarterly*, 6(4), 279-293.
- Arter, J., Spandel, V., Culham, R., & Pollard, J. (1994). *The impact of training students to be self-assessors of writing*. Paper presented at the Annual Meeting of the American Educational Research Association. New Orleans, LA.
- 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. In D. Laveault (Éd.), *Les pratiques d'évaluation en éducation* (pp. 151-165). XV^e session d'étude de l'Association pour le développement de la mesure et de l'évaluation en éducation. ADMEE-Canada.
- Bachor, D., & Anderson, J. (1994). Elementary teachers' assessment practices as observed in the province of British Columbia. *Assessment in Education*, 1(1), 63-93.
- Barnes, S. (1985). Study of classroom pupil evaluation : the missing link in teacher education. *Journal of Teacher Education*, 36(4), 46-49.
- Baron, J. (1991a). Performance assessment : Blurring the edges of assessment, curriculum, and instruction. In G. Kulm & S. Malcom (Eds.), *Science assessment in the service of reform* (pp. 247-266). Washington : American Association for the Advancement of Science.
- Baron, J. (1991b). Strategies for the development of effective performance exercises. *Applied Measurement in Education*, 4(4), 305-318.
- Bateson, D. (1990). Measurement and evaluation practices of British Columbia science teachers. *The Alberta Journal of Educational Research*, 36(1), 45-51.
- Bateson, D. (1994). Psychometric and philosophic problems in "authentic" assessment : Performance tasks and portfolios. *The Alberta Journal of Educational Research*, 40(2), 233-245.
- Baxter, G., & Shavelson, R. (1994). Science performance assessments : Benchmarks and surrogates. *International Journal of Educational Research*, 21, 279-297.
- Bercier-Larivière, M. (1994). *La motivation des élèves de 6^e, 7^e et 8^e année et leur perception des pratiques d'évaluation des apprentissages*. Thèse de maîtrise non publiée, Université d'Ottawa, Ottawa.

Bercier-Larivière, M., & Forgette-Giroux, R. (1995). L'évaluation des apprentissages telle que perçue par des élèves de 6^e, 7^e et 8^e année. *Mesure et évaluation en éducation*, 18(1), 37-58.

Bercier-Larivière, M., & Forgette-Giroux, R. (1998). *L'évaluation des élèves et la motivation scolaire. Apprentissage et socialisation*, 18(1 - 2).

Berlak, H. (1992). Toward the development of a new science of educational testing and assessment. In H. Berlak et al., *Toward a new science of educational testing & assessment* (pp. 181-206). Albany : New York Press.

Bertrand, R., & Tremblay-Desrochers, M. (1985). Analyse comparative de trois modèles d'instruments diagnostiques en mathématiques. *Mesure et évaluation en éducation*, 8(3), 5-40.

Biggs, J. (1995). Assessing for learning : Some dimensions underlying new approaches to educational assessment. *The Alberta Journal of Educational Research*, 41(1), 1-17.

Bloom, B. Hastings, T. & Madaus, G. (1971). *Handbook on formative and summative evaluation of student learning*. New York : McGraw-Hill.

Bloom, B., Madaus, G., & Hasting, T. (1981). *Evaluation to improve learning*. Toronto : McGraw-Hill.

Blumenfeld, P., Marx, R., Soloway, E., & Krajcik, J. (1996). Learning with peers : From small group cooperation to collaborative communities. *Educational Researcher*, 25(8), 37-40.

Bogdan, R. C., & Biklen, S. K. (1992). *Qualitative research for education. An introduction to theory and methods* (2nd édition). Toronto : Allyn and Bacon.

Bonniol, J.-J., & Vial, M. (1997). *Les modèles de l'évaluation. Textes fondateurs avec commentaires*. Paris : De Boeck Université.

Brown, A. L., & Campione, J. C. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons : Integrating cognitive theory and classroom practice* (pp. 229-270). Cambridge, MA : MIT Press.

Bruer, J. T. (1994). *Schools for thought. A science of learning in the classroom*. Boston : Massachusetts Institute of Technology.

Caouette, C. (1992). *Si on parlait d'éducation*. Montréal : VLB Éditeur.

Cardinet, J. (1985) Des instruments d'évaluation pour chaque fonction. *Mesure et valuation en éducation* 8(1-2), 45-118.

Cardinet, J., & Laveault, D. (1996). Dix années de travaux européens et nord-américains sur l'évaluation : quelles lignes de force? *Mesure et évaluation en éducation*, 18(3), 1-25.

- Carter, K. (1986). Test-wiseness for teachers and students. *Educational Measurement : Issues and Practice*, 5(4), 20-23.
- Carver, R. (1974). Two dimensions of tests. *American Psychologist*, 29(7), 512-518.
- Champagne, A., & Newell, S. (1992). Directions for research and development : Alternative methods of assessing scientific literacy. *Journal of Research in Science Teaching*, 29(8), 841-860.
- Chase-Furman, G. (1994). Outcome-based education and accountability. *Education and Urban Society*, 26(4), 417-437.
- Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data. Complementary research strategies*. London: Sage Publications.
- Commission royale sur l'éducation (1994). Pour l'amour d'apprendre : Rapport de la Commission royale sur l'éducation. Toronto : Imprimerie de la reine pour l'Ontario.
- Conseil supérieur de l'éducation. (1992). *Évaluer les apprentissages au primaire : Un équilibre à trouver*. Québec : Conseil supérieur de l'éducation.
- Cox, K. I. (1994). *What 'counts' in english class? A study of classroom assessment practices among California high school teachers*. Doctoral Thesis. Riverside : University of California. Dissertation Abstracts International, DAI-A 56/03, p. 851.
- Cox, K. I. (1995). *What counts in english class? Selected findings from a statewide study of California high school Teachers*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Cronbach, L. J. (1969). Validation of educational measures. In P. H. DuBois (Ed.), *Proceedings of the invitational conference on testing problems* (pp. 35-52). Princeton : Educational Testing Service.
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.
- Darling-Hammond, L. (1994). Symposium : Equity in educational assessment. *Harvard Educational Review*, 64(1), 1-30.
- Dassa, C., (1988). L'intégration du diagnostic pédagogique aux apprentissages scolaires : de la théorie à la voie informatique. *Mesure et évaluation en éducation*, 11(1), 7-26.
- Dassa, C., & Gaudreau, J. (1989). Limite de l'évaluation formative auprès des futurs maîtres : Le dilemme entre l'excellence et la compétence, *Mesure et évaluation en éducation*, 12(1), 23-39.
- Davinroy, K. H. (1994). "How does my teacher know what I know?" *Third graders' perceptions of math, reading, and assessment*. Paper presented at the Annual Meeting of the American Educational Research Association. New Orleans.

- Deci, E., & Ryan, R. (1985). *Intrinsic motivation and self-determination in human behavior*. New York : Plenum Press.
- DeLandsheere, G. (1992). *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris : Presses Universitaires de France.
- Dempster, F. (1992). Using tests to promote learning : A neglected classroom resource. *Journal of Research and Development in Education*, 25(4), 213-217.
- Denzin, N. K. (1988). Triangulation. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement : An international handbook* (pp. 511-513). Oxford : Pergamon Press.
- DePascale, C. (1990). *Factors influencing the cognitive emphasis of classroom tests*. Doctoral Thesis, University of Minnesota, Dissertation Abstracts International, DAI-A 51/05, p. 1587.
- Desrosiers, P., Godbout, P., & Marzouk, A. (1992). Des pistes pour soutenir la transformation des pratiques évaluatives. In D. Laveault, *Les pratiques d'évaluation en éducation* (pp. 139-150). XV^e session d'étude de l'Association pour le développement de la mesure et de l'évaluation en éducation. ADMEE-Canada.
- Doyon, C., & Juneau, R. (1991). *Faire participer l'élève à l'évaluation de ses apprentissages*. Laval : Éditions Beauchemin.
- Dunbar, S., Koretz, D., & Hoover, H. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education*, 4(4), 289-303.
- Elliot, S. (1991). Authentic Assessment : An introduction to a neobehavioral approach to classroom assessment. *School Psychology Quarterly*, 6(4), 273-278.
- Erickson, F. (1986). Qualitative methods in research on teaching. In V. Shulman & M. Wittrock, *Handbook of research on Teaching* (3rd edition, pp. 119-161). New York : Macmillan.
- Ferrara, S., Goldberg, G., & Mctighe, J. (1995). *Ways in which teachers communicate learning targets, criteria, and standards for performance to their students*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.
- Fitzpatrick-Reardon, S., Scott, K., & Verre, J. (1994). Symposium : Equity in educational assessment. *Harvard Educational Review*, 64(1), 1-4.
- Fleming, M., & Chambers, B. (1983). Teacher-made tests : Windows to the classroom. In W. E. Hathaway (Ed.), *Testing in the schools* (pp. 29-47). San Francisco : Jossey-Bass.
- Fontana, A. & Frey, J. (1994). Interviewing : The art of science. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research*. (pp. 361-376). Thousand Oaks (CA) : Sage.

Forgette-Giroux, R., Simon, M., & Bercier-Larivière, M. (1996). Les pratiques d'évaluation des apprentissages en salle de classe : Perceptions des enseignants et des enseignantes. *Revue canadienne de l'éducation*, 21(4), 384-395.

Fournier, C., & Laveault, D. (1992). Liens entre anticipation, autoévaluation et résultat à un examen de rendement scolaire. In D. Laveault (Éd.), *Les pratiques d'évaluation en éducation* (pp. 49-61). XV^e session d'étude de l'Association pour le développement de la mesure et de l'évaluation en éducation. ADMEE-Canada.

Fraser, B. J. (1986). *Classroom environment*. London : Croom Helm.

Fraser, B., & Tobin, K. (1991). Combining qualitative and quantitative methods in classroom environment research. In B. Fraser & H. Walberg (Eds.), *Educational environments. Evaluation, antecedents and consequences* (pp. 271-292). New York : Pergamon Press.

Frederiksen, J., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(9), 27-32.

Gagné, F., & Thouin, M. (1991). L'évaluation formative des apprentissages en orthographe et l'attitude des élèves à l'égard de l'évaluation. *Mesure et évaluation en éducation*, 14(1), 5-15.

Garcia, G. E., & Pearson, P. D. (1994). Assessment and Diversity. *Review of Research in Education*, 20, 337-391.

Gardner, H. (1992). Assessment in context : The alternative to standardized testing. In B. Gifford & M. O'Connor, *Changing assessments : Alternative views of aptitude, achievement, and instruction* (pp. 77-120). Boston : Kluwer.

Giasson, J. (1995). *La lecture. De la théorie à la pratique*. Montréal : Gaëtan Morin.

Gipps, C. (1994a). *Beyond testing : Towards a theory of educational assessment*. Washington: Falmer Press.

Gipps, C. (1994b). Quality in teacher assessment. In W. Harlen (Ed.), *Enhancing quality in assessment* (pp. 71-86). London : Chapman.

Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Philadelphia : Open University Press.

Glaser, R., Lesgold, A., & Lajoie, S. (1984). Toward a cognitive theory for the measurement of achievement. In R. Ronning, J. Glover & J. Conoley : *The influence of cognitive psychology on testing*, (Vol. 3, pp. 41-85). Buros-Nebraska Symposium on Measurement and Testing. London : Lawrence Erlbaum Associates.

Glaser, R., & Silver, E. (1994). *Assessment, testing and instruction : Retrospect and Prospect*. National Center for Research on Evaluation, Standards, and Student Testing. CSE Technical Report 379, Los Angeles.

Good, T. L., & Brophy, J. E. (1994). *Looking in classrooms* (6th edition). New York : Harper Collins.

Gordon, B., Engelhard, G., Gabrielson, S., & Bernknopf, S. (1996). Conceptual issues in equating performance assessments : Lessons from writing assessment. *Journal of Research and Development in Education*, 29(2), 81-88.

Gravel, L. (1991). *L'évaluation du rendement de l'élève dans les écoles élémentaires et secondaires de langue française de la région Nord-Est*. Document non publié, Université d'Ottawa.

Gullickson, A. (1984). Teacher perspectives of their instructional use of tests. *Journal of Educational Research*, 77(4), 244-248.

Haertel, E. (1992). Performance measurement. In M. Alkin, *Encyclopedia of Educational Research* (6th edition, pp. 984-989). Toronto : Macmillan.

Hambleton, R., & Murphy, E. (1992). A psychometric perspective on authentic measurement. *Applied Measurement in Education*, 5(1), 1-16.

Hammersley, M. (1993). *Educational research : Current issues*. London : Chapman.

Hammersley, M. (1990). *Classroom ethnography : empirical and methodological essays*. Toronto : OISE Press.

Hardy, R. (1996). Performance assessment : Examining the costs. In M. B. Kane & R. Mitchell (Eds.), *Implementing performance assessment : promises, problems, and challenges* (pp. 107-117). Mahwah NJ : Lawrence Erlbaum Associates.

Harlen, W. (Ed.) (1994). *Enhancing quality in assessment*. London : Chapman.

Harmon, M. (1991). Fairness in testing : Are science education assessments biased? In G. Kulm & S. Malcom (Eds.), *Science assessment in the service of reform* (pp. 31-54). Washington : American Association for the Advancement of Science.

Harnisch, D. (1994). Performance assessment in review : New directions for assessing student understanding. *International Journal of Educational Research*, vol. 21, 341-350.

Harnisch, D., & Mabry, L. (1993). Issues in the development and evaluation of alternative assessments. *Journal of Curriculum Studies*, 25(2), 179-187.

Herman, J. (1992). What research tells us about good assessment. *Educational Leadership*, 49(8), 74-78.

Herman, J., Aschbacher, P., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria (VA) : ASCD.

Herman, J., & Dorr-Bremme, D. (1984). *Teachers and testing : Implications from a national study*. Center for the Study of Evaluation. LA : California University, Centre for the study of Evaluation.

Herman, J. L. & Klein, D. C. D. (1996). Evaluating equity in alternative Assessment : An illustration of opportunity-to-learn issues. *The Journal of Educational Research*, 89(4), 246-256.

Higgins, N., & Rice, E. (1991). Teachers' perspectives on competency-based testing. *Educational Technology Research and Development*, 39(3), 59-69.

Hipps, J. (1993). *Trustworthiness and authenticity : Alternate ways to judge authentic assessments*. Paper presented at the Annual Meeting of the American Educational Research Association. Atlanta.

Horowitz, R. (1995). A 75-year legacy on assessment : Reflections from an interview with Ralph W. Tyler. *The Journal of Educational Research*, 89(2), 68-75.

Hould, B. (1992). Comment redonner à mes étudiants de l'autonomie et du pouvoir sur leurs apprentissages. In D. Laveault (Éd.), *Les pratiques d'évaluation en éducation* (pp. 87-99). XV^e session d'étude de l'Association pour le développement de la mesure et de l'évaluation en éducation. ADMEE-Canada.

Jaeger, R. (1989). Certification of student competence. In R. Linn, *Educational Measurement* (3rd edition, pp. 485-514). National Council on Measurement in Education and American Council on Education. New York : Macmillan Pub. Co.

Jonnaert, P., & Laveault, D. (1994). Évaluation de la familiarité de la tâche : quelle confiance accorder à la perception de l'élève? *Revue des sciences de l'éducation*, 20(2), 271-291.

Keeves, J. P. (1988). Models and Model building. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement : An international handbook* (pp. 559-566). Oxford : Pergamon Press.

Kelly, B. et al. (1988). *Découvertes Mathématiques 5*. Éditions Accord, distribué par Addison-Wesley. 406 pages, ISBN 2-920868-08-X.

Kirst, M. (1991). Interview on assessment issue with Lorrie Shepard. *Educational Researcher*, 20(2), 21-23.

Kulm, G. (1993). *A theory of classroom assessment and teacher practice in mathematics*. Paper presented at the Annual Meeting of the American Educational Research Association. Atlanta, GA.

Kulm, G. (1994). *Mathematics assessment. What works in the classroom*. San Francisco : Jossey-Bass.

Kulm, G., & Stuessy, C. (1991). Assessment in science and mathematics education reform. In G. Kulm & S. Malcom (Eds.), *Science assessment in the service of reform* (pp. 89-96). Washington : American Association for the Advancement of Science.

Kusch, J. W. (1995). *Teaching in a surveillance mode: a case study of how methods course syllabi and practicum classrooms construct assessment as classroom control*. Doctoral Thesis, University of Wisconsin-Madison. Dissertation Abstract International, DAI-A 56/05, p. 1645.

- Lane, S. (1993). The conceptual framework for the development of a mathematics performance assessment instrument. *Educational Measurement : Issues and Practice*, 12(2), 16-23.
- Lane, S., & Glaser, R. (1995). Assessment in the service of learning. In L. W. Anderson (Ed.), *International encyclopedia of teaching and teacher education* (2nd edition, pp. 279-282). Cambridge U. K. : University Press.
- Leder, G. (1992). *Assessment and learning of mathematics*. Victoria, Australia : Australian Council for Educational Research.
- Legendre, R. (1993). *Dictionnaire actuel de l'éducation* (2e édition). Montréal : Guérin.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills : Sage Publications.
- Linn, R. (1989). *Educational Measurement* (3rd edition). National Council on Measurement in Education and American Council on Education. New York : Macmillan Pub. Co.
- Linn, R. (1992). Achievement testing. In M. Alkin, *Encyclopedia of Educational Research* (6th edition, Vol. 1, pp. 1-12), Toronto : Macmillan.
- Linn, R., Baker, E., & Dunbar, S. (1991). Complex, performance-based assessment : Expectations and validation criteria. *Educational Researcher*, 20(8), 15-21.
- Linn, R., & Burton, E. (1994). Performance-based assessment : Implications of task specificity. *Educational Measurement : Issues and Practice*, 13(1), 5-8, 15.
- Lussier, D. (1992). *Évaluer les apprentissages, dans une approche communicative*. Paris : Hachette.
- Lussier, D., & Turner, C. (1995). *Le point sur ... L'évaluation en didactique des langues*. Anjou : Centre Éducatif et Culturel.
- Macdonald, W. J. (1993). *Student assessment in the elementary intermediate visual arts program : A closer look at teacher assessment methods*. Doctoral Thesis, Toronto : University of Toronto. Dissertation Abstracts International, DAI-A 54/09, p. 3299.
- Massé, P. (1992). *Méthodes de collecte et d'analyse de données en communication*. Sillery : Presses de l'Université du Québec.
- McComas, W. F. (1989). The issue of effective and valid student evaluation. *Education and Urban Society*, 22(1), 72-82.
- McLean, L. (1990). Time to replace classroom test with authentic measurement. *The Alberta Journal of Educational Research*, 36(1), 78-84.
- McMorris, R., & Boothroyd, R. (1993). Tests that teachers build : An analysis of classroom tests in science and mathematics. *Applied Measurement in Education*, 6(4), 321-342.

Mehrens, W. (1991). Using performance assessment for accountability purposes. *Educational Measurement, Issues and Practice*, 11(1), 3-9, 20.

Messick, S. (1989a). Meaning and values in test validation : The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11.

Messick, S. (1989b). Validity. In R. Linn, *Educational Measurement* (3rd edition, pp. 13-103). National Council on Measurement in Education and American Council on Education. New York : Macmillan Pub. Co.

Messick, S. (1992). Validity of test interpretation and use. In M. C. Alkin, *Encyclopedia of Educational Research* (6th edition, pp. 1487-1495), Toronto : Macmillan.

Meyer, C. (1992). What's the difference between "authentic" and "performance" assessment? *Educational Leadership*, 49(8), 39-41.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd edition). London : Sage.

Ministère de l'Éducation de l'Ontario. (1976). *L'évaluation du rendement de l'élève*. Documentation à l'usage des enseignants.

Ministère de l'Éducation du Québec (1985a). *L'évaluation formative*. Série Éléments de domimologie, fascicule 3, Québec.

Ministère de l'Éducation du Québec (1985b). *L'évaluation pédagogique, une démarche*. Série Éléments de domimologie, fascicule 1, Québec.

Ministère de l'Éducation et de la Formation de l'Ontario (1995a). Communiqué. Le ministre Snobolen dépose des projets de loi créant l'Ordre des enseignantes et des enseignants et permettant l'administration de tests aux élèves de la province. 14 décembre 1995. Internet, site du MÉFO.

Ministère de l'Éducation et de la Formation de l'Ontario (1995b). *Le programme d'études commun. Politiques et résultats d'apprentissage de la 1^{re} à la 9^e année*. Imprimerie de la Reine pour L'Ontario.

Ministère de l'Éducation et de la Formation de l'Ontario (1996). Déclaration à l'Assemblée législative du ministre de l'Éducation et de la Formation John Snobelen. Diffusion du bulletin provincial normalisé. 9 octobre 1996. Internet, site du MÉFO.

Ministère de l'Éducation et de Formation de l'Ontario (1997a). Le curriculum de l'Ontario de la 1^{re} à la 8^e année. Français. Toronto : Imprimeur de la reine pour l'Ontario.

Ministère de l'Éducation et de Formation de l'Ontario (1997b). Le curriculum de l'Ontario de la 1^{re} à la 8^e année. Mathématiques. Toronto : Imprimeur de la reine pour l'Ontario.

Ministère de l'Éducation et de Formation de l'Ontario (1997c). Notes pour une allocution de M. John Snobelen ministre de l'Éducation et de la formation annonce de la restructuration de l'éducation. 13 janvier 1997, Internet, site du MÉFO.

Mislevy, R. (1992). *Linking educational assessments. Concepts, issues, methods, and prospects*. Princeton, Educational Testing Service.

Mislevy, R. (1993). Foundations of a new test theory. In N. Frederiksen, R. Mislevy & I. Bejar, (Eds.) *Test theory for a new generation of tests* (pp. 19-39). Hillsdale, NJ : Lawrence Erlbaum Associates.

Mitchell, R. (1992). *Testing for learning : How new approaches to evaluation can improve american schols*. New York : Free Press.

Moss, P. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62(3), 229-258.

Moss, P. (1994). Can there be validity without reliability? *Educational Researcher*, 23(2), 5-12.

Moss, P. (1995). Themes and variations in validity theory. *Educational Measurement : Issues and Practice*, 14(2), 5-13.

Moss, P. (1996). Enlarging the dialogue in educational measurement : Voices from interpretive research traditions. *Educational Researcher*, 25(1), 20-28.

Mucchielli, A. (1996). *Dictionnaire des méthodes qualitatives en sciences humaines et sociales*. Paris : Armand Colin

Nicholson, D., & Anderson, J. (1993). A time and place for observations : Talking with primary teachers about classroom assessment. *The Alberta Journal of Educational Research*, 39(3), 363-374.

Nitko, A. (1989). Designing tests that are integrated with instruction. In R. Linn (Ed.), *Educational measurement* (3rd edition, pp. 447-473). New York : American Council of Education/Macmillan.

Oescher, J., & Kirby, P. C. (1990). *Assessing teacher-made tests in secondary math and science classrooms*. Paper presentend at the meeting of the National Council on Measurement in Education, Boston.

O'Neil, J. (1992). Putting performance assessment to the test. *Educational Leadership*, 49(8), 14-19.

Ouelette, L.-M. (1989). *Modèle interactif d'évaluation formative et relation de formation*. Thèse de doctorat, non-publiée. Montréal : Université de Montréal.

Parker, E. L., Armengol, R., Brooke, L. B., Carper, K., Cronin, S., Denman, A., Irwin, P., McGunnigle, J., Pardini, T., & Kurtz, N. (1995). Teacher's Choices in Classroom Assessment (Reading Assessment). *Reading Teacher*, 48(7), 622-24.

Parker, R. E. (1992). *Mathematical power for all kids : Aligning classroom instruction with mathematics reform goals*. Doctoral Thesis, University of Oregon. Dissertation Abstracts International, DAI-A 53/08, p. 2724.

Peddie, R. (1992). *Beyond the norm? An introduction to standards-based assessment*. Wellington : New Zealand Qualifications Authority.

Pellegrino, J. (1992). Commentary : Understanding what we measure and measuring what we understand. In B. Gifford & M. O'Connor, *Changing assessments : Alternative views of aptitude, achievement, and instruction* (pp. 275-300). Boston : Kluwer.

Pilcher, J. K. (1994). The value-driven meaning of grades. *Educational Assessment*, 2(1), 69-88.

Poisson, Y. (1990). *La recherche qualitative en éducation*. Sillery : Presses de l'Université du Québec.

Popham, J. (1990). *Modern educational measurement : a practitioner's perspective* (2nd ed.). Toronto : Prentice-Hall.

Popham, J. (1994). The instructional consequences of criterion-referenced clarity. *Educational Measurement : Issues and Practice*, 13(4), 15-18, 30.

Poupart, J., Deslauriers, J.-P., Groulx, L.-H., Laperrière, A., Mayer, R., Pires, A. (1997). *La recherche qualitative. Enjeux épistémologiques et méthodologiques*. Montréal : Gaëtan Morin.

Pourtois, J. P., & Desmet, H. (1988). *Épistémologie et interprétation en sciences humaines*. Liège : Pierre Mardaga.

Pryor, E. L. (1992). *Case studies of three first-grade teachers' assessment and evaluation of literacy (grading)*. Doctoral Thesis, Kent State University. Dissertation Abstracts International, DAI-A 56/08, p. 2669.

Redding, N. (1992). Assessing the big outcomes. *Educational Leadership*, 49(8), 49-51.

Quellmalz, E. (1991). Developing criteria for performance assessments : The missing link. *Applied Measurement in Education*, 4(4), 319-331.

Resnick, L. (1994). Performance puzzles. *American Journal of Education*, 102(4), 511-526.

Rheinberg, F. (1983). Achievement evaluation : A fundamental difference and its motivational consequences. *Studies in Educational Evaluation*, 9, 185-194.

Rieben, L. (1988). Un point de vue constructiviste sur la pédagogie de maîtrise. In M. Huberman (Éd.), *Assurer la réussite des apprentissages scolaires? Les propositions de la pédagogie de la maîtrise* (pp. 127-154). Paris : Delachaux et Niestlé.

Rogers, T. (1991). Educational assessment in Canada : Evolution or extinction? *The Alberta Journal of Educational Research*, 37(2), 179-192.

Romberg, T. (1992). Assessing mathematics competence and achievement. In H. Berlak et al. *Toward a new science of educational testing & assessment* (pp. 23-52). Albany : State University of New York Press.

Rope, J. F. (1994). *Identifying differences in testing attitudes, practices, and concerns of teachers in outcome and nonoutcome-based elementary schools*. Doctoral Thesis, Iowa State University. Dissertation Abstracts International, DAI-A 56/02, p. 425.

Ross, J. L. (1994). *The alignment of teacher classroom assessment with a state performance assessment program*. Doctoral Thesis, University of Maryland College Park. Dissertation Abstracts International, DAI-A 55/10, p. 3169.

Rowley, J. T. (1988). *A study of how teachers plan their lessons : Student needs vs. maintaining activity flow*. Thèse de doctorat. University of San Francisco. Dissertation Abstracts, DAI-A 50/07, p. 1925.

Rowntree, D. (1987). *Assessing students : How shall we know them?* London : Kogan Page.

Satterly, D. (1994). Quality in external assessment. In W. Harlen (Ed.), *Enhancing quality in assessment* (pp. 53-70). London : Chapman.

Sawyer, M. H. (1994). *Professional development and educational reform : a study of changes in teachers and classrooms during literature portfolio implementation*. Doctoral Thesis, State University of New York at Albany. Dissertation Abstracts International, DAI-A 55/09, p. 2752.

Scallon, G. (1988a). *L'évaluation formative des apprentissages. La réflexion*. Québec : Presses de l'Université Laval.

Scallon, G. (1988b). Plaidoyer pour une méthodologie instrumentée d'évaluation formative. *Mesure et évaluation en éducation*, 11(1), 43-55.

Schwager, M. T. (1994). *Paradigms of assessment : What is authentic assessment in your world view?* Doctoral Thesis. Riverside : University of California. Dissertation Abstracts International, DAI-A 55/12, p. 3792.

Schwager, M., & Carlson, J. (1994). Building assessment cultures. Teacher perceptions and school environments. *Education and Urban Society*, 26(4), 390-403.

Shavelson, R., Baxter, G., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, 30(3), 215-232.

Shavelson, R., Baxter, G., & Pine, J. (1991). Performance assessment in Science. *Applied Measurement in Education*, 4(4), 347-362.

Shepard, L. (1989). Why we need better assessments. *Educational Leadership*, 46(7), 4-9.

Shepard, L. (1991). Psychometricians' beliefs about learning. *Educational Researcher*, 20(7), 2-16.

Shepard, L. (1992a). Commentary : What policy makers who mandate tests should know about the new psychology on intellectual ability and learning. In B. Gifford & M. O'Connor (Eds.), *Changing assessments : Alternative views of aptitude, achievement, and instruction* (pp. 301-328). Boston : Kluwer.

Shepard, L. (1992b). Uses and abuses of testing. In M. Alkin, *Encyclopedia of Educational Research* (6th edition, Vol. 4, pp. 1477-1485). Toronto : Macmillan.

Shepard, L. (1993). Evaluating test validity. *Review of Research in Education*, 19, 405-450.

Shepard, L. et al., (1996). Effects of introducing classroom performance assessments on student learning. *Educational Measurement : Issues and Practice*, 15(3), 7-18.

Simmons, R. (1994). The horse before the cart : Assessing for understanding. *Educational Leadership*, 51(5), 22-23.

Simon, M., Forgette-Giroux, R., & Bercier-Larivière, M. (1993). *L'évaluation des apprentissages des élèves*. Rapport d'enquête menée auprès d'enseignants du secondaire du Conseil des Écoles Séparées Catholiques de Stormont, Dundas et Glengarry.

Smith, C. (1992). *Different hats, different questions*. Assessment audiences/questions /answers. Paper presented at a Conference of School Principals, New York.

Smith, K. (1990). *Let your students write their own tests*. Paper presented at the Annual Meeting of the International Association of Teachers of English as a Foreign Language. Dublin, Ireland.

Stake, R. E. (1994). Case studies. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (pp. 236-247). London : Sage.

Stiggins, R. (1990). Toward relevant classroom assessment research agenda. *The Alberta Journal of Educational Research*, 36(1), 92-97.

Stiggins, R. (1991a). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-539.

Stiggins, R. (1991b). Facing the challenges of a new era of educational assessment. *Applied Measurement in Education*, 4(4), 263-273.

Stiggins, R. (1994). *Student-centered classroom assessment*. Toronto : Maxwell Macmillan.

Stiggins, R., Bridgeford, N., Conklin, N., & Brody, C. (1985). *A feeling for the student : An analysis of the art of classroom assessment*. Northwest Regional Educational Laboratory, Portland.

Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands. Investigating the practices of classroom assessment*. New York : State University of New York Press.

Stiggins, R., Miller-Griswold, M., & Reed-Wikelund, K. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement*, 26(3), 233-246.

- Stone-Wiske, M. (1994). How teaching for understanding changes the rules in the classroom. *Educational Leadership*, 51(5), 19-21.
- Sugrue, B. (1995). A theory-based framework for assessing domain-specific problem-solving ability. *Educational Measurement : Issues and Practice*, 14(3), 29-35.
- Szetela, W., & Nicol, C. (1992). Evaluating problem solving in mathematics. *Educational Leadership*, 49(8), 42-45.
- Tardif, J. (1993). L'évaluation dans le paradigme constructiviste. In R. Hivon, *L'évaluation des apprentissages. Réflexions, nouvelles tendances et formation* (pp. 27-56). Sherbrooke : Éditions du CRP.
- Tardif, J. (1994). L'évaluation du savoir-lire : une question de compétence plutôt que de performance. In J.-Y. Boyer, J.-P. Dionne et P. Raymond, *Évaluer le savoir-lire* (pp. 69-102). Montréal : Éditions Logiques.
- Taylor, C. (1994). Assessment for measurement or standards : The peril and promise of large-scale assessment reform. *American Educational Research Journal*, 31(2), 231-262.
- Thouin, M. (1993). L'évaluation des apprentissages en mathématiques : Une perspective constructiviste. *Mesure et évaluation en éducation*, 16(1-2), 47-64.
- Tobin, K., Butler-Kahle, J., & Fraser, B. (Eds.) (1990). *Windows into science classrooms*. New York : Palmer Press.
- Toumeur, Y., Noël, B., & Depover, C. (1989). Evaluation formative et structures d'accueil. *Mesure et évaluation en éducation*, 12(2-3), 119-129.
- Tousignant, R., & Morissette, D. (1990). *Les principes de la mesure et de l'évaluation des apprentissages* (2e édition). Boucherville : Gaëtan Morin.
- Trahan, M., & Dassa, C. (1978). La courbe normale en éducation : postulat, constante naturelle ou erreur scientifique. *Revue des sciences de l'éducation*, 4(1), 71-80.
- Traub, R. (1990). Assessment in the classroom : What is the role of research? *The Alberta Journal of Educational Research*, 36(1), 85-91.
- Traub, R., & Maraun, M. (1994). La mesure en éducation au Canada : Le passé, le présent et l'avenir. *Mesure et évaluation en éducation*, 17(2), 21-47.
- Van der Maren, J.M. (1995). *Méthodes de recherche pour l'éducation*. Montréal : Les Presses de l'Université de Montréal.
- Van Nieuwenhoven, C., & Jonnaert, P. (1992). Approcher les représentations des enseignants du primaire à propos de l'évaluation formative. In D. Laveault, *Les pratiques d'évaluation en éducation* (pp. 125-138). XV^e session d'étude de l'Association pour le développement de la mesure et de l'évaluation en éducation. ADMEE-Canada.
- Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46(7), 41-47.

- Wiggins, G. (1992). Creating tests worth taking. *Educational Leadership*, 49(8), 26-33.
- Wiggins, G. (1993). *Assessing student performance*. San Francisco : Jossey-Bass Publishers.
- Wiley, D., & Haertel, E. (1996). Extended assessment tasks : Purposes, definitions, scoring, and accuracy. In M. B. Kane & R. Mitchell (Eds.), *Implementing performance assessment : promises, problems, and challenges* (pp. 61-89). Mahwah NJ : Lawrence Erlbaum Asstes.
- Wilson, L. (1993). *What gets graded is what gets valued*. Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
- Wilson, R. (1990). Classroom processes in evaluating student achievement. *The Alberta Journal of Educational Research*, 36(1), 4-17.
- Wilson, R., & Rees, R. (1990). The ecology of assessment : Evaluation in educational settings. *Canadian Journal of Education*, 15(3), 215-228.
- Winfield, L. F. (1995). Performance-based assessments : Contributor or detractor to equity? In M. T. Nettles & A. L. Nettles (Eds.), *Equity and excellence in educational testing and assessment* (pp. 221-241). Boston : Kluwer.
- Wolf, K. (1993). From informal to informed assessment : Recognizing the role of the classroom teacher. *Journal of Reading*, 36(7), 518-523.
- Wolf, D., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well : Investigating new forms of student assessment. *Review of Research in Education*, 17, 31-71).
- Zahorik, J. (1994). Teacher evaluating behavior. In T. Husen & T. Postlewaite (Eds.), *Intnrtional Encyclopedia of Education* (2nd edition, Vol. 10, pp. 6012-6015). Toronto : Pergamon.

Annexe A

Assentiment du Comité de déontologie de l'Université d'Ottawa



Université d'Ottawa · University of Ottawa

Faculté d'éducation

Faculty of Education

le 13 décembre 1996

Madame Micheline Bercier-Larivière,
Université d'Ottawa
Faculté d'éducation

Chère Madame Bercier-Larivière,

Le Comité de déontologie a examiné votre projet de recherche intitulé "Élaboration et validation d'un modèle d'évaluation des apprentissages" et lui donne sa pleine approbation, pour la durée d'un an.

Avec nos meilleurs vœux de succès


Aline Giroux, présidente.

Annexe B

Guide de la première entrevue

Guide de la première entrevue

-
1. Pourquoi croyez-vous que les résultats d'évaluation de la situation que nous avons décrite et dont vous venez de lire le résumé situent bien les élèves dans leur développement, en d'autres mots, que ces résultats sont justes?

 2. En général, y a-t-il d'autres conditions que vous surveillez pour vous assurer d'arriver à des résultats justes pour chacun des élèves lorsque vous faites des évaluations?

 3. Y a-t-il des occasions où vous croyez que la note ne rend pas justice à l'élève? Si oui, pouvez-vous nous en parler?

Annexe C

Synopsis de la première entrevue

Situation d'évaluation # 1

Christine

Synopsis de l'entrevue # 1

Cette année, Christine dit être une enseignante choyée. Son groupe ne comporte que 18 élèves, tous de 5e année. Lorsque nous discutons de l'exactitude des résultats obtenus à cette épreuve, elle dit avoir passé tout le temps nécessaire à l'apprentissage de la multiplication et s'être assurée que tous recevaient l'attention dont ils avaient besoin. Les mini-tests avaient pour but d'encourager les élèves à apprendre leurs tables de multiplication par coeur. Grâce à leurs résultats quotidiens, ils pouvaient juger de leur progrès. En plus des leçons de groupe, les travaux quotidiens faits en classe pendant les 6 semaines qu'a duré l'apprentissage des multiplications, ont permis de recevoir toute l'aide individualisée nécessaire.

Les élèves ont été avisés de cette épreuve une semaine à l'avance. L'épreuve est tirée du manuel de classe (Découverte 5) et présente les problèmes de la même manière que les exercices quotidiens. Le test a eu lieu le matin; Christine a lu les questions à haute voix et répondu aux questions avant que chacun se mette à l'ouvrage. Vers la fin de la période, elle a circulé dans les allées et répondu aux questions des personnes qui n'avaient pas encore terminé leur travail.

Christine ne croit pas que des élèves ont pu mal réussir au test à cause d'une mauvaise maîtrise de l'algorithme de multiplication ou qu'ils ne savaient pas comment arrondir les nombres qu'ils devaient ensuite multiplier. Il y a déjà plusieurs années que les élèves apprennent à arrondir. De plus, ils en ont fait des exemples au tableau, juste avant l'épreuve. Aussi, d'après son expérience, les élèves qui ne maîtrisent pas l'algorithme de la multiplication sont ceux qui, la plus part du temps, n'apprennent pas leurs tables de multiplication par coeur. Il n'y avait donc pas lieu de vérifier ces habiletés séparément. À cette étape-ci, c'était l'ensemble de ces habiletés qui était évalué.

En général, pour s'assurer que les résultats situent bien les élèves dans leur apprentissage, Christine trouve important de varier les stratégies d'évaluation. Ce ne sont pas tous les élèves qui ont de la facilité à s'exprimer, elle aime donc utiliser les choix

multiples, les vrais ou faux, les tirets à remplir et les mots à apparier. Elle ne fait jamais d'évaluation le lundi; elle fait toujours une révision les jours avant un test et quelques fois même juste avant le test. Elle trouve que les élèves se sentent ainsi rassurés; une fois sécurisés, ils sont mieux disposés à faire un test. Il lui arrive aussi d'aider des élèves en difficulté d'apprentissage à écrire leurs réponses.

D'après son expérience, les résultats qui ne rendent pas justice aux élèves résultent de tâches avec des directives ambiguës, à une mauvaise stratégie d'évaluation ou à une évaluation précoce, comme il arrive parfois lorsqu'arrive un bulletin et que les élèves n'ont pas vraiment eu le temps d'apprendre la matière.

Situation d'évaluation # 2

Kevin

Synopsis de l'entrevue # 1

Kevin est confiant que les résultats obtenus situent bien ses élèves dans leur apprentissage de la théorie musicale. Ces notions ont été travaillées en classes, avec des exercices à faire en devoir. Ils ont été corrigés en classe et les difficultés ont été expliquées. Les questions de l'épreuve étaient identiques à celles des exercices. Une bonne indication que les résultats obtenus à cette épreuve sont justes est que ce sont ceux qui jouent bien de la flûte qui ont le mieux réussi l'épreuve.

Même si la cette matière n'est pas directement rapporté dans le bulletin, Kevin trouve important que ces résultats soient représentatifs de l'apprentissage des élèves; il croit que la maîtrise de cette matière est essentielle pour jouer de quelque instrument que ce soit. Ce serait peine perdue de travailler la technique si l'élève ne sait pas lire la feuille de musique.

Généralement, pour favoriser l'obtention de résultats justes, Kevin préfère administrer les épreuves le matin, évitant autant que possible le vendredi, surtout l'après-midi. Il prévient ses élèves à l'avance et il leur propose des exercices semblables au test, avec questions formulées de la même manière. Il fait parfois des exemples au tableau juste avant le test. Il trouve important de ne pas précipiter les choses; il faut travailler la matière autant qu'il le faut avant de faire une évaluation.

Il lui est arrivé de penser qu'une note ne rend pas justice aux capacités d'un élève, en particulier lorsque celui-ci a dû s'absenter quelques jours, manquant ainsi des leçons importantes. Il arrive aussi que des élèves aient de la difficulté à formuler une réponse, même s'ils connaissent le sujet. L'état émotif d'un élève lorsqu'il fait une évaluation est aussi important. Par exemple, les confrontations à la récréations et la mauvaise humeur semblent particulièrement indisposer les élèves. Dans ces cas, Kevin accorde parfois plus de valeurs à d'autres sources d'informations pour juger de la maîtrise des habiletés évaluées, afin que la note représente mieux le développement de l'élève.

Situation d'évaluation # 3

Marie

Synopsis de l'entrevue # 1

Lorsque j'ai menée cette évaluation de lecture, je croyais avoir réuni les conditions nécessaires pour bien situer les élèves dans leur développement. J'ai soigné la mise en situation pour éveiller la curiosité des élèves, j'ai choisi un texte qui me semblait intéressant et qui comportait à la fois du nouveau (production de film) et du connu (directeur d'école). L'exercice est tiré du manuel scolaire mais j'ai recomposé les questions parce que l'an passé, les élèves avaient eu de la difficulté avec le vocabulaire et la disposition des questions (annexe c). L'organisation du texte avec des sous-titres facilitait le repérage des parties du texte touchées par le questionnaire. Les élèves connaissaient l'importance de cette activité qui comptait pour le bulletin.

Après coup, lorsque j'ai cherché à comprendre pourquoi les résultats n'étaient pas aussi bons que je m'y attendais, je me suis rendue compte qu'ils ne sont pas habitués d'être aussi autonome dans leur travail, pour la lecture en particulier. D'habitude j'anime l'activité, toute la classe participe et les élèves répondent aux questions oralement. Étant donné l'autonomie exigée pour cette évaluation, le texte était trop long pour certains élèves, qui se sont découragés. De plus, à cause d'un contre-temps, l'activité a eu lieu l'après-midi alors que normalement on fait ce genre de travail le matin, lorsque les élèves sont reposés.

Par la suite, j'ai fait deux autres évaluations à partir d'exercices qui ne nécessitent pas l'écriture pour démontrer la compréhension du texte (annexes D et E). La proportion de réussites était semblable, sauf pour un élève qui a de graves difficultés d'apprentissage; à chacun de ces deux autres exercices, il a obtenu un résultat très supérieur à celui de l'exercice précédent.

Pour obtenir des résultats qui situent bien les élèves dans leur développement, j'exige peu d'autonomie de leur part dans les activités d'évaluation. Par exemple en science, ils réussissent mieux quand je lis le texte avec eux, que je lis les questions et qu'ils répondent en choisissant la bonne réponse ou en n'écrivant que quelques mots. Ainsi,

l'épreuve ne porte que sur les connaissances reliées à la matière évaluée et non pas la lecture et l'écriture. J'essaye aussi de garder la tranquillité dans la classe, tant que des élèves n'ont pas fini leur évaluation; j'écris des directives au tableau pour occuper ceux qui ont terminé avant les autres.

Il m'est déjà arrivé aussi d'être impressionnée par la performance d'un élève lorsqu'il lit à haute voix et de m'apercevoir par la suite qu'il ne comprend pas ce qu'il lit. J'ai constaté aussi qu'en mathématiques beaucoup des épreuves tirées de Découvertes 3 ne vont chercher que l'application de procédures, sans que l'élève ait à penser; par exemple, les additions sont regroupées ensembles et les soustractions aussi. Il y a très peu de résolution de problèmes, ce qui irait chercher la capacité de penser de l'élève. Enfin, quand je dois évaluer dans le domaine des arts (danse, art visuel, musique) il m'arrive de ne pas me sentir assez compétente pour juger si le résultat rend justice à l'élève.

Situation d'évaluation # 4

Matilde

Synopsis de l'entrevue # 1

Les résultats obtenus suite à l'évaluation des présentations orales situent bien les élèves dans leur développement pour les raisons suivantes. J'évaluais l'oral, j'ai donc demandé aux élèves de parler spontanément, naturellement à partir d'un plan, non pas d'écrire. Les critères d'évaluation étaient établis d'avance et étaient les mêmes pour tout le monde. Le thème de la musique pouvait être traité de plusieurs manières, il laissait donc de la place aux intérêts personnels de chacun. Les élèves étaient limités au même laps de temps pour faire leur présentation.

Pour chaque élève, j'ai rempli une fiche d'évaluation sur laquelle j'ai écrit des commentaires pour appuyer la note attribuée. L'élève a toujours la possibilité d'en discuter s'il n'est pas d'accord avec mon jugement. La note qui apparaît au bulletin est une moyenne de plusieurs évaluations. Toutefois, suite à cette expérience, je vais rajouter un critère à la fiche d'évaluation pour tenir compte de la gestuelle, de l'expression orale et du contact visuel, pour renforcer l'importance du naturel de la présentation.

Pour favoriser des résultats qui sont justes, je suis prête à faire concessions. Si un élève vit des moments difficiles (santé ou émotions) et qu'il m'en fait part, je lui donne la possibilité de remettre l'échéance à plus tard. Il m'est aussi arrivé d'enregistrer des présentations orales pour éviter les distractions qui m'empêchent de me concentrer sur l'évaluation. Suite à une évaluation orale, j'aime bien pouvoir revenir sur le jugement que j'ai posé, après en avoir évalué d'autres, c'est pourquoi je ne remets pas aux élèves leur fiche d'évaluation tout de suite après leur présentation. Quand il y a un texte écrit remis en plus de la présentation orale, je m'y réfère pour ce qui est des notes attribuées à la grammaire et à la structure des phrases. Je fais attention aussi pour ne pas donner des évaluations les jours d'activités spéciales ou quand les élèves sont fatigués, comme le vendredi en fin de journée.

Il m'arrive d'avoir l'impression qu'une note ne rend pas justice à l'élève. Ce peut être suite à des difficultés personnelles qui perturbent un élève; ça arrive qu'il me le dise après coup

plutôt qu'avant. Il arrive aussi que des élèves manquent la classe et ne reçoivent pas l'enseignement; le résultat n'est pas ce qu'il aurait normalement été sans ces absences. Il arrive que des élèves ne remettent pas un travail, ce qui diminue leur moyenne pour le terme. J'ai déjà remarqué aussi que si je fais des évaluations formatives en début d'apprentissage et que je compte ces notes-là, j'arrive parfois à des notes qui ne font pas justice aux élèves. Il m'est arrivé aussi d'utiliser une grille d'évaluation qui ne me permettait pas de noter correctement les présentations (annexe B). Certains élèves très gênés ou mal à l'aise devant la classe perdent tous leurs moyens. Enfin, moi aussi il m'arrive d'être plus sévère à certains moments, selon mon humeur ou les événements de la journée. C'est là que le temps de réflexion est utile pour réajuster mon jugement, avant de remettre la fiche d'évaluation à l'élève.

Situation d'évaluation # 5

Marie-Chantal

Synopsis de l'entrevue # 1

Je crois que les résultats d'évaluation des exposés oraux sont justes d'abord parce que l'activité a lieu à un moment propice. Le lundi matin, ils veulent souvent raconter leur fin de semaine à leur compagnons de classe. Aussi, il n'y a que les élèves qui veulent bien présenter qui présentent; les élèves qui n'ont rien à dire ne sont pas forcés de performer. Puisque l'activité revient chaque semaine, ils ont éventuellement l'occasion, au cours d'un trimestre, de présenter quand ça leur convient.

Les résultats d'évaluation des exposés oraux sont exprimés sous forme de commentaires constructifs (positifs ou négatifs), positifs ou négatifs, et se rapportent directement à l'exposé. L'élève qui présente est le premier à relever les erreurs s'il est conscient qu'il en a fait. Que ce soit les élèves ou moi-même qui remarquions un anglicisme dans le discours, l'élève qui vient de présenter se rappelle du fait et il y a plusieurs témoins de l'incident. Si un commentaire n'était pas justifié, il serait vite corrigé. De plus les critères d'évaluation sont établis depuis le début de l'année scolaire, les élèves ont donc eu le temps de les intégrer. Aussi, leur participation à l'évaluation des autres élèves leur donne la chance d'apprendre en suggérant des améliorations à celui qui présente. quand j'ai à mettre une note sur le bulletin, elle vient généralement de plusieurs présentations.

Pour m'assurer d'arriver à des résultats qui situent bien les élèves dans leur développement, j'aime bien, quand c'est possible, évaluer plus d'une fois, avec des stratégies différentes. Ceci donne aux élèves plus de chances de démontrer ce qu'ils savent, en particulier les élèves qui ont de la difficulté à s'exprimer, surtout quand il faut composer une réponse. Aussi, l'élève a jusqu'au bulletin pour démontrer qu'il maîtrise ou pas quelque chose, donc je trouve important de lui donner la chance de se reprendre ou d'améliorer son résultat, il n'a qu'à venir m'en parler. Il y a aussi des moments de la journée qui conviennent mieux à l'évaluation, comme en début d'avant-midi ou en début d'après-midi; j'ai remarqué aussi que la faim ou la chaleur excessive nuit à la concentration.

Il arrive que nos questions ne soient pas formulées clairement ou ne soient pas dans un langage qu'ils comprennent. L'absentéisme de certains élèves nuit parfois à l'apprentissage et par la suite à l'évaluation car il est difficile de reprendre 2 ou 3 jours de classe. De plus, le moment de préparer des bulletins ne coïncide pas toujours rarement avec le temps où on est prêt à évaluer toutes les matières.

Quand il m'arrive de constater que des élèves n'ont pas des résultats à la hauteur de mes attentes, je m'interroge d'abord sur mon enseignement. Est-ce toute la classe ou un nombre restreint d'élèves? Ceux-ci ont-ils vécus des dérangements récemment? Quelle partie de l'épreuve fait défaut? Ont-ils manqué de temps? Je m'interroge moins quand les résultats sont supérieurs à mes attentes. Aussi, il y a des jours où on est plus positif et d'autres où on est moins sûr de soi. Quand quelque chose nous tracasse, on a plus de difficulté à se concentrer. Personnellement, quand je suis fatiguée, je suis plus sévère. Il vaut mieux alors remettre la correction au lendemain.

Annexe D

Clé de codification des protocoles verbaux, première entrevue

Tableau 5

Clé de codification des protocoles verbaux, première entrevue.

Catégories	Sous-catégories	Codes	Exemples
Habileté	Bien dissocier les habiletés visées	Hab	.. si par exemple en français ou en mathématiques je demande une explication pour résoudre un problème ou sur un sujet quelconque, des fois l'élève peut avoir de la difficulté à s'exprimer. (Kevin, E ₁ , ligne 64)
	Intéressante ou motivante	Ta/Mot	.. dans l'activité qui est décrite, c'est une activité qui porte sur la musique puis en 7e année c'est un thème qui intéresse ces élèves-là. (Matilde, E ₁ , ligne 8)
Tâche	Construction technique	Ta/Tech	.. mes évaluations habituellement ont une partie où ils doivent donner une définition ou expliquer du vocabulaire ... ensuite j'ai toujours une partie vrai ou faux, une partie choix multiples, une partie remplir les tirets, une partie relier la colonne de droite à celle de gauche ... (Christine, E ₁ , ligne 45)
	Longueur de la tâche	Ta/Long	Le texte, trop long peut-être pour certains élèves (Marie, E ₁ , ligne 49)
	Niveau de difficulté de la tâche	Ta/Dif	Il y avait une question qui était plus générale, où la réponse ils pouvaient la trouver un peu partout dans le texte et ils ont trouvé ça difficile. (Marie, E ₁ , ligne 44)
	Adaptée aux élèves	Ta/Ad	Donc, je fais même de l'évaluation orale pour certains élèves que je sais qu'ils connaissent la réponses mais qui ont pas été capables de me la donner (Christine, E ₁ , ligne 88)
	Nouvelle mais familière	Ta/NF	.. les questions de l'épreuve étaient les mêmes questions que j'avais présentées sur les deux travaux qui précédaient (Kevin, E ₁ , ligne 7)

Tableau 5 (suite)

Clé de codification des protocoles verbaux, première entrevue.

Catégories	Sous-catégories	Codes	Exemples
Lien avec apprentissage	Mise en situation comme pour apprentissage	A/Sit	Je pense que la mise en situation était assez soignée. ... Il y a avait des mots au tableau, on a discuté à propos des métiers qu'on pouvait faire au cinéma. Aussi à propos du travail d'un directeur d'école, ... (Marie, E ₁ , ligne 1)
	Intégration des critères d'évaluation	A/Crit	... l'élève qui doit évaluer l'autre nécessairement apprend aussi parce que ça lui permet de mieux saisir les critères d'évaluation puisque lui doit évaluer. (Marie-Chantal, E ₁ , ligne 11)
	Forces et faiblesses	A/FF	C'est des commentaires. C'est, exemple, tu aurais dû parler plus fort ou je te félicite parce que tu as parlé fort. (Marie-Chantal, E ₁ , ligne 31)
	Aide disponible pour précisions	A/Préc	Ensuite, s'il y a des questions, je réponds aux questions, à mesure (Kevin, E ₁ , ligne 60)
	Apprentissage préalable	A/Préal	... j'ai pu prendre du temps pour m'assurer que tout le monde comprenait les notions enseignées (Christine, E ₁ , ligne 7)
	Reprise possible	A/Rep	.. si l'élève ne réussit pas bien puis veut s'améliorer, admettons le test est au mois d'avril puis le bulletin est au mois de juin, bien il a jusqu'au mois de juin pour revenir me voir et me dire : est-ce que tu pourrais me rétester sur telle partie du test? (Marie-Chantal, E ₁ , ligne 53)

Tableau 5 (suite)

Clé de codification des protocoles verbaux, première entrevue.

Catégories	Sous-catégories	Codes	Exemples
Transparence	Contenu précisé	Tr/Cont	Alors la révision se fait toujours : voici, ça c'est important, ça c'est moins important, ça lisez-le rapidement, ça ... (Christine, E ₁ , ligne 82)
	Critères définis	Tr/Crit	C'est déterminé dès le début de l'année. On définit les critères d'une bonne présentation orale. Et puis on leur dit comme : éviter les anglicismes, parler fort, regarder ceux à qui on parle, tout ça. (Marie-Chantal, E ₁ , ligne 2)
	Participation de l'élève à l'évaluation	Tr/Part	.. tout le groupe évalue la personne, bien plutôt le discours de la personne qui vient de présenter. (Marie-Chantal, E ₁ , ligne 8)
	Forces et faiblesses	Tr/FF	.. l'enfant savait où il se situait par rapport à ses multiplications et moi je le savais en même temps. (Christine, E ₁ , ligne 12)
	Prévenir les élèves	Tr/Prev	.. aussi j'essaie d'avoir aucune surprise, je préviens mes élèves (E ₁ , Kevin, ligne 39)
	Retour sur l'évaluation	Tr/Ret	... on a revu les erreurs en classe, à la suite (Kevin, E ₁ , ligne 98)

Tableau 5 (suite)

Clé de codification des protocoles verbaux, première entrevue.

Catégories	Sous-catégories	Codes	Exemples
Circonstances	Déroulement habituel	C/Dér	Avant qu'on leur donne le test, je leur donne toujours une période de révision, un 10-15 minutes où ils peuvent me poser des questions ... et puis ensuite je fais la lecture des questions. ... À la fin, avant de remettre le test, je relis les questions avec eux, répondre aux questions qu'ils ont, puis ensuite ils commencent le test. Donc il y a quand même un déroulement qui se fait (Christine, E., ligne 65)
	Moment de la journée/semaine	C/Mom	Si c'est vendredi, dernière période, c'est certain que les élèves sont plus fatigués, ils ont hâte de s'en aller à la maison, c'est pas un bon moment pour donner une évaluation (Matilde, E., ligne 66)
	Absentéisme	C/Abs	Aussi si l'élève a manqué beaucoup de jours en classe et des certaines directives. Reprendre ça dans son temps libre, c'est pas toujours évident. (Kevin, E., ligne 70)
	Temps alloué	C/Temps	... par exemple en mathématiques tu te dis est-ce que c'est parce que j'ai pas donné assez de temps? (Marie-Chantal, E., ligne 102)
	Environnement	C/Env	J'essaye de voir à ce que les conditions soient favorables dans la classe au niveau du silence et éviter le va-et-vient, surtout à la fin de l'évaluation (Marie, E., ligne 83)
	État d'esprit de l'élève	C/Esp	Aussi, si à la récré ou pendant l'heure du dîner il a une confrontation avec des amis ou qu'il se sent pas bien, il est malade ou il est fâché ... (Kevin, E., ligne 74)
Autre			Bien je pense qu'il faut avoir quand-même une certaine connaissance. Si tu juges une technique de dessin, bien toi, il me semble que tu devrais maîtriser jusqu'à un certain point ces techniques-là, pour pouvoir en juger. Donc être apte à la tâche. C'est pareil pour la danse. Je ne m'y connais pas du tout, je peux apprécier, mais après ça ... à savoir juger ... vraiment ...

Annexe E

Bilan de codification des protocoles verbaux, première entrevue

Tableau 6

Bilan de codification des protocoles verbaux, première entrevue.

209

Catégories	Sous-catégories/Code	Christ	Kevin	Marie	Matilde	M-Chant
Habileté	Dissocier habiletés visées (H/Vis)		✓	✓✓✓✓ ✓✓✓✓	✓✓	✓
Tâche	Motivante (Ta/Mot)			✓	✓	✓
	Construction (Ta/Tech)	✓		✓	✓	✓
	Longueur (Ta/Long)		✓	✓		
	Difficulté (Ta/Dif)	✓		✓		
	Adaptée (Ta/Ad)	✓			✓	
	Nouvelle mais familière (Ta/NF)	✓	✓✓	✓		✓
Lien avec apprent.	Mise en situation (A/Sit)			✓		
	Intégration des critères (A/Crit)					✓
	Forces et faibl. (A/FF)					✓
	Aide disponible (A/Préc)	✓	✓	✓		
	Apprentissage préalable (A/Préal)	✓✓✓	✓✓✓		✓	✓✓
	Reprise (A/Rep)					✓✓
Transp	Contenu précisé (Tr/Cont)					✓
	Critères définis (Tr/Crit)				✓	✓
	Participation (Tr/Part)					✓
	Forces et faibl. (Tr/FF)	✓			✓	✓
	Prévenir (Tr/Prev)		✓	✓		
	Retour (Tr/Ret)		✓			✓
Circonst.	Déroulement (C/Dér)	✓	✓	✓✓		✓
	Moment (C/Mom)	✓✓	✓	✓	✓✓	✓✓
	Temps alloué (C/Temps)					✓
	Absentéisme (C/Abs)		✓		✓	✓✓
	Environnement (C/Env)			✓		✓
	État d'esprit (C/Esp)		✓✓		✓✓✓	✓✓✓
Autres		4	9	3	10	5

Annexe F

Guide de formation des participants

UNE CONCEPTION DE L'ÉVALUATION DES
APPRENTISSAGES
EN SALLE DE CLASSE

POUR UN RÉSULTAT
D'ÉVALUATION
JUSTE

document de travail

© Micheline Bercier-Larivière, Ottawa, 1997

Attention : Reproduction strictement interdite

TABLE DES MATIÈRES

	page
Introduction	
L'apprentissage et son évaluation	
Fonctions de l'évaluation	
 Chapitre 1 - L'évaluation des apprentissages en salle de classe	
1. Pertinence de la tâche par rapport aux habiletés visées	
a) Les critères de performance	
b) Le choix d'une tâche appropriée	
c) La bonne combinaison habileté-tâche	
2. Cohérence avec l'activité pédagogique	
3. Transparence de l'activité évaluative	
4. Absence de désavantage circonstanciel	
a) l'équité	
b) la motivation	
c) la collaboration	
 Chapitre 2 - Élaboration d'un schème de notation	
1. Spécificité des critères choisis	
2. Niveaux ou échelles de notation	
3. Échelles numériques et échelles descriptives	
4. Schèmes de notation holistiques et analytiques	

page

- a) Schème holistique
- b) Schème analytique

Chapitre 3 - Deux situations d'évaluation situant le concept de justesse

A - Évaluation en vue de situer des élèves de 1^{re} année dans leur apprentissage de l'addition

B - Activité d'apprentissage de résolution de problème et autoévaluation

Feuillet de l'élève

Analyse de l'activité selon le concept de justesse

Chapitre 4 - Thèmes d'intérêt

A - Évaluer les travaux faits en groupe

B - Évaluer les attitudes face à une matière

Introduction

Les enseignants ont besoin de balises pour guider leurs pratiques d'évaluation des apprentissages en salle de classe, afin de produire les résultats les plus justes possible. Ces balises doivent être cohérentes avec la nature des apprentissages et s'adapter aux types formatif et sommatif d'évaluation.

L'apprentissage et son évaluation

L'apprentissage est un changement intérieur qui n'est pas observable directement; il faut provoquer l'apparition de comportements qui révèlent l'existence de cet apprentissage, afin de pouvoir en apprécier la qualité ou l'étendue. Parce que l'élève apprend constamment, son savoir évolue sans cesse, rendant les résultats d'évaluation vite périmés.

Dans sa conception moderne, l'apprentissage est complexe et en constante évolution. Il ne peut être qu'estimé, toujours de manière imprécise et toujours selon le contexte dans lequel il est observé. Son évaluation mène à des résultats pleinement significatifs uniquement lorsqu'il est regardé et analysé dans sa globalité.

Les apprentissages sont multidimensionnels et complexes, non pas linéaires. Il faut développer les habiletés de pensée supérieures, des savoirs transdisciplinaires et des stratégies métacognitives. Ces

savoirs doivent être intégrés, appliqués en situations appropriées et extérieurement manifestables.

Fonctions de l'évaluation

L'évaluation formative sert à apporter des informations utiles pour nourrir le processus d'apprentissage, comme, entre autres, d'identifier la manière personnelle de chacun d'apprendre. La théorie du constructivisme accorde un rôle actif à l'apprenant, valorise le rattachement des nouvelles notions aux connaissances déjà acquises.

Pour ce faire, la fonction formative de l'évaluation est beaucoup plus importante que la fonction sommative.

L'évaluation sommative survient plutôt en fin d'étape. Elle a pour but de confirmer ou de certifier l'acquisition des habiletés développées durant une période d'apprentissage donnée.

La performance est à son meilleur quand les buts sont connus, les exigences précisées et que les attentes sont explicites. La motivation, l'effort et l'estime de soi jouent un rôle déterminant dans l'apprentissage. Celui-ci comporte un volet social et le travail en groupe est bénéfique.

Chapitre 1

L'évaluation des apprentissages en salle de classe

L'évaluation des apprentissages basée sur le concept de justesse des résultats convient bien au sens moderne de l'apprentissage. Les résultats d'évaluations menées par l'enseignant en salle de classe, au moyen d'outils variés, pour réguler les apprentissages ou reconnaître les acquis, sont considérés justes s'ils situent correctement un apprenant dans son apprentissage, identifient ses forces et ses faiblesses, ce qu'il maîtrise ou ne maîtrise pas. Cette conception de l'évaluation devrait donner des résultats sensés, représentatifs, exacts, précis, signifiants et vrais, entachés du moins d'erreur possible. Elle devrait donc maximiser la confiance qu'il est possible d'avoir dans les résultats d'évaluation pour ce qui est de situer un élève dans son développement.

Il existe plusieurs points critiques pour produire des résultats d'évaluation justes; ces points sont souvent interdépendants, rendant difficile d'intervenir sur un seul d'entre eux à la fois. Dans ce document, ils sont regroupés en quatre ensembles :

- Pertinence de la tâche d'évaluation par rapport aux habiletés visées,
- Cohérence avec l'activité pédagogique,

- Transparence de l'activité évaluative,
- Absence de désavantage circonstanciel.

Tous ces points sont retenus pour leur contribution à l'assurance de résultats d'évaluation plus justes. Chacun d'eux sont discutés en détail dans les pages qui suivent.

1. Pertinence de la tâche par rapport aux habiletés visées

Toute activité pédagogique exige d'établir les **habiletés** visées par les apprentissages. Ce sont ces habiletés qui déterminent les activités pédagogiques à mener en salle de classe. Les habiletés peuvent être regroupées de plusieurs façons. Voici celle proposée par Stiggins :

- a) posséder des connaissances théoriques
- b) appliquer des connaissances théoriques
- c) raisonner
- d) réaliser une tâche en utilisant divers savoirs (*skills*)
- e) avoir des sentiments ou des dispositions

(Stiggins 1994, p. 67, traduction libre)

Selon Stiggins, ces catégories sont hiérarchiques : savoir raisonner

nécessite d'appliquer des connaissances factuelles et ces deux habiletés sont nécessaires pour démontrer un certain savoir-faire (*skills*). Produire une oeuvre nécessite des connaissances, du raisonnement et des habiletés alors que l'affectivité se développe à travers la réussite dans les autres types d'habiletés. Les habiletés complexes sont des habiletés qui intègrent des habiletés de divers niveaux de pensée, allant jusqu'aux habiletés supérieures de la pensée (analyse, synthèse et évaluation).

Dans la philosophie des programmes d'études récents, la mémorisation de faits isolés a été remplacée par la connaissance de sources d'informations variées et de la manière d'y accéder. Bien qu'il soit presque impossible d'enseigner et d'apprendre sans un minimum de mémorisation de la part de l'élève, ces acquisitions devraient être justifiées par leur importance et leur utilité pour effectuer des tâches plus complexes.

Les nombreuses taxonomies des objectifs pédagogiques (comme celles du domaine cognitif, Bloom et al. et du domaine affectif, Krathwohl et al.) peuvent faciliter l'identification des habiletés visées. Voici rassemblés ici des verbes associés à des habiletés de divers niveaux de complexité dans chacune de ces taxonomies.

Taxonomie des objectifs du domaine cognitif	
1. Acquisition de connaissances	apprendre, connaître, acquérir, énumérer, cocher, dire, souligner, identifier, reconnaître, définir, repérer.
2. Compréhension	comprendre, donner des exemples, décrire, dire en ses propres mots, illustrer, distinguer, estimer, ordonner, discuter, transformer, interpréter, réarranger, différencier.
3. Application	présenter, exprimer, utiliser, transférer, réagir, relier, suivre, calculer, fabriquer, expérimenter, explorer, créer, démontrer, employer, classer, organiser.
4. Analyse	expliquer, établir des liens, modéliser, distinguer, détecter, déduire.
5. Synthèse	prédire, suggérer, résumer, combiner, organiser, synthétiser, classer, déduire, formuler, modifier.
6. Évaluation	justifier, juger, argumenter, contraster, évaluer.

Taxonomie des objectifs du domaine affectif	
1. Réception	écouter, accepter, choisir
2. Réponse	confier, offrir spontanément, pratiquer, communiquer, se conformer, encourager.
3. Valorisation	aider, démontrer, nier, protester, débattre, argumenter.
4. Organisation	apprécier, formuler, comparer, définir, harmoniser.
5. Caractérisation par une valeur ou un système de valeurs	réviser, contribuer, changer, compléter, éviter, diriger, résoudre, résister.

Une fois l'habileté principale définie, il faut déterminer si celle-ci en comprend d'autres, ce qui est normalement le cas d'une habileté complexe. La résolution de problèmes, par exemple, comporte les habiletés de raisonner, d'appliquer des procédures mathématiques et de communiquer, toutes trois en interaction. Faire un exposé oral, quand à lui, est composé d'un contenu à transmettre, de l'expression verbale et de la gestuelle qui l'accompagne.

Il est important de reconnaître les habiletés distinctes impliquées dans une habileté complexe, pour ne pas les confondre avec d'autres habiletés semblables et de tirer de l'évaluation des informations utiles pour améliorer cet apprentissage. Les composantes ou les dimensions importantes d'une habileté complexe demeurent liées entre elles; prises isolément, elles n'ont aucune valeur pour juger de l'atteinte de l'habileté complexe.

a) Les critères de performance

Un critère est une caractéristique d'après laquelle on porte un jugement d'appréciation. Des critères bien choisis sont propres à l'activité en milieu réel et sont applicables à une variété de tâches dans le domaine. Ils peuvent s'ajuster au niveau de développement des évalués et leur cheminement scolaire. Leur utilisation facilite les jugements et les décisions pour lesquelles il a été estimé nécessaire d'entreprendre une démarche d'évaluation. D'une tâche à l'autre, les critères demeurent toujours adaptés à l'habileté identifiée au départ.

L'établissement de ces critères, quoique ardue, est aussi essentielle à la mise au point d'activités d'apprentissage qu'à l'évaluation. Même si une habileté complexe comporte plusieurs habiletés plus simples, il n'est toutefois pas obligatoire que son évaluation tienne compte, à chaque fois, de toutes ces composantes. Cet aspect de l'évaluation

étant plutôt complexe, elle sera abordée de manière détaillée au chapitre 2.

b) Le choix d'une tâche appropriée

La tâche est le stimulus qui devrait faire apparaître des comportements révélant l'état des habiletés ciblées par l'évaluation. Dans le cas d'habiletés complexes, la tâche doit mettre à profit l'interaction des habiletés plus simples. Malgré des recherches poussées, l'équivalence entre deux situations d'apprentissage est très difficile à définir.

En fait, une tâche d'évaluation appropriée est aussi une bonne activité d'apprentissage de l'habileté en question. Elle doit être choisie pour sa pertinence à mettre équitablement en valeur les habiletés ciblées, avec une marge minimale d'erreur. Il va de soi que cette tâche doit aussi être définie clairement, dans un langage compréhensible pour l'évalué et ne pas comporter de fautes techniques.

Plusieurs auteurs se sont penchés sur les caractéristiques à considérer lors du choix d'une tâche ou d'une situation d'évaluation. En plus d'être cohérente avec les grands buts de l'éducation, d'être complexe, sensée et engageante, elle devrait aussi être signifiante et utile. Sa difficulté devrait convenir aux évalués et les résultats obtenus

devraient être généralisables aux contenus examinés. De plus, il est important de tenir compte de son coût (temps, énergie, argent), de ses conséquences et de son caractère équitable.

Le recours à une situation authentique réduit le degré d'inférence au sujet des capacités évaluées parce qu'elles sont directement observables; les situations authentiques peuvent donc favoriser des résultats plus justes, à condition d'être choisies correctement. D'une part, l'observation directe d'un élève au travail peut dans certains cas fournir des renseignements inégalables mais, d'autre part, plus une tâche d'évaluation consomme de temps, plus le nombre en sera limité.

Le niveau d'authenticité d'une tâche peut varier, allant de la simulation en salle de classe à l'application réelle de la tâche, mais elle demeurera toujours plus près des habiletés visées qu'une épreuve papier-crayon. Elle est donc plus susceptible de générer des résultats justes.

c) La bonne combinaison «habileté-tâche»

L'agencement optimal des habiletés visées et le type de tâche choisi pour les faire ressortir s'avère crucial; c'est le fondement de l'obtention de résultats d'évaluation justes. Des lacunes à ce niveau-ci de la démarche d'évaluation mettraient en péril la justesse des résultats, peu

importe la rigueur des étapes subséquentes. Voici réunies dans un tableau les combinaisons «habileté - tâche» possibles et celles qui sont les plus favorables à une évaluation juste. Les cases grisées sont celles qui sont les plus susceptibles de mener à un jugement éclairé parce qu'elles nécessitent le moins grand degré d'inférence. Toutefois, chaque situation doit être examinée de près, parce que toutes comportent des risques. Les cases laissées vides représentent des combinaisons moins sûres. Par exemple, les questions à réponses choisies ne sont pas nécessairement les plus appropriées pour évaluer des attitudes, en particulier si le répondant a conscience de l'effet qu'il veut donner. Par ailleurs, l'entrevue pour évaluer le rappel de connaissances factuelles est un gaspille d'énergie, à moins que l'élève ait des difficultés à écrire ou ne puisse pas lire.

Une fois la tâche choisie, il est ensuite possible d'établir le schème d'interprétation des résultats avec les indices qui permettent de situer le niveau de la performance quant à chacun des critères établis précédemment. Contrairement aux critères, les indicateurs sont habituellement spécifiques à la tâche choisie. Ce sujet sera abordé séparément, dans un autre chapitre, étant donné sa particularité.

Tableau 1

Combinaisons des habiletés évaluées et des stratégies adoptées (inspiré de Stiggins, 1994)

	Choisir une réponse	Composer une réponse	Démontrer par une action (<i>performance</i>)	Échanger verbalement
Connaissances théoriques	excellent jumelage		pas approprié	pas judicieux
Application de connaissances théoriques		excellent jumelage	peu approprié	pas judicieux
Raisonnement	pas approprié	excellent jumelage	bon jumelage	bon jumelage
Réalisation utilisant divers savoirs (<i>skills</i>)	pas approprié	pas approprié	excellent jumelage	peu approprié
Sentiments ou dispositions			bon jumelage	excellent jumelage

2. Cohérence avec l'activité pédagogique

Pour être intégrée à l'apprentissage, l'activité d'évaluation proposée à l'apprenant doit s'insérer convenablement dans l'activité pédagogique et arriver au moment opportun. Bien que toute situation d'évaluation doive présenter un aspect nouveau, il est essentiel que les apprenants soient familiers avec le genre de situation d'évaluation qui leur est proposée. Non seulement doivent-ils connaître les critères qui guident l'évaluation de leur performance, mais avoir déjà eu l'occasion de les appliquer. Cet aspect sera plus longuement discuté lorsqu'il sera question de la transparence de l'activité évaluative.

L'évaluation étant menée dans le but de servir l'apprentissage ou de reconnaître les acquis, les résultats seront exprimés en termes appropriés, pour être éventuellement réinvestis dans l'activité pédagogique, selon le cas. Un schème de notation analytique permet d'apprécier la démarche d'apprentissage de l'élève, son processus de pensée, la reconnaissance de la zone de construction du développement; les résultats seront alors exprimés en termes de forces et de faiblesses, selon les critères choisis, prêts à être utilisés dans une démarche d'apprentissage subséquente. Par ailleurs, une évaluation en fin d'étape, donc de type sommatif, aura plutôt recours à un schème holistique, avec un jugement global quant à l'atteinte de l'habileté développée.

Seules les activités d'apprentissage vécues précédemment par les élèves permettent de déterminer avec certitude à quelles habiletés la tâche d'évaluation fait réellement référence. Par exemple, il est impossible pour une personne qui n'a pas vue ce à quoi les élèves ont été exposés lors de l'apprentissage, de déterminer si une tâche d'évaluation requiert le rappel de connaissances ou le raisonnement. Une épreuve pouvant être réussie avec des stratégies d'apprentissage élémentaires, comme des automatismes, alors qu'elle vise des habiletés plus complexes est mal préparée. C'est «le danger de s'appuyer uniquement sur des objectifs définis au moyen de comportements observables» (Thouin, 1993, p. 50). Si l'enseignant est davantage préoccupé d'atteindre le résultat final que de développer les apprentissages qui mèneront éventuellement aux comportements souhaités, il court-circuite la démarche pédagogique.

3. Transparence de l'activité évaluative

La transparence d'une démarche d'évaluation se concrétise par la participation, la discussion, et l'appropriation par l'élève des divers aspects de l'évaluation de ses apprentissages. Selon sa maturité et son apprentissage de la responsabilité à cet égard, l'élève pourra aider à identifier les critères d'évaluation, les indicateurs de performance et expérimenter leur application à ses activités d'apprentissage; l'autoévaluation et l'évaluation par les pairs à des fins formatives sont

particulièrement propices à cette prise en charge. Présenter aux élèves et commenter des exemples concrets de travaux de diverses qualités peuvent aussi aider à communiquer ces attentes et favoriser le transfert de ces critères à d'autres activités d'apprentissage.

Tout en favorisant une communication enseignant-élève plus ouverte, ces mesures de transparence peuvent réduire la méfiance et l'anxiété face à une situation d'évaluation, éliminant ou réduisant ainsi des sources importantes d'erreurs. La transparence contribue à générer des résultats plus représentatifs de ce que l'élève peut ou ne peut pas faire parce que celui-ci sait précisément ce qui est attendu de lui. Les indicateurs appuient aussi le jugement porté sur les réalisations des élèves, favorisant leur traitement objectif, constant et équitable de la part de l'enseignant.

Une activité évaluative transparente donne aussi l'occasion à l'élève d'exprimer son opinion sur le jugement posé par l'enseignant. L'utilisation d'un schème de notation est tout aussi utile à l'élève qu'à l'enseignant; il réduit la subjectivité et dirige l'attention sur les éléments importants de l'apprentissage. Tout en contribuant à responsabiliser l'élève, la transparence se refléterait aussi sur la motivation de l'élève.

4. Absence de désavantage circonstanciel

Malgré les nombreuses mises en garde décrites jusqu'à maintenant, il existe encore d'autres menaces à la justesse des résultats d'évaluation. Même s'il n'est pas possible de tout contrôler, il faut au moins être conscient de ce qui peut fausser les résultats obtenus.

Par exemple, à cause de l'absence de limites de temps, les devoirs faits à la maison peuvent témoigner plus validement des capacités d'un élève. Par contre, puisqu'il n'est pas possible d'affirmer hors de tout doute que l'élève a fait le travail lui-même, cette stratégie peut réduire la justesse des résultats.

L'activité évaluative doit aussi rassembler des conditions favorables à l'élève; le bruit excessif, les limites de temps, la fatigue, des événements excitants peuvent nuire à une réalisation optimale et contribuer à l'instabilité de la performance observée.

a) L'équité

Il faut aussi être conscient que la stratégie choisie peut être plus favorable à certains élèves qu'à d'autres, dépendant de leurs caractéristiques personnelles. Par exemple, l'entrevue peut créer de l'anxiété chez des élèves timides et ainsi les défavoriser, comparativement à des élèves sociables et sûrs d'eux; de la même

manière, un élève qui écrit avec difficulté sera plus à l'aise face à un questionnaire à réponses choisies plutôt qu'à réponses élaborées. Il va de soi aussi que cette tâche ne comporte pas de biais culturel ou sexiste.

b) La motivation

Des résultats d'évaluation représentant vraiment les capacités d'un élève sont tirés d'une tâche signifiante pour lui, compatible avec ses intérêts et justifiant d'investir le temps et l'énergie qu'elle réclame.

c) La collaboration

Dans le cas où des élèves collaborent souvent en classe pour des activités d'apprentissage, il serait juste de recourir, à l'occasion, à des situations semblables pour l'évaluation. Afin d'être juste envers chaque membre de l'équipe, la contribution de chacun doit être bien documentées.

La qualité optimale d'une situation d'évaluation étant difficile à atteindre, un résultat insatisfaisant devrait mener à conclure en l'incapacité de démontrer la maîtrise plutôt qu'à la non-maîtrise des habiletés évaluées, parce que l'apprentissage n'est pas nécessairement en cause.

Les quatre composantes de la justesse des résultats d'évaluation sont donc :

- la pertinence de la tâche par rapport aux habiletés visées,
- la cohérence avec l'activité pédagogique,
- la transparence de l'activité évaluative et
- l'absence de désavantage circonstanciel.

L'établissement des bonnes habiletés à évaluer est la base d'une démarche saine; le choix d'une stratégie susceptible de bien les mettre en valeur lui est toutefois indispensable. Les trois autres composantes contribuent à mettre en lumière les apprentissages réalisés. Si l'une de ces composantes est de moins bonne qualité, l'image finale de l'apprentissage évalué en sera moins juste. La figure 1, page suivante, tente d'opérationnaliser ces quatre composantes.

POUR UN RÉSULTAT D'ÉVALUATION JUSTE	
1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées - Tâche et critère(s) d'évaluation congruents avec les diverses habiletés visées - Tâche signifiante, adaptée, claire, techniquement correcte, nouvelle mais familière	
2. Cohérence avec l'activité pédagogique - Activités d'apprentissage préalables avec mise en valeur des mêmes critères d'évaluation - Mise en situation appropriée et aide disponible - Schème approprié au type d'évaluation et permettant le réinvestissement des résultats - Possibilité de reprise	3. Transparence de l'activité évaluative - Participation de l'élève à toutes les étapes de la démarche d'évaluation - Schème d'évaluation familier aux élèves - Résultats communiqués en termes de forces et de faiblesses - Retour sur l'activité d'évaluation en classe
4. Absence de désavantage circonstanciel - conditions propices à la concentration - collaboration contrôlée - procédure équitable et habituelle - temps amplement suffisant - présence pendant l'apprentissage préalable	

Figure 1. Schématisation des composantes du concept de justesse des résultats.

Chapitre 2

Élaboration d'un schème de notation

Dans le chapitre 1, l'un des aspects importants de l'élaboration d'un schème de notation a été abordé, soit l'établissement des critères d'évaluation d'une tâche. Ici il sera question de leur degré de spécificité. Autres aspects importants d'un schème de notation sont les niveaux de performance, les indicateurs de niveaux et les échelles de notation.

Un schème de notation contribue à uniformiser la notation de la performance des élèves. De plus il force à reconnaître les aspects jugés importants dans l'apprentissage. Il peut communiquer plus explicitement les attentes aux élèves et leur fournir une rétroaction précise sur leurs habiletés évaluées.

L'élaboration de schème de notation est complexe et les premières tentatives en ce sens ont avantage à se faire en collaboration avec une personne d'expérience. Les premiers exercices peuvent être longs et pénibles, parce qu'ils forcent à préciser par écrit ses attentes de manière non équivoque, bien avant d'arriver à la correction des travaux des élèves. Les schèmes de notation s'avèrent aussi très rentables pour préciser les activités d'apprentissage et d'évaluation appropriées, pour donner de la transparence au processus d'évaluation et pour assurer sa cohérence avec l'apprentissage. En

fait, ils rendent l'enseignement plus efficace.

1. Spécificité des critères choisis

Dans sa forme la plus élémentaire, le schème de notation peut être un simple corrigé dans lequel les réponses attendues sont indiquées, accompagnées de leur valeur en points. Ces schèmes de notation structurent l'attribution des points et contribuent, par conséquent, à un traitement semblable de toutes les copies examinées. Cependant, ils ne remplissent bien leur fonction que pour les évaluations d'habiletés simples, comme la mémorisation de faits, l'application d'automatismes et le respect de consignes simples. En fait, leur application se limite aux questions à choix de réponses et à réponses courtes, de même qu'à noter la présence ou l'absence d'éléments jugés importants (listes de vérification). Ils ne comportent pas de jugement, si ce n'est lors de la préparation de la tâche, en déterminant ce qui est valable d'évaluer, et dans la détermination de ce qui constitue ou non une bonne réponse.

Ces schèmes rudimentaires ne sont toutefois d'aucune utilité pour qualifier et décrire une performance en termes de forces et de faiblesses, en vue de réinvestir cette information dans l'activité pédagogique subséquente. L'exploration du processus de pensée et le développement des habiletés complexes nécessitent des structures plus élaborées.

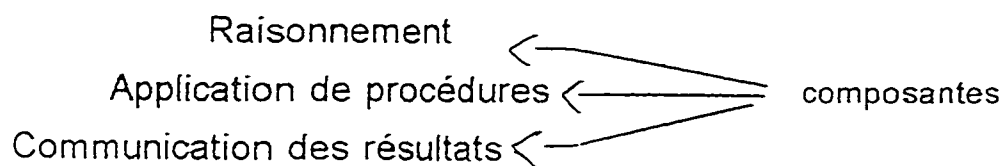
Dans la documentation, il est actuellement très difficile de trouver des directives au sujet de l'élaboration des schèmes de notation; nous sommes ici aux frontières de l'évolution des pratiques évaluatives et encore très peu d'experts osent se prononcer sur cette partie délicate de la tâche d'évaluation. Les écrits à ce sujet sont rares; par conséquent, les principes énoncés ici sont en pleine évolution et susceptibles de changer.

La précision de l'information générée par une démarche d'évaluation dépend du niveau de spécificité des critères. Dans le cas d'une évaluation formative, laquelle s'intéresse davantage au processus qu'au produit, les critères seront choisis pour faire ressortir toutes les composantes du processus qui pourraient exiger de l'aide pour progresser; ils seront donc relativement nombreux et spécifiques. Si l'évaluation est sommative, les critères seront moins nombreux, plus englobant, puisqu'un niveau de spécificité aurait peu d'utilité, ce type d'évaluation survenant habituellement en fin d'étape.

Voici deux exemples de différents niveaux de spécificité des critères d'évaluation, tous deux portant sur la résolution de problèmes. Selon que l'on veut examiner avec plus ou moins de détails la performance de l'élève, le nombre de composantes varie.

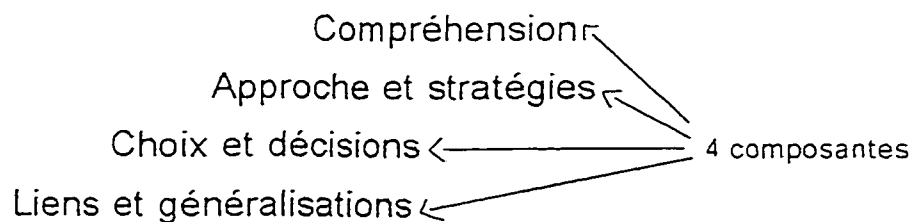
Dans le premier encadré, il y a trois composantes ou critères d'évaluation alors que dans le deuxième il y en a quatre. L'un n'est pas mieux que l'autre, seulement, l'un peut être plus approprié en certaines circonstances.

Résolution de problème <— habileté complexe

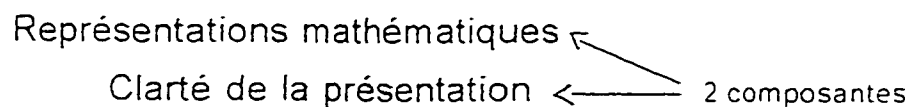


Résolution de problèmes

Résoudre le problème $\leftarrow$ 1^{re} habileté complexe



Communiquer la solution $\leftarrow$ 2^e habileté complexe



Les composantes indiquées dans le deuxième encadré permettent de cibler plus exactement les composantes à améliorer, tout en permettant de reconnaître les forces de l'élève. Cependant, plus il y a de détails, plus l'examen de la tâche est long.

Les composantes plus spécifiques conviennent particulièrement bien en début d'apprentissage, parce qu'elles permettent d'identifier les difficultés précises de l'élève et les aspects qu'il faut travailler davantage. Par la suite, il est possible d'évaluer avec moins de détails, pour éventuellement arriver à une évaluation plus globale, en fin d'apprentissage.

Bien que le nombre de composantes à examiner dans une démarche d'évaluation doive tenir compte de la progression dans l'apprentissage, il est aussi important de respecter le niveau de maturité des élèves. Un grand nombre de critères peut mieux convenir à l'évaluation de problèmes plus complexes ou à un niveau scolaire élevé, alors que les critères sont gardés à leur minimum avec les très jeunes élèves. Rappelons encore une fois qu'il est possible, occasionnellement, de se concentrer sur l'évaluation d'un seul critère, en particulier au début de l'apprentissage, lorsque qu'il faut diagnostiquer des difficultés très précises. Il est toutefois important de choisir une situation qui mette en oeuvre toutes les composantes de l'habileté.

2. Niveaux ou échelles - de notation, de rendement ou de performance

Les niveaux sont les catégories selon lesquelles les travaux ou les performances seront classées. Par exemple, le Bulletin provincial de l'Ontario comporte les quatre niveaux suivants : Excellent - Bon - Satisfaisant - Doit être amélioré. Parce que ces niveaux comportent à la fois une appréciation ou un jugement de la qualité du rendement pour chacune de ces catégories, elles sont aussi appelées échelles de rendement ou échelle d'évaluation. Cependant les niveaux peuvent aussi ne pas comporter d'appréciation, comme les catégories 1, 2, 3 et 4. Enfin, certaines échelles comportent à la fois des valeurs numériques et des appréciations qualitatives. Par exemple, le niveau 1 correspond à un travail de «novice», le niveau 2 à un travail «satisfaisant», le niveau 3 à un travail «très bien» et le niveau 4 à un travail «exceptionnel».

Un autre aspect à considérer dans les niveaux ou échelles adoptées, c'est le nombre d'échelons désirés. Tel que discuté au début de ce document, plus il y a de divisions ou de niveaux, plus le classement sera difficile; plus le classement est raffiné, plus il comporte de risque de commettre des erreurs. Aussi, plus la classification est fine, plus la tâche d'évaluation doit permettre de discriminer avec confiance. L'apprentissage est difficilement quantifiable et son classement, par critère et par niveau de performance, est difficile à défendre, en

particulier s'il y a plusieurs échelons. Voici quelques exemples de niveaux ou d'échelles de notation ou d'appréciation.

1	2	3	4
---	---	---	---

A Excellent	B Bon	C Satisfaisant	D Doit s'améliorer
----------------	----------	-------------------	-----------------------

A Excellent	B Très bien	C Satisfaisant	D Insatisfaisant	E Inacceptable
----------------	----------------	-------------------	---------------------	-------------------

(E) Exceptionnel	(T) Très bien	(S) Satisfaisant	(I) Insatisfaisant
---------------------	------------------	---------------------	-----------------------

Expert	Compétent	Novice
--------	-----------	--------

Élevé	Moyen	Faible
-------	-------	--------

Exceptionnel	Supérieur	Acceptable	Limité	Rudimentaire
--------------	-----------	------------	--------	--------------

A+	A	A-	B+	B	C
----	---	----	----	---	---

3. Échelles numériques et échelles descriptives

Un schème de notation présente habituellement un certain nombre de niveaux de performance, lesquels peuvent être décrits ou qualifiés (schèmes descriptifs) ou simplement gradués au moyen de chiffres. En voici deux exemples, le premier avec échelle numérique, le deuxième avec échelle descriptive.

a) échelle numérique

Résolution de problèmes			
	très bien	bien	pauvre
A - Compréhension du problème	3	2	1
B - Solution	3	2	1
C - Explication	3	2	1

b) échelle descriptive

Résolution de problèmes			
	Très bien	Bien	Pauvre
A - Compréhension du problème	Excellente compréhension	Ne saisit pas toutes les données	aucun signe de compréhension du problème
B - Solution	Bonne	Acceptable	Essai mais sans succès
C - Explication	Claire et complète	Incomplète	Incompréhensible

4. Schèmes de notation holistiques et analytiques

Tout comme pour les niveaux de performance à inclure dans un schème de notation, il est nécessaire de s'interroger sur le degré de précision du jugement à rendre. S'agit-il de comparer une performance à des normes établies, de rapporter le niveau de développement d'un élève à ses parents ou de documenter le développement des habiletés afin de générer des informations utiles à son apprentissage? Dans chaque cas le schème de notation sera adapté au motif pour lequel la démarche est entreprise.

Rappelons simplement qu'un critère d'évaluation est une caractéristique d'après laquelle on porte un jugement d'appréciation. Lorsque la performance est détaillée par critères, le schème est dit

«analytique»; si la performance est appréciée globalement, le schème est dit «holistique». Le premier est utilisé pour fournir des renseignements détaillés sur la performance d'un élève, afin d'identifier ses forces et ses faiblesses, ou pour diagnostiquer une difficulté précise; ce schème fournit des informations pouvant être réinvesties dans l'activité pédagogique, amenant l'élève à progresser dans son apprentissage. Il est souvent plus pertinent en début d'apprentissage parce qu'il convient bien à l'évaluation formative et à l'examen du processus d'apprentissage.

Par ailleurs, le schème holistique comporte un jugement global, sans le décrire par composantes ou par critères. Il est davantage utilisé pour exprimer une appréciation sur un produit plutôt que sur le processus. Il est davantage utilisé en fin d'apprentissage, pour confirmer ou certifier la maîtrise.

Voici deux exemples de schèmes de notations, tous deux portant sur la résolution de problèmes; l'un est analytique et l'autre est holistique. Bien entendu, les composantes d'un schème peuvent varier à l'infini. Leur nombre peut être relativement élevé ou limité, mais l'important est qu'il serve bien l'usage qu'on veut en faire et qu'on se sente à l'aise de l'utiliser tel qu'il est conçu.

On remarque qu'avec tous les schèmes présentés jusqu'ici, le résultat d'évaluation n'est pas exprimé en «points». Un des avantages

reconnus des schèmes de notation descriptifs est de décrire une performance plutôt que de la quantifier; en n'indiquant pas de note en des termes classiques (pourcentages), on mettrait l'accent sur la qualité de la performance et la possibilité de s'améliorer plutôt que d'obtenir une note moyenne. Cependant, dans le cas où les notes sont encore accordées en pourcentages, ou que les élèves sont habitués aux notes chiffrées, il est possible de rattacher des valeurs en points aux divers critères utilisés. En effet, il serait possible de répartir les valeurs en points selon les critères et leur importance relative, et de répartir ces points selon les divers niveaux de performance. Ceci pourrait constituer un compromis acceptable pour une période de transition ou pour ajuster une structure traditionnelle.

a) Schème de notation holistique pour la résolution de problèmes

Niveau 1 : Performance rudimentaire

Ne saisit pas le problème; n'utilise pas de stratégie particulière, pas de terme mathématique; sans diagramme; comporte quelques calculs avec erreurs, explication incompréhensible.

Niveau 2 : Performance insatisfaisante

Saisit quelques éléments du problème; stratégie partiellement correcte; ébauche d'un diagramme; quelques termes mathématiques, pas nécessairement bons; fait quelques erreurs de calcul; explication incomplète; ne termine pas le problème.

Niveau 3 : Performance satisfaisante

Comprend globalement le problème; identifie les principaux éléments en jeux; bonne stratégie; explication compréhensible, comporte un schéma et des termes mathématiques appropriés; bonne argumentation.

Niveau 4 : Performance exceptionnelle

Comprend très bien le problème; identifie tous les éléments impliqués, selon leur importance; stratégie efficace; argumentation solide et convaincante; réponse complète, claire et cohérente; schéma simple et clair; explications soignées.

b) Schème de notation analytique pour la résolution de problèmes

Critères	Note accordée			
	1	2	3	4
Compréhension de la tâche	Ne comprend pas du tout	Comprend partiellement	Comprend le problème	Maîtrise parfaitement le problème
Démarche	Inappropriée	Appropriée à certaines occasions	Convient	Efficace et raffinée
Solutions	Arrive à un résultat	Arrive à une réponse correcte	Trouve plus d'une solution	Trouve et explique plus d'une solution
Termes utilisés	Aucun terme mathématique	Utilise un terme mathématique approprié	Utilise deux termes mathématiques appropriés	tous les termes mathématiques sont appropriés
Représentations mathématiques	Aucune	En utilise mais ne sont pas nécessairement bonnes	Utilisation convenable	Utilise les représentations mathématiques efficacement
Présentation	Incompréhensible	Quelques éléments clairs	Assez clair pour voir qu'il a compris	Bien organisé Complet

Chapitre 3

Deux situations d'évaluation situant le concept de justesse

Afin d'illustrer comment s'applique le concept de justesse à l'élaboration d'une démarche d'évaluation, deux exemples de situations sont présentées. La première pourrait être une démarche en vue d'examiner un aspect très spécifique de l'apprentissage de l'addition chez des élèves de première année. Il s'agit d'une évaluation formative en vue de mieux cibler l'activité d'apprentissage à venir. Le deuxième exemple consiste en une activité d'apprentissage à laquelle se rattache une démarche d'autoévaluation. Selon les contraintes imposées, cette activité pourrait convenir aussi bien à des élèves de 3^e année qu'à des élèves de 8^e année.

A - Évaluation en vue de situer des élèves de 1^{re} année dans leur apprentissage de l'addition

Il s'agit d'une évaluation formative; l'information recueilli permettra de situer chaque élève afin de mieux intervenir pour l'aider à progresser. Ce pourrait aussi être en vue de diagnostiquer le niveau de développement des élèves d'une classe avant d'aborder l'enseignement de cette notion mathématique. Ce pourrait aussi être pour regrouper les élèves de niveaux semblables pour des activités.

1 a) Habileté visée : additionner deux nombres à un chiffre. Il s'agit d'une habileté simple.

b) Tâche : Il faut choisir une tâche qui permette de déterminer quelle stratégie utilise l'élève. Il faut donc d'abord établir quelles sont les stratégies possibles pour l'élève.

Il pourrait utiliser du matériel concret, semi concret ou le faire mentalement. S'il le fait mentalement, il pourrait soit commencer toujours par le premier nombre, soit toujours commencer par le nombre le plus grand.

Il n'est donc pas pertinent de donner une tâche où seul le résultat obtenu sera utilisé comme indicateur car peu importe la stratégie utilisée, il pourrait arriver à la bonne réponse. C'est le processus qui nous intéresse, non pas le produit. Il serait préférable de l'observer à la tâche, en mettant à sa dispositions des ressources (cubes, papier, crayon).

Pour que la tâche choisie laisse place à toutes les stratégies possibles, il est essentiel que le premier nombre qu'elle présente soit plus petit que le deuxième, pour voir quelle nombre l'élève prendra en premier. Par exemple $2 + 4$.

$$2 + 4 =$$

- a) Utilise des objets pour compter
 - b) Dessine une représentation des nombres
 - c) Compte mentalement de 2 à 6
 - d) Compte mentalement de 4 à 6.
-

Pour que la tâche soit signifiante pour l'élève, il serait préférable de contextualiser cette tâche d'addition en la traduisant en termes de bonbons, de gommes à mâcher, de billes ou de tout autre petits objets appréciés de ces élèves.

2. Cohérence avec l'activité pédagogique

On pourrait écrire la tâche sur un papier, en s'assurant de la présenter comme les élèves y sont habitués, c'est à dire de haut en bas ou de gauche à droite, selon ce à quoi ils ont été exposés jusqu'à maintenant. Il faudrait aussi mettre à sa disposition toutes les ressources dont il dispose habituellement en classe pour résoudre un tel problème.

Selon l'intention de départ, cette activité pourrait se situer en cours d'apprentissage, pour diagnostiquer ceux qui ont besoin d'attention individuelle pour progresser. Ce pourrait aussi être pour identifier ou se situent des élèves avant d'entreprendre l'apprentissage de ce concept.

Il faudrait s'assurer que les élèves vivent régulièrement des situations semblables en salle de classe lorsqu'il font des activités d'apprentissage. De cette manière ils ne seront pas intimidés par la démarche d'évaluation, parce qu'elle ressemblera aux activités auxquelles il sont habitués.

3. Transparence de l'activité évaluative

Demander à chacun des élèves individuellement de faire la tâche. Expliquer qu'il doit s'y prendre comme à l'habitude, ou adopter la manière de faire la plus efficace. Lui dire qu'il doit montrer ce qu'il sait faire et l'inviter à penser tout haut, pour mieux comprendre ce qui se passe dans sa tête.

L'élève devrait avoir été observé au travail et invité à penser tout haut à maintes reprises. Si l'élève est invité à penser tout haut que lorsqu'il est évalué, il se sentira peut-être menacé. Il est important qu'il sache que c'est la seule manière de savoir ce qu'il pense en travaillant et de l'aider à apprendre.

4. Absence de désavantage circonstanciel

Choisir un moment où l'élève est en forme; éviter les moments où il est fatigué, affamé ou excité et ne peut pas se concentrer. Il ne

faudrait pas presser l'élève et s'assurer qu'il comprend bien la tâche qu'il doit faire.

Discussions :

L'activité décrite ne rejoint pas nécessairement tous les aspects des composantes du concept de justesse présenté à la figure 1, simplement parce que certaines sont jugées non pertinentes. Par exemple, le schème de notation peut difficilement être familier à des élèves de 1^{re} année; il n'est pas non plus pertinent que les élèves aient fait de l'autoévaluation de tâches d'apprentissage semblables, ni qu'il y ait un retour en salle de classe sur cette activité. Ce sont plutôt dans les activités d'apprentissage suivantes que ces résultats seront réinvestis pour favoriser le développement de l'élève.

Le schème de notation de la performance de chacun des élèves se résume ici aux quatre stratégies possibles pour résoudre l'addition. Elles sont ici identifiées comme des choix de réponses possibles dans le seul but de pouvoir noter où se situe chacun des élèves à une date précise. Éventuellement, l'enseignant pourrait répéter l'activité pour voir s'il y a progrès et noter à nouveau où se situe l'élève dans les stratégies utilisées.

Comme il n'est pas absolument certain de respecter parfaitement les composantes du modèle et que l'activité prend peu de temps, il serait

préférable de répéter l'exercice à quelques jours d'écart. Le résultat en serait encore plus juste.

B - Activité d'apprentissage de résolution de problème et autoévaluation

Les pages 27 à 31 sont des feuilles à reproduire, prêtes à être distribuer aux élèves pour cette activité. Il s'agit d'une situation de résolution de problème avec un grand nombre de solutions possibles et sa solution ne relève pas directement d'habiletés mathématiques. Elle comporte plusieurs étapes et fait appel à de nombreuses habiletés, dont celles de travailler en groupe et de communiquer les résultats. Elle est susceptible d'être bien reçue, même par les mathophobes ou les élèves faibles en maths. Le volet d'évaluation de cette activité est formatif. L'analyse de cette activité d'apprentissage, dans les pages suivantes, en fait ressortir les points saillants permettant d'arriver à des résultats justes lors de l'évaluation de résolutions de problèmes semblables.

PROBLÈME À RÉSOUDRE

Invente une manière de coder pour identifier ce qui appartient à chacun des élèves de la classe, comme ses livres, sa chaise, sa boîte à goûter, ses travaux, etc., sans utiliser de lettres ou de chiffres.

- N. B. Tu dois pouvoir coder toi-même les objets
 Le code doit être simple et efficace
 Chaque élève doit avoir son code personnel
 Les élèves doivent pouvoir reconnaître leur code facilement

Exemple : Utiliser toutes les combinaisons qu'il est possible d'obtenir à partir de deux dés à jouer. Puisqu'il existe 36 possibilités différentes, ce pourrait servir à identifier 36 objets identiques appartenant à 36 personnes différentes.

Annie	• •	Marie	• • • •	Joseph	• • • •
Jean	• • •	Chantale	• • • • •	Geneviève	• • • • •
Paul	• • • •	Dominique	• • • • •	Ève-Marie	• • • • •
Karim	• • • • •	Marie-Josée	• • • • •	Monique	• • • • •
Benoît	• • • • •	Joey	• • • • •	Renée	• • • • •
Lise	• • • • •	Clémence	• • • • •	Marielle	• • • • •
Lucie	• • • • •	Micheline	• • • • •	Johanne	• • • • •
Mathieu	• • • • •	Éric	• • • • •	Étienne	• • • • •

Suggestions : Inventer plusieurs codes et choisir ensuite le meilleur.
 Tu peux utiliser le dos de cette feuille pour te pratiquer.

2. En groupe de trois ou quatre personnes, explique aux autres membres la manière que tu as choisi de coder et discute avec eux pour déterminer s'il est possible d'inventer une autre manière encore plus efficace. Une fois la meilleure manière identifiée, dessine le code de chaque élève de la classe.

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

3. Donne trois raisons pourquoi tu trouves que cette manière de coder est la meilleure?

a) _____

b) _____

c) _____

4. Fais une liste du matériel nécessaire pour réaliser ce projet.

5. Aimerais-tu qu'on utilise des codes pour identifier les objets appartenant aux élèves de la classe, pourquoi?

6. Qu'as-tu trouvé le plus difficile dans ce travail?

7. Qu'as-tu aimé faire dans ce travail?

8. Comment évalues-tu la qualité de ton travail?

Mets un crochet dans la case qui indique la qualité de ton travail. Si tu n'es pas certain, demande à un autre élève de t'aider.

	insatisfaisant	satisfaisant	très bien	exceptionnel
Invente un code (#1)				
Participe au groupe (#2)				
Dessine de manière soignée les codes (#1, #2)				
Analyse le code proposé par le groupe (#3)				
Démontre un esprit pratique (#4)				
Réfléchit à l'activité (#5, #6, #7)				

Veux-tu faire des commentaires?

Analyse de l'activité de résolution de problèmes selon le concept de justesse

Il s'agit d'une situation d'apprentissage d'une habileté complexe à faire seule ou en groupe. Elle n'est pas reliée directement à une discipline en particulier mais comporte un aspect de **Modélisation**. Il s'y greffe une activité d'autoévaluation de l'élève.

1 a) Habileté visée : Comporte les habiletés suivantes :

- Comprendre le problème
- Trouver une stratégie et l'appliquer
- Collaborer avec un groupe
- Analyser la solution proposée
- Communiquer ses résultats
- Créer des régularités

b) tâche choisie : Voir pages 37 à 41.

Cette tâche ne comporte pas beaucoup de chiffres, ce peut être plus attirant pour les élèves qui n'aiment pas les mathématiques, en particulier la résolution de problèmes. Les problèmes pour lesquels il existe plusieurs bonnes réponses ou peuvent être résolus de multiples façons sont particulièrement indiqués pour les activités de groupe.

2. Cohérence avec l'activité pédagogique

Cette activité d'apprentissage fait partie d'un ensemble de stratégies d'enseignement de la résolution de problème. Les étapes de la résolution de problèmes sont affichées dans la classe. À maintes

occasions, l'enseignant démontre à ses élèves comment on aborde un problème, en modélisant la démarche à entreprendre. Il peut conscientiser les élèves au nombre de problèmes qu'ils doivent résoudre quotidiennement, valoriser le développement de cette habileté et faire ressortir son utilité, pas seulement à l'école mais dans la vie de tous les jours. Les élèves sont invités à se référer aux modèles proposés lorsque l'occasion se présente.

Toute activité formelle de résolution de problèmes entreprise en classe devrait mettre en valeur les habiletés qu'elle nécessite. Ce pourrait être en utilisant un schème de notation, pour faire ressortir les caractéristiques qui permettent de juger de la qualité de la démarche. L'évaluation formative devrait permettre d'apprécier le progrès de l'élève à longue échéance parce qu'il s'agit d'une habileté complexe, donc lente à développer.

3. Transparence de l'activité évaluative

Expliquer aux élèves que le problème sert à voir comment ils s'y prennent pour résoudre un problème. Le plus important est de trouver une solution convenable au problème. Le schème de notation descriptif porte l'élève à regarder la qualité de son travail plutôt qu'à chercher les fautes et à amasser des points.

L'autoévaluation permet de valoriser les aspects importants de la résolution de problème. S'assurer que les élèves comprennent les critères et les inviter à en suggérer d'autres. Garder des traces de l'évolution de cette habileté longue à maîtriser.

4. Absence de désavantage circonstanciel

Bien distinguer la capacité à travailler en groupe de celle de résoudre des problèmes. En évaluant chacune séparément il est possible de diagnostiquer plus précisément certaines sources de difficulté.

Il faut être conscient que la créativité a pu être réprimée chez certains enfants. Dans ce cas l'imagination peut être lente à s'activer de nouveau.

La performance d'un élève peut être affectée par la difficulté à discuter en groupe, à cause d'un manque de confiance en soi. Cependant, cette stratégie peut stimuler l'intérêt d'autres élèves.

Les discussions sont enrichissantes mais consomment beaucoup de temps. S'assure que les élèves disposent d'assez de temps pour bien faire le travail.

Discussion

En classe, encourager les idées originales et la recherche de solutions, même si elles ne sont pas les meilleures. Perpétuer l'idée qu'un problème peut avoir plusieurs bonnes solutions. L'important n'est pas toujours d'avoir la meilleure solution mais plutôt de résoudre le problème de manière satisfaisante.

Amener les élèves à reconnaître des situations de vie qui nécessitent l'habileté de résoudre des problèmes.

Selon l'expérience des élèves et leur maturité, ils connaissent les critères jugés importants pour évaluer l'habileté à résoudre des problèmes.

Chapitre 4

Thèmes d'intérêt

A - Évaluer les travaux faits en groupe

Le travail en groupe est une stratégie d'enseignement. Elle est reconnue pour favoriser les échanges qui forcent le raisonnement et autres habiletés supérieures de la pensée, elle crée des situations forçant les relations interpersonnelles et la considération des idées des autres. Elle peut aussi contribuer à améliorer l'attitude face au travail scolaire.

Le bon fonctionnement d'un groupe de travail nécessite les **habiletés** suivantes : écouter les autres, respecter leurs idées, partager la charge de travail, partager ses idées, s'entraider.

Quatre aspects du travail de groupe sont déterminants pour son bon fonctionnement : Composition du groupe, tâche à accomplir, but de la stratégie, évaluation de la performance. Ils affectent tous le succès de l'activité.

a) Collaboration : le groupe fonctionne mieux lorsque la tâche à effectuer comporte plusieurs solutions possibles.

b) Entraide : Il faut que ceux qui peuvent aider les autres veulent bien le faire et que ceux qui ont besoin d'aide veulent bien le dire et sachent comment dire ce qu'ils ne comprennent pas..

c) Évaluation : Si une seule note est attribuée au groupe, elle est rarement méritée par tous les membres. Du point de vue motivation, une note commune réduit l'impression de contrôle que chacun a de sa performance, réduisant aussi sa motivation intrinsèque. Si chaque membre du groupe reçoit une note, la collaboration peut sembler une perte de temps. Dans les deux cas, les rapports interpersonnels à l'intérieur du groupe peuvent en souffrir.

d) Composition du groupe : La collaboration de personnes qui ont des habiletés très différentes peut créer des frictions. Les élèves faibles ne sont pas considérés comme des atouts.

Évaluation des travaux de groupes

Le travail de groupe a habituellement deux buts : apprendre un certain contenu et apprendre à travailler en groupe. Donc, deux aspects peuvent être évalués : la participation au groupe et la performance (individuelle et collective). Il est difficile de se pencher sur ces deux aspects en même temps. On pourrait donc choisir l'un ou l'autre aspect d'après la pertinence de la tâche, le contexte et l'importance du contenu travaillé.

Pour évaluer la participation au groupe, l'autoévaluation et l'évaluation par les pairs semblent appropriées. Si l'équipe note chacun de ses membres ouvertement, en comparant le travail de chacun et en obtenant l'assentiment des autres, la notation sera probablement réfléchie et juste

Pour valoriser le travail de groupe sans pénaliser personne, il est possible aussi d'assigner des tâches connexes au contenu travaillé en groupe à chacun des participants et les noter individuellement. Ce peut être par exemple de résumer individuellement sa perception du travail effectué en groupe; pour évaluer la performance, il serait probablement plus juste d'évaluer des copies individuelles. Cela force chacun à trouver profit dans la participation au groupe.

RÉFÉRENCES

- Blumenfeld, P., Marx, R., Soloway, E., & Krajcik, J. (1996). Learning with peers : From small group cooperation to collaborative communities. *Educational Researcher*, 25(8), 37-40.
- Herman, J., Aschbacher, P., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria (VA) : ASCD.
- Kulm, G. (1994). *Mathematics assessment. What works in the classroom*. San Francisco : Jossey-Bass.
- Leder, G. (1992). *Assessment and learning of mathematics*. Victoria, Australia : Australian Council for Educational Research.

[illegible]

Critères	E Exceptionnel	T Très bien	S Satisfaisant	I Insatisfaisant
1 - Participation	Participe et réagit positivement aux propos des autres	Donne son opinion et l'explique	Donne une idée mais ne peut la défendre	Regarde les autres travailler
2 - Contribution	Prend des initiatives pour que le travail soit efficace	Fait sa part de travail	Porte attention Fait un minimum d'effort	Sort du sujet ou ne fait rien
3 - Communication des résultats	Présente avec aisance et originalité	S'exprime clairement Veut que les autres comprennent	S'adresse aux autres mais est difficile à suivre	Refuse de s'adresser au groupe-classe
4 - Produit final	Remet un travail soigné et original	Remet un travail bien fait	Remet un travail de qualité acceptable mais sans plus	Ne termine pas le travail

B - Évaluer les attitudes face à une matière

Il faudrait d'abord définir ce qu'est une attitude positive et la distinguer d'autres composantes affectives comme intérêt, motivation, engagement, valeurs.

Les attitudes sont du domaine affectif.

Une attitude est une disposition intérieure acquise à l'égard de soi-même ou de son environnement, qui incite à une manière d'être ou d'agir favorable ou défavorable à un certain objet psychologique. Une attitude est basée sur le système de valeur de la personne. Elle est plus stable que l'opinion, plus profonde, plus complexe et plus irrationnelle. Les enfants peuvent changer favorablement d'attitudes lorsqu'ils prennent plaisir à leur travail ou qu'ils en voient l'utilité (Legendre, 1993).

L'évaluation des attitudes est importante parce que le développement d'attitudes positives est nécessaire à l'épanouissement personnel et au bon fonctionnement en société. Ce qui est important de situer, c'est sa direction (positive ou négative) et son intensité (forte ou faible), pour pouvoir ensuite apprécier un changement suite à une intervention appropriée.

Il est important aussi d'avoir une attitude positive à l'égard de l'apprentissage scolaire parce elle contribue à améliorer la performance. L'élève ayant une attitude positive apprend et travaille mieux.

Une attitude ne peut changer sur demande. On peut influencer l'attitude des élèves en «modélisant» une attitude positive et en valorisant ceux qui la démontre. Elle doit être observée sur une période prolongée parce qu'elle est lente à changer. La désirabilité est un obstacle sérieux à l'évaluation des attitudes. Il est donc important d'exprimer clairement que les attitudes se reflètent dans de nombreux petits gestes qui ne peuvent pas tromper. Donner des exemples concrets de tels comportements observés dans la classe à mesure qu'ils se produisent.

Éthiquement, il serait répréhensible d'évaluer des attitudes ou autres dispositions affectives qui ne sont pas concernées par le contenu du programme d'enseignement ou de ne pas intervenir dans le but de les améliorer.

Stratégies d'intervention, pour améliorer l'attitude à l'égard des mathématiques par exemple :

- Accorder de l'importance au plaisir que l'élève ressent en faisant des activités d'apprentissage.
- Les inviter à inventer des problèmes.
- Porter attention au contenu, pour tenir compte de leurs intérêts personnels, et parler de choses qui les intéressent.
- Choisir des activités tirées de la vraie vie, où ils ont non seulement à calculer mais à manipuler du matériel concret.

- Intégrer le contenu mathématique à d'autres matières. Ceci met en valeur l'utilité des mathématiques et risque d'intéresser à la fois ceux qui aiment les mathématiques et ceux qui aiment la matière intégrée.
- Relier le contenu étudié à l'importance de développer cette habileté.
- Faire prendre conscience que les mathématiques font partie de la vie de tous les jours. Donner des exemples de situations où les mathématiques ont été utiles, en particulier en dehors de la classe. Ceci pourrait constituer un exercice : l'élève fait une liste des situations où il a eu recours à des habiletés apprises en mathématiques pour effectuer une activité ou une tâche en dehors de la salle de classe, pendant une période de temps fixée. Répéter l'exercice à quelques occasions, afin de voir si l'élève progresse.

Le journal mathématiques est une façon de jauger, sur une période à long terme, l'évolution des attitudes.

Inviter l'élève à exprimer ses dispositions face aux mathématiques (oralement ou par écrit), ce qu'il aime et n'aime pas et surtout, dire pourquoi. Cependant, il ne faudrait pas juger l'élève sur ce qu'il écrit mais la confiance qu'il démontre envers l'enseignant et l'effort qu'il a mis à révéler ses positions. Se servir du contenu que pour avoir des pistes sur la manière d'améliorer les dispositions de l'élève et pouvoir apprécier un changement éventuel d'attitude.

RÉFÉRENCES

- Giasson, J. (1995). *La lecture. De la théorie à la pratique*. Montréal : Gaëtan Morin.
- Good, T., & Brophy, J. (1990). *Educational psychology* (4th edition). New York : Longman.
- Kulm, G. (1994). *Mathematics assessment. What works in the classroom*. San Francisco : Jossey-Bass.
- Leder, G. (1992). *Assessment and learning of mathematics*. Victoria, Australia : Australian Council for Educational Research.
- Legendre, R. (1993). *Dictionnaire actuel de l'éducation*. Montréal : Guérin.
- Stiggins, R. (1994). *Student-centered classroom assessment*. Toronto : Maxwell Macmillan.

Annexe G

Plans d'évaluation des situations 6 à 10

Plan d'évaluation, situation # 6

1. Évaluation formative

Par le biais de rencontres individuelles avec chaque équipe, l'enseignant vérifie chaque étape de la démarche scientifique et stimule la progression du projet. Il fait des suggestions, propose des corrections, tant du point de vue du français que de l'approche scientifique. Il voit ainsi au développement optimal des habiletés visées par le projet. Les observations pertinentes sont notées au dossier anecdotique des élèves.

Un modèle du Guide et cahier de l'élève (annexe E), de même qu'un présentoir complété sont affichés dans la classe pour guider les élèves dans leur projet.

2. Présentation orale en salle de classe

À la date fixée, les équipes présentent leurs travaux de recherche devant la classe. Pour ce faire, il préparent un présentoir tel que recommandé à l'avant-dernière page du Guide et cahier de l'élève. Ce guide contient aussi, à la dernière page, une liste des critères selon lesquels le présentoir sera évalué, de même que l'expression orale et le contenu à traiter lors de la présentation. Cette liste de critères est reprise à l'annexe B.

3. Travail d'équipe

Les élèves travaillent en équipe de deux. Normalement, ils partagent la charge de travail équitablement. Afin d'assurer la bonne entente tout en respectant les intérêts et les talents de chacun, la responsabilité de rapporter toute situation insatisfaisante à cet égard est confiée à l'élève. Après chaque période de travail en commun, les participants sont invités à remplir un billet sourire (annexe C) et à le remettre à l'enseignante. En signant leur nom et en encerclant la figure illustrant le mieux l'entente qui règne au sein de leur équipe, les élèves peuvent signaler les mauvais fonctionnements de l'équipe. Ce sont surtout les situations insatisfaisantes qui retiennent l'attention; les cas litigieux sont négociés par l'enseignante le plus tôt possible et une note est inscrite au dossier anecdotique des élèves en cause. Après la présentation du projet, les co-équipiers discutent et s'accordent mutuellement une note, selon le même barème que les cartons-

notes.

4. Épreuve

Suite à la foire scientifique, les élèves subissent une épreuve papier-crayon à répondre sur le questionnaire (annexe D). Elle fait plus appel à la réflexion qu'à la mémorisation et devrait permettre de se prononcer sur les habiletés qui n'ont pas été touchées par les autres volets du plan d'évaluation. C'est la seule partie où des notes sont accordées, pour un total de 40 points, traduites ensuite dans le barème de couleur des cartons-notes.

plus de 35 points - Bleu - exceptionnel

entre 28 et 35 points - Vert - très bien

entre 20 et 27 points - Jaune - satisfaisant

moins de 20 points - Rouge - insatisfaisant

5. Rétrospective

Lors de la remise des copies de l'épreuve, l'enseignant communique aux élèves les réponses attendues et discute de toute méprises à ce sujet. Les échecs, s'il en existe, sont suivis individuellement. Puisque l'évaluation de cette activité a eu recours à certaines stratégies nouvelles pour les élèves, ils sont invités à faire part de leurs impressions à ce sujet.

Toutes les notes cumulées sont consignées dans le Dossier de notation du projet scientifique (annexe F), lequel comporte un dossier anecdotique pour chacun des élèves permettant de noter toute observation pertinente en cours d'apprentissage ou d'évaluation.

Plan d'évaluation, situation # 7

1. Épreuve

après une mise en situation appropriée, un collage de texte est distribué aux élèves sous le titre *En toute sécurité!* (annexe A) Il est accompagné d'un questionnaire (annexe B). L'épreuve papier-crayon, à répondre sur le questionnaire, fait appel à la capacité de synthétiser l'idée principale d'un texte, à faire des inférences et à établir des liens entre le contenu de la lecture et ses expériences personnelles. Des points sont attribués pour les réponses, pour un total de 40 points. Le résultat est comptabilisé sous les rubriques *Repère de l'information* et *Inférences*. La note totale est accompagné d'un classement selon les quatre couleurs : bleu (36 points et +; exceptionnel), vert (28-35 points; très bien), jaune (20-

2. Rétrospective

Lors de la remise des copies de l'épreuve, l'enseignant communique aux élèves les réponses attendues et discute de toute autres formulations possibles. Puisque cette activité de lecture diffère des activités habituelles de lecture, les élèves sont invités à réagir à l'activité.

Plan d'évaluation, situation # 8

1. Évaluation formative

Cette activité a pour but de diagnostiquer de manière systématique la maîtrise du concept de multiplication, la maîtrise des diverses stratégies (compter par bonds, répéter des additions et utiliser une autre multiplication), la capacité de résoudre une multiplication en situation de résolution de problème par une stratégie appropriée et de communiquer sa réponse.

2. Exercices de consolidation

L'enseignante note au dossier anecdotique des élèves (annexe B) des observations sur leur apprentissage, qu'il s'agisse de forces ou des faiblesses. Dans le cas de difficultés liées à la multiplication en situation de résolution de problème, il faudra déterminer s'il s'agit de l'identification de l'information pertinente, du choix d'une stratégie appropriée, ou de la compréhension du problème lui-même. Selon le besoin individuel de chacun, elle entreprend d'apporter l'aide individuelle et assigne d'autres exercices spécifiques à la difficulté détectée, lesquels sont choisis dans la banque d'exercices apparaissant à l'annexe C.

3. Évaluation sommative

Étant donné l'importance pour certains élèves de 3e année de pouvoir compter sur une aide extérieure pour persister à la tâche, ce feuillet d'exercices (annexe D) a une place prévue pour laisser des traces de l'aide apportée. Ceci permet d'utiliser au maximum l'aspect « apprentissage » de l'activité d'évaluation, tout en permettant de documenter le degré d'autonomie de chacun des élèves, élément jugé révélateur dans la détermination du niveau de performance d'un élève en bas âge. La marge de droite sert à documenter l'aide apportée et ces traces sont utilisées lors de l'attribution de notes.

L'évaluation sommative est une épreuve papier-crayon. Pour l'élève, le résultat est exprimé sous trois rubriques : choisir l'information pertinente, maîtriser le concept de multiplication et résoudre des multiplications en situation de problème. Pour le bulletin,

il se limite à une seule rubrique, maîtriser le concept de multiplication en situation de problèmes.

4. Bilan individuels

Après avoir vérifié les copies des élèves pour s'assurer qu'ils aient trouvé une réponse correcte à chacun des problèmes imposés, l'enseignante remet les copies aux élèves et leur remet une feuille bilan (annexe B). À partir de traces laissées par l'enseignante lors de cette revue sommaire de la pertinence des réponses et des indications au sujet de l'aide apportée durant l'évaluation, chacun peut faire le «compte» de ses forces et de ses faiblesses. Encore une fois, c'est à partir d'une échelle en 4 points, comme celle utilisée pour le bulletin, que les élèves résumeront leur performance. Le travail de l'élève se classe dans un des quatre niveaux, selon le barème suivants :

bleu - travail exceptionnel

vert - très bon travail

jaune - travail satisfaisant

rouge - travail insatisfaisant

Après avoir terminé son bilan, l'élève l'apporte à l'enseignante. Elle explique à l'élève le sens de ce bilan et prend en note la performance de chacun.

5. Rétrospective

Les épreuves sont remises et revues en classe; les élèves peuvent poser des questions. Cette manière d'évaluer étant légèrement différente de la manière habituelle, les élèves sont invités à commenter cette manière de faire.

Plan d'évaluation, situation # 9

1. Présentation orale en salle de classe

À une date fixée, les équipes présentent leur machine aux autres élèves de la classe. Le contenu à traiter lors de ces présentations, de même que le temps alloué et le mode d'évaluation apparaissent à la dernière page du feuillet de l'élève. On y retrouve aussi les critères d'évaluation de ces présentations. Ce document est repris à l'annexe B.

2. Travail d'équipe

Les élèves travaillent en équipe de 4 élèves. Chaque équipe reçoit un formulaire d'évaluation de la participation des membres (annexe C) et doit le remplir à la fin de l'activité. Ils doivent s'évaluer en groupe et s'entendre sur la note à attribuer à chacun.

3. Compte rendu du projet

Chaque élève est responsable de rendre compte personnellement du déroulement du projet en remplissant le feuillet de l'élève. Étant donné la nouveauté d'un tel projet, ce rapport sera évalué globalement pour attribuer une seule note, selon les critères établis dans le feuillet de l'élève.

4. Rétrospective

Après la notation de toutes les composantes du plan d'évaluation, les travaux seront retournés aux élèves. L'enseignante fait un retour sur l'activité et donne l'occasion aux élèves d'exprimer leur impressions suite à un tel projet. Elle répond aux questions et fait les réajustements nécessaires.

Toutes les notes cumulées sont consignées dans le **Dossier de notation du projet de robotique (annexe D)**, lequel comporte un dossier anecdotique pour chacun des élèves permettant de noter toute observation pertinente en cours d'apprentissage ou d'évaluation.

Plan d'évaluation, situation # 10

1. Habitudes de travail

Afin de valoriser les bonnes habitudes de travail et d'encourager l'autonomie des élèves dans la recherche d'idées, les élèves sont invités à évaluer eux-mêmes les comportements suivants une fois qu'ils auront terminé le projet : initiatives, manipulation, rangement, utilisation du temps et persévérance. Des tableaux seront affichés dans la classe (pièce-témoin) et chacun des élèves s'attribue une note en utilisant le symbole à côté de l'indicateur qui convient le mieux (pièce-témoin C).

2. Présentation du modèle

Lorsque les 4 ou 5 élèves travaillant simultanément au défi technologique ont terminé leur modèle ou sont arrivés à échéance, ils se rassemblent auprès de l'enseignante pour présenter leur réalisation. Chacun décrit sommairement le prototype conçu et le fonctionnement de l'appareil. Il explique le rôle du moteur et des engrenages, fait fonctionner l'appareil, compte le nombre de rotations de l'hélice en 20 secondes et explique comment il pourrait changer la vitesse de l'hélice.

L'enseignante demande aux autres membres du groupe de commenter le modèle présenté sur les éléments suivants :

- a) Structure : Forme générale, solidité, matériaux, détails (hélice à l'arrière, patins)
- b) Proportion dans la longueur des pales et la longueur de l'appareil, dans la taille de l'habitacle et de l'appareil, et la relation taille-poids.
- c) Fonctionnement : l'efficacité du fonctionnement (le moteur actionne l'hélice, la vitesse et engrenages).

3. Plans des modèles

L'enseignante évalue globalement la qualité des deux schémas remis par chacun des élèves selon les critères suivants : propreté, clarté et pertinence

4. Rétrospective

Lorsque tous élèves ont terminé le projet, l'enseignante remet les plans aux élèves et discute de la performance générale des élèves. Elle les invite à faire part de leurs impressions sur le déroulement du projet et la manière de noter les travaux, étant donné que l'approche est légèrement différente de l'habituelle.

Toutes les notes cumulées sont consignées dans le Dossier de notation du projet d'hélicoptère (annexe D), lequel comporte un dossier anecdotique pour chacun des élèves permettant de noter toute observation pertinente lors de l'activité d'apprentissage et de l'évaluation.

Annexe H

Analyse de la justesse des résultats d'évaluation, situations 6 à 10

Foire scientifique, 5 ^e et 6 ^e années	
1. Pertinence de la fiche d'évaluation par rapport aux habiletés visées Habilité visée : Connaître la méthode de recherche scientifique pour résoudre un problème, ce qui comprend : Énoncer clairement un problème ou une question de recherche, choisir une méthode appropriée pour résoudre un problème et l'appliquer, expérimenter des liens de cause à effet et les décrire, suivre les étapes d'une méthode de recherche normalisée, démontrer précision, minutie, persistance, persévérance, créativité, honnêteté, exactitude et rigueur, apprécier les avantages d'une méthode de recherche, travailler en équipe de deux personnes, selon les exigences du projet et partager l'information. Tâche : Préparer un présentoir; communiquer à la classe l'essentiel de la démarche de recherche explorée; faire une épreuve sur le processus de recherche. Les élèves ont une marge de liberté pour choisir le sujet à étudier (motivant); les diverses tâches ont été clairement définies. Les activités mettent en valeur la débrouillardise et les bonnes habitudes de travail plutôt que le bon fonctionnement de l'expérience. Critères : Présentoir (lisibilité, disposition, propreté); présentation (expression orale, contenu); travail d'équipe (collaboration); épreuve (compréhension de la démarche scientifique); ce sont des critères propres au milieu authentique La qualité du français n'a pas été évaluée mais on a évalué la contribution à l'équipe	
2. Cohérence avec l'activité pédagogique L'activité d'évaluation est distincte de l'activité d'apprentissage mais est en soi une bonne activité d'apprentissage Lien étroit entre l'apprentissage et l'évaluation grâce au Guide de l'élève bien préparé La mise en situation est stimulante Les élèves ont l'habitude des présentations orales et du travail en équipe et sont familiers avec la démarche scientifique L'enseignant est disponible pour de l'aide Résultats exprimés sous forme de profil, donc plus facile à réinvestir Mise en commun favorise l'apprentissage Conception de l'épreuve nouvelle, formule de notation nouvelle L'épreuve ne peut pas être bien réussie avec des automatismes seulement	3. Transparence de l'activité évaluative Modèle d'un présentoir pour communiquer les attentes Critères établis d'avance mais remis quelques jours plus tard Utilisation d'un dossier anecdotique, Notation sous forme de profil, permet à l'élève de connaître ses forces et ses faiblesses, résultats exprimés par code de couleur (4 niveaux) correspond à un classement ou une appréciation Retour sur l'activité quand tous ont fini la tâche. Les notes sont communiquées et discutées avec les élèves. La note La manière de noter favorise l'intégration des critères, un premier pas vers la prise en charge de son évaluation
4. Absence de désavantage circonstanciel Présenter devant la classe peut indisposer certains élèves Amplement de temps pour compléter l'épreuve La grande partie des activités évaluatives a lieu dans la classe Le choix d'un partenaire a été fait avec beaucoup de sérieux	

Figure 5. Analyse de la justesse des résultats d'évaluation, situation 6

Lecture, 5e et 6e années	
1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées Habiletés visées : Comprendre l'information importante d'un texte, c'est à dire dégager l'idée principale d'un texte, faire des inférences et établir des liens entre le contenu de la lecture et ses expériences personnelles Tâche trop longue Directives de l'enseignant biaise le type de réponses attendues Pas de notes attribuées pour le français, la propreté ou autre habiletés connexes.	
2. Cohérence avec l'activité pédagogique La lecture est faite à haute voix; l'élève peut répondre à plusieurs questions sans avoir à lire le texte Pas de mise en situation, ni de page titre Discussion très limitée Le niveau actuel des élèves est très hétérogène, difficile d'avoir une tâche adaptée au niveau de tous La longueur du texte rend l'activité plus difficile Certaines questions sont complexes	3. Transparence de l'activité évaluative Les élèves savaient à l'avance qu'ils allaient être évalués Le code de couleur transmet une appréciation du résultat Retour sur l'activité en classe
4. Absence de désavantage circonstanciel Fin d'année, les élèves ont fait plusieurs évaluations rapprochées Plusieurs dérangements durant l'activité d'évaluation Manque d'espace, la proximité des élèves permet d'échanger des réponses	

Figure 6. Analyse de la justesse des résultats d'évaluation, situation 7

Multiplication, 3e année	
1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées Habiletés : Dans un contexte de résolution de problèmes, effectuer des multiplications, c'est-à-dire démontrer sa maîtrise du concept de multiplication, reconnaître l'information pertinente, appliquer diverses stratégies pour résoudre un problème de multiplication et communiquer sa réponse La situation d'évaluation n'est pas très authentique, même si on a cherché à utiliser des problèmes signifiants pour ces élèves La tâche est définie en termes assez concis	
2. Cohérence avec l'activité pédagogique La tâche a été expliquée par l'enseignante, comme pour les activités d'apprentissage Les élèves ont accès à des pages de calendrier et à des tableaux des 100 premiers nombres et autre matériel concret, comme lors des activités d'apprentissage Les élèves sont assez familiers avec ce genre de problèmes; ils réussissent à faire l'exercice sans aide L'activité d'évaluation est une bonne tâche d'apprentissage Les habiletés à démontrer sont les mêmes que ceux de l'exercice de revue Le vocabulaire et les consignes sont simplifiées au maximum	3. Transparence de l'activité évaluative Les élèves savent que c'est une évaluation finale, dernière avant les bulletins et qu'ils doivent laisser des traces de leur raisonnement Retour sur l'activité en classe, après qu'elle est complétée, à revoir avec les parents Les élèves connaissent le barème du bulletin, barème utilisé pour noter les copies Marie écrit beaucoup de commentaires sur la copie pour situer les difficultés de l'élève en fonction des diverses habiletés évaluées
4. Absence de désavantage circonstanciel L'évaluation s'est déroulée en 4 séances, permettant aux élèves de donner à chaque fois leur maximum Quelques élèves ont été absents L'arrivée de la fin de l'année scolaire ne permettait pas de poursuivre la période d'apprentissage avant d'évaluer.	

Figure 7. Analyse de la justesse des résultats d'évaluation, situation 8

Robotique, 8e année	
1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées Habiletés visées : Confronté à un défi, réaliser un projet technologique en suivant la démarche établie (exploration, construction, évaluation, modification). Ceci inclue : comprendre le problème présenté et ses paramètres, imaginer une solution possible, planifier la construction, construire un prototype en appliquant les concepts de levier et de plan incliné, évaluer la performance du prototype, le modifier si nécessaire, travailler en équipe, démontrer de bonnes habitudes de travail et présenter la synthèse du travail devant la classe. La tâche est à effectuer en groupe, c'est plus stimulant que de travailler seul. Elle intègre des notions étudiées en science (levier et plan incliné). Ce n'est pas une situation familière, ni une tâche authentique. Elle ne justifie pas nécessairement d'investir toute l'énergie nécessaire à sa réalisation. Les critères mettent en valeur la débrouillardise et les bonnes habitudes de travail plutôt que le bon fonctionnement de l'appareil Critères d'évaluation : habitudes de travail, travail d'équipe, prototype, travail écrit, présentation	
2. Cohérence avec l'activité pédagogique L'enseignante ne peut pas aider les élèves Pas d'activité d'apprentissage préalable - c'est la première du genre cette année Résultats exprimés sous forme de profil La situation d'évaluation est elle même très pédagogique	3. Transparence de l'activité évaluative Les critères sont nouveaux Les élèves n'ont pas reçu les critères d'évaluation en même temps que la tâche Utilisation d'un dossier anecdotique Code de couleur communique explicitement l'appréciation L'activité d'évaluation a été une première dans l'apprentissage de la responsabilisation des élèves face à leur évaluation Elle a permis d'intégrer les critères d'évaluation Retour sur l'activité en classe
4. Absence de désavantage circonstanciel Les paramètres établis constituaient un très grand défi Le travail effectué en équipe de 4 élèves n'a pas nécessairement donné l'occasion à chacun de démontrer son plein potentiel La situation n'a pas forcé tous et chacun à développer l'habileté du design technologique Cette tâche étant facilement stéréotypée, les filles auraient peut-être dû être regroupées ensemble	

Figure 8. Analyse de la justesse des résultats d'évaluation, situation 9

Projet technologique, 4e et 5e année	
4. Absence de désavantage circonstanciel La mise en situation force les filles à s'impliquer dans la conception du projet. Le temps alloué était suffisant Collaboration permise est documentée Les élèves effectuant l'activité sont en retrait; pendant l'activité la classe se poursuit. L'élève peut être tenté d'essayer de suivre ce qu'il manque pendant qu'il effectue ce projet. L'enseignante est moins disponible. Évaluation à l'heure du dîner peu réduire l'intérêt des élèves	
2. Cohérence avec l'activité pédagogique Peu de temps consacré à l'apprentissage des habiletés ciblées (design technologique) La mise en situation est stimulante, accompagnée de matériel inspirant (jouets, livres, fiches Dacta-Légo) Les bonnes habitudes de travail ont été valorisées dans la plupart des situations d'apprentissage Résultats exprimés sous forme de profil, donc plus facile à réinvestir Mise en commun à la fin du projet est formative	3. Transparence de l'activité évaluative Les exigences de l'évaluation sont transmises aux élèves en même temps que la tâche Utilisation d'un dossier anecdotique, schème descriptif avec profil, permet à l'élève de connaître ses forces et ses faiblesses, résultats exprimés par code de couleur Retour sur l'activité quand tous ont fini L'autoévaluation incite les élèves à prendre leur évaluation en main Les élèves doivent lire toutes les descriptions pour choisir celle qui convient le mieux, ils ne le font pas nécessairement avec sérieux
1. Pertinence de la tâche d'évaluation par rapport aux habiletés visées Habiletés visées : Confronté à un défi, réaliser un projet technologique dans lequel le mouvement d'un moteur électrique est transmis à une hélice, c'est-à-dire comprendre le problème présenté et ses paramètres, imaginer une solution possible, planifier la construction, construire un prototype en appliquant les paramètres, évaluer la performance du prototype, le modifier si nécessaire et démontrer de bonnes habitudes de travail. La mise en situation est stimulante toute en étant familière. Elle n'est pas une tâche authentique et ne justifie pas nécessairement d'investir toute l'énergie nécessaire à sa réalisation. Les critères mettent en valeur la débrouillardise et les bonnes habitudes de travail plutôt que le bon fonctionnement de l'appareil. Critères : habitudes de travail, prototype, plan	

Figure 9. Analyse de la justesse des résultats d'évaluation, situation 10

Annexe I

Guide de l'entrevue finale

Guide de l'entrevue finale

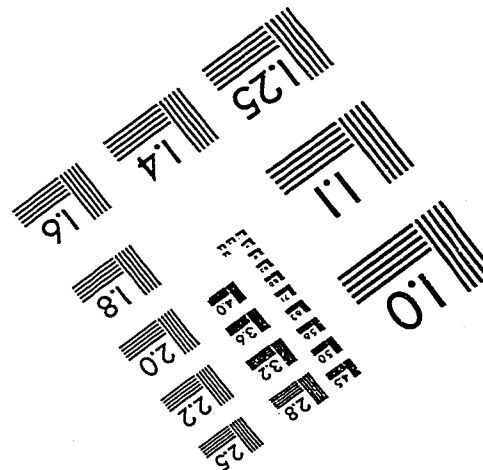
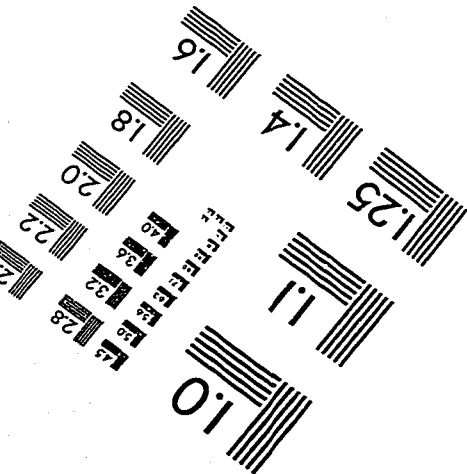
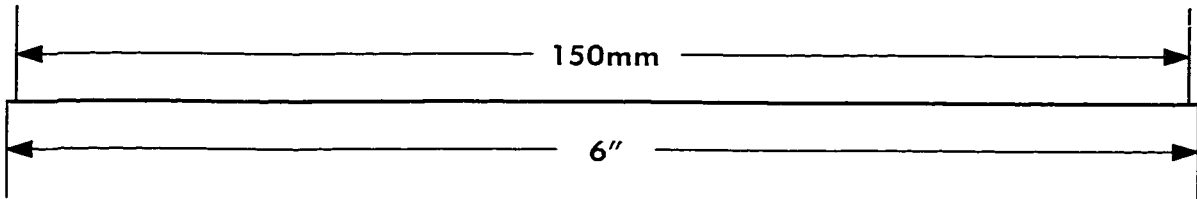
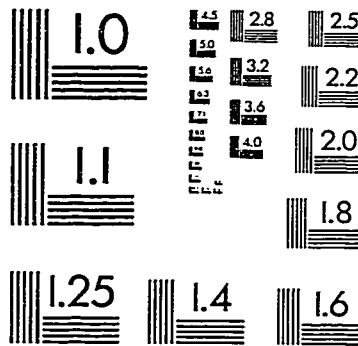
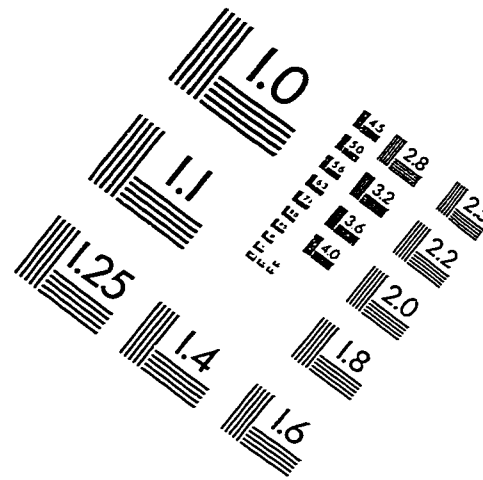
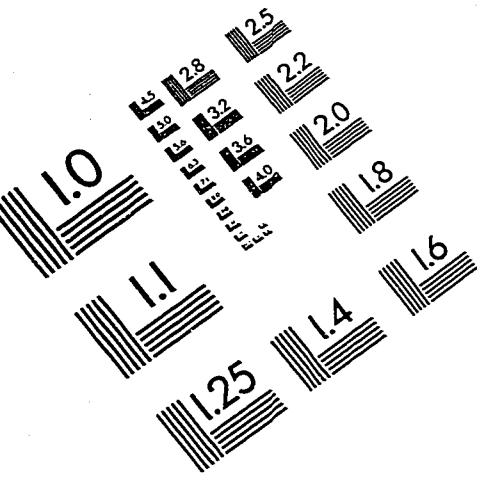
Rappeler au participant qu'il peut arrêter l'entrevue en tout temps.

1. Pourquoi avez-vous accepté de participer à cette recherche?
2. Aviez-vous des attentes quand vous avez accepté de participer et si oui, lesquelles?
3. Cette expérience a-t-elle eu des répercussions sur votre manière d'enseigner ou d'évaluer ou sur d'autres aspects de votre vie à l'école?
4. En vous référant aux schémas du concept de justesse que voici,
 - a) pouvez-vous estimer chaque composante pour son utilité à fournir des résultats justes?
 - b) pour le réalisme de ces composantes, c'est-à-dire, combien trouvez-vous réaliste de pouvoir les respecter ou en tenir compte dans la vie en salle de classe?
 - c) pouvez-vous suggérer des améliorations aux composantes ou à l'ensemble, pour rendre le concept plus approprié?
 - d) croyez-vous intégrer ces composantes à votre pratique future?
5. Croyez-vous qu'il y ait une ou des composantes particulièrement difficiles à mettre en pratique? Pourquoi?
6. D'après vous, y a-t-il un type d'enseignant qui pourrait bénéficier plus particulièrement d'une expérience comme celle que vous avez vécue dans ce projet?

7. Si je devais refaire cette activité avec d'autres enseignants, quelle modification suggèreriez-vous à la procédure?

8. Aimerez-vous rajouter quelque chose au sujet de cette recherche?

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE, Inc.
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved