

Rapport de recherche
31 août 2001

Création d'épreuves étalonnées
En relation avec les nouveaux socles de compétences
Pour l'enseignement fondamental

Rapport final
2^{ème} année

Recherche en éducation n° 67/00
Communauté française de Belgique

Rey B. : Professeur et promoteur de la recherche
Carette V., Defrance A., Kahn S., : chercheurs ULB
Deroisy A., Schoeters D., Marcoux G., : stagiaires

Université Libre de Bruxelles
Service des Sciences de l'Éducation
Avenue F.D. Roosevelt, 50
1050 Bruxelles
U.L.B. tél. 02/650 56 87 ou 54 58 ou 45 89

Nous tenons à remercier vivement

Les nombreux élèves

Les directeurs

Les enseignants

Les inspecteurs de la province de Namur

Pour leur participation active

Introduction

Ce rapport présente les résultats et les analyses, qui en découlent, de notre deuxième année de recherche portant sur la création d'épreuves étalonnées en relation avec les nouveaux Socles de compétences pour l'enseignement fondamental.

Cette recherche débutée en septembre 1999 s'inscrit directement dans le processus d'application du décret du 24 juillet 1997 définissant les Missions prioritaires de l'Enseignement Fondamental et de l'Enseignement Secondaire et organisant les Structures propres à les atteindre. Ce décret a entraîné la mise en place de groupes de travail qui ont reçu la mission de rédiger « les Socles de compétences ». Ces Socles de compétences publiés et distribués depuis le mois de mai 1999 sont comme l'indique le décret « missions » un « référentiel présentant de manière structurée les compétences de base à exercer jusqu'au terme des huit premières années de l'enseignement obligatoire et celles qui sont à maîtriser à la fin de chacune des étapes de celles-ci parce qu'elles sont considérées comme nécessaires à l'insertion sociale et à la poursuite des études »

Comme indiqué à l'article 19 du décret, à la suite de la publication des Socles de compétences, il est créé une *Commission des outils d'évaluation* relatifs à ceux-ci. Cette commission est chargée de produire des batteries d'épreuves d'évaluation étalonnées et correspondant aux socles de compétences. Celles-ci seront diffusées, « à titre indicatif » auprès de tous les établissements organisés ou subventionnés par la Communauté française.

C'est dans ce cadre institutionnel que s'inscrit cette recherche. Celle-ci a comme objet de proposer des épreuves à cette commission qui prendra la décision de les publier.

La première année de recherche nous a conduit à construire des épreuves en référence à un cadre conceptuel analysant les notions d'évaluation et de compétence. Deux épreuves construites suivant un modèle en trois phases ont été administrées auprès d'un échantillon d'environ 200 élèves de troisième année primaire et de 200 élèves de sixième année. L'analyse des résultats semble indiquer la pertinence de notre modèle et confirmerait notre hypothèse de départ : les compétences proposées dans les socles sont actuellement peu maîtrisées par les élèves et qu'il faudra un temps certain avant que les enseignants modifient leur pratique de façon à amener leurs élèves à une véritable maîtrise de celles-ci. Cette hypothèse, énoncée dès le début de la recherche, a soulevé des questions sur le caractère normatif ou critériel de l'étalonnage. Pour cette raison, nous avons opté pour des épreuves à vocation diagnostique.

Début septembre 2000, nous avons rencontré la commission des outils d'évaluation à qui nous avons présenté les résultats et les conclusions de notre première année de recherche. Celle-ci n'ayant pas fait de demandes particulières pour la suite de notre travail, le comité d'accompagnement nous a demandé de consacrer la deuxième année de recherche à tester d'autres épreuves construites sur notre modèle. Notre objectif principal est de confirmer ou d'infirmer les résultats obtenus lors de la première année. En effet, des résultats statistiquement semblables entre les six épreuves testées – trois pour le cycle 5-8 et trois pour le cycle 10-12 - permettraient de confirmer notre modèle.

Au delà de cet objectif principal, nous avons tenté d'apporter des éléments de réponses à des questions que la première année de recherche a suscitées. Ainsi, nous voulions tester l'influence du milieu social, l'influence de la dynamique de l'établissement, l'influence de l'enseignant et l'influence du vécu de l'élève, sur les performances obtenues par les élèves lorsqu'ils sont mis devant une situation complexe demandant un transfert de procédures automatisées. Pour confirmer notre modèle et pour tenter d'apporter quelques éléments de réponses à ces questions, nous avons administré quatre nouvelles épreuves – deux

destinées au cycle 5 – 8 et deux destinées au cycle 10-12- à un échantillon conséquent – environ 800 élèves de 45 classes de 19 écoles différentes - respectant certains critères et nous avons interrogé, par l'intermédiaire de questionnaires, des élèves, les enseignants et les directeurs. Parallèlement, dans le cadre d'une formation réalisée auprès d'instituteurs et de directeurs à l'Institut Supérieur Pédagogique de Namur, nous avons proposé à ceux-ci de réaliser et de tester des épreuves suivant notre modèle. Les résultats de leur travaux seront analysés et nous permettrons d'émettre un avis sur la faisabilité d'une telle démarche de formation.

Ce rapport est divisé en trois parties distinctes. La première présente le cadre conceptuel sur lequel nous nous sommes reposés pour construire les six épreuves. La seconde partie présente les objectifs de notre deuxième année de recherche et la méthodologie mise en place. Ce sera l'occasion de rappeler succinctement les résultats de la première année qui ont suscités un ensemble de questions auxquelles nous avons tenté d'apporter des éléments de réponse. Dans la troisième partie, nous présentons sur base de l'analyse de résultats statistiques et de résultats qualitatifs des éléments de réponses aux différentes questions suscitées. Enfin, nous terminerons par une conclusion qui tentera de faire une synthèse de nos réflexions et des questions suscitées par ces deux années de recherche.

Calendrier de la recherche pour l'année 2000 - 2001

Septembre 2000	Rencontre avec la commission des outils d'évaluation Présentation du rapport terminal de la première année Définition des objectifs de recherche de la deuxième année Finalisation des quatre épreuves
Octobre 2000	Construction de l'échantillon et prises de contact avec les écoles Finalisation des quatre épreuves avant administration
Novembre 2000	Construction de l'échantillon et prise de contact avec les écoles Rédaction des grilles de correction Construction des questionnaires
Décembre 2000	Construction de l'échantillon et prise de contact avec les écoles Rencontre avec nos partenaires de l'inspection Rédaction des grilles de correction Construction des questionnaires Préparation de l'outil statistique informatisé
Janvier 2001	Passation des épreuves dans les classes Préparation de l'outil statistique informatisé Formation auprès d'instituteurs et directeurs en fonction
Février 2001	Passation des épreuves dans les classes Ramassage des épreuves Correction des épreuves Rédaction du rapport intermédiaire Formation auprès d'instituteurs et directeurs en fonction
Mars 2001	Correction des épreuves Dépouillement des questionnaires Rencontre avec nos partenaires de l'inspection
Avril 2001	Correction des épreuves Dépouillement et analyse des questionnaires Adaptation de l'outil statistique informatisé Encodage des résultats Traitement des résultats
Mai 2001	Traitement des résultats Rencontres avec les écoles participant à la recherche Analyse des épreuves réalisées par des instituteurs et directeurs dans le cadre de la formation
Juin 2001	Traitement des résultats Rencontres avec les écoles participant à la recherche
Juillet 2001	Traitement des résultats Modification de certaines parties des épreuves
Août 2001	Rédaction du rapport terminal

Partie 1 : Le cadre conceptuel

Nous avons élaboré nos épreuves au regard d'un cadre de références construit sur une analyse du concept d'évaluation et du concept de compétence. Ce cadre de références a été présenté dans les rapports précédents, mais nous pensons devoir le rappeler pour permettre aux nouveaux lecteurs de mieux comprendre notre démarche.

L'analyse du contexte de l'enseignement – à la lecture du décret « Missions » - en Communauté française nous a amené à émettre une hypothèse quant à la faisabilité d'une évaluation certificative de compétences telles que présentées dans les socles de compétences. Celle-ci nous a elle-même conduit à prendre position par rapport à la notion d'étalonnage des épreuves. Ainsi, suite à cette analyse, comme nous le développerons ci-dessous, nous avons privilégié des épreuves diagnostiques mises à la disposition des enseignants plutôt que des épreuves certificatives étalonnées.

Après un rappel succinct du contexte de la recherche qui a suscité notre hypothèse de travail, nous développerons le cadre de références qui nous a amené à construire nos épreuves suivant un modèle en trois phases.

1.1. Contexte de la recherche

1.1.1. Introduction

L'enseignement de la Communauté française de Belgique vit depuis quelques années au rythme des réformes. De l'enseignement secondaire « Rénové » à la mise en place des cycles dans l'enseignement fondamental en passant par la création des socles de compétences, ces différentes réformes reposent sur une volonté politique de lutter contre un enseignement sélectif caractéristique d'une tradition scolaire. Le taux de retard important et progressif enregistré dans nos écoles de la première année primaire à la dernière année du secondaire est un élément concret soulignant la propension de notre enseignement à sélectionner (Crahay, 1996). Face à cette tradition, qui d'ailleurs n'a rien à voir avec la qualité des études, contrairement à ce que certains pensent, les instances politiques éducatives ont initié durant ces 30 dernières années des expériences cherchant à transformer cette école de « la sélection » en une école de « la formation ».

Au niveau de l'école fondamentale, les cinq dernières années pourraient être considérées comme celles de la concrétisation d'une réflexion parsemée d'expériences depuis près de trente ans. Structurellement, l'école d'aujourd'hui est invitée à se modifier et à répondre à des exigences nouvelles : travail en équipe, travail en cycles, évaluation formative, travail interdisciplinaire...

En résumé, mais sans trahir l'esprit du décret de juillet 1997 définissant les Missions prioritaires de l'Enseignement fondamental et de l'enseignement Secondaire et organisant les Structures propres à les atteindre, il est demandé aux enseignants de l'école primaire et des deux premières années du secondaire de s'organiser entre eux au sein de cycles pour permettre à chaque élève d'atteindre au terme de ceux-ci un ensemble de compétences définies depuis mai 1999 dans un document s'intitulant « socles de compétences ».

L'introduction de la notion de compétences, comme nous le développerons ci-dessous, entraîne des nouvelles exigences au niveau des pratiques pédagogiques et de l'évaluation. Pour soutenir les enseignants dans cette réforme, le décret de juillet 97 a prévu de proposer aux enseignants des épreuves d'évaluation étalonnées qu'ils pourront utiliser pour les aider à évaluer leurs élèves en référence à ces nouvelles exigences.

C'est dans ce contexte que s'inscrit la présente recherche

Notre première année de recherche nous a conduit à construire un dispositif d'évaluation original que nous proposons de décrire et d'analyser. Les résultats de cette première année nous ont amenés à dégager quelques principes qui peuvent s'articuler dans un modèle. La deuxième année a été consacrée à tester, auprès d'un échantillon construit suivant des critères plus précis, quatre nouvelles épreuves rédigées au regard de notre modèle.

1.1.2. L'organisation et les réformes de l'enseignement fondamental en Communauté française de Belgique

Comme nous venons de l'énoncer brièvement, l'enseignement de la Communauté française est en pleine mutation. Vivant dans une société post-industrielle où le secteur des services est devenu la principale source de travail, l'école a dû abandonner progressivement sa politique méritocratique. Ainsi, aujourd'hui, l'école a comme rôle défini de former chaque et tous les élèves à la maîtrise d'un ensemble de compétences indispensables à son intégration et à son épanouissement dans une société citoyenne. Pour casser la tradition « de sélection » solidement ancrée dans notre culture, les responsables politiques de la Communauté française de Belgique ont décidé, en se référant à des expériences pilotes, de transformer structurellement l'école en imposant la mise en place des cycles.

Depuis septembre 2000, les classes des élèves de 5 ans à 8 ans (troisième maternelle à deuxième primaire) doivent répondre aux critères du cycle. A partir de 2005, ce sont celles des élèves de 8 à 10 ans et de 10 à 12 ans qui devront y répondre. Concrètement, le cycle modifie la structure temporelle traditionnelle calquée sur un cursus que l'on pourrait appeler annuel.

Les critères se résument en trois points : non redoublement au sein du cycle, respect du rythme d'apprentissage de chaque enfant en pratiquant l'évaluation formative et la pédagogie différenciée et obligation pour chaque équipe éducative d'essayer d'amener tous les élèves à maîtriser à la fin du cycle un ensemble de compétences énoncées explicitement dans « les socles de compétences » et qui sont la seule référence quant à l'évaluation. Les équipes pédagogiques doivent s'organiser pour y répondre mais elles gardent une liberté pédagogique et organisationnelle importante.

Elles doivent cependant tenter de s'inscrire dans des orientations pédagogiques qu'une analyse détaillée du décret « missions » met en relief. Ainsi, le terme compétence comme nous le développerons de manière approfondie ci-après, défini comme « aptitude à mettre en œuvre un ensemble organisé de savoirs, de savoir-faire et d'attitudes permettant d'accomplir un certain nombre de tâches » repose sur l'idée que l'apprentissage est réel lorsque l'apprenant est capable d'utiliser des savoirs ou savoir-faire dans des tâches diverses. Sans nul doute, cette perspective n'est pas neutre et est influencée par le courant appelé « constructiviste ». En effet, des auteurs comme Tardif (1992), De Vecchi (1992), soulignent dans les caractéristiques de ce courant l'importance de mettre l'apprenant devant des situations complexes susceptibles de permettre à celui-ci de modifier ou de compléter ses représentations. L'apprentissage de ce fait est actif et constructif.

Les socles de compétences définis comme « référentiel présentant de manière structurée les compétences de base à exercer jusqu'au terme des huit premières années de l'enseignement obligatoire et celles qui sont à maîtriser à la fin de chacune des étapes de celles-ci parce qu'elles sont nécessaires à l'insertion sociale et à la poursuite des études » sont présentés comme le contrat de base entre l'école et la société. Devenu l'unique référence à laquelle les programmes des différents réseaux devront se référer, les socles de compétences inscrits dans une politique de cycles sont l'enjeu véritable de l'école de demain bousculée par les besoins d'une société en évolution constante.

1.2. L'hypothèse de travail

Le décret de juillet 97 a prévu pour soutenir la réforme de proposer aux équipes éducatives des épreuves étalonnées construites en référence aux socles de compétences. Pour ce faire, on a confié cette tâche à une commission « des outils d'évaluation ». Pour aider celle-ci dans la rédaction et l'étalonnage de ces épreuves, des recherches ont été allouées à des universités. C'est dans ce cadre que la présente recherche se situe.

Dès le début de notre recherche, en se référant à la notion de compétences telles que définies dans les socles, nous avons émis l'hypothèse que « les compétences proposées dans les socles sont actuellement peu maîtrisées par les élèves et qu'il faudra un temps certain avant que les enseignants modifient certaines pratiques amenant leurs élèves à une véritable maîtrise de celles-ci ». Cette hypothèse repose sur cette notion de compétence qui devrait se traduire dans la réalité par une volonté de former les élèves à la résolution de tâches leur demandant une intégration de compétences ou de procédures trop souvent isolées.

Cette volonté est à vrai dire une révolution en elle-même, car elle demande à l'école d'arrêter de faire le pari du transfert qui en fin de compte ne permet qu'à certains élèves, en général privilégiés, de réussir. De ce fait, nous pensons que le respect des socles de compétences demande un changement de pratiques pédagogiques et que les socles avant d'être le référentiel de compétences à respecter devraient dans un premier temps être un outil au service du changement.

Cette hypothèse, comme elle semble se vérifier après nos deux années de travail, met en avant des difficultés méthodologiques importantes quant à l'étalonnage des tests. Un étalonnage – test normatif - réalisé sur les compétences actuelles des élèves risque d'entraîner la création de tests éloignés des compétences attendues par le décret. Un étalonnage – test critérié – mesurant l'écart entre les compétences attendues par le décret et les compétences réelles des enfants peuvent, si les épreuves servent à une évaluation uniquement certificative, entraîner un échec massif des élèves.

Face à ce constat et pour éviter cet écueil, nous avons pris le parti, après une analyse du concept « évaluation » et du concept « compétence » de créer des épreuves à vocation diagnostique évaluant des compétences.

1.3. Le cadre conceptuel

Deux notions sont à la base de notre cadre conceptuel : le concept d'évaluation et le concept de compétences.

1.3.1. le concept d'évaluation

1.3.1.1. le cadre des enjeux de l'évaluation

Toute évaluation est dans notre esprit *diagnostique*. Que l'évaluation soit à tendance certificative ou à tendance formative, l'objectif de celle-ci est de permettre d'établir un diagnostic des compétences ou des processus de l'élève dans les différents domaines de l'apprentissage. En d'autres termes, l'évaluation n'a de sens dans notre esprit que si elle apporte des *informations* pertinentes à l'apprenant lui permettant de réguler ses apprentissages. Dans cette perspective, la plupart des évaluations sommatives sont pauvres en informations.

Evaluation diagnostique

Formative	Certificative
------------------	----------------------

curseur

	Privé		Public	
Fonction	dialogue	Faire le point	Dresser un bilan	certifier
Quand ?	Pendant l'enseignement	Avant et après l'apprentissage	A des moments durant le cycle	En fin de cycle
Comment ?	Dialoguer l'évolution de l'apprentissage	Tests	Tests	Tests - étalonnés -
Base de Référence	L'apprenant	L'apprenant Socles de compétences	L'apprenant Socles de compétences	Socles de compétences
Communication	Dialogue particulier	Porte Folio	Bulletin	Certificat Diplôme attestation
Responsabilité	L'enseignant et l'élève	L'équipe d'enseignants	L'équipe d'enseignants	Pouvoir Organisateur
	Axer sur les compétences et les processus (métacognition)		Axer sur les compétences	

Ce tableau définissant « notre cadre de référence » est construit sur deux pôles opposés mais en même temps complémentaires : l'évaluation formative et l'évaluation certificative. Pour bien comprendre notre tableau, nous l'avons construit en imaginant que suivant différents facteurs, l'évaluation serait à *tendance* certificative ou à *tendance* formative. Dans notre esprit, plus on se dirige vers la gauche, plus on est dans une évaluation à tendance formative. Plus on se dirige vers la droite, plus on se situe dans une évaluation à tendance certificative. C'est l'idée de curseur. Le critère différenciant l'évaluation certificative de l'évaluation formative est le côté public ou privé de celle-ci.

1.3.1.2. Description du cadre

L'évaluation certificative

Toute évaluation rendue publique par la voix de documents est à notre sens à tendance certificative. En effet, ces différentes publications sont en général utilisées lorsqu'une décision est à prendre en fin de cycle. De ce fait, nous estimons que le terme « bulletin d'évaluation formative » fortement utilisé dans les communautés éducatives est une contradiction.

L'évaluation la plus certificative : certifier

Formatif		Certificatif	
Privé		Public	
dialogue	Faire le point	Dresser un bilan	certifier

A la fin de chaque cycle, l'équipe d'enseignants du cycle sous la responsabilité de son pouvoir organisateur prend une décision. Ils certifient publiquement que l'élève maîtrise ou ne maîtrise pas les compétences attendues en fin de cycles et définies dans les socles de compétences. Cette certification doit reposer sur un bilan diagnostic des compétences attendues car l'équipe d'enseignants doit pouvoir justifier son choix. Pour ce faire, le gouvernement de la Communauté française *proposera* d'ici juin 2003 des épreuves étalonnées qui devaient aider les équipes d'enseignants à prendre leurs décisions.

De ce fait, à la fin de chaque cycle, des épreuves devraient être organisées pour évaluer l'acquisition ou non des compétences.

L'évaluation un peu moins certificative : dresser un bilan

Formatif		Certificatif	
Privé		Public	
dialogue	Faire le point	Dresser un bilan	certifier

Durant le cycle, l'équipe d'enseignants réalise à certains moments du cycle des « bilans » tentant de déterminer l'état de « santé » de chaque élève. Ces bilans ont deux objectifs. Le premier est de pouvoir communiquer aux parents un ensemble d'informations sur les difficultés rencontrées par leur enfant. Le second est de pouvoir récolter des informations objectives et publiques, durant le cycle, qui pourront être utilisées à la fin de celui-ci lorsqu'une décision devra être prise. En effet, il serait sans nul doute extrêmement dangereux de prendre une décision en se référant uniquement à l'épreuve de fin de cycle. Ces bilans sont « des arrêts sur image ».

L'évaluation formative

Une évaluation est formative lorsqu'elle est privée. Cette prise de position repose sur l'idée en quelque sorte philosophique que l'apprentissage concerne l'apprenant et l'enseignant et que pour que celui-ci soit efficace, il est nécessaire de construire un climat de confiance personnalisé. Trop souvent, les enseignants ont tendance à « publier » tous les « faits et gestes des élèves » ce qui va à l'encontre d'une vision de l'apprentissage « constructiviste » où le rôle de l'erreur est essentiel. Dans ces circonstances, l'enfant va à l'école pour réussir

et non pour apprendre. Développer sa motivation intrinsèque est une priorité. Un système de publications de résultats intensifs renforce à nos yeux la motivation extrinsèque.

L'évaluation la plus formative

Formatif		Certificatif	
Privé		Public	
dialogue	Faire le point	Dresser un bilan	certifier

L'évaluation la plus formative est *le dialogue particulier* qu'entretient l'enseignant avec l'élève. Durant l'apprentissage, l'enseignant doit amener chaque élève à dialoguer. Ce dialogue *privé* individuel est indispensable à la prise de conscience par l'apprenant de ses propres processus d'apprentissage au regard de ses « performances ». Ce dialogue particulier que de nombreux enseignants utilise régulièrement et intuitivement doit permettre à l'enseignant de mieux comprendre les processus d'apprentissage des élèves et d'adapter ses interventions didactiques en conséquence.

La gestion mentale proposée par « La Garanderie » et la P.N.L. (Programmation neuro-linguistique) fort appréciées par de nombreux enseignants rejoignent à nos yeux cette préoccupation. Même si les fondements théoriques de ces approches sont fortement critiquées par les universitaires, elles permettent néanmoins aux enseignants de s'intéresser aux élèves en référence à certains critères, discutables, mais précis.

La métacognition peut nous apporter certains outils pour l'amélioration de ce dialogue particulier.

Selon le dictionnaire *Robert*, « méta » intervient dans les néologismes scientifiques (métalangue, métamathématique...) et signifie « *ce qui dépasse, ce qui englobe ... un objet de pensées, une science...* ».

De manière plus particulière, selon Gombert, on utilise le terme « métacognition » pour désigner :

« *le domaine qui regroupe :*

les connaissances introspectives conscientes qu'un individu a de ses propres états et processus cognitifs ;

les capacités que l'individu a de délibérément contrôler et planifier ses propres processus cognitifs en vue de la réalisation d'un but ou d'un objectif déterminé. »

A cet égard, les auteurs s'accordent à reconnaître l'existence d' « opérateurs métacognitifs de base » susceptibles de donner lieu à des stratégies métacognitives plus ou moins complexes. L'utilisation de ces opérateurs par l'enseignant peut, d'après nous, améliorer le dialogue entre lui et l'apprenant. Ce sont :

L'explicitation par l'élève de ses propres processus cognitifs . Il s'agit de la capacité à identifier, décrire, expliciter les démarches cognitives (et leurs produits) utilisées pour accomplir une tâche. Concrètement, l'enseignant fait parler ses élèves, individuellement ou collectivement, sur des tâches accomplies ou à accomplir.

L'anticipation des résultats attendus ou des stratégies à mettre en œuvre pour y parvenir.

La décentration qui consiste à se placer dans la perspective d'autrui. Elle doit permettre de faire apparaître à l'élève qu'il existe pour une même tâche des fonctionnements cognitifs différents d'individu à individu.

L'auto-évaluation effectuée par l'individu lui-même et portant sur ses propres comportements cognitifs et sur les facteurs qui les influencent. Cette auto-

évaluation n'a du sens qu'à partir du moment où l'élève a construit avec l'enseignant des critères lui permettant cette auto-évaluation.

Lors des apprentissages, demander aux élèves d'explicitier, d'anticiper, de se décentrer ou de s'auto-évaluer devrait leur permettre d'améliorer leurs connaissances métacognitives et de favoriser à long terme leurs apprentissages.

La métacognition inscrite dans le cadre de l'évaluation (la plus) formative montre l'importance de ce dialogue pédagogique. Celui-ci doit être privé, mais il peut être soit individuel, soit collectif. En effet, un dialogue au sein de la classe reposant sur les opérateurs décrits ci-dessus rejoint d'après nous la perspective d'un enseignement auto-socio-constructiviste « interactif ».

L'évaluation « un peu moins » formative : faire le point

Formatif		Certificatif	
Privé		Public	
dialogue	Faire le point	Dresser un bilan	certifier

Dialoguer avec l'apprenant pour tenter de développer ses capacités métacognitives et essayer d'améliorer à long terme ses « performances », tel est l'enjeu de l'évaluation formative. Cependant, la réalité quotidienne de la classe demande une certaine objectivation. Il est en effet peu réaliste d'envisager un dialogue quotidien et individuel avec tous les élèves et il est nécessaire pour réguler les apprentissages en classe de se construire des outils orientés vers une pratique de classe qui reste un lieu d'apprentissage collectif. Faire le point sur l'apprentissage des élèves en tentant d'objectiver les difficultés ou les facilités des élèves est, dans ce sens, important et pour y arriver la construction d'outils d'évaluation considérés comme supports à l'apprentissage est utile.

Il est cependant indispensable de rester dans le caractère privé et de ne pas publier systématiquement les résultats obtenus lors de ces évaluations.

Ces outils d'évaluation doivent être au service des apprentissages et doivent permettre à l'enseignant et à chaque élève d'apporter des informations utiles permettant une régulation de ceux-ci. Ces informations recueillies peuvent être utiles pour dialoguer avec l'élève (évaluation la plus formative), mais aussi pour permettre à l'enseignant de construire la suite de ses séquences didactiques. Diagnostiquer est l'objectif de ces outils.

1.3.1.3. Construisons des épreuves diagnostiques

Dans notre cadre de références, que l'on soit dans une orientation certificative ou formative, il est indispensable que l'évaluation soit diagnostique.

Evaluer les performances des élèves de 8 ans et de 12 ans dans la résolution de situations complexes est l'objectif des épreuves que nous avons construites. Cependant, l'hypothèse que peu d'élèves étaient, à vrai dire, capable d'y arriver nous a conduit à construire des épreuves tentant d'analyser qualitativement et quantitativement les performances des élèves invités à résoudre des tâches complexes.

L'idée générale, en référence à notre cadre de référence, est que ces épreuves peuvent dans l'avenir être utilisées à plusieurs fins. Elles peuvent, en effet, être utilisées par les enseignants pour certifier leurs élèves en fin de cycle avec la difficulté qu'aucun seuil de maîtrise n'a été actuellement été déterminé, ou elles peuvent être aussi utilisées de manière plus formative. De fait, un enseignant dans sa classe sur base de nos épreuves peut

réaliser un diagnostic intéressant quant aux compétences de chacun de ses élèves mis devant des situations complexes interdisciplinaires.

1.3.2. Le concept de compétence

1.3.2.1. L'élément essentiel de la notion de compétence : l'accomplissement efficace d'une tâche

La compétence est toujours conçue comme la disposition que possède un individu humain à accomplir une tâche déterminée. Ainsi cette référence à une tâche est au centre de la définition qu'en donne le décret « missions ».

On retrouve cette idée dans tous les textes utilisant la notion de compétence, même quand le mot "tâche" est absent. Ainsi Philippe Perrenoud (1997) la définit comme "une capacité d'agir efficacement dans un type défini de situation". Il est clair que l'expression "agir efficacement" renvoie ici à l'accomplissement d'une tâche, c'est-à-dire d'une modification par l'individu de son environnement (au sens large de l'environnement matériel ou conceptuel).

Or parler de tâche, ce n'est pas parler de *comportement* ni d'*opération*. La notion de comportement, notamment, comme d'ailleurs celle d'opération (chez Piaget), relève de la psychologie.

La notion de tâche, au contraire, est une notion du sens commun. Une tâche a un sens et une fonction au sein des activités socialement reconnues d'une culture donnée. Elle a une utilité dans le domaine technique, dans celui de la production et des échanges ou dans un des domaines : domestique, politique, social, artistique, scolaire, sportif, etc.

Une tâche a donc une finalité. C'est cette finalité qui constitue son unité et du même coup celle de la compétence. Car sur le plan **psychologique et physiologique**, une compétence est foncièrement hétérogène : elle peut bien mettre en jeu des connaissances déclaratives, des connaissances procédurales, des automatismes, des raisonnements, des données retenues en mémoire de travail, des schèmes sensori-moteurs, des savoirs, des savoir-faire, des attitudes ou n'importe quel assemblage de ces éléments. Elle est hétérogène par ses constituants et homogène par sa finalité technico-sociale ou scolaire.

1.3.2.2. Deux intérêts de la définition de la compétence comme aptitude à accomplir une tâche

La lecture de la littérature fait apparaître que cette définition de la compétence par référence à une ou plusieurs tâches présente deux intérêts qui la légitiment :

La définition de la compétence par référence à une tâche évite la décomposition et la perte de sens.

Notons d'abord que les compétences peuvent se définir à différents niveaux.

On peut parler de compétence en rapport avec un métier : la compétence de l'horticulteur, la compétence du médecin. Il s'agit alors de la disposition à accomplir une tâche entendue comme la globalité d'une profession.

Mais une telle macro-compétence peut se détailler en compétences de niveaux inférieurs (par exemple pour le médecin : établir un diagnostic ou prescrire), lesquelles à leur tour peuvent se décomposer en micro-compétences (prendre le pouls, exécuter une palpation du foie).

Il y donc un emboîtement des compétences et les compétences globales se laissent décomposer en compétences qui renvoient à des tâches partielles.

Mais cette décomposition a une limite : il faut que la tâche reste une tâche, c'est-à-dire une unité fonctionnelle renvoyant à une finalité.

Si on décompose au-delà, on n'a plus que des "comportements", c'est-à-dire ce que le béhaviorisme, dans sa recherche d'une objectivité scientifique maximale, définissait comme "toute réaction musculaire ou glandulaire de l'organisme étudié". La finalité est alors évacuée.

On trouve la même difficulté du côté de la psychologie cognitive, dans une optique mentaliste cette fois, lorsqu'elle décompose une activité en opérations mentales dont certaines sont en dessous du seuil de conscience. Une telle analyse est évidemment tout à fait intéressante sur le plan de la compréhension scientifique des mécanismes mentaux engagés par l'exécution d'une tâche. Mais il y aurait un risque à vouloir entraîner des apprenants à des opérations qui ne seraient que les segments d'une tâche et dont aucune, prise en elle-même, ne renverrait à une finalité. Car si précisément ce sont des opérations élémentaires, le lien qu'elles ont avec la globalité de la tâche dans laquelle elles s'insèrent peut être très obscur aux yeux du novice.

Par suite, l'apprenant peut éventuellement douter de la nécessité d'exercices parcellaires dont il ne voit pas la signification et se demander si le formateur ne l'égaré pas dans des efforts inutiles. Va-t-il donner du sens à une activité qui, à elle seule, ne renvoie à aucune fonctionnalité ? Il peut décider finalement de passer outre à l'injonction ou de n'investir dans cet exercice là que le minimum de temps et d'énergie.

L'intérêt de la notion de compétence, en tant que disposition à une tâche ayant un sens dans l'univers humain, c'est qu'elle est un garde-fou contre de telles dérives.

La définition de la compétence comme disposition à accomplir une tâche conduit à la mise en activité des apprenants

Une tâche, en effet, est une disposition à un "*faire*" et non pas à des connaissances, des représentations, des états physiques ou mentaux. Bien entendu, cette action peut être aussi bien une action intellectuelle qu'une action physique. Dans ce dernier cas, elle portera sur des objets matériels (par exemple : cintrer un tube de cuivre), dans le premier cas sur des objets idéels ou, si l'on préfère, symboliques (par exemple : effectuer une soustraction, rédiger un texte). En fait toutes les tâches sont à la fois intellectuelles et physiques. Ce qui importe ici, c'est que ce sont des actions, c'est-à-dire qu'elles ont toujours une finalité.

Qu'une compétence soit une disposition à un "*faire*", voilà qui permet d'abord d'instituer une saine hygiène didactique. Son usage doit inciter les enseignants à réfléchir à ce qu'ils veulent que les élèves sachent faire à l'issue d'une leçon ou d'un cours.

Mais, en plus, la rédaction en termes de compétences de ce qui doit être visé par l'action pédagogique de l'enseignant, devrait inciter celui-ci à multiplier les situations pédagogiques où les élèves doivent être en activité (cette activité pouvant être aussi bien intellectuelle que manuelle).

Surtout, l'introduction de la notion de compétence, en tant qu'elle est une capacité à accomplir une tâche, permet d'introduire une vision renouvelée du savoir. On peut soutenir en effet qu'un véritable savoir n'est pas la simple mémorisation d'un ensemble d'énoncés, mais la capacité de l'élève de donner du sens à cet ensemble. Or une des principales voies pour que l'élève donne sens au savoir, c'est qu'il soit capable d'en faire usage (Rey, 1998). Il peut en faire un usage pragmatique, en dehors de l'école ou dans des activités fonctionnelles. Mais il peut aussi en faire un usage au sein même du savoir pour résoudre des problèmes. En ce sens tout vrai savoir est une compétence.

Si l'on veut voir dans la connaissance une véritable activité intellectuelle de l'élève et non pas seulement un fait de psittacisme, il faut admettre que toute connaissance comporte en elle-même des compétences qui lui sont inhérentes.

Cet aspect des connaissances, très visible dans le domaine de l'école, se retrouve dans le champ du savoir scientifique en train de se faire. Car celui-ci se développe progressivement en réponse aux difficultés et aux problèmes qui surgissent en son sein. Par suite les concepts y sont élaborés en vue de résoudre ces problèmes. C'est particulièrement remarquable en mathématiques, (cf. par exemple les nombres irrationnels). Mais c'est le cas aussi des autres sciences : en biologie, par exemple, même si des cellules végétales sont observées dès le 17^{ème} siècle, la notion de cellule n'apparaît véritablement comme modèle explicatif qu'au 19^{ème} siècle pour répondre à un certain nombre de questions qu'on se pose sur la fécondation et la reproduction. Un concept scientifique ne se définit pas seulement par les relations logiques qu'il entretient avec les autres concepts de la théorie, il est aussi un outil intellectuel servant à accomplir des tâches explicatives que se donne la science. En ce sens, comprendre ce concept, c'est détenir une compétence.

C'est d'ailleurs par référence à cette instrumentalité des concepts scientifiques que se développe la didactique constructiviste et même plus précisément la didactique de la "situation-problème". Elle consiste à proposer aux élèves un problème qu'ils ne sont pas en état de résoudre au moyen de leurs connaissances actuelles, mais pour lequel ils ont à inventer une notion nouvelle, celle-là même que l'enseignant veut leur faire acquérir à ce moment-là. Notons qu'une telle démarche est très ambitieuse puisqu'elle consiste à faire retrouver le savoir par les élèves eux-mêmes et qu'elle n'y réussit pas toujours. Mais même quand les élèves n'ont pas été capables de reconstruire seuls le savoir attendu, celui-ci, au moment où il est délivré par l'enseignant, leur apparaît d'emblée comme instrument pour résoudre le problème auquel ils se sont affrontés sans succès : la connaissance nouvelle est immédiatement compétence. En ce sens l'introduction des socles de compétences est une très forte incitation, si elle est bien comprise, à l'introduction d'une pédagogie active. Mais elle n'élimine en aucune manière l'effort pour faire accéder les élèves à des savoirs. Elle lui est, au contraire, consubstantielle.

1.3.2.3. Trois types de compétences

Cette définition de la compétence par référence à une ou plusieurs tâches à accomplir incite à distinguer plusieurs types de compétences en fonction d'une part du caractère simple ou complexe de la tâche et d'autre part de la plus ou moins grande familiarité de la tâche pour l'individu.

Entraîner des élèves à exécuter une opération (ou une suite d'opérations) en réponse à une question donnée ne pose pas de problème particulier. Il est ainsi possible d'obtenir d'élèves de l'enseignement fondamental qu'ils exécutent les opérations nécessaires pour effectuer une soustraction, les opérations nécessaires pour accorder le verbe au sujet, celles nécessaires pour mettre à la forme interrogative une phrase affirmative, etc.

Pour rendre compte du processus d'apprentissage qui permet d'acquérir de telles "compétences", le béhaviorisme propose un modèle explicatif : c'est le renforcement qui lie l'exécution d'une réponse à l'apparition d'un stimulus. Il faut cependant noter que la plupart des compétences ainsi apprises à l'école consistent en des opérations *mentales* et ne se traduisent en comportements observables que d'une manière très indirecte. Malgré cette réserve, l'apprentissage d'une compétence est bien le processus qui lie une opération ou un ensemble d'opérations à une situation ou une famille de situations. Ainsi, en première analyse, la compétence n'est rien d'autre que ce lien.

Toutefois, cette conception de ce qu'est une compétence peut paraître un peu trop restrictive. Quand, dans le langage courant, on dit que quelqu'un est compétent dans un domaine, on veut dire beaucoup plus et notamment :

- qu'il possède toute une gamme de ces compétences élémentaires que nous venons d'indiquer ;
- qu'il est capable de **choisir**, dans ce répertoire de procédures automatisées, celle qui convient à un problème ou une situation.

Mais on peut faire encore un pas de plus et considérer qu'une compétence digne de ce nom consiste, en réponse à une tâche inédite et complexe, à **choisir** et **combiner plusieurs** des procédures élémentaires ci-dessus évoquées. En ce sens une authentique compétence est la capacité à répondre à des situations complexes et inédites par une combinaison nouvelle de procédures connues ; et non pas seulement à répondre par une procédure stéréotypée à un signal préétabli.

On peut donc, au total, distinguer trois types de compétences, tout en restant au sein de la définition de la compétence que propose le décret « missions » :

1. Savoir exécuter une opération (ou une suite prédéterminée d'opérations) en réponse à un signal (qui peut être, en classe, une question, une consigne, ou une situation connue et identifiable sans difficulté ni ambiguïté) ; nous parlerons alors de "**procédure**" ou encore de "**compétence élémentaire**". Dans la littérature nord-américaine on parle souvent à ce propos d'habiletés.

2. Posséder toute une gamme de ces compétences élémentaires et savoir, dans une situation inédite, choisir celle qui convient ; là une interprétation de la situation (ou comme nous le dirons plus bas un "cadrage" de la situation) est nécessaire ; nous parlerons donc de "**compétence élémentaire avec cadrage**".

3. Savoir choisir et combiner correctement plusieurs compétences élémentaires pour traiter une situation nouvelle et complexe. Nous parlerons alors de "**compétence complexe**".

1.3.3. Une épreuve à 3 phases

Comme nous l'avons déjà énoncé, il s'agit de construire des épreuves évaluant les compétences telles qu'on les trouve définies dans les socles. D'après l'analyse ci-dessus, il faut entendre par compétences en premier lieu des compétences complexes : l'enjeu principal d'une telle évaluation est de contrôler si les élèves sont capables, face à une tâche complexe, de choisir, de combiner, voire d'adapter les procédures élémentaires qu'ils possèdent.

Mais si l'on veut que l'évaluation soit un instrument de maîtrise des processus d'apprentissage tant pour les maîtres que pour les élèves, il convient qu'elle ait un caractère diagnostique. Pour cette raison, on ne peut se contenter d'un repérage dichotomique de la réussite ou de l'échec à une tâche complexe. Il faut que les enseignants aient un instrument qui leur permette de saisir, autant que faire se peut, ce qui a pu entraîner les difficultés éventuelles rencontrées par leurs élèves. Une des voies possibles dans ce but est de voir si les élèves sont capables, face à une tâche simple mais nouvelle pour eux, de mobiliser la procédure connue d'eux qui convient à cette tâche. Enfin il faut également vérifier si les élèves savent accomplir les procédures simples, auxquelles ils ont été entraînés en classe, lorsqu'on le leur demande dans les conditions semblables à celles dans lesquelles ils les ont apprises.

Par suite de ces trois exigences, nous proposons que les épreuves d'évaluation en rapport avec les socles de compétences comprennent systématiquement trois phases.

- **Phase 1** : On demande aux élèves d'accomplir une tâche complexe, exigeant le choix et la combinaison d'un nombre significatif de procédures qu'ils sont censés posséder à la fin d'un cycle (deuxième et sixième primaires). Cette tâche renvoie à une situation fonctionnelle, ce qui correspond à l'esprit du décret et des socles de compétences qui appellent à des situations d'apprentissage porteuses de sens pour les élèves. Cette tâche est en outre systématiquement pluridisciplinaire.

- **Phase 2** : On propose à nouveau aux élèves la même tâche. Mais cette fois, la tâche complexe est découpée en tâches élémentaires dont les consignes sont explicites et qui sont présentées dans l'ordre où elles doivent être accomplies pour parvenir à la réalisation de la tâche complexe globale. Mais il appartient à l'élève, pour chacune de ces tâches élémentaires, de déterminer la procédure à mettre en œuvre parmi celles qu'il est censé posséder.

- **Phase 3** : On propose aux élèves une série de tâches simples décontextualisées, mais dont les consignes sont celles qui sont utilisées ordinairement dans l'apprentissage des procédures élémentaires qu'on propose à l'école : effectuer une soustraction ; écrire des mots ; accorder un verbe avec un sujet, etc.

Comme il est facile de le voir, ces trois phases renvoient aux trois degrés de complexité des compétences que nous avons présentés ci-dessus :

La phase 1 évalue la compétence au sens plein, c'est-à-dire l'aptitude à saisir dans une situation nouvelle et complexe les traits pertinents qui indiquent qu'il y a lieu de choisir et de combiner d'une manière originale des procédures connues. Il s'agit donc d'une "compétence complexe", au sens que nous avons donné à cette expression. En d'autres termes, il s'agit d'une résolution de problème.

La phase 2 évalue l'aptitude à choisir la procédure qui correspond à une tâche simple mais inédite ; il s'agit là de ce que nous avons appelé "compétence élémentaire avec cadrage de la situation".

Enfin, la phase 3 permet d'évaluer des procédures de base ou compétences élémentaires et leur degré d'automatisation chez les élèves.

L'épreuve ainsi conçue doit permettre d'opérer une évaluation diagnostique, respectant les spécificités de la notion de compétence.

1.4. Conclusions

Proposer des épreuves en trois phases est l'originalité de notre modèle. Ces épreuves destinées aux enseignants à titre informatif ont une volonté diagnostique. Reposant sur l'hypothèse qu'actuellement peu d'élèves sont capables de résoudre des situations complexes demandant la recherche de démarches, nous avons décidé de pas étalonner nos épreuves. Comme nous l'avons déjà énoncé : un étalonnage normatif risquerait au regard de notre analyse du concept de compétences tel que proposé dans le décret « Missions » de nous amener à corriger les épreuves, en conséquence celles-ci, n'évalueraient plus réellement des compétences et un étalonnage critérié pourrait entraîner un échec massif des élèves.

Sans trahir l'esprit du décret soulignant explicitement le caractère informatif des épreuves devant être proposées par la commission des outils d'évaluation, nous avons opté pour des épreuves diagnostiques en trois phases. Selon nous, l'enseignant proposant ce type d'épreuves à ses élèves pourra déterminer, pour les matières entrant en ligne de compte dans l'épreuve proposée, d'une manière plus précise les difficultés ou non de ses élèves.

Ainsi, par exemple, il risque de rencontrer des élèves éprouvant des difficultés lors de la phase 1, mais n'ayant pas de réelles difficultés en phase 2 et en phase 3. D'autres constats sont naturellement possibles.

La suite de ce rapport décrit la mise à l'épreuve de ce modèle et les constats que nous avons pu retirer de notre expérimentation.

Deuxième partie : Objectifs, méthodologie et questionnement

Comme nous le développerons dans le point 1 de cette deuxième partie, le modèle en trois phases tel que décrit dans la partie précédente a été une première fois mis à l'épreuve lors de la première année de recherche. Deux épreuves ont été créées – une destinée au cycle 5-8 et l'autre au cycle 10-12- et pré-testées auprès d'un échantillon occasionnel d'environ 200 élèves par épreuve.

Les résultats de ce pré-test ont permis certains constats confirmant dans une certaine mesure notre hypothèse de travail et soulignant la cohérence et la pertinence de notre modèle. Ces résultats et constats interprétés avec toute la prudence nécessaire ont suscité un ensemble de questions. Pour tenter d'y répondre et pour observer et comparer des résultats issus d'autres épreuves administrées à d'autres élèves, nous avons décidé, avec l'accord de notre comité d'accompagnement, de proposer quatre nouvelles épreuves à un échantillon construit suivant différents critères. Ce sont les résultats de ces épreuves, au regard des résultats obtenus la première année, qui devraient apporter des éléments de réponses à nos questions.

2.1. La méthodologie et les résultats de la première année

2.1.1. La construction des épreuves

Sur le modèle décrit ci-avant, nous avons construit une épreuve destinée à la fin de la deuxième primaire et une autre destinée à la fin de la sixième primaire.

En premier lieu, nous avons choisi une tâche complexe. (phase 1)

Ainsi nous avons imaginé de demander aux élèves de sixième année, sur base d'un ensemble de documents rassemblés dans un portefeuille de lecture, d'organiser une excursion à Bruxelles pour une classe fictive. Trois sous-tâches doivent être réalisées : le choix de la formule la moins chère, l'organisation de l'horaire de la journée et la rédaction d'une lettre destinée à la réservation d'une visite guidée dans un musée.

Pour les deuxièmes années, nous avons imaginé de leur demander d'organiser un spectacle à l'école. Deux sous-tâches doivent être réalisées : calculer le nombre de chaises nécessaires pour les personnes venant assister au spectacle et réaliser une affiche d'invitation.

Pour chacune des sous-tâches, suite à un premier prétest auprès d'une classe, nous avons créé des échelles d'évaluation déterminant de manière quantitative la performance des élèves. De ce fait, les élèves de sixième année peuvent avoir un résultat compris entre 0 et 6 et ceux de troisième année, un résultat entre 0 et 8.

En parallèle, pour déterminer les compétences mises en jeu au regard des socles de compétences, les chercheurs ont analysé les deux épreuves en les décomposant.

Ensuite, nous avons repris les différentes épreuves et sous-tâches et nous les avons décomposées en items (Phase 2). Pour résoudre les tâches, les élèves ne doivent plus choisir la démarche mais résoudre les items les amenant à la résolution de la tâche.

Pour la correction, chaque item représente 1 point. Trois types d'items sont apparus : des items faisant appel uniquement à la lecture des documents, des items faisant appel à des opérations et des items faisant appel à de la lecture et à des opérations. Pour ne pas pénaliser les élèves se trompant dès le début de l'épreuve, nous avons construit un système de correction s'adaptant aux réponses erronées des élèves.

Enfin, nous avons construit des épreuves faisant appel à des procédures normalement automatisées en français et en calcul. Les différents items choisis, suite à l'analyse des chercheurs, correspondent à des procédures devant être maîtrisées par les élèves pour réussir les phases 2 et 1

2.1.2. Passation des épreuves

Dans un souci de récolter aussi bien des informations sur la pertinence de notre modèle d'épreuves (à trois phases) que sur les épreuves particulières et dans l'objectif de construire et tester un outil statistique susceptible d'être généralisé, nous avons administré ces épreuves auprès de 193 élèves du cycle 5 –8 (11 classes) et de 214 du cycle 10-12 (10 classes) ce qui représente un nombre important d'élèves dans une phase de prétest. Nous avons essayé de sélectionner des classes offrant des critères sociaux et géographiques différents. Mais le caractère occasionnel de l'échantillon nous oblige, cependant, à être prudent dans l'interprétation des résultats.

Pour des raisons d'organisation du calendrier, nous n'avons pu effectuer cette opération qu'en janvier 2000 et comme, à cette période de l'année, les élèves de deuxième sont loin d'avoir atteint les compétences qu'ils maîtrisent six mois plus tard, il nous a semblé plus judicieux de tester l'épreuve qui leur était destinée auprès d'élèves de troisième.

Toutes les phases 1 ont été administrées par les chercheurs durant le mois de janvier auprès des classes des écoles volontaires suivant une procédure commune et à des moments de la journée précis. Cet investissement nous semblait essentiel pour éviter des interventions d'enseignants durant la passation. Les phases 2 et 3 ont été administrées par les instituteurs dans la semaine qui ont suivi la phase 1.

Pour éviter le biais important d'une correction réalisée par les instituteurs et après un entraînement sur une trentaine de copies, les chercheurs se sont attelés à la correction de l'ensemble des épreuves.

2.1.3. Les résultats et constats

Nous nous permettons d'insister sur le fait que le caractère occasionnel de l'échantillon de notre prétest nous oblige à être prudent dans l'interprétation des résultats. Cependant, un certain nombre d'informations importantes émergent des données présentées ci-dessous.

Tableau 1

		Cycle 5 -8		Cycle 10-12	
		Echantillon 193 élèves		Echantillon 214 élèves	
Phase 1		Résultats	Ecart-type	Résultats	Ecart-type
	Résultats moyens obtenus pour les tâches	6.31/16 <u>39.5 %</u>	4.19	6.36/ 18 <u>35.3 %</u>	3.69
Phase 2	Résultats moyens obtenus pour les tâches	8.71 / 16 <u>54.43 %</u>	4.2	16.84/ 30 <u>56.1 %</u>	7.02
Phase 3	Résultats moyens obtenus pour les tâches	<u>78.13 %</u>	16.87	53.85/ 80 <u>67.3 %</u>	11.19
	Résultats français	Lecture : 85.3 % Expr. écrite :44.39%	0.39 25.09	29.33 / 40 73.3 %	7.7
	Résultats Math.	81.78 %	18.57	12.26 /20 61.3 %	4.7

2.1.3.1. La hiérarchie entre les phases des épreuves

La lecture du tableau 1 souligne **une hiérarchie entre les trois phases**. Ainsi, les résultats moyens en phase 1 sont très faibles : 39.5 % pour les élèves du cycle 5 – 8 et 35.3 % pour ceux du 10-12.

Les résultats en phase 2 sont meilleurs qu'en phase 1 : 54.43 % pour le cycle 5 – 8 et 56.1 % pour le cycle 10-12..

Enfin, les résultats en phase 3 sont de loin les meilleurs avec une moyenne de 78.13 % pour les élèves du cycle 5 –8 et de 67.3 % pour ceux du cycle 10 – 12.

Ces résultats, même si nous devons rester prudents vu les modalités de correction différentes entre les phases, semblent indiquer que les élèves, devant des tâches complexes nécessitant la création d'une démarche, éprouvent plus de difficultés que lorsque ces tâches sont décomposées. De même, les résultats semblent indiquer clairement que les élèves sont de manière très significative plus performant dans la résolution de procédures isolées et normalement automatisées que lorsqu'ils sont amenés à choisir la procédure dans une situation inédite.

Bien entendu, les résultats obtenus en phase 1 pourraient être interprétés comme signifiant que nos épreuves sont "trop difficiles". Mais en réalité il convient de distinguer entre plusieurs causes de cette difficulté :

Certaines tâches partielles de notre épreuve destinée au cycle 10-12 ont été mal exécutées, à la fois dans la phase 1 (où elles se trouvaient intriquées dans un problème complexe), et dans la phase 2 (où elles étaient présentées une par une aux élèves). On peut considérer qu'il s'agit alors de tâches hors de portée des élèves de l'échantillon.

En revanche, certaines tâches ne sont exécutées que par une minorité d'élèves dans le cadre de la phase 1, alors qu'elles sont beaucoup plus largement réussies en phase 2. Il semble alors qu'une majorité d'élèves éprouvent des difficultés plus grandes devant un problème complexe que lorsqu'il est décomposé en tâches partielles. C'est donc, dans ce cas, l'affrontement à la complexité ou, en d'autres termes, la situation de résolution de

problème qui constitue la difficulté pour beaucoup d'élèves. Il en va de même de la différence de réussite entre la phase 2 et la phase 3 qui fait apparaître que, si beaucoup d'élèves maîtrisent des procédures élémentaires, ils ont beaucoup plus de difficulté à repérer dans une situation nouvelle l'opportunité de les mettre en œuvre. L'analyse des écarts-types (indice de dispersion autour de la moyenne) le confirme : il augmente de la phase 3 à la phase 2 et de celle-ci à la phase 1. Cela signifie que la résolution de problème creuse les écarts entre les élèves.

Cette difficulté peut-elle être jugée exagérée ? Il faut s'entendre sur la norme par rapport à laquelle on juge "facile" ou "difficile" une épreuve. Construire des épreuves d'évaluation qui tiennent compte des socles de compétences pour la fin de deuxième et de sixième primaires, c'est trouver un outil qui permette d'évaluer la performance de chaque élève par rapport aux exigences formulées dans les socles. C'est dans cette optique que nos épreuves ont été construites.

Tableau 2 : Fréquences cumulées cycle 10 - 12

Pourcentage d'élèves ayant obtenu au moins	10 - 12	10 - 12	10 - 12	Total
	Phase 1	Phase 2	Phase 3	Total
90 - 100	0	0,46	8,41	0,46
80 - 90	1,86	7,47	32,24	4,67
70 - 80	6,54	19,15	57,47	15,8
60 - 70	15,88	40,18	68,69	31,3
50 - 60	29,9	55,14	81,3	55,14
40 - 50	35,98	69,15	90,18	73,83
30 - 40	48,59	80,84	93,92	85,98
20 - 30	71,96	91,12	98,13	97,19
10 - 20	95,32	96,72	99,06	99,53
0 - 10	100	100	100	100

Tableau 3 : Fréquence cumulée cycle 5 - 8

Pourcentage d'élèves ayant obtenu au moins	5 - 8	5 - 8	5 - 8	Total
	Phase 1	Phase 2	Phase 3	Total
90 - 100	6,21	16,06	28,49	6,73
80 - 90	10,88	23,83	58,03	15,54
70 - 80	12,95	29,01	74,61	28,49
60 - 70	21,24	45,59	87,56	44,55
50 - 60	36,26	61,65	92,74	64,76
40 - 50	42,48	66,32	95,85	79,79
30 - 40	61,65	79,79	98,9	88,60
20 - 30	71,5	82,38	98,96	95,33
10 - 20	87,04	91,19	99,48	100
0 - 10	100	100	100	100

2.1.3.2. La confirmation de notre hypothèse

Les résultats repris dans les tableaux 2 et 3 font apparaître la difficulté de beaucoup d'élèves à cadrer des situations complexes et à choisir, pour ces situations originales, les procédures adéquates. Ainsi, seulement 10.88 % des élèves du cycle 5-8 et 1.86 % des élèves du cycle 10-12 obtiennent un résultat égal ou supérieur à 80 % à la phase 1. Ils sont 36.26 % pour l'épreuve 5 - 8 et 29.9 % pour l'épreuve 10-12 à avoir obtenu un résultat égal ou supérieur à 50 % en phase 1. Ces résultats n'ont rien de surprenant. Il confirme notre hypothèse de départ à savoir que les compétences de résolution de problème qu'exigent les *Socles de compétences* sont actuellement peu maîtrisées par les élèves et qu'il faudra un temps certain avant que les enseignants modifient certaines pratiques de façon à amener leurs élèves à une véritable maîtrise de celles-ci.

Ce constat renforce l'intérêt d'une épreuve diagnostique et pose la question de la pertinence d'un étalonnage, qu'il soit normatif ou critérié.

2.1.3.3. La cohérence de l'épreuve et l'importance des procédures automatisées.

Tableau 4 : Corrélations cycle 10 – 12

	Phase 1	Phase 2	Phase 3
Phase 1		0.59	0.42
Phase 2	0.59		0.48
Phase 3	0.42	0.48	

Tableau 5 : Corrélations cycle 5 – 8

	Phase 1	Phase 2	Phase 3
Phase 1		0.57	0.58
Phase 2	0.57		0.61
Phase 3	0.58	0.61	

Pour examiner le lien et la dépendance entre les différentes phases, nous avons réalisé un ensemble de calculs corrélationnels qui nous permis de réaliser les constats suivants.

D'une manière remarquablement uniforme, nous avons pu observer un taux très significatif de corrélation entre les résultats des différentes phases, ainsi qu'entre les résultats des compétences de chaque matière de la phase 3 avec les résultats globaux des deux autres phases. C'est le signe de la grande cohérence de nos épreuves.

Cette cohérence est renforcée par la similitude de résultats entre les deux épreuves qui, et ceci a son importance, ont été construites par des chercheurs différents suivant le modèle imaginé.

Les résultats présentés dans les tableaux 4 et 5 montrent le lien important existant entre les différentes phases. Sur base de ces corrélations, un ensemble d'affirmations peuvent être énoncées.

- Plus on est performant en phase 2 et plus on est performant en phase 1 et réciproquement.
- La majorité des élèves forts en phase 2 sont moyens en phase 1.
- La majorité des élèves moyens en phase 2 sont faibles en phase 1.
- Les élèves faibles en phase 2 restent faibles en phase 1.

- Plus on est performant en phase 3 et plus on est performant en phase 2
- La différence de résultats entre un élève moyen et un élève fort en phase 3 augmente de manière significative entre ces deux élèves à la phase 2, mais ce constat n'est pas observé entre la phase 2 et la phase 1 pour ces mêmes élèves.
- Un pourcentage élevé d'élèves sont considérés comme forts en phase 3 (63 % pour les 10 – 12 et 68 % pour les 5-8), mais seulement la moitié de ceux-ci sont également forts en phase 2. (33 % pour les 10-12 et 32 % pour les 5-8)
- 1/3 des élèves du cycle 5 – 8 et du cycle 10-12 sont forts en phase 2 et en phase 3

- Un élève fort à la phase 1 est d'office fort à la phase 3.
- La différence de résultats entre un élève moyen et un élève fort en phase 3 augmente de manière significative à la phase 1
- Plus de 80 % des élèves forts en phase 3 sont faibles en phase 1.

Les analyses corrélationnelles montrent clairement que les procédures automatisées sont une condition nécessaire à la réussite des autres phases. La capacité de choisir la bonne procédure est également nécessaire pour la résolution de tâches complexes. Ceci confirme la hiérarchie entre les phases de nos épreuves.

Il est cependant très important de souligner que si l'acquisition des procédures automatisées et la capacité à choisir la bonne procédure sont des conditions nécessaires à la réussite de tâches complexes, elle n'est toutefois pas la seule condition comme l'indique clairement les résultats proposés.

Ces constats posent naturellement une question essentielle. Pourquoi des élèves, lors d'une résolution de problèmes, sont-ils capables de choisir et d'utiliser la bonne procédure et d'autres non ? En d'autres termes : pourquoi des élèves sont-ils capables de transférer leurs procédures automatisées et d'autres non ?

2.1.4. Conclusions de la première année de recherche

Sur base de notre analyse de la notion de compétences qui s'inscrit dans la définition de compétences telles que proposées par le décret « Missions », nous avons construit un dispositif d'évaluation de compétences.

Le très net écart de réussite entre la phase 1, la phase 2 et la phase 3, semble confirmer notre modèle théorique qui distingue trois niveaux de compétences : les compétences complexes, les compétences élémentaires avec cadrage de la situation et les simples procédures. Il fonde du même coup notre dispositif d'évaluation en trois phases et sa supériorité par rapport à des épreuves d'évaluation couramment utilisées aujourd'hui qui disent offrir des tâches complexes, mais qui, en fait, les décomposent en questions partielles et qui correspondent donc à notre phase 2.

Notre hypothèse quant à la difficulté rencontrée actuellement par les élèves à résoudre individuellement des tâches complexes s'est révélée exacte. Ceci renforce notre idée qu'un étalonnage, qu'il soit normatif ou critérié, n'a de sens pour le moment que si les résultats sont utilisés dans une fin diagnostique.

Les épreuves ont montré leur caractère diagnostique et ont été accueillies favorablement auprès des enseignants qui ont collaboré à notre pré-test. Ainsi, ces derniers au regard des résultats aux épreuves ont pu clairement, pour leur classe et pour chacun de leurs élèves, réaliser un diagnostic faisant intervenir les performances de ceux-ci dans les trois phases. Par exemple, un enseignant a pu déterminer que l'élève x maîtrisait les procédures automatisées, était capable de les utiliser dans des situations inédites limitées, mais éprouvait des difficultés devant un tâche complexe nécessitant la mise en place d'une démarche.

Tous ces constats demandent naturellement d'être confirmés. L'échantillon occasionnel et le nombre limité d'épreuves exigent une interprétation prudente et conditionnelle de ces conclusions. Comme nous le développerons ci-dessous, la confirmation ou non de ces constats fut l'objectif premier de notre deuxième année de recherche.

2.2. Objectifs et méthodologie de la deuxième année de recherche

2.2.1. Objectifs de la deuxième année de recherche

Si l'on administre d'autres épreuves créées suivant notre modèle, à d'autres élèves, obtiendrons-nous des résultats permettant de confirmer les résultats observés lors de la première passation ?

Au delà de cette question principale, nous avons essayé lors de cette deuxième phase de la recherche d'apporter des réponses aux questions suivantes. Celles-ci sont la synthèse des nombreuses interrogations suscitées lors de la première année.

Ces questions sont :

1. Existe-t-il réellement une influence du milieu social de l'école sur les compétences des élèves à résoudre des tâches complexes ?
2. Y a-t-il une influence de la dynamique de l'équipe pédagogique d'une école sur les compétences des élèves à résoudre des tâches complexes ?
3. Les modalités d'administration des épreuves ont-elles une influence sur les résultats des élèves ?
4. Le niveau de réflexion de l'enseignant sur les réformes en cours exerce-t-il une influence sur les compétences des élèves à résoudre des tâches complexes ?
5. Quels sont les facteurs de réussite à des épreuves mesurant l'acquisition de compétences et quelles sont les conditions d'apprentissage les plus adaptées à cette acquisition ?
6. Des enseignants, après une formation, sont-ils capables de s'appropriier le modèle proposé et de créer et d'analyser des épreuves construites suivant celui-ci ? Comment perçoivent-ils cette tâche ?

Dès maintenant, nous devons annoncer que nous n'avons pas la prétention de répondre de manière exhaustive à toutes ces questions. En effet, chacune d'entre elles exigerait d'être l'objet d'une recherche approfondie avec une méthodologie spécifique. Nous espérons seulement apporter des éléments susceptibles d'alimenter la réflexion portant sur l'évaluation des compétences.

Pour tenter de confirmer notre modèle et pour apporter des éléments de réponses aux questions énoncées ci-dessus, nous avons administré auprès d'un échantillon, respectant différents critères, quatre nouvelles épreuves. 45 classes de 19 écoles dépendant des cantons scolaires de la province de Namur ont été sélectionnées ce qui représente un échantillon de plus de 1000 élèves. Sur deux années, plus de 1200 élèves auront été évalués. En plus des épreuves, nous avons élaboré trois questionnaires :

1. Un questionnaire destiné aux enseignants concernés par nos évaluations évaluant leurs degrés d'informations et d'adhésion vis-à-vis des réformes en cours et de manière plus précise vis-à-vis de la réforme des cycles. Un questionnaire a été également proposé aux directions des 20 écoles de notre échantillon.
2. Un court questionnaire destiné aux élèves de sixième année qui devrait permettre d'estimer, en comparant leurs réponses aux résultats obtenus aux épreuves, le poids de leur vécu personnel et scolaire sur leur performance.
3. Un court questionnaire destiné aux enseignants de sixième année essayant de déterminer leurs pratiques didactiques et les contenus effectivement abordés avec leurs élèves et nécessaires à la réalisation des tâches.

Parallèlement, une formation, coordonnée par un membre de notre équipe et destinée à des instituteurs ou directeurs en fonction devrait permettre d'apporter des éléments de réponses à la question portant sur la possibilité pour les enseignants de s'approprier facilement notre modèle.

2.2.2. Méthodologie

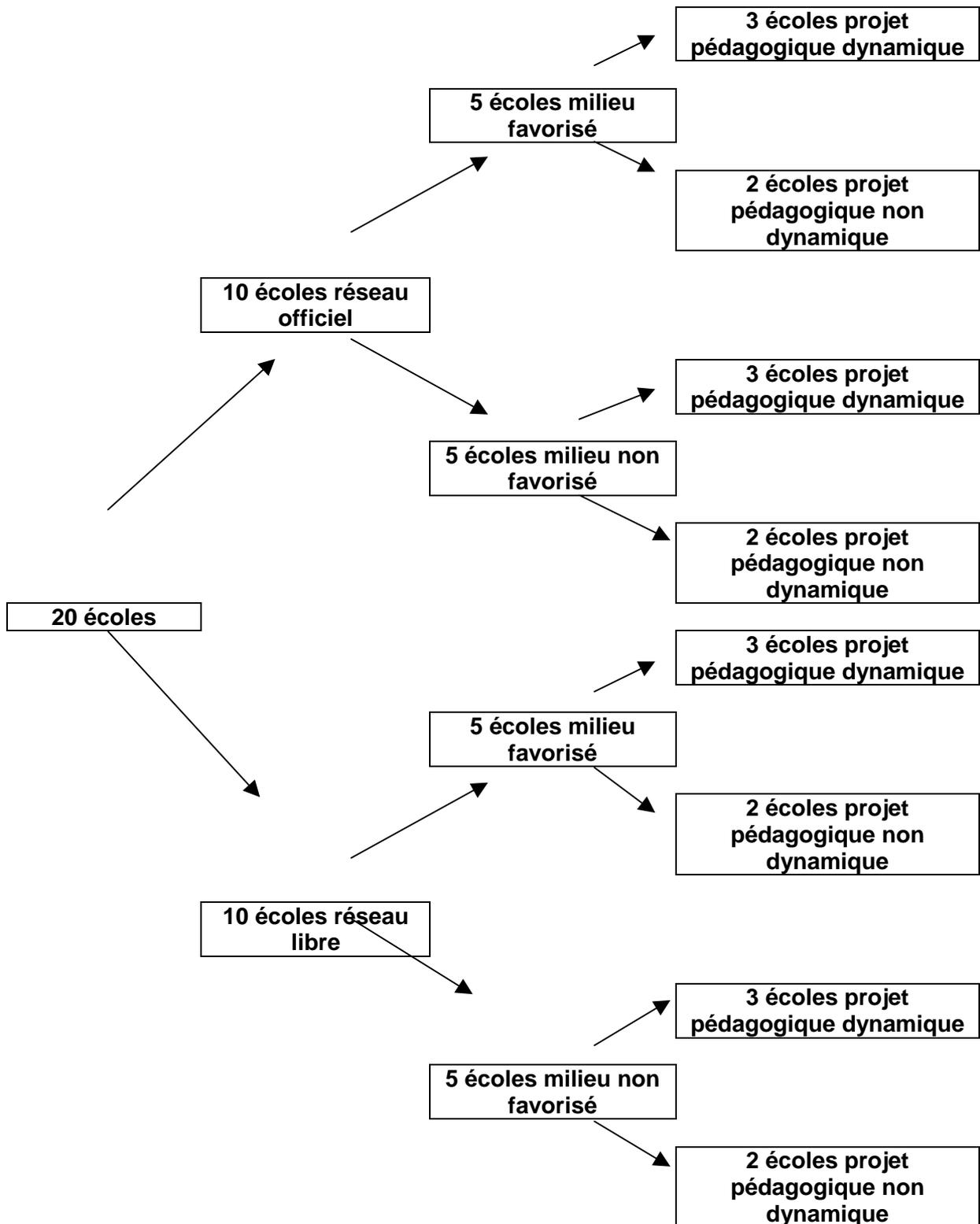
2.2.2.1. Une collaboration efficace avec les inspecteurs, des directeurs et des enseignants des cantons scolaires de la Province de Namur.

Pour rejoindre nos différents objectifs et dans le but de confronter nos objets de recherche avec différents acteurs du monde éducatif, nous avons mis en place une collaboration avec les inspecteurs de la province de Namur. Ceux-ci ont accepté avec enthousiasme de participer à notre recherche.

Avec leur aide, sur base de différents critères, nous avons construit un échantillon de 20 écoles. Outre la construction de l'échantillon et l'aide importante qu'ils nous ont apportée dans nos prises de contact avec les écoles, nous avons programmé des rencontres avec eux pour débattre de notre travail. Leur aide efficace, nous a permis de rentrer en contact avec les écoles et de construire une collaboration active avec elles. Ce mode d'investigation, même s'il nous demande un travail considérable de relation, de communication d'informations, de réunions et dans certains cas d'organisation de conférences pédagogiques, est très riche en terme de récolte d'informations.

2.2.2.2. La construction de l'échantillon

2.2.2.2.1. les critères



Pour l'élaboration de notre échantillon, trois critères, en relation avec notre questionnaire, ont été retenus :

1. **Le réseau** : officiel ou libre
2. **Le milieu social des écoles** : favorisé ou défavorisé
3. **Le projet pédagogique de l'école** : dynamique ou non dynamique

Les écoles considérées comme « défavorisées » correspondent aux écoles appelées « ZEP ». Elles sont considérées par la Communauté française, sur base de différents critères, comme des écoles devant être soutenues en raison du niveau social des élèves les fréquentant.

Le critère « projet pédagogique de l'école » est quant à lui beaucoup plus subjectif ; car pour le déterminer, nous nous sommes reposés sur la connaissance que les différents inspecteurs ont de leurs écoles. Le choix de ces écoles s'est réalisé sur l'impression de chaque inspecteur cantonal. Nous aurons l'occasion de discuter de ce critère subjectif reposant sur des impressions.

20 écoles ont été sélectionnées et 19 ont réellement participé. Pour des raisons d'ordre déontologique, nous ne citerons pas le nom de ces écoles. Cependant, il nous semble intéressant de souligner que parmi ces différentes écoles, nous trouvons des écoles urbaines, rurales et « mixtes ».

2.2.2.2. L'élaboration de l'échantillon

Dans chacune des 19 écoles de notre échantillon, une classe de troisième primaire et 1 classe de sixième primaire ont été sélectionnées.. Suite à la demande des directions et/ou des enseignants de cinq écoles, des épreuves ont été administrées dans deux classes d'une même année. De ce fait, notre échantillon final était composé de 22 classes de troisième primaire et de 23 classes de sixième. Ceci représente plus de 900 élèves testés

L'échantillon a été subdivisé en deux. Nous avons administré à 10 classes de chaque cycle (5-8 et 10-12) les épreuves que nous avons nommées A et aux 10 autres classes celles que nous avons nommées B. Les épreuves A sont « les murs de la classe » pour le cycle 10-12 et « la visite de St Nicolas » pour le cycle 5-8. Les épreuves B sont « Le premier mai » pour le cycle 10-12 et « le jardinage » pour le cycle 5-8. Les classes dédoublées à la demande des écoles ont reçu les mêmes épreuves que la classe de leur établissement initialement sélectionnée.

Pour évaluer le biais de l'administration des épreuves par les enseignants, nous avons constitué trois modes de passation. Dans cinq écoles, nous avons administré nous-même les trois phases. Dans dix écoles, nous avons administré la 1^{ère} phase et les enseignants ont administré les phases 2 et 3. Enfin dans les quatre dernières écoles, ce sont les enseignants qui ont administré les trois phases. Les épreuves, dans les sept classes ne faisant pas partie de l'échantillon initial, ont été administrées par les enseignants. Chaque sous-échantillon comprend des écoles respectant nos différents critères différents. Nous essayerons par analyse statistique de déterminer si la passation des épreuves par des enseignants constituent un biais réel ou non.

Pour la raison identique à celle invoquée lors de notre première année de recherche, nous avons fait le choix d'administrer les épreuves destinées aux élèves du cycle 5-8 à des élèves de 3^{ème} primaire. Pour rappel, la raison est que dans le cadre de notre recherche s'étalant de septembre à août, nous sommes obligés d'administrer nos épreuves avant le mois de février si nous voulons traiter les résultats correctement. La différence de maîtrise de compétences entre un enfant de deuxième primaire au mois de janvier et celle d'un enfant de la même année au mois de juin pouvant s'avérer, à nos yeux, importante, nous avons préféré évaluer des élèves de troisième primaire. Nous aurions pu administrer nos épreuves à des troisièmes primaires au mois d'octobre pour être plus proche de la période où celles-ci

devraient normalement être utilisées. Pour des raisons organisationnelles, mais également pour réaliser des comparaisons entre nos résultats de l'année dernière et ceux de cette année, nous avons fait le choix d'administrer nos épreuves au mois de janvier comme lors de notre première année de recherche.

En résumé, voici le dispositif que nous avons mis en place.

Ecole	Réseau	Social	Pédag.	Epreuves	Condition de passation	Nbre de Classes 5 - 8	Nbre de Classes 5 - 8
1	Libre	-	+	A	Phase 1	1	1
2	Libre	-	-	A	Phase 1	1	1
3	Libre	-	+	B	Phase 1,2,3	1	1
4	Libre	-	+	A	Phase 1	2	2
5	Libre	-	-	B	Phase 1	1	1
6	Libre	+	+	B	Enseignant	1	2
7	Libre	+	-	B	Phase 1	1	1
8	Libre	+	+	B	Phase 1,2,3	1	1
9	Libre	+	-	A	Phase 1,2,3	2	1
10	Libre	+	+	A	Phase 1	2	2
11	Officiel	-	+	B	Phase 1	1	1
13	Officiel	-	+	B	Phase 1,2,3	2	1
14	Officiel	-	+	A	Enseignant	1	1
15	Officiel	-	-	A	Enseignant	1	1
16	Officiel	+	+	A	Phase 1	1	1
17	Officiel	+	-	A	Phase 1	1	1
18	Officiel	+	+	A	Phase 1,2,3	1	1
19	Officiel	+	+	B	Phase 1	1	1
20	officiel	+	-	B	Enseignant	1	1
						23	22

Légende :

Phase 1 = Phases 1 administrées par les chercheurs et Phases 2 et 3 par les enseignants

Enseignant = Les trois phases administrées par les enseignants

Phases 1,2,3 = Les trois phases administrées par les chercheurs

2.2.2.2.3. L'intérêt de cet échantillon

Bien que cet échantillon ne puisse pas encore être considéré comme représentatif, il a néanmoins été construit pour permettre de répondre à certaines questions que nous nous posons. En effet, par analyse statistique, le traitement des résultats devrait nous aider à recueillir des informations intéressantes quant à l'influence du milieu social et à l'influence du projet pédagogique des écoles sur les performances des élèves aux différentes phases.

Il nous permettra également de comparer les épreuves (A et B) entre elles d'une manière globale, mais également en tenant compte des critères que sont : le réseau, le niveau social et le projet pédagogique des écoles.

Enfin, il devrait nous aider à évaluer le biais possible de la passation des épreuves par les enseignants.

2.2.3.3. Les épreuves : élaboration, passation et traitement statistique

2.2.3.3.1. Présentation des épreuves

Les quatre épreuves ont été conçues, lors de notre première année de recherche, suivant le modèle que nous avons rappelé dans la première partie du présent rapport. Elles ont cependant subi quelques modifications suite à une première expérimentation auprès de quelques élèves.

Pour chaque épreuve, trois tâches spécifiques sont proposées aux élèves.

Les murs de la classe

Tâche 1 : rédiger une lettre à la directrice pour demander l'autorisation de repeindre la classe

Tâche 2 : calculer les surfaces à peindre et le nombre de pots de peinture nécessaires.

Tâche 3 : rédiger un bon de commande pour le matériel nécessaire.

Le premier mai

Tâche 1 : calculer le gain moyen par élève de la classe de 6^{ème} et le comparer avec le gain moyen de la classe de 5^{ème}.

Tâche 2 : choisir le matériel susceptible d'être acheté avec l'argent récolté.

Tâche 3 : Préparer une élocution par écrit sur le premier mai.

La visite de St Nicolas

Tâche 1 : Calculer le nombre de pièces de 1 franc récoltées par la classe et les échanger avec des pièces de valeur supérieure.

Tâche 2 : Choisir les cadeaux pour St Nicolas, Père Fouettard et l'âne.

Tâche 3 : Faire son portrait par écrit.

Le jardinage :

Tâche 1 : Choisir les légumes pouvant être semés à partir du mois de mars et récoltés avant les grandes vacances.

Tâche 2 : Dessiner le plan du jardin suivant des consignes données.

Tâche 3 : Réaliser un panneau de classification de légumes.

Dans les quatre épreuves, les élèves ont un portefeuille de documents leur donnant des renseignements nécessaires à la réalisation des tâches. Par rapport aux deux premières épreuves administrées l'année dernière, nous avons introduit le portefeuille de lecture au niveau du cycle 5-8 et nous avons diminué considérablement le nombre d'informations destinées aux élèves du cycle 10-12. Ces modifications auront peut-être un impact sur les résultats des élèves.

Les épreuves sont présentées dans le deuxième volume de ce rapport ainsi que les compétences qu'elles évaluent.

2.2.3.3.2. Les modalités et critères de correction

Toutes les corrections seront réalisées par notre équipe. Ce travail extrêmement important et contraignant est nécessaire pour respecter le critère de fidélité lors de nos analyses statistiques.

Un travail de correction collective permettant d'atteindre un accord entre chercheurs a été réalisé.

Pour chaque épreuve, une grille de correction précise a été construite. Les principes de construction ont été les mêmes pour les quatre épreuves et peuvent s'énoncer en ces termes.

Pour la phase 1

Chaque tâche est évaluée sur une échelle à 9 niveaux de 0 à 8.

Les critères ont été choisis suivant une logique quantitative mais permet également lors de l'analyse d'apporter des éléments qualitatifs.

Ainsi, pour toutes les tâches :

Le niveau 8 correspond aux conditions pour que la tâche soit réussie.

Le niveau 0 correspond « à la feuille blanche ».

Le niveau 1 correspond à un essai sans rapport réel avec la tâche.

Le niveau 4 correspond à la moitié de la tâche réussie. Ce critère a été déterminé pour chaque épreuve par les chercheurs.

Le niveau 2 correspond à un entrée dans la tâche.

Méthodologiquement, cette manière d'évaluer les tâches de la phase 1 devrait nous permettre une comparaison entre les épreuves suffisamment rigoureuse.

Lors de la correction, si un élève se trompe en début de réalisation – dans les tâches appelées en « cascade », nous corrigeons en tenant compte de sa ou de ses premières réponses.

Pour la phase 2

En phase 2, nous rencontrons deux systèmes de cotation.

Soit l'élève doit réaliser la même tâche qu'en phase 1 et on lui donne des renseignements supplémentaires devant l'aider à y parvenir, soit l'élève doit compléter un ensemble d'items l'amenant aux résultats attendus.

Dans le premier cas, nous reprenons le système de cotation de la phase 1 à savoir la notation suivant des critères sur une échelle à 9 niveaux.

Dans le second cas, nous obtenons un score correspondant aux nombres d'items réussis. Ces items sont subdivisés suivant le type de démarche à mettre en œuvre pour les résoudre, à savoir : items de lecture, items de connaissance ou items d'opérations. L'analyse de ces items devraient, comme l'année dernière, nous apporter des informations quant aux différents poids de ces démarches dans la résolution des tâches. Elle permet à l'enseignant, dans une perspective diagnostique, de souligner les facilités ou les difficultés des élèves.

Pour la phase 3.

Chaque item est noté de manière binaire. Soit 0, soit 1.

2.2.3.3.3. L'analyse statistique

Pour confirmer, infirmer ou nuancer les constats de notre première année de recherche et apporter des éléments de réponses à certaines de nos questions, nous avons réalisé différentes analyses statistiques.

a. Pour les épreuves

- Pour l'ensemble des épreuves

1. Calcul des moyennes et des écarts types aux différentes phases,
2. Corrélations entre les différentes phases – tableaux de contingence à trois variables,
3. Résultats aux différentes tâches de la phase 1,
4. Résultats aux différentes tâches de la phase 2,
5. Comparaison entre les résultats aux tâches de la phase 1 et à celles de la phase 2,

- Pour chaque classe

6. Calcul des moyennes et des écarts types aux différentes phases,
7. Corrélations entre les différentes phases – tableaux de contingence à trois variables,
8. Résultats aux différentes tâches de la phase 1,
9. Résultats aux différentes tâches de la phase 2,

- Comparaison des moyennes, écarts types et corrélations entre les phases

10. entre les classes,
11. entre les classes de même niveau social,
12. entre les classes de milieu social différent,
13. entre les classes de niveau « de projet pédagogique » équivalent,
14. entre les classes de niveau « de projet pédagogique » différent,
15. entre les classes de même milieu social et de niveau de « projet pédagogique » équivalent,
16. entre les classes de même milieu social et de niveau de « projet pédagogique » différent,

b. Pour l'ensemble des épreuves de même cycle

17. Calcul des moyennes et des écarts types aux différentes phases,
18. Corrélations entre les différentes phases – tableaux de contingence à trois variables,

- Comparaison des moyennes, écarts types et corrélations entre les phases

19. entre les classes,
20. entre les classes de même niveau social,
21. entre les classes de milieu social différent,
22. entre les classes de niveau « de projet pédagogique » équivalent,
23. entre les classes de niveau « de projet pédagogique » différent,
24. entre les classes de même milieu social et de niveau de « projet pédagogique » équivalent,
25. entre les classes de même milieu social et de niveau de « projet pédagogique » différents

c. Comparaison entre chaque épreuve de même cycle

26. des moyennes, écarts types et corrélations sur les résultats globaux aux épreuves,
27. des moyennes, écarts types et corrélations sur les résultats par niveau social,
28. des moyennes, écarts types et corrélations sur les résultats par niveau du projet pédagogique...

Ces différents résultats devraient répondre aux questions suivantes :

- Retrouve-t-on dans les six épreuves des résultats comparables ?
 - Les coefficients de corrélation sont-ils semblables d'une épreuve à l'autre ?
 - Peut-on observer une influence du milieu social sur les performances des élèves ?
 - Peut-on observer une influence du projet pédagogique de l'établissement sur les performances des élèves ?
 - A niveau social équivalent et à niveau « du projet pédagogique » équivalent, obtient-on des résultats statistiquement comparables ?
 - Les différences de résultats susceptibles d'être observées et dépendant du niveau social et du niveau « du projet pédagogique » se retrouvent-elles dans les différentes épreuves ?
 - Y a-t-il des classes qui à niveau social et pédagogique équivalents obtiennent des résultats statistiquement supérieures aux autres ?
- ...

Pour répondre à ces différentes questions, nous avons adapté l'outil statistique construit et utilisé l'année dernière, en le généralisant de telle sorte qu'il puisse fonctionner avec une certaine souplesse sur les différentes épreuves. Il a été élaboré sur le logiciel Excel. De plus, les études qui visent plusieurs facteurs simultanément ont été traitées sur le programme statistique SPSS.

2.2.4. Les questionnaires

La réflexivité de l'enseignant pouvant se traduire par sa connaissance et l'analyse des réformes en cours, exerce-t-elle une influence sur les compétences des élèves à résoudre des tâches complexes ?

Le poids des apprentissages vécus par l'élève aussi bien en ce qui concerne les compétences que les moyens didactiques mis en œuvre exerce-t-il une influence sur les compétences des élèves à résoudre des tâches complexes ?

Pour répondre à ces questions, trois questionnaires ont été élaborés.

2.2.4.1. Analyse de « la réflexivité » de l'enseignant

La réflexivité de l'enseignant par rapport à ses conceptions de l'école et ses pratiques quotidiennes est un objet de recherche actuellement fort débattu. Par l'intermédiaire d'un questionnaire, nous essayons de déterminer si la réflexivité de l'enseignant, par rapport aux réformes mises en place, pourrait être influente sur les performances de ses élèves mis en face de tâches complexes. Le questionnaire, élaboré dans le cadre du mémoire de licence par monsieur Schoeters¹, a été adapté pour tenter d'apporter des éléments de réponses à cette question.

Celui-ci, présenté dans l'annexe de ce rapport, s'intéresse aux connaissances et aux analyses des enseignants par rapport à la mise en place des cycles. Estimant que cette transformation de structure est la partie la plus visible et la plus concrète de la réforme décrétée en juillet 1997, il nous a semblé opportun d'analyser les conceptions de l'enseignant par rapport au concept de cycle. Ce questionnaire fait d'ailleurs référence à d'autres concepts tels que l'évaluation, les pratiques didactiques ...

Sur base des résultats obtenus par les élèves de chaque classe, nous avons essayé de déterminer, en tenant compte du niveau social des écoles, s'il existe des points communs

¹ Schoeters Didier, (2001) *Le travail en cycle dans l'enseignement fondamental . Evolution, état des lieux et perspectives* Mémoire de fin de licence en sciences de l'éducation ULB. (à paraître)

dans la perception du cycle entre les enseignants des classes obtenant des performances semblables.

Si nous acceptons l'hypothèse régulièrement avancée qu'un enseignant efficace est un enseignant réflexif, nous pourrions observer un lien statistique entre le niveau d'information et d'analyse de la réforme par les enseignants et les performances de leurs élèves à nos épreuves.

Nous insistons sur le fait que cette approche est tout à fait exploratoire. Quels que soient les résultats, nous ne pourrions pas établir une relation causale entre conception du cycle par l'enseignant et performances de ses élèves à nos épreuves. Nous pourrions, toutefois, souligner l'intérêt ou non d'investiguer de manière plus approfondie dans cette voie.

2.2.4.2. Le travail d'appropriation du modèle par des enseignants

Dans le cadre d'un module de formation organisé et animé par monsieur Vincent Carette à l'Ecole Supérieure de Pédagogie de Namur et portant sur l'évaluation des compétences, il a été proposé aux participants de construire et de tester une épreuve suivant notre modèle en trois phases. Les participants composés d'instituteurs primaires et maternels et de quelques directeurs ont proposé de ce fait une épreuve originale accompagnée de commentaires critiques.

L'objectif de ce module était de vérifier que des enseignants étaient en mesure de construire des épreuves conformes à notre modèle. Pour ce faire, un questionnaire tentant d'évaluer cette faisabilité fut distribué à la fin du module à chaque participant.

2.2.4.3. Les facteurs de réussite à des épreuves mesurant l'acquisition de compétences et les conditions d'apprentissage les plus adaptées à cette acquisition.

Au delà des résultats globaux et de l'influence hypothétique de l'enseignant dit réflexif, nous nous sommes attelés durant toute cette recherche à réfléchir sur les facteurs possibles de réussite à des épreuves mesurant l'acquisition de compétences ainsi qu'aux conditions d'apprentissage les plus adaptées à cette acquisition. Cette perspective nous a amené à poser un certain nombre de questions et à énoncer quelques hypothèses en réponse à celles-ci.

Pour tenter de vérifier ces hypothèses développées ci-dessous, nous avons comme intention de pratiquer une analyse qualitative, de manière systématique et détaillée, des copies des élèves et de dépouiller deux questionnaires. Le premier interrogeait les enseignants de sixième sur leur pratique didactique et sur l'état d'acquisition de certaines matières. Le second questionnait leurs élèves sur leurs apprentissages.

Dès à présent, nous devons annoncer, pour ne pas donner de faux espoir au lecteur, que nous n'avons pas été en mesure de vérifier ces hypothèses. En effet, méthodologiquement, les courts questionnaires destinés aux élèves et aux enseignants de sixième n'ont pas répondu à nos attentes. Très courts, trop courts sans doute, ils n'ont permis ni l'un, ni l'autre d'apporter des éléments probants aux questions posées. Quant à l'analyse systématique des épreuves, nous avons dû par manque de temps et de moyen la remettre à plus tard.

Toutefois, même si nous ne sommes pas en mesure de proposer des résultats vérifiant nos hypothèses quant aux facteurs de réussite à des épreuves mesurant l'acquisition de compétences et aux conditions d'apprentissage les plus adaptées à leur acquisition, ils nous semble opportun à ce stade du rapport de rappeler notre questionnement, les hypothèses suscitées par celui-ci et la méthodologie que nous avons envisagée et qui ne s'est pas révélée probante.

Ce questionnement non exhaustif met en évidence un ensemble de pistes de recherche devant nous conduire à répondre à une question essentielle en sciences de l'éducation : la question du transfert des apprentissages.

2.4.3.1. Les opérations cognitives qui entrent dans la mise en œuvre d'une compétence

Comme nous l'avons fait remarquer dans le rapport de recherche pour l'année 1999-2000 et comme nous l'avons rappelé ci-dessus, une compétence véritable ne saurait se borner à savoir exécuter une procédure apprise en réponse à un signal prédéterminé. Il y a compétence au sens plein lorsque le sujet est capable de déterminer, devant une situation nouvelle pour lui, la procédure à choisir parmi celles qu'il connaît ; et, mieux encore, lorsqu'il est capable de choisir et de combiner plusieurs procédures en réponse à une situation inédite. C'est cette remarque qui nous a amenés à distinguer, à côté de la simple procédure, deux autres degrés de compétences (compétence élémentaire avec cadrage et compétence complexe). Une étude des *Socles de compétence* fait apparaître que les compétences qui y sont consignées relèvent, dans leur grande majorité, de ces deux derniers degrés.

Dans ces degrés de compétence (les seuls, à nos yeux, qui méritent vraiment le nom de compétence), il y a un moment décisif qui est le choix d'une ou plusieurs procédures convenant à une situation nouvelle, ou encore la mise en rapport de procédures connues du sujet avec la situation, la mise en concordance, l'adéquation, la mesure du bon escient, etc. Les enseignants savent par expérience que c'est là qu'est la difficulté majeure à laquelle se heurtent beaucoup d'élèves. Dès lors qu'il est docile, on peut entraîner n'importe quel élève à exécuter une procédure (arithmétique, linguistique, grammaticale, orthographique, etc.) en réponse à une consigne à laquelle il est habitué. En revanche, les choses se compliquent dès lors que l'élève doit déterminer lui-même quelle est la procédure à mettre en œuvre devant une situation inhabituelle.

Les écarts très significatifs que nous avons constatés, l'an dernier, dans les résultats des élèves entre les phases 1, 2 et 3 de nos épreuves, confirment d'une manière objective l'intuition des enseignants sur ce point.

Comme nous l'avons déjà succinctement évoqué dans le rapport pour l'année 1999-2000, d'autres auteurs ont remarqué avant nous le caractère problématique de l'application d'une procédure connue à une situation nouvelle.

Ainsi, Meirieu et Develay (1992), reprenant en cela des notions empruntées à la didactique des mathématiques (Noirfalise, 1991), estiment qu'une compétence doit comprendre non seulement la maîtrise d'une opération (ce que nous appelons une procédure), mais aussi la connaissance de la famille de situations à laquelle elle s'applique.

Mais ils font remarquer aussitôt qu'il est souvent difficile pour le novice de percevoir d'une manière adéquate l'étendue de cette famille de situations, c'est-à-dire le champ d'application de cette opération. Ils prennent l'exemple du théorème de Pythagore. Traduit en terme de compétence, ce théorème permet, quand on connaît la mesure de deux côtés d'un triangle rectangle, de calculer la mesure du troisième.

Or, parmi les élèves du secondaire qui débute dans son usage, certains ont tendance à vouloir l'utiliser chaque fois que, dans un problème de géométrie, il est question d'un triangle rectangle (même si par exemple c'est l'aire du triangle que l'on veut connaître). En revanche, des élèves (parfois les mêmes) peuvent ne pas voir qu'il y a lieu d'utiliser le théorème de Pythagore, lorsqu'on a à découvrir la mesure de la diagonale d'un carré dont on connaît le côté. Dans le premier cas, il y a une surestimation du champ d'application du théorème et une sous-estimation dans le second.

Préciser aux élèves quel est le champ d'application d'une procédure à laquelle on les a entraînés est à l'évidence une précaution pédagogique tout à fait utile et ce serait certainement un progrès que les enseignants le fassent systématiquement. Mais, comme on le voit sur l'exemple du théorème de Pythagore, cette pratique pédagogique ne résoudra pas toutes les difficultés d'application pour les élèves. Car, même si l'élève connaît la famille de situations à laquelle convient une procédure (en ayant identifié quelques-uns de leurs caractères spécifiques), il lui restera à saisir en quoi une situation qui est nouvelle pour lui appartient ou n'appartient pas à cette famille.

La seule manière d'éviter cette difficulté serait de décrire à l'élève et de lui faire apprendre la liste exhaustive des situations dans laquelle telle procédure s'applique. Mais il est clair que cela est impossible. En outre, même si ça l'était, ce ne serait pas souhaitable, car l'ambition d'un enseignement digne de ce nom est que l'élève soit capable de déterminer par lui-même dans quelle situation il doit mettre en œuvre une opération.

Il faut donc reconnaître qu'il y a, dans la mise en œuvre d'une compétence, un moment indispensable et spécifique où le sujet doit interpréter la situation, pour y saisir les éléments qui vont lui permettre de la rattacher à une famille de situations qui relèvent d'une procédure. On ne peut nier ni l'existence ni l'importance de cette activité cognitive.

En effet cette interprétation ne se borne pas à repérer, dans la situation nouvelle, des indicateurs apparents. Reprenons l'exemple du théorème de Pythagore : la présence visible ou explicite d'un triangle rectangle n'est pas un bon indicateur du fait que la situation appartient à la famille de celles qui relèvent de ce théorème. Les véritables indicateurs sont multiples, complexes (il faut non seulement qu'il y ait un triangle rectangle, déjà tracé ou à construire, mais que connaissant les mesures de deux côtés on ait besoin de la troisième) et surtout ne sont pas toujours apparents. Si les indicateurs ne sont pas faciles à voir dans le cas d'exercices scolaires relativement codifiés et traditionnels, ils le seront encore moins si on a affaire, non plus à des exercices artificiels, mais aux situations de la vie réelle, extra-scolaires.

A vrai dire, toute situation comporte une infinité de caractères. C'est le sujet qui doit sélectionner ceux qui lui paraissent pertinents et négliger les autres. Il ne s'agit donc pas d'un repérage, mais plutôt d'une sélection. Et sans doute cette présentation est-elle encore insuffisante. Car le sujet ne sélectionne pas parmi des éléments préexistants qui seraient tous également visibles. Il a à faire émerger les éléments pertinents dans un ensemble indistinct et non délimité. Il s'agit en effet d'abord de circonscrire dans l'espace et dans le temps ce qui aux yeux du sujet fait partie de la situation. C'est pour cela que nous parlons, pour désigner cette activité cognitive, du "cadrage" de la situation.

Une infinité de cadrages sont possibles devant un même segment de la réalité. Mais certains cadrages sont considérés, à l'école, comme plus adéquats.

La question est alors de savoir quelles sont les conditions qui favorisent chez les élèves le choix du cadrage scolairement convenable. Sur ce problème nous sommes passés par deux étapes successives de réflexion, que nous allons relater ci-dessous.

2.2.4.3.2. Première étape de réflexion : procédures de base, habitude d'exercices spécifiques, familiarité.

L'étude des résultats des élèves aux phases 1 et 2 de nos épreuves nous a orienté, durant l'année 1999-2000, vers un constat et deux hypothèses :

2.2.4.3.2.1. Un constat

Le large pré-test opéré l'an dernier montre clairement que la maîtrise des procédures de base qui entrent en jeu dans la résolution de problèmes complexes est à la fois radicalement nécessaire et nettement insuffisante. Il nous restera à vérifier que les résultats obtenus au test opéré durant la présente année 2000-2001, confirment ce fait. Cette vérification se fera par comparaison statistique des résultats aux phases 1 et 2 avec ceux de la phase 3 (qui renvoie uniquement à des procédures de base).

2.2.4.3.2.2. Une première hypothèse

Mais l'étude fine des réponses par classe nous a conduit à faire l'hypothèse que l'habitude de certains exercices de la part d'enseignants pouvait avoir un effet non négligeable sur la réussite des élèves dans la phase 1 de nos épreuves. En effet lorsque, dans cette phase 1, on invite les élèves à préparer une élocution ou à consulter une documentation, on peut faire l'hypothèse que la réussite est supérieure dans les classes dans lesquelles l'enseignant a fait systématiquement pratiquer ces exercices.

Il s'agit donc de savoir si les élèves ont une pratique courante d'exercices tels que les élocutions, les résumés, les rédactions de lettres, les consultations de documentations, l'usage de bons de commande, les calculs de prix, la construction d'affiches, de mode d'emploi, de règles du jeu, etc.

Notons que ces activités sont déjà plus complexes que les procédures de base (comme l'effectuation d'une soustraction, l'accord du verbe avec le sujet ou bien encore la reprise d'un nom par un pronom, etc.). Mais il convient de préciser ici que notre distinction entre procédures (premier degré), compétences avec cadrage (deuxième degré) et compétences complexes (troisième degré) n'a pas un caractère absolu. C'est pour un individu donné qu'une compétence donnée est une procédure automatisée, ou nécessite un cadrage de la situation. Une compétence qui exige d'un débutant un cadrage de la situation parce qu'elle est nouvelle pour lui (donc de deuxième degré) pourra être une procédure automatisée pour l'expert qui, lui, a rencontré de nombreuses fois ce type de situations. L'élaboration d'une élocution représente une compétence complexe pour les élèves qui ne l'ont jamais ou presque jamais pratiquée : l'élève doit interpréter une situation nouvelle et trouver parmi les procédures qu'il maîtrise celles qui conviennent à cette situation. Elle sera en revanche, sinon une procédure tout à fait automatisée, du moins une activité plus routinière pour celui qui en a une pratique habituelle.

Par suite, il nous a paru intéressant, dans la démarche de la présente année, de faire remplir un questionnaire d'une part par les élèves, et d'autre part par les enseignants des classes concernées, pour savoir si telle activité en jeu dans l'épreuve faisait l'objet d'une pratique courante dans cette classe.

Ainsi dans le questionnaire destiné aux élèves et concernant "les murs de la classe", nous demandions :

- As-tu déjà écrit une lettre ?
- En classe, écris-tu souvent des rédactions, des textes ?

Et nous avons demandé aux instituteurs de ces classes :

- Proposez-vous souvent des exercices de rédaction aux élèves ?
- Avez-vous abordé la rédaction de la lettre avec vos élèves ?
- Les élèves ont-ils été amenés à compléter un bon de commande ?
- Les élèves sont-ils amenés souvent à consulter des documents de référence ?

Dans le questionnaire en rapport avec l'épreuve sur "le 1er mai", nous demandions aux élèves :

- As-tu déjà préparé une élocution en classe ?
- En classe, écris-tu souvent des rédactions, des textes ?

- As-tu appris à faire un résumé ?
- En classe, réalises-tu beaucoup de travaux demandant de rechercher de l'information dans des documents ?

En rapport avec cette même épreuve, nous demandons aux instituteurs :

- Proposez-vous souvent des exercices de rédaction aux élèves ?
- Préparez-vous souvent des élocutions avec vos élèves ?
- Les élèves présentent-ils parfois des élocutions préparées à la maison ?
- Les élèves sont-ils amenés souvent à consulter des documents de référence ?

Nous espérons que ces questionnaires nous permettent de mieux évaluer l'effet de l'habitude de certains exercices sur la réussite à nos épreuves.

Notons que si les résultats du questionnaire nous avaient permis de confirmer cette hypothèse, elle pourrait donner lieu à deux interprétations différentes :

La première serait qu'il est souhaitable que les enseignants multiplient la pratique d'exercices tels que ceux que nous venons d'énumérer et qu'on pourrait appeler "exercices fonctionnels".

Mais la deuxième remarque que l'on peut faire à ce sujet est qu'il est impossible qu'un instituteur initie systématiquement ses élèves aux très nombreux exercices possibles qui entrent dans cette catégorie. En outre, quand un élève réussit à la phase 1 de nos épreuves parce qu'entre autres exigences, il y retrouve une activité à laquelle il a été entraînée systématiquement, cette phase de notre épreuve ne joue plus son rôle d'évaluation d'une compétence de troisième degré. Pour cet élève là, la compétence évaluée est proche d'une procédure de base. L'aspect de découverte d'une situation nouvelle et de sélection des meilleurs moyens d'y répondre n'existe plus.

Du coup, lorsque cela se produit, il s'introduit dans notre évaluation un biais. Grâce aux questionnaires indiqués ci-dessus, nous pouvons réduire ce biais. La question est de savoir s'il serait opératoire de l'introduire dans les épreuves standardisées que la Commission des outils d'évaluation proposera aux enseignants. En outre, on aperçoit là une des limites des évaluations externes : les résultats risquent de refléter, non pas toujours l'acquisition de compétences complexes, mais les types d'exercices que l'enseignant a jugé bon de faire pratiquer d'une manière répétée.

Dans le même ordre d'idées, la mesure de l'acquisition de compétences complexes peut être biaisée par la plus ou moins grande insistance avec laquelle l'enseignant a traité de certains contenus. Dans les trois épreuves qui concernent les élèves de 12 ans, on peut présumer que des élèves pouvaient être plus à l'aise si leur instituteur avait longuement insisté sur les pourcentages, les calculs sur les nombres décimaux, les calculs de moyenne, etc. Certaines parties du questionnaire destiné aux enseignants tentent de repérer ces faits éventuels. Par exemple :

- Avez-vous travaillé les mesures de surface ?
- Avez-vous travaillé les opérations sur les nombres décimaux ?
- Avez-vous abordé la notion de pourcentage avec vos élèves ?
- Avez-vous abordé la notion de gain moyen avec vos élèves ?

2.2.4.3.2.3. Une deuxième hypothèse

Une étude plus qualitative des réponses des élèves nous a amené à avancer l'hypothèse suivante : la familiarité des élèves avec le contexte évoquée dans l'épreuve jouerait un rôle positif dans la réussite. Ou, plus précisément, l'absence de familiarité aurait souvent un effet négatif.

Ainsi dans l'épreuve proposée aux élèves de sixième année et renvoyant à des compétences de troisième degré, il était question du métro bruxellois. Néanmoins, aucune connaissance spécifique à ce métro n'était nécessaire pour réussir l'épreuve en dehors de celles que l'on fournissait à l'élève. Or il apparaissait nettement que certains élèves habitant des zones rurales et n'ayant jamais pris le métro, avaient éprouvé des difficultés à opérer les cadrages nécessaires au traitement du problème posé.

C'est un phénomène assez bizarre, dès lors que les connaissances sur le contexte ne jouent pas de rôle opératoire dans l'effectuation de l'épreuve. Il peut éventuellement s'expliquer par le fait que l'évocation d'un contexte inconnu de certains élèves, engendre pour ceux-ci une zone d'opacité qui retient leur attention, au préjudice des éléments du problème qui sont importants pour sa résolution. On a peut-être là une occupation d'une partie de la mémoire de travail de l'élève, qui rend plus difficile l'appréhension des données pertinentes.

Le psychologue Patrick Mendelsohn (1990) aboutit à une conclusion comparable, lorsqu'il fait remarquer que, dans un nombre significatif d'expérimentations, les sujets ne voient pas l'identité de structure entre un problème qu'ils savent résoudre et un problème nouveau, tant qu'ils ne se sont pas familiarisés avec le contexte de ce dernier. Car il y a une hiérarchie dans le traitement de l'information. Les informations de surface concernant le contexte de la tâche sont traitées en priorité et, si elles ne sont pas automatisées chez un sujet donné, elles utilisent toute la capacité disponible de mémoire de travail. Mendelsohn écrit : "L'identité structurale entre deux problèmes n'est admise qu'à partir du moment où la mémoire de travail est libérée de la gestion des informations contextuelles propres à chaque version de la tâche".

Or cette connaissance des données contextuelles propres à une situation nouvelle, peut exister chez certains élèves et pas chez d'autres. Certains ont déjà pris le métro, d'autres ont déjà vu leurs parents entreprendre de repeindre leur logement, d'autres sont familiers des transactions commerciales, d'autres participent volontiers aux bricolages domestiques, etc.

Dans ce cas, on a affaire encore à un biais dans l'évaluation des compétences. Mais cette fois, il ne tient pas uniquement à la pratique pédagogique de l'instituteur, mais aussi aux habitudes de chaque famille et à l'histoire personnelle de chaque élève. On aura donc là un biais d'autant plus pernicieux qu'il n'opère pas au niveau d'une classe, mais au niveau des individus.

Aussi, dans l'opération de test de nos épreuves, nous avons tenté de le contrôler par certaines questions posées aux élèves dans le questionnaire évoqué ci-dessus.

Aux élèves qui devaient traiter l'épreuve portant sur la peinture des murs de la classe, nous demandions :

- As-tu déjà réalisé des travaux de peinture ?
- Es-tu déjà allé dans un magasin de bricolage ?
- Dans ta famille, bricole-t-on souvent ?
- As-tu déjà vu des peintres travailler ?
- Tes parents sont-ils des commerçants ou travaillent-ils dans le bâtiment ?

De même, à ceux qui devaient s'affronter à l'épreuve où il était question de vendre du muguet le jour du 1er mai, nous posions des questions comme :

- Fais-tu souvent les courses avec tes parents ?
- Pourquoi le 1er mai est-il un jour de congé ?
- Dans ton école, as-tu déjà réalisé des projets pour gagner de l'argent ?

Pour conclure l'exposé des hypothèses émises dans notre première étape de réflexion, il est intéressant de noter qu'elles portent sur des facteurs (l'habitude de certains exercices, la familiarité avec le contexte) qui pourraient être favorables à la réussite des élèves à nos

épreuves. Mais une telle réussite ne constituerait pas un indicateur de la maîtrise de compétences complexes. On serait ramené au phénomène, certes intéressant, mais insuffisant au regard de l'inspiration des *Socles de compétences*, selon lequel un élève arrive à réaliser une tâche lorsqu'il y a été entraîné et lorsqu'elle est présentée d'une manière qui ne le déconcerte pas. Or ce que nous voulons savoir, c'est ce qui peut aider un élève à analyser une situation dont il n'a pas l'habitude.

2.2.4.3.3. Deuxième étape de réflexion : une certaine visée sur la réalité

Puisque les conditions favorables que nous venons de dégager pour la réussite à nos épreuves ne sont pas tout à fait des conditions d'acquisition de véritables compétences complexes, il faut relancer la réflexion. Comme nous allons le voir, cette poursuite du questionnement rejoint un certain nombre de faits qui ressortent de l'analyse fine des réponses des élèves, dans le test que nous avons entrepris durant l'année 2000-2001.

La question qui se pose est de savoir comment il se fait que certains élèves sont capables d'interpréter des situations qui soient vraiment nouvelles pour eux, c'est-à-dire dont ils n'ont pas déjà l'habitude. Qu'est-ce qui permet de sélectionner dans les multiples traits de la situation, ceux qui sont pertinents et qui vont amener le sujet à choisir, parmi les procédures qu'il possède, celles qui conviennent ?

2.2.4.3.3.1. Une erreur : essai d'interprétation

Une erreur fréquente des élèves dans une de nos épreuves va nous mettre sur la voie. Dans l'épreuve consacrée à la peinture des murs de la classe, les élèves doivent trouver le nombre de pots de peinture nécessaires. On leur donne, dans ce but, les dimensions de la classe (longueur, largeur, hauteur) ainsi que des fenêtres et de la porte. Le calcul de la surface à peindre n'est pas difficile, mais il a une relative complexité, puisqu'il faut soustraire de l'aire des murs, celle des fenêtres et celles des parties (porte notamment) sur lesquelles on mettra une peinture de qualité différente. Or il est remarquable qu'un nombre significatif d'élèves n'ont pas cherché à calculer cette surface, mais ont "estimé" intuitivement le nombre de pots nécessaires, et cela, pour la très grande majorité d'entre eux, en sous-estimant cette quantité.

Du point de vue de nos attentes et plus généralement de celles de l'école, ce traitement de la situation n'est pas satisfaisant. Mais il faut reconnaître que c'est une solution qui n'est pas pragmatiquement aberrante. Peut-être même est-ce la plus souvent pratiquée lorsque des non-professionnels décident de repeindre leur appartement. Sur la base de leur intuition ou de leur expérience, ils opèrent une estimation de la quantité de peinture nécessaire. Généralement, ils préfèrent faire cette estimation par défaut plutôt que par excès : si, en cours de travaux, il s'avère qu'on manque de peinture, il sera toujours temps de retourner au magasin pour en acheter. Une pareille pratique peut être inspirée par la volonté d'éviter un calcul de surface jugé fastidieux, mais aussi par une certaine prudence. La quantité de peinture nécessaire pour une surface donnée est approximative : il se peut que la peinture soit, à l'usage, plus "couvrante" que ne l'indique le producteur ; il se peut qu'on s'aperçoive que, finalement, une seule couche suffit, etc. Si la compétence est la capacité à s'acquitter efficacement d'une tâche ou d'un ensemble de tâches, alors cette façon de faire est indéniablement compétente.

Nous disions, plus haut, qu'une compétence complexe comprend nécessairement une opération cognitive qui consiste à sélectionner, dans la multitude des caractéristiques d'une situation, les traits pertinents. Il faut ajouter maintenant que plusieurs sélections sont possibles, correspondant à plusieurs manières de voir la réalité : dans une optique scolaire, la quantité de peinture à acheter passe par le calcul précis de l'aire à peindre ; dans une optique pragmatique, ce sont d'autres considérations qui viendront au premier plan.

2.2.4.3.3.2. Résolution de problème et problématisation

Michel Fabre (1999) fait remarquer qu'il y a une différence notable entre "résoudre un problème" et "problématiser". Dans le premier cas, les données pertinentes ont déjà été sélectionnées dans l'énoncé du problème : c'est le cas le plus ordinaire dans les exercices scolaires traditionnels ; l'élève n'a plus qu'à appliquer à ces données les procédures convenables. Dans le deuxième cas, il s'agit de problématiser, c'est-à-dire, à partir d'une situation non cadrée, de construire le problème, d'en inventer à la fois les termes et la question qu'on se pose. Devant la situation de devoir repeindre les murs d'une salle, on peut se poser la question de savoir quelle quantité de peinture il faut prévoir. Mais on peut aussi se poser la question de savoir dans quel ordre chronologique on va agir et décider de se procurer d'abord une quantité raisonnable de peinture et de commencer le travail, puis d'aviser. Dans cette deuxième manière de construire le problème, la question de la mesure des surfaces à peindre passe au second plan et même peut disparaître.

2.2.4.3.3.3. Différentes visées

On voit donc que ce cadrage de la situation qui intervient dans l'exercice d'une compétence complexe n'est pas seulement, comme nous l'avons dit plus haut, la sélection des éléments pertinents de la situation. Car dire cela, c'est présupposer que les éléments de la situation préexistent. Il serait plus juste de dire qu'ils sont à construire. Le sujet qui aborde une situation complexe construit le problème qu'il se pose, en fonction du sens qu'il donne à la situation. Ce sens dépend de son expérience et de son projet, bref de la visée qu'il a sur la situation. On peut avoir une visée pragmatique, une visée plus savante, voire théoricienne, une visée intuitive, une visée esthétique, une visée émotive, etc. Ce n'est pas la situation qui, par ses caractères supposés préétablis, détermine le sens qu'elle a pour le sujet et donc la mise en œuvre de telle ou telle procédure. C'est l'inverse : c'est le sujet qui en fonction du sens qu'il donne à la situation va faire émerger des caractères qui permettront la mise en œuvre de procédures. Le sens d'une situation ne vient pas de la situation en elle-même, elle vient du sujet et de la visée qu'il a sur elle.

Y a-t-il des visées meilleures que d'autres ? C'est difficile à dire. Ce qui est sûr c'est que l'école essaie d'inculquer une visée particulière, qu'on pourrait appeler la visée instruite. Il n'est pas aisé de préciser, d'une manière exhaustive, en quoi consiste cette vision instruite. On pourrait avancer qu'elle consiste à préférer ce qui est systématique à ce qui est ponctuel, à préférer ce qui est général à ce qui est anecdotique, à préférer ce qui est réflexif à ce qui est spontané, à préférer l'anticipation et la prévision à l'improvisation, le durable au fugitif (et donc souvent l'écrit à l'oral), le rationnel à l'irrationnel, etc.

Cette question de la spécificité de la visée instruite sur le monde amène à revoir ce que nous avons dit, ci-dessus, de la familiarité avec le contexte comme aide à la réussite des épreuves.

Dans certains cas, comme celui que nous avons évoqué, la connaissance de certaines données de contexte peut libérer la capacité attentionnelle. Mais dans d'autres cas, la familiarité avec le contexte peut égarer les élèves en contribuant à leur faire adopter une visée de la vie pratique, plutôt qu'une visée savante. Ainsi dans l'épreuve qui concernait le projet de faire pousser des légumes (épreuve destinée aux élèves de fin du cycle 5-8), on demandait aux élèves de donner au jardin deux formes triangulaires. Or un certain nombre d'entre eux ont cru nécessaire de dessiner, sur le mode réaliste, des choux, des carottes, et autres légumes. Cette volonté de représentation réaliste, qui pouvait témoigner d'une connaissance familière du jardinage, induisait une visée pragmatique ; alors que la visée instruite qui était attendue aurait consisté à ne retenir des alignements de légumes que leur forme géométrique

Mais l'essentiel serait surtout de savoir quelles sont les conditions favorables pour que les élèves intériorisent cette visée à l'école. On pourrait présumer que la maîtrise de procédures de base guide le sujet vers un type de questionnement et un cadrage de la situation. Par exemple le fait de savoir calculer des surfaces peut contribuer à induire l'idée qu'il serait intéressant de pouvoir calculer d'emblée la quantité de peinture nécessaire, avant de commencer le travail.

Cependant, il semble que cela ne suffise pas, car dans notre échantillon certains élèves qui savent calculer des surfaces déterminent cependant d'une manière purement intuitive le nombre de pots de peinture à acheter.

Plus encore, on peut voir que chez des élèves, la maîtrise de certaines procédures de base, loin d'induire la visée attendue sur la situation, les amène à des erreurs. Ainsi, un nombre significatif d'élèves, profitant du fait qu'on leur donne les trois dimensions de la salle, calculent le volume. Tout se passe comme si se mettait en action, d'une manière irrépressible, l'algorithme suivant : "Si je connais la longueur, la largeur et la hauteur, alors je calcule le volume". Le problème est que le volume ne sert à rien dans l'approche de la situation proposée.

Il semble donc qu'il faille, non pas la mise en œuvre d'une ou de plusieurs procédures de base, mais une rencontre entre des procédures et une visée sur une situation. Ou encore, pour le dire autrement, il faudrait que la maîtrise des procédures ne soit pas de l'ordre de l'application mécanique, mais induise, d'une manière dynamique, une visée sur le monde. Par exemple, il faudrait que la procédure "savoir effectuer une multiplication" ne consiste pas à l'effectuer quand le sujet en reçoit la consigne, mais devienne un instrument pour inventorier les situations. Il faudrait qu'elle devienne une sorte de regard multiplicatif, un effort constant pour trouver dans la réalité ce qui pourrait être éclairé ou facilité par cette opération. De même, savoir reconnaître un pronom n'est pas intéressant si l'élève ne met en œuvre cette procédure que lorsqu'on le lui demande explicitement. C'est en revanche une visée qui oriente le regard sur la parole et l'écrit, si elle s'intègre dans un effort de pronominalisation ou plus généralement d'anaphore qui, lui-même, prend place dans la volonté de se faire comprendre d'autrui.

2.2.4.3.3.4. Trois hypothèses

Il n'existe pas actuellement, à notre connaissance, de travaux qui éclairent véritablement le mécanisme d'acquisition de la visée instruite sur la réalité ou plus généralement le processus d'interprétation par un sujet de situations nouvelles pour lui. Nous voudrions émettre à ce sujet trois hypothèses. Les moyens et le temps que nous disposons pour cette recherche ne nous ont pas permis d'opérer le dépouillement systématique des réponses à la phase 1, qui nous aurait permis de confirmer ou d'infirmer nos hypothèses, mais elles sont intéressantes dans la perspective de recherches ultérieures.

2.2.4.3.3.4.1. Le travail en projet

Il s'agit finalement que les procédures qu'on fait apprendre à l'école ne restent pas des rituels scolaires que l'individu n'intègre en aucune manière à sa vision du monde. Si l'on veut que ces procédures soient animées d'un dynamisme et entrent dans la visée que l'élève aura sur la réalité, il peut être utile que ces procédures soient acquises à l'occasion d'un projet que l'élève ait fait sien. Car le projet qu'a un individu joue un rôle majeur dans le sens qu'il donne à une situation et donc dans le mécanisme de cadrage des situations.

Pour cette raison, nous faisons l'hypothèse qu'une pédagogie du projet est favorable à l'acquisition de compétences complexes, parce qu'elle donne l'habitude aux élèves de voir

les procédures apprises à l'école comme des instruments pour réaliser des projets qui leur tiennent à cœur et donc comme des outils pour le comprendre.

Pour éviter toute ambiguïté, précisons que nous appelons pédagogie du projet la pratique qui consiste pour l'enseignant à mobiliser les élèves sur un projet, qui ne soit pas artificiel et intra-scolaire, mais qui débouche sur une production ayant valeur en dehors de l'école (monter un spectacle, publier un journal, préparer un voyage, écrire un roman ou un guide touristique, tourner un film, etc.). Dès lors que le projet est lancé, les élèves se rendent compte par eux-mêmes que sa réalisation exige qu'ils acquièrent certains savoirs et savoir-faire. Non seulement le projet est un facteur de motivation pour les apprentissages, mais en plus ce qui est appris dans ce contexte prend d'emblée aux yeux de élèves la figure d'instruments pour comprendre la réalité et agir sur elle.

Pour vérifier cette hypothèse, nous posons, dans le questionnaire destiné aux instituteurs, la question suivante : - Travaillez-vous souvent par projet ?

2.2.4.3.3.4.2. Les situations-problèmes ou "défis"

Malgré son intérêt évident et éprouvé, la pédagogie du projet comporte un certain nombre de difficultés et de risques qui incitent à dire qu'elle ne devrait pas occuper une place exclusive. En effet, lorsque des enfants ou des adolescents sont profondément impliqués dans une réalisation qui a un sens socialement reconnue, c'est-à-dire lorsqu'ils sont pris dans la dynamique d'un projet, ils fonctionnent dans une logique d'efficacité. Ce qu'ils veulent, c'est arriver, par le chemin le plus court et le plus économique, au résultat attendu. Leur attention est donc focalisée sur les moyens de la réalisation de leur projet, mais il n'y a pas de raison qu'ils s'intéressent aux justifications rationnelles de ces moyens. L'étude des justifications ne peut apparaître que comme une perte de temps et un détour inutile. Le risque est grand que les élèves ne retiennent que des recettes qui leur permettent, au plus court, d'arriver à leur fin. L'action réussie permet de rapporter une procédure à une situation. Elle n'invite aucunement à comprendre pourquoi c'est cette procédure qui convient.

Ce qu'on risque d'éliminer de l'école, en pratiquant une pédagogie du projet trop systématique et exclusive, ce sont les savoirs. Mais on comprend que c'est aussi une visée instruite sur les réalités qui risque d'être ainsi minorée. Tout en restant une pratique pédagogique irremplaçable, la pratique du projet doit être contrôlée par l'enseignant de façon à éviter cette dérive.

Pour cette raison, nous sommes amenés à nous demander si une pratique un peu particulière du projet ne pourrait pas jouer un rôle favorable à la construction de la visée instruite : un projet qui ne débouche pas sur une efficacité extra-scolaire, mais qui ait pour but de résoudre un problème interne aux savoirs scolaires.

C'est ce qu'on trouve dans le dispositif de la situation-problème. Il s'agit, on le sait, de proposer aux élèves une tâche scolaire dont on sait qu'ils ne sont pas en état de la réaliser avec l'état actuel de leurs connaissances, mais qui est proche de cet état et dont ils peuvent comprendre l'énoncé. La situation est clairement présentée comme n'étant pas une tâche destinée à les évaluer, mais comme un défi. C'est d'ailleurs sous ce nom qu'elle est parfois connue chez les enseignants.

On peut espérer d'abord qu'il y ait un pouvoir motivant de ce type de tâche. Dès 1987, Meirieu faisait remarquer qu'une des conditions de l'apprentissage, c'est que l'enseignant sache "créer l'énigme" (Meirieu, 1987). En outre, c'est une manière très intéressante d'introduire auprès de élèves de nouveaux éléments de savoir. Car ceux-ci leur apparaissent d'emblée comme des instruments conceptuels qui permettent de résoudre un problème auquel ils se sont affrontés. Le savoir est donc immédiatement présenté comme compétence et notamment comme ce qui va servir à une visée.

C'est pour opérer un premier repérage de la validité de cette hypothèse que nous avons posé aux enseignants (dans le questionnaire qui leur est destiné) la question suivante : "Mettez-vous souvent les élèves devant des défis ou situations-problèmes ?"

Il n'est pas exclu, toutefois, que la pratique de la situation-problème soit mieux comprise par les enfants venant de milieux culturellement favorisés. Il se pourrait que la visée instruite et donc la possibilité de développer des compétences complexes soient socialement discriminatoires. C'est un point sur lequel le dispositif de recherche actuel ne nous donne pas d'éléments fiables, mais nous pourrions le placer au centre de recherches ultérieures.

2.2.4.3.3.4.3. Le travail en groupes

Les bénéfices cognitifs des interactions sociales ont souvent été signalés. Ils l'ont été par exemple par Bruner qui a mis en évidence le rôle de l'étayage et des rapports avec l'adulte dans les premiers apprentissages de l'enfant (Bruner, 1983, 1987). Ils l'ont été aussi par les spécialistes de psychologie sociale de Genève et Neuchâtel qui ont travaillé particulièrement sur les interactions entre pairs (Doise et Mugny, 1981 ; Perret-Clermont et Nicolet, 1988). Ceux-ci ont repris le concept de conflit cognitif par lequel Piaget décrivait la contradiction interne que peut ressentir un enfant et qui va déboucher sur une rééquilibration majorante, assurant ainsi le progrès cognitif. Ce qu'ont montré Doise, Mugny, Perret-Clermont et Nicolet, c'est que ces conflits cognitifs pouvaient être hâtés par des discussions entre enfants se considérant comme égaux (conflits socio-cognitifs).

Il s'agit, dans tous les cas, du développement d'une intuition déjà formulée par Vigotski (1985) dans les années trente, selon laquelle l'enfant accéderait toujours à une fonction intellectuelle en deux étapes : dans la première, il serait capable d'accomplir l'opération avec les autres et dans la dynamique d'une action collective ; dans la seconde, il deviendrait capable de l'accomplir seul.

On peut donc s'attendre à ce que le travail en groupes, s'il est correctement conduit, joue d'abord sur la motivation des élèves en engendrant un effet d'engagement dans la tâche ou comme on dit parfois "d'enrôlement". D'autre part on peut imaginer que les discussions entre pairs sur un problème posé provoquent, par correction mutuelle, un ajustement progressif de la visée sur la réalité, une approximation de plus en plus exacte de ce qu'il est intéressant de prendre en compte dans une situation inédite.

Bien sûr, c'est une hypothèse qui serait à vérifier et qui exigerait une recherche spécifique, car la recherche actuelle, impliquant une passation individuelle de nos épreuves, ne nous apporte aucun indice sur ce point. Il s'agirait de voir si une situation de travail collective favorise le cadrage des situations et donc le développement de compétences complexes. On pourrait même imaginer un dispositif où l'on pourrait mesurer pour un même élève sa performance à une de nos épreuves en situation de travail individuelle et l'apport du travail en groupe sur sa performance à une épreuve du même type.

Un tel travail permettrait de mesurer objectivement l'impact du travail en groupe sur l'acquisition de compétences complexes. Mais il permettrait aussi d'affiner l'instrument d'évaluation proposée par la présente recherche. Car une des limites de cet instrument est qu'il évalue d'une manière individuelle la performance à des tâches complexes qui sont, dans de nombreuses classes, pratiquées en travail collectif. Il y a donc un décalage entre les conditions de l'apprentissage et les conditions de l'évaluation, décalage qui a souvent été déploré par les instituteurs que nous avons rencontrés.

Troisième partie : présentation et analyse des résultats

Cette troisième partie sera consacrée à la présentation et à l'analyse des résultats provenant de la passation des quatre nouvelles épreuves administrées durant notre deuxième année de recherche. Cette analyse devrait nous permettre d'établir si les constats énoncés la première année sont confirmés ou infirmés. Par la suite, notre analyse se focalisera sur les questions de l'influence sociale, de l'influence du projet pédagogique et de l'influence du mode de passation des épreuves sur les performances des élèves.

Ensuite, nous analyserons le questionnaire, adressé aux enseignants et aux directions concernés par notre épreuve, qui tentait d'apporter des éléments de réponses quant à l'influence ou non sur les performances des élèves de l'adhésion et de l'information de ceux-ci par rapport au concept de cycles.

Enfin, nous ferons une analyse de la faisabilité de l'appropriation de notre modèle en trois phases par des enseignants ayant participé à un module de formation dont l'objectif était la création d'épreuves suivant les principes développés durant cette recherche.

3.1. Résultats et analyse des épreuves

3.1.1. Introduction

Comme nous l'avons énoncé précédemment, l'objectif principal de cette deuxième année de recherche était, en administrant quatre nouvelles épreuves à un échantillon d'élèves, de confirmer, d'infirmier ou de nuancer les constats établis la première année sur base de l'analyse de résultats aux deux premières épreuves.

Pour rappel, quatre constats importants ont été dressés. : la hiérarchie des phases, la confirmation de notre hypothèse sur la difficulté des élèves à affronter des tâches complexes, la cohérence des épreuves et l'importance de la maîtrise des procédures automatisées sur les résultats aux deux autres phases.

Au-delà de cette analyse, nous avons construit un échantillon pour tenter de récolter des informations sur l'influence du milieu social, à l'influence du « projet pédagogique » et à l'influence du mode de passation sur les performances des élèves.

3.1.2. Avant toute chose

Les résultats et les constats qui en découlent proviennent d'une analyse statistique des données recueillies dans l'échantillon. Ce dernier a été construit avec soin mais nul n'est à l'abri d'un effet négatif dû au choix aléatoire des données. Sa construction a été établie pour pouvoir rendre compte de l'influence de différents facteurs: milieu social différent, types de passation des épreuves, existence d'un projet pédagogique positif. Il n'est donc pas à proprement parler, représentatif de toute la population des élèves de la Communauté Française. De plus, il faut rester attentif au fait que les données recueillies et analysées proviennent de quatre épreuves dont les compétences évaluées ont été choisies arbitrairement, que ces épreuves ont été corrigées suivant une procédure spécifique et de ce fait contestable ... Enfin, n'oublions pas que toute interprétation de résultats statistiques est conditionnée par les hypothèses sous lesquelles l'échantillon a été choisi.

Enfin, pour ne pas alourdir la lecture de ce rapport, l'ensemble des résultats statistiques qui ont servi notre étude ne seront pas présentés. Ils sont naturellement à la disposition de toute personne qui souhaiterait les consulter. . Nous nous limitons dans la suite à présenter les résultats qui illustrent directement nos propos et qui se veulent une synthèse de l'ensemble des résultats récoltés.

3.1.3. Analyse des résultats globaux : la confirmation des différents constats

3.1.3.1. La confirmation de l'hypothèse

Tableau 1 : résultats globaux des échantillons 1999-2000 et 2000 - 2001

1216 élèves	Phase 1		Phase 2		Phase 3					
	%	Ecart Type	%	Ecart type	Calcul		Français		Global	
					%	Ecart Type	%	Ecart type	%	Ecart Type
Echantillon complet 1999 – 2000 407 élèves	32.63	22.12	56.38	26.6	71	23.31	73.72	19.31	73.5	18.27
Echantillon complet 2000-2001 809 élèves	46.26	20.96	56.34	24.47	63.66	20.28	66.69	20.28	65.46	17.69
Cycle 5 –8 Echantillon 1999 – 2000 193 élèves	39.44	26.15	53.97	30.45	81.78	18.57	74.14	19.41	78.13	16.67
Cycle 5 – 8 Echantillon 2000-2001 427 élèves	47.18	21.69	57.13	23.06	72.18	21.89	62.33	19.17	67.22	17.02
Cycle 10-12 Echantillon 1999 – 2000 214 élèves	26.48	15.37	59.29	20.75	61.29	23.49	73.34	19.26	69.32	18.67
Cycle 10-12 Echantillon 2000-2001 382 élèves	46.24	20.09	55.46	25.96	54.35	19.8	71.45	20.41	63.55	18.22

La note moyenne obtenue par les élèves en phase 1 dans notre second échantillon montre que la majorité des élèves devant résoudre des tâches complexes leur demandant la mise en place de démarches de résolution sont en réelles difficultés. Ce constat semble confirmer, comme l'année dernière, notre hypothèse. Celle-ci était, pour rappel, que les compétences de résolution de problème qu'exigent les *Socles de compétences*, sont actuellement peu maîtrisées par les élèves et qu'il faudra un temps certain avant que les enseignants modifient certaines pratiques de façon à amener leurs élèves à une véritable maîtrise de celles-ci.

Les résultats de notre deuxième échantillon sont cependant significativement meilleurs que ceux du premier. Ainsi, les 809 élèves de notre deuxième échantillon comprenant la passation de quatre épreuves obtiennent en phase 1 une moyenne de 46.26 %. Elle n'était que de 32.63 % pour les deux épreuves administrées la première année. Deux raisons peuvent être avancées pour expliquer cette différence. La première est que la première tâche de l'épreuve 10-12 de la première année où l'on demandait aux élèves de choisir la formule d'excursion la moins chère s'était avérée trop difficile. Suite à l'analyse des résultats, nous l'avons d'ailleurs modifiée. La seconde raison de cette différence observée repose peut-être sur la conception des grilles de correction. En effet, pour les quatre

épreuves administrées la seconde année, nous avons utilisé la même échelle de mesure répartie de 0 à 8 en respectant des principes quantitatifs les plus objectifs. Ainsi, comme nous l'avons déjà développé, le niveau 8 correspondait pour les tâches des quatre épreuves à une réussite complète, le niveau 0 à une feuille blanche, le niveau 1 à un essai de la tâche ayant aucun rapport avec ce qui était demandé, le niveau 5 correspondait quantitativement à la moitié de la tâche ... Ces critères n'avaient pas été explicitement utilisés lors de la création des deux premières épreuves.

Tableau 2 : Résultats détaillés aux trois phases des échantillons 1999-2000 et 2000-2001

Proportion d'élèves ayant obtenu au moins	1999 - 2000			Proportion d'élèves ayant obtenu au moins	2000 - 2001		
	Phase 1	Phase 2	Phase 3		Phase 1	Phase 2	Phase 3
90	1.97	11.9	18.92	90	1.19	5	3.99
80	2.95	20.96	46.93	80	4.65	19	23.58
70	3.93	35.98	66.58	70	15.39	37	46.78
60	8.85	48.44	78.85	60	25.18	52	66.49
50	21.38	62.61	88.7	50	46.06	64	80.93
40	29.24	72,24	94.1	40	63.25	75	70.72
30	45.95	82.15	96.81	30	76.49	85	95.1
20	63.64	88.1	99.02	20	89.5	91	98.71
10	86.49	93.48	99.51	10	95.11	94	99.74

Les résultats du tableau 2 confirment nos propos. Ainsi, si l'on considère uniquement la phase 1, on remarque dans l'échantillon 2000-2001 que seulement 4.65 % des élèves obtiennent un résultat de 80 % et qu'ils ne sont que 46.06 % à obtenir un résultat égal ou supérieur à 50 %. Nous avons pu constater, dans une moindre mesure que l'année dernière, qu'une grande majorité d'élèves dans cette phase de résolution de problèmes étaient en difficulté.

3.1.3.2. La hiérarchie des phases

La lecture des deux premiers tableaux conforte notre idée d'une hiérarchie entre les trois phases.

En effet, aussi bien dans notre premier échantillon que dans notre second, les élèves sont meilleurs en phase 2 lorsqu'on leur propose des tâches décomposées nécessitant une aptitude à choisir la procédure qui correspond à une tâche simple mais inédite, qu'en phase 1 lorsqu'on leur demande, pour les mêmes tâches, de saisir dans une situation nouvelle les traits pertinents qui indiquent qu'il y a lieu de choisir et de combiner d'une manière originale des procédures connues. Comme lors de notre premier échantillon, ils sont systématiquement meilleurs dans la phase 3 qui évalue les procédures de base normalement automatisées nécessaires aux autres phases.

Les résultats en phase 2 (tableau 1) du premier et second échantillons, que ce soit pour ceux obtenus par les élèves du 5-8 ou ceux obtenu par les élèves du 10-12 sont remarquablement semblables. Ainsi, la moyenne obtenue lors de la première année était de 56.38 %. Il est pour le nouvel échantillon de 56.34 %. Les résultats présentés par cycle confirme cette homogénéité. La proportion d'élèves (tableau 2) ayant obtenu 80 % était de

20.96 % la première année et est de 19 % cette année. 62.21 % des élèves du premier échantillon avaient obtenu un résultat égal ou supérieur à 50 %. Ils sont 64 % cette année.

Ces résultats issus d'un échantillon comprenant un nombre important d'élèves sont très concordants en phase 2 et donneraient une cohérence à la construction de celle-ci.

Enfin les résultats en phase 3 restent significativement supérieurs à ceux obtenus aux autres phases. Les résultats enregistrés cette année – 65.46 % - sont moins bons que ceux de l'année dernière – 73.5 % - mais restent cependant en conformité avec les résultats que nous pouvions attendre.

3.1.3.3. La cohérence de l'épreuve

Comme l'année dernière, nous avons examiné le lien de dépendance entre les différentes phases. Pour ce faire, nous avons réalisé un ensemble de calculs corrélationnels qui nous permettent de rédiger les constats suivants qui recourent, en partie, ceux énoncés la première année.

Tableau 3 : corrélation entre phases pour l'échantillon global 2000-2001

	Phase 1	Phase 2	Phase 3
Phase 1		0.464	0.489
Phase 2	0.464		0.627
Phase 3	0.489	0.627	

Les différentes corrélations calculées entre les 3 phases affichent un taux très significatif. Ceci indique qu'il existe un lien important entre les différentes phases. Ceci renforce l'idée d'une cohérence du modèle basé sur le rapport entre les différentes phases.

Ces résultats et l'analyse des tableaux de contingence nous permettent de réaliser un ensemble de constats fort semblables à ceux énoncés la première année.

Pour lire nos tableaux de contingence, il faut comprendre que nous avons arbitrairement découpé les résultats de notre échantillon en trois groupes. Le premier comprend le groupe dit des « forts » qui ont obtenu pour la ou les phases observées plus de 65 %. Le deuxième groupe reprend les élèves ayant des résultats compris entre 35 % et 65 % et nous l'appelons « moyens » et le troisième est constitué des élèves ayant obtenu un résultat inférieur à 35 % que nous avons appelé « faibles ».

	Phase 1	Phase 2	Phase 3
Bon	22 %	44 %	56%
Moyen	47 %	35 %	36%
Faible	31 %	21%	7%

Le croisement entre ces résultats à trois niveaux de deux phases nous donne donc neuf catégories. Les forts dans les deux phases, les forts dans une phase et moyen dans une autre, les forts dans un phase et faibles dans l'autre ...

Les résultats affichés dans les tableaux de contingence reprennent le pourcentage d'élèves appartenant à chaque catégorie. Par exemple, si nous observons le tableau de contingence entre la phase 1 et la phase 2, nous constatons que 17 % des élèves sont forts en phase 1 et en phase 2. Ils sont 4 % à être moyen en phase 2 et forts en phase 1 et aucun élève n'est faible en phase 2 et fort en phase 3.

Voici ces tableaux de contingence et les diagrammes qui y sont associés:

Tableau 4 : tableaux de contingence

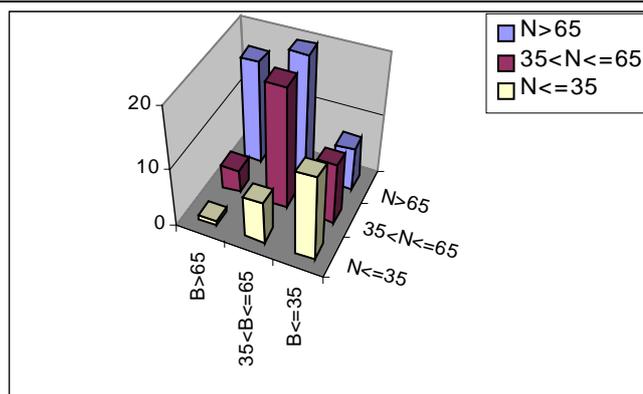
Phase 1 (B) / Phase 2 (N)

B>65
35<B<=65
B<=35

N>65
17
20
7

35<N<=65
4
20
10

N<=35
1
7
14



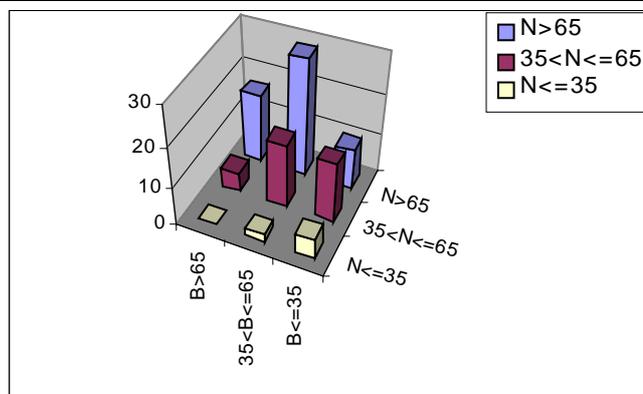
Phase 1 (B) / Phase 3 (N)

B>65
35<B<=65
B<=35

N>65
17
29
10

35<N<=65
5
16
15

N<=35
0
2
5



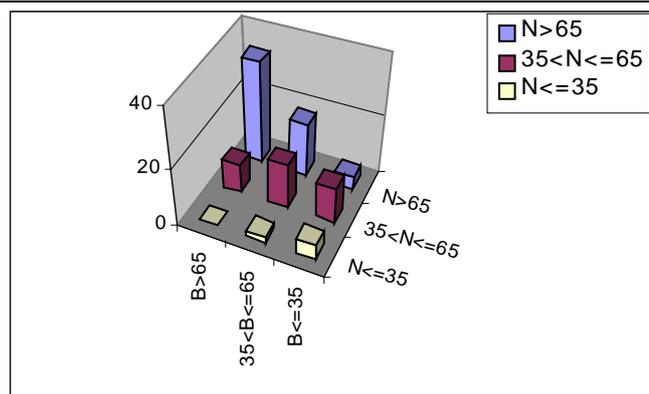
Phase 2 (B) / Phase 3 (N)

B > 65
35 < B ≤ 65
B ≤ 35

N > 65
34
18
4

35 < N ≤ 65
10
15
12

N ≤ 35
0
2
5



Voici maintenant les constats que nous pouvons énoncer :

- Plus on est performant en phase 2 et plus on est performant en phase 1.
- La majorité des élèves forts en phase 1 le sont également en phase 2 et en phase 3. Ils représentent 17 % de l'échantillon. Toutefois, 5 % des élèves ne respectent pas cette règle.
- 38 % des élèves forts en phase 2 restent forts en phase 1. 45 % d'entre – eux deviennent moyens et 16 % faibles.
- 59 % des élèves moyens en phase 2 restent moyens en phase 1 et environ 30 % des élèves moyens deviennent faibles en phase 1.
- La moitié des élèves forts en phase 2 et en phase 3 – 34 % de l'échantillon - ne le sont plus en phase 1.
- 40 % des élèves forts en phase 3 ne le sont plus en phase 2.
- La majorité des élèves forts en phase 2 le sont également en phase 3. Ils représentent 34 % de l'échantillon. 22 % de ces élèves - 10 % de l'échantillon - ne respectent pas cette règle.
- Plus on est performant en phase 3 et plus on est performant en phase 2

- 60 % des élèves forts en phase 3 le sont également en phase 2. Ils ne sont plus que 30 % en phase 1 à être considérés comme forts.
- Environ 70 % des élèves forts en phase 3 ne le sont plus en phase 1. 51 % sont moyens et 19 % sont faibles.
- Aucun élève faible en phase 3 n'est fort en phase 2 ou en phase 1

Les analyses corrélationnelles de l'échantillon de la deuxième année nous permettent de redire que les procédures automatisées sont une condition nécessaire à la réussite des autres phases et que la capacité de choisir la bonne procédure est également nécessaire pour la résolution de tâches complexes. On constate également une « perte de compétences » d'une phase à l'autre en prenant comme point de départ de notre analyse la phase 3.

La différence importante observée dans l'analyse de ce second échantillon par rapport au premier se situe dans le fait qu'une majorité d'élèves moyens en phase 2 le reste en phase 1 alors que l'on avait observé dans le premier échantillon que la majorité d'entre eux devenaient faibles. Cette différence s'explique par le fait que les résultats obtenus en phase 1 pour les raisons déjà développées, sont meilleurs. Sinon, les autres résultats confortent notre première analyse.

La forte liaison entre les différentes phases, la comparaison des résultats entre le premier et le second échantillon et les résultats fort semblables obtenus par les élèves du cycle 5 – 8 et ceux du 10-12 seraient des signes de la cohérence de nos épreuves.

3.1.4. Les résultats par épreuve : observation de différences

Les analyses des résultats de chaque épreuve sont-elles conformes aux résultats globaux et sont-elles cohérentes entre elles?. Les résultats présentés dans le tableau 5 montrent certaines différences sur lesquelles il nous semble important de s'arrêter.

3.1.4.1. Constats

En phase 1, on observe clairement une augmentation des résultats des épreuves du second échantillon (épreuves 2,3,5 et 6) Ainsi, les résultats varient pour celui-ci entre 43.29 % et 50,68 %. Des explications quant aux raisons susceptibles d'expliquer cette augmentation ont été avancées ci - avant. Le pourcentage d'élèves ayant obtenu un résultat égal ou supérieur à 80 % dans les phases 1 des différentes épreuves varie entre 2.78 % et 7.25 %. A 50 %, les résultats sont compris entre 39.35 % et 53.62 %.

En phase 2, on observe des résultats assez contrastés allant de 49.19 % pour l'épreuve « Les murs de la classe » à 63.45 % pour celle « le premier mai ». Sur les quatre épreuves du nouvel échantillon, trois obtiennent des résultats supérieurs à ceux enregistrés l'année dernière. L'épreuve « les murs de la classe » obtient un résultat significativement inférieur qui influence considérablement la moyenne du nouvel échantillon. Dans la phase 2 de ces épreuves, le pourcentage d'élèves ayant obtenu au moins 80 % varie entre 16.9 % et 22.03 % et le pourcentage de ceux ayant obtenu au moins 50 % varie entre 39.35 % et 53.62 %.

Tableau 5 : résultats globaux pour chaque épreuve

Cycle 5 - 8												
	Phase 1				Phase 2				Phase 3			
	%	Ecart type	Réussite à 50 %	Réussite à 80 %	%	Ecart type	Réussite à 50 %	Réussite à 80 %	%	Ecart type	Réussite à 50 %	Réussite à 80 %
1	Une invitation au théâtre											
	39.44	26.1	31.61	6.22	53.97	30.45	57.51	26.42	78.13	16.67	92.75	58.03
2	St Nicolas											
	44.09	21.56	42.74	2.99	56.98	22.82	64.76	22.03	63.63	18.2	77.78	22.22
3	Le jardinage											
	50.68	21.35	53.62	7.25	57.29	23.39	64.65	20.2	71.32	14.56	89.42	29.1

Cycle 10 -12												
	Phase 1				Phase 2				Phase 3			
	%	Ecart type	Réussite à 50 %	Réussite à 80 %	%	Ecart type	Réussite à 50 %	Réussite à 80 %	%	Ecart type	Réussite à 50 %	Réussite à 80 %
4	L'excursion à Bruxelles											
	26.48	15.37	12.15	0	59.29	20.75	68.75	14.38	69.32	18.67	85.05	36.92
5	Les murs de la classe											
	43.29	18.23	39.35	2.78	49.19	28.17	52.58	16.9	68.08	17.36	85.58	31.25
6	Le premier mai											
	47.56	21.92	49.72	6.08	63.45	20.26	77.25	18.56	57.77	17.71	69.33	9.2

En phase 3, les résultats avoisinent les 70 % exceptés ceux obtenus par les élèves ayant réalisé l'épreuve « La visite de St Nicolas » et surtout ceux ayant répondu à celle appelée « le premier mai ». Plus inquiétant est le résultat des élèves qui ont passé l'épreuve « le premier mai » qui en phase 3 est inférieur au résultat de la phase 2. Dans la phase 3 des différentes épreuves 22.2 % et 31.25 % des élèves obtiennent un résultat égal ou supérieur à 80 % excepté les élèves qui ont participé à l'épreuve « Le premier mai » dont seulement 9.2 % ont un résultat égal ou supérieur à ces 80 %. A 50 %, les résultats se situent entre 69.33 % et 89.42 %.

Si on calcule la progression de la phase 1 à la phase 2 pour chaque épreuve, nous constatons également des différences allant de 6 % à 23 %.

Les résultats décrits ci-dessus montrent clairement des différences significatives entre les épreuves. Même si la confirmation de notre hypothèse reste d'actualité et que la hiérarchie est observée – excepté pour l'épreuve « le premier mai »-, nous estimons important d'analyser celles-ci..

3.1.4.2. Les résultats aux épreuves sont statistiquement différents

Une analyse de la variance univariée des épreuves de notre deuxième échantillon en phase 1 et en phase 2 démontre statistiquement que les résultats obtenus aux différentes épreuves sont significativement différents au niveau 0.5 pour les deux phases. Ce constat n'a rien d'étonnant car il aurait été quasiment suspect d'obtenir des résultats statistiquement équivalents pour des épreuves évaluant des compétences différentes. Le contenu, la longueur des tâches proposées, la familiarité des élèves avec celles-ci et d'autres facteurs influencent le niveau de difficultés des épreuves.

Ces résultats statistiques confirment d'ailleurs des impressions récoltées lors de la passation des tests dans les écoles. Ainsi, nous avons rapidement constaté que l'épreuve « Les murs de la classe » et l'épreuve « La visite de St Nicolas » étaient plus difficiles que les deux autres. D'une manière assez heureuse, nous avons pu constater en réalisant cette analyse de la variance que les résultats aux deux épreuves dont nous avons l'impression d'être plus difficiles étaient statistiquement équivalents ainsi que les résultats aux épreuves nous semblant les plus faciles.

Il est important de souligner que ce hasard a sans nul doute influencé les résultats globaux par cycle dont nous vantions les similitudes. Deux épreuves difficiles au cycle 5-8 et deux faciles au cycle 10-12 auraient sans doute compromis ce constat.

Des résultats différents d'une épreuve à l'autre, du moins pour les phases 1 et 2, semblent compréhensibles. Ce qui l'est moins, ce sont les différences de progression dans les six épreuves proposées entre la phase 1 et la phase 2 observées et les différences de résultats entre les épreuves en phase 3. Nous nous proposons d'analyser celles-ci.

3.1.4.3. Analyse des phases 1

Les tests statistiques nous ont montré une différence significative entre les épreuves. En phase 1, même si les résultats obtenus ne modifient pas le constat que la majorité des élèves sont en difficulté lorsqu'ils sont confrontés à des tâches complexes, nous pouvons néanmoins observer des différences de résultats dans la comparaison des épreuves entre-elles. Si l'on compare chaque tâche de chaque épreuve avec les autres, nous observons des différences de résultats très significatives. Ainsi, si l'on tient compte des résultats des douze tâches proposées dans les quatre épreuves, nous pouvons lire dans le tableau 6 des résultats variant de 63.31 % de moyenne à 24.04 %.

Sur base des résultats repris dans le tableau 6 et le tableau 7, nous allons tenter d'émettre des hypothèses pour justifier la difficulté ou la facilité d'une tâche proposée en phase 1.

Tableau 6 : classement des tâches de la mieux réussie à la moins bien réussie

Tâche 1	les murs de la classe	Rédaction d'une lettre	63.31 %
Tâche 3	St Nicolas	Le portrait	57.32 %
Tâche 2	Le jardinage	Le plan du jardin	56.94 %
Tâche 1	Le premier mai	Calcul du gain	54.21 %
Tâche 2	Le premier mai	Choix du cadeau	53.8 %
Tâche 1	Le jardinage	Choix des graines	51.99 %
Tâche 2	St Nicolas	Choix du cadeau	50.91 %
Tâche 3	Le jardinage	La liste des légumes	43.12 %
Tâche 3	Les murs de la classe	Rédaction du bon de commande	36.57 %
Tâche 3	Le premier mai	L'élocution par écrit	34.67 %
Tâche 2	Les murs de la classe	Calcul du nombre de pots de peinture	29.98 %
Tâche 1	St Nicolas	Calcul du nombre de pièces	24.04 %

Les deux tâches les mieux réussies sont des tâches de rédaction en général familières dans les classes. Pour la rédaction de la lettre, nous pouvons constater que seulement 7 % des élèves ne sont pas rentrés dans la tâche. Ils sont 17 % pour la rédaction du portrait. Il est intéressant de constater que ce sont deux tâches de rédaction, ne faisant pas appel à la lecture de documents, qui obtiennent les meilleurs résultats.

Les autres tâches faisant appel uniquement à des compétences de lecture et de rédaction sont « L'élocution par écrit » et « La liste des légumes ». Une caractéristique commune de ces deux tâches est l'importance donnée à la prise d'informations dans des documents annexes. Dans la première, des compétences en lecture, en synthèse et en rédaction sont nécessaires. Dans la seconde, ce sont des compétences en lecture et en classement de l'information qui sont demandés. Nous observons clairement que ces tâches ont été bien moins bien réussies que les premières. 30 % des élèves pour l'élocution par écrit ne sont pas rentrés dans la tâche et 35 % pour la liste des légumes.

Les deux tâches les moins bien réussies sont des tâches de résolution de problèmes demandant la prise pertinente d'informations dans des documents et la combinaison de concepts mathématique de numération, de grandeurs et de géométrie (dans le cas du calcul du nombre de pots de peinture). Ces deux tâches ont également comme caractéristiques d'être longues. 75 % des élèves pour le calcul du nombre de pots de peinture sont rentrés dans la tâche mais seulement 15 % arrivent à accomplir la tâche à moitié. Une difficulté de cette tâche semble également être la difficulté des opérations. 15 % des élèves n'ont rien répondu.

Pour le calcul du nombre de pièces, 53 % des élèves ne rentrent pas dans la tâche. La non compréhension de celle-ci par les élèves semble être la raison principale de son résultat très faible. En effet, il est intéressant de constater que 17 % de l'échantillon n'a rien répondu.

Le calcul du gain de l'épreuve du premier mai est également à nos yeux une tâche de résolution de problèmes demandant la prise pertinente d'informations dans des documents. Ce qui la différencie des deux précédentes est qu'elle présente un aspect assez couramment rencontré dans les classes et ne fait appel qu'à des concepts de numération. Ceci explique, peut-être, son résultat significativement supérieur aux deux autres. 17 % des élèves ne sont cependant pas entrés dans la tâche.

Tableau 7 : résultats phase 1 par tâches

Proportion d'élèves qui ont obtenu au moins	Les murs de la classe			Le premier mai		
	Tâche 1	Tâche 2	Tâche 3	Tâche 1	Tâche 2	Tâche 3
8	9.26	1.39	1.39	5.52	17.68	4.97
7	32.87	1.85	12.04	29.28	25.41	12.15
6	35.65	4.63	25.46	40.33	35.91	12.71
5	69.91	7.41	36.11	47.51	35.91	19.89
4	76.39	15.74	40.28	56.91	62.98	25.41
3	88.89	45.83	51.39	75.14	79.56	47.51
2	94.44	76.85	56.94	82.87	82.87	70.72
1	99.07	86.11	68.98	96.13	90.06	83.98
0	100	100	100	100	100	100

Proportion d'élèves qui ont obtenu au moins	St Nicolas			Le jardinage		
	Tâche 1	Tâche 2	Tâche 3	Tâche 1	Tâche 2	Tâche 3
8	0	11.54	19.23	5.8	5.8	6.28
7	1.28	19.66	24.79	34.78	5.8	18.36
6	3.85	35.9	47.44	34.78	43.48	28.99
5	6.41	38.03	53.85	35.75	51.69	34.78
4	12.82	64.1	63.68	64.73	79.71	40.58
3	36.32	69.66	78.63	71.98	85.99	56.52
2	47.86	81.62	82.48	75.36	86.47	64.25
1	83.76	86.75	88.46	92.75	96.62	95.17
0	100	100	100	100	100	100

La rédaction du bon de commande est une tâche dont les résultats des élèves ne sont guère brillants. Très longue, faisant appel à des prises d'informations dans des documents et à des opérations, elle présente sans doute pour la majorité des élèves un caractère inédit. En effet, 43 % des élèves ne sont pas rentrés dans la tâche.

Le choix de cadeau dans l'épreuve du premier mai et de St Nicolas sont des tâches demandant une prise d'informations dans des documents et des opérations normalement peu complexes. Une caractéristique de ces deux tâches est qu'elles étaient assez courtes. Il est intéressant de constater que pour ces deux tâches fort semblables adressées à des élèves de cycles différents, nous obtenons des résultats forts proches.

Le plan du jardin et le choix des graines sont des tâches qui nous semblaient complexes. Le plan du jardin demande une connaissance de concepts géométriques, une lecture rigoureuse des consignes et une logique d'esprit. Le choix des graines demande une compétence en lecture et une maîtrise de certains concepts de temps. Ce ne sont des

tâches pas trop longues. Comparées aux autres tâches, les résultats obtenus sont bons. Cependant, nous devons souligner que pour la tâche « le choix des graines », l'élève était invité à sélectionner 3 noms de légumes parmi une liste de 6 légumes sans justifier sa démarche. Le hasard a peut-être faussé les résultats. Pour le plan du jardin, la grille de correction est sans doute généreuse.

Enfin il est important de constater que trois tâches sur 4 obtiennent des résultats situés vers le fond du classement. Nous pouvons supposer, lors de la réalisation des épreuves, que des élèves ont vécu un sentiment de lassitude une fois arrivé à cette dernière tâche.

Les commentaires que nous venons d'énoncer, sans être exhaustifs, permettent d'émettre des hypothèses tentant d'expliquer les différences de difficultés observées d'une tâche à l'autre. La longueur de la tâche, le type de tâches, la familiarité avec celles-ci, la complexité des opérations, la prise d'informations dans des documents, les grilles de correction, la place de la tâche dans l'épreuve sont autant de facteurs susceptible d'influencer la complexité de la tâche.

3.1.4.4. Analyse des phases 2

Les résultats en phase 2 varient également d'une épreuve à l'autre. Ainsi les élèves du cycle 10-12 ayant passé l'épreuve « Les murs de la classe » ont obtenu une moyenne de 49.19 % et ceux qui ont été confrontés à l'épreuve « Le premier mai » obtiennent une moyenne de 63.45 %. Pour essayer de comprendre cette différence, il est important de rappeler que les tâches présentées en phase 2 sont les mêmes tâches que celles proposées en phase 1. La différence fondamentale entre ces deux phases est qu'en phase 1 l'élève est obligé de construire sa démarche alors qu'en phase 2, soit il est explicitement guidé par une décomposition de la tâche en items, soit un ensemble d'indications supplémentaires lui sont fournies.

Ce rappel indique clairement que les tâches des phases 2 ne peuvent être analysées qu'au regard de leur tâche 1 respective. Ce n'est donc pas le résultat global qui nous intéresse au premier plan pour comparer les tâches de nos phases 2, mais la différence de résultats entre une tâche en phase 1 et cette même tâche en phase 2. C'est ce que nous avons appelé le gain.

Le tableau 8 nous montre des différences importantes dans le gain entre les deux phases. Il est compris entre un gain négatif - 0.52 % pour la tâche de rédaction de la lettre de l'épreuve « Les murs de la classe » à un gain positif de 22.8 % pour la tâche du calcul du nombre de pots de peinture de cette même épreuve.

L'analyse des tâches et des grilles de correction nous permet d'émettre une hypothèse pour essayer d'expliquer ce constat. En effet, cette analyse nous a permis de dégager trois types de construction de tâches.

- La moitié des tâches proposent aux élèves une décomposition systématique de la tâche en items. Le résultat obtenu par l'élève correspond aux nombres d'items réussis.
- 1/3 des tâches proposées aux élèves sont les tâches proposées en phase 1. La grille de correction est la même. La différence entre les deux phases de ces tâches se situent dans le fait qu'en phase 2, un ensemble d'informations supplémentaires devant les aider dans leur démarche leur sont proposées.
- 2 tâches (1/6 des tâches) sont ce que l'on pourrait appelé des tâches mixtes. Un ensemble d'items sont proposés aux élèves avant la réalisation de la tâche qui est identique à celle proposée en phase 1. Le résultat obtenu provient de l'addition des items réussis et de la note obtenue au regard de la grille de correction (échelle de 0 à 8) utilisée en phase 1.

Tableau 8 : Résultats aux tâches de la phase 1 et de la phase 2

	Phase 1		Phase 2			Gain	Corrélation
	%	Ecart type	%	Ecart type	% 3 tâches Poids iden ² .	En %	
St Nicolas							
1. Tâche 1 Calcul du nombre de pièces	24.04	6.54	45.81	26.29		21.77	0.356
2. Tâche 2 Choix du cadeau	50.91	10.52	69.34	29.16		18.43	0.327
3. Tâche 3 Le portrait	57.32	10.96	59.03	33.32		1.71	0.349
Ensemble des tâches	44.09	21.56	56.98	22.82	58.06	12.89 ou 13.97	0.431
Le jardinage							
4. Tâche 1 Le choix des graines	51.99	10.61	58.07	22.07		6.08	0.26
5. Tâche 2 Le plan du jardin	56.94	8.08	65.73	23.88		8.79	0.212
6. Tâche 3 La liste des légumes	43.12	10.48	52.96	32.23		9.84	0.551
Ensemble des tâches	50.68	21.35	57.29	23.39	58.92	6.61 ou 8.24	0.0529
Les murs de la classe							
7. Tâche 1 Rédaction de la lettre	63.31	8.18	62.79	28.02		- 0.52	0.521
8. Tâche 2 Calcul du nombre de pots	29.98	6.5	52.78	21.82		22.8	0.587
9. Tâche 3 Rédaction du bon de commande	36.57	11.08	45.34	39.4		8.77	0.46
Ensemble des tâches	43.29	18.23	49.19	28.17	53.63	5.9 ou 10.34	0.575
Le premier mai							
10. Tâche 1 Calcul du gain	54.21	9.99	72.43	19.24		18.22	0.575
11. Tâche 2 Choix du cadeau	53.8	10.56	53.97	34.93		0.17	0.625
12. Tâche 3 L'élocution par écrit	34.67	9.28	41.47	32.56		6.8	0.283
Ensemble des tâches	47.56	21.92	63.45	20.26	55.95	15.89 ou 8.39	0.543

² Le pourcentage « 3 phases poids identique » représente la moyenne des résultats aux trois tâches lorsque l'on donne le même poids aux trois tâches.

Sur base de ce groupement (tableau 9), nous pouvons constater de manière évidente que le type de construction des tâches proposées en phases 2 a une influence sur les gains obtenus. Ainsi les tâches proposant une décomposition item par item sont celles qui obtiennent le gain moyen le plus élevé. Les deux tâches utilisant un système de cotation « mixte » obtiennent un gain moyen quasiment deux fois moins élevé que dans notre premier groupement. Enfin, les tâches où l'élève reçoit une tâche en phase 2 totalement identique à celle proposée en phase 1 et pour laquelle on utilise la même grille de correction obtiennent un gain minime.

Tableau 9 : Moyenne des gains entre la phase 1 et 2

	Tâches concernées	Moyenne du gain en %
Correction items par items	1,2,4,6,8,10	16.19
Correction identique phase 1	3,7,11,12	2.04
Correction « mixte »	5,9	8.78

Les corrélations que nous avons calculées entre la phase 1 et la phase 2 pour chacune des tâches sont significativement élevées. Pour toutes les tâches, il existe un lien important.

Les tâches décomposées en phase 2 où l'élève ne doit construire aucune démarche de résolution sont les mieux réussies. Ces tâches sont d'après nous celles qui correspondent le mieux à la définition de la phase 2 telle que proposée dans notre modèle. Il semblerait que les tâches demandant la mise en place d'une démarche, même restreinte, posent des problèmes.

Pourtant, même si les tâches décomposées sont celles qui correspondent mieux à la définition de la phase 2 telle que proposée dans notre modèle, nous n'avons pas pu éviter, pour des tâches d'expression par exemple, l'utilisation des autres types d'élaboration.

3.1.4.5. Analyse des phases 3

Des différences de résultats au phase 3 entre les six épreuves sont également constatées. Les résultats globaux de ces phases varient pour celles-ci entre 57.77 % et 78.13 %. Ces variations sont encore plus importantes si l'on se penche sur les résultats en calcul qui se situent entre 47.82 % et 81.78 %.

Tableau 10 : résultats en phase 3

	français		Calcul		Global	
	%	Ecart type	%	Ecart type	%	Ecart type
Épreuve 1	Une invitation au théâtre					
	74.14	19.41	81.78	18.57	78.13	16.67
Epreuve 2	La visite de St Nicolas					
	64.85	18.72	61.55	22.29	63.63	18.2
Epreuve 3	Le jardinage					
	59.44	19.33	84.33	13.54	71.32	14.56
Epreuve 4	L'excursion à Bruxelles					
	73.34	19.26	61.29	23.49	69.32	18.67
Epreuve 5	Les murs de la classe					
	73	19.85	59.47	18.56	66.08	17.36
Epreuve 6	Le premier mai					
	69.47	20.99	47.82	19.44	57.77	17.71

Une hypothèse pouvant expliquer ces différences est sans doute le contenu des items proposés. En effet, les items des phases 3 évaluant des procédures normalement automatisées ont été construits au regard des tâches proposées dans les phases 1 et 2. De ce fait, les procédures demandées sont différentes d'une épreuve à l'autre et cela peut expliquer des niveaux de difficultés différents. Très précisément, l'analyse des épreuves en calcul montre que les élèves qui ont répondu à l'épreuve « Le premier mai » ont extrêmement mal répondu aux quatre items demandant l'utilisation de pourcentage ou de fractions. Ces items ont été uniquement présentés à ce groupe d'élèves.

Les résultats observés en phase 3 sont, exception faite de l'épreuve « Le premier mai », supérieurs aux résultats obtenus en phase 2. Cependant, nous pouvons constater que ces résultats ne sont pas extrêmement brillants. La raison de ceci, qui a d'ailleurs soulevé un débat au sein de notre équipe, est que nous avons pris le parti de proposer des items évaluant des procédures automatisées devant normalement être utilisées dans la réalisation des phases 1 et 2, mais ne correspondant pas toujours au réel niveau de difficulté rencontré dans la réalisation de celles-ci. En d'autres termes, de nombreux items proposés en phase 3 demandaient une maîtrise supérieure des procédures automatisées que celles normalement utilisées dans la réalisation des phases 1. Ce choix peut s'expliquer par notre volonté, dans un souci d'analyse des relations entre les phases, de discriminer les élèves entre eux. Notre dispositif nous a permis d'observer très clairement que c'était bien les élèves qui maîtrisaient mieux les procédures automatisées proposées qui obtenaient les meilleurs résultats en phase 1 et en phase 2. Les corrélations calculées entre les résultats en calcul et les résultats en phase 2 (0.347) - ainsi qu'entre les résultats en français et ces mêmes résultats en phase 2 (0.67) présentent un taux très significatif à moins de 1 %. Ceci indique un lien important entre les résultats en français et en calcul et ceux de la phase 2. Des épreuves plus faciles en phase 3, mais correspondant mieux à la philosophie de notre modèle, ne nous auraient peut être pas permis d'énoncer ce constat, à nos yeux, essentiel.

Toutefois, suite à nos différents échanges, nous sommes arrivés à la conclusion, qu'il était préférable, si d'aventure d'autres épreuves construites selon nos principes devaient voir le jour, de proposer des items en phase 3 dont la difficulté correspondrait le plus possible aux procédures automatisées réellement mises en jeu dans les autres phases. C'est pour cette raison que nous avons décidé de modifier certains items des différentes phases 3

Enfin, il est important de signaler que l'analyse par item de la phase 3 en calcul de l'épreuve « le premier mai » a mis en évidence que 8 items sur les 20 avaient été très mal réussis. L'épreuve s'est avérée trop discriminatoire.

3.1.4.6. Conclusions des analyses par épreuve

Les analyses des épreuves, sans remettre en cause les constats réalisés sur l'échantillon global et les échantillons par cycle, ont mis en évidence des différences dans les résultats obtenus à chaque épreuve. Nous avons voulu, le plus objectivement possible, tenter de comprendre celles-ci et les constats ou hypothèses que nous avons formulés nous apportent des éléments nouveaux dans la gamme de conseils que nous pourrions énoncer pour rédiger rigoureusement des épreuves diagnostiques.

Ainsi, pour le choix des tâches en phase 1, il est important de tenir compte du type de tâche, de sa longueur, de sa familiarité, ... avec les activités proposées en classe. En phase 2, il est important d'analyser les résultats en référence au système de correction utilisé. Enfin, en phase 3, il nous semble plus judicieux de proposer des items dont la difficulté correspond aux procédures automatisées réellement rencontrées dans les autres tâches.

Dans une perspective diagnostique, les différents constats énoncés devraient permettre aux enseignants d'analyser les épreuves – que ce soit les nôtres ou celles qu'ils auraient construites – en tenant compte de nos constats.

3.1.5. Les influences

A la suite de notre première année de recherche, des questions quant aux influences possibles sur les résultats obtenus par les élèves ont été soulevées. Nous pouvons les résumer en ces termes :

1. Existe-t-il réellement une influence du milieu social de l'école sur les compétences des élèves à résoudre des tâches complexes ?
2. Y a-t-il une influence de la dynamique de l'équipe pédagogique d'une école sur les compétences des élèves à résoudre des tâches complexes ?
3. Les modalités d'administration des épreuves ont-elles une influence sur les résultats des élèves ?

Pour tenter d'apporter des éléments de réponses à celles-ci, nous avons construit un échantillon reprenant les critères suivants : réseau, milieu social, orientations pédagogiques de l'école. Au sein de chaque catégorie, nous avons administré les épreuves suivant des procédures différentes : chercheur, chercheur et enseignant, enseignant seul .

Voici des éléments de réponses que cette recherche, en tenant compte de ses limites méthodologiques, apporte.

3.1.5.1. Existe-t-il réellement une influence du milieu social sur les compétences des élèves à résoudre des tâches complexes ?

Tableau 11 : résultats par milieu social

809 élèves	Phase 1		Phase 2		Phase 3					
	%	Ecart Type	%	Ecart type	Calcul		Français		Global	
	%	Ecart Type	%	Ecart type	%	Ecart Type	%	Ecart type	%	Ecart Type
Milieu social favorisé	47.75	21.8	58.39	23.97	65.25	22.49	69.44	19.24	67.65	22.49
Milieu social défavorisé (ZEP)	44.24	19.61	53.55	24.9	64.47	22.9	62.91	21.08	62.46	18.24

3.1.5.1.1. Premiers constats

Les résultats affichés dans les tableaux 11 et 12 indiquent que les élèves des milieux favorisés obtiennent des résultats supérieurs aux élèves issus des milieux défavorisés. Cette différence tourne aux alentours de 5 % en phase 2 et en phase 3 et est d'environ 3 % en phase 1. Le pourcentage d'élèves fréquentant une école favorisée obtenant une note égale ou supérieure à 80 % est de 5.39 %. Il est de 3.65 % pour les élèves fréquentant une école « défavorisée »

50 % des élèves du milieu favorisé obtiennent un résultat égal ou supérieur à 50 %. Ils ne sont que 40.73 % pour le milieu défavorisé. La lecture du tableau montre si l'on subdivise notre échantillon en trois groupes (fort : + de 65 % ; moyen : entre 35% et 65 % et faible : - de 35 %) que la différence en phase 1 entre les deux groupes se creusent réellement auprès des élèves moyens. En phase 2 et 3, les écarts se creusent dès « les 80 % ».

Tableau 12 :

Proportion d'élèves ayant obtenu au moins	Social +			Proportion d'élèves ayant obtenu au moins	Social -		
	Phase 1	Phase 2	Phase 3		Phase 1	Phase 2	Phase 3
90	1.45	5.39	4.01	90	0.84	4.69	3.98
80	5.39	23.28	28.73	80	3.65	14.37	19.27
70	17.43	40.52	53.23	70	12.64	33.72	37.92
60	29.88	54.74	71.94	60	48.82	48.68	59.02
50	50	67.89	84.19	50	40.73	58.94	76.45
40	65.56	75	93.76	40	60.11	70.09	86.54
30	78.22	81.47	96.21	30	74.16	79.77	93.58
20	89.21	93.1	98.89	20	89.89	88.27	98.47
10	93.78	98.49	99.78	10	96.91	95.31	99.69

3.1.5.1.2. Analyse statistique des phases 1 et 2**En phase 1****Tableau 13 :**

	Social +	social -
Effectif		
Résultat global sur 100	47.8	44.2
Ecart type	21.8	19.6
	distribution normale	distribution normale
Réussite complète	0.21% (4.53)	0
Au moins 2 tâches de réussies	4.36% (20.4)	1.97 % (13.9)
Au moins 1 tâche de réussite	21.58 % (41.18)	14.89 % (35.65)
Réussite presque complète	1.87 %	0.28 %
Au moins 2 tâches presque réussies	12.2 %	11 %
Au moins 1 tâche presque réussie	43.3 %	37.84 %

L'analyse des résultats globaux indique une différence significative entre les deux groupes . Cette différence se retrouve en général, si on analyse les résultats entre tâches. La différence est la plus élevée chez les élèves qui ont réussi complètement au moins une tâche.

Si on considère que les élèves sont « brillants » s'ils ont réussi au moins 2 tâches, on constate que les élèves brillants sont en même proportion pratiquement dans les 2 groupes, par contre l'écart entre les milieux favorisés et défavorisés est plus élevé chez les élèves moins forts.

Ces élèves sont plus nombreux en milieu défavorisé, ce sont ceux qui sont probablement le plus influencés par le milieu social.

Par contre, les élèves qui n'ont rien réussi (3% en milieu favorisé ,contre 6% en milieu défavorisé) sont plus nombreux dans le milieu favorisé.

Social + 50% des élèves ont minimum 50%	Social - 50% des élèves ont minimum 45% 40,7% ont minimum 50%
--	---

En partant de l'hypothèse que les caractéristiques cognitives sont identiques des deux côtés, 9% des élèves de milieu défavorisé sont en échec alors qu'ils ne l'auraient probablement pas été s'ils venaient d'un milieu favorisé, ce qui fait approximativement entre 4 et 5 % de l'effectif total.

En phase 2

	Social +	social -
Effectif		
Résultat global sur 100	58,30	53,55
Ecart type	23,97	24,90
Réussite complète	0,22 %	0,59 %
Au moins 2 tâches de réussies	5,60 %	4,11 %
Au moins 1 tâche de réussite	29,96 %	24,90 %

Les élèves issus de milieux dits favorisés réussissent mieux dans l'ensemble que ceux issus de milieux défavorisés. Cependant, il y a légèrement plus d'élèves de milieux défavorisés qui ont réussi parfaitement toutes les tâches (non significatif).

Il est intéressant de noter que l'écart entre les milieux sociaux est sensiblement le même dans les deux phases.

3.1.5.1.3. Conclusions

En phase 1, il existe une différence statistiquement significative entre les deux milieux. Elle est cependant limitée et est due surtout aux résultats des élèves obtenant des résultats entre 35 % et 65 %. La différence plus importante en phase 3 entre les deux milieux laissent supposer qu'une proportion non négligeable d'élèves du milieu social défavorisé et maîtrisant les procédures automatisées réussit à obtenir des résultats considérés comme bons en phase 1.

Même si une différence statistique a été calculée, la différence intrinsèque entre les deux niveaux s'est révélée moins importante que nous l'avions imaginée.

3.1.5.2. Y a-t-il une influence de la dynamique de l'équipe pédagogique d'une école sur les compétences des élèves à résoudre des tâches complexes ?

Les résultats affichés dans le tableau 13 ont de quoi surprendre. En effet, en phase 1, les écoles avec un projet pédagogique considéré comme positif ont des résultats identiques à celles considérées comme ayant un projet pédagogique moins positif. En phase 2 et 3, ce sont les écoles présentant les projets pédagogiques les moins positifs qui obtiennent les meilleurs résultats.

Tableau 14

809 élèves	Phase 1		Phase 2		Phase 3					
					Calcul		Français		Global	
	%	Ecart Type								
Projet -	46.24	22.78	58.52	25.96	66.5	22.54	70.08	19.48	68.78	17.15
Projet +	46.27	19.92	55.16	23.57	62.14	22.7	64.87	20.48	63.69	17.74

Ces résultats nous demandent de préciser la manière dont les écoles ont été sélectionnées. Comme nous l'avons présenté dans la partie méthodologique, ce sont les inspecteurs du canton de Namur, qui, lors d'une réunion entre eux, ont donné les indications sur le caractère dynamique ou non de telle école. Les critères de sélection retenus n'ont pas été réellement définis et la sélection a reposé sur l'interprétation personnelle par chaque inspecteur de ce qu'est un projet pédagogique positif. Ce mode de sélection est naturellement totalement contestable et peu rigoureux. Pour cette raison, nous n'avons pas trouvé opportun de proposer une analyse plus approfondie des résultats.

Le côté contestable de la définition de ce facteur ne nous empêche pourtant pas d'être interpellé. En effet, il est quand même surprenant que l'addition de résultats d'écoles, considérées comme ayant un projet pédagogique positif, par un ensemble d'inspecteurs ayant une formation pédagogique de qualité et une connaissance approfondie du terrain, ne soit pas significativement supérieur à l'addition de résultats provenant d'écoles considérées comme n'ayant pas ou peu de projet pédagogique. Nous aurions pu nous attendre, du moins en phase 1 ou 2 à observer une différence en faveur de l'échantillon d'écoles considéré comme pratiquant un projet pédagogique intéressant.

Pour nuancer ces résultats sans avoir la prétention de les expliquer, un fait intéressant que nous avons eu l'occasion d'observer lors de la passation dans les classes doit être rapporté. Dans une école fréquentée par des élèves issus d'un milieu favorisé, fonctionnant dans une pédagogie en cycles à base de projets, nous avons pu observer une démobilisation complète des élèves face aux épreuves proposées. Cette anecdote montrerait que le statut de l'évaluation pourrait être un facteur important dans la réussite ou non à des épreuves externes. C'est peut-être parce qu'ils ne sont pas ou guère habitués à ce type de tâches scolaires et parce qu'au contraire ils ont l'habitude de travailler « sur la réalité » qu'ils n'ont pas saisi l'intérêt de nos épreuves..

3.1.5.3. Les modalités d'administration des épreuves ont-elles une influence sur les résultats des élèves ?

Tableau 15 :

809 élèves	Phase 1		Phase 2		Phase 3					
					Calcul		Français		Global	
	%	Ecart Type								
Chercheur tout	47.76	21.11	60.57	23.03	68.3	18.88	72.86	18.02	70.74	14.12
Chercheur phase1	43.02	20.05	55.27	24.18	61.17	24.44	65.55	20.28	63.96	18.89
Enseignant	49.55	21.14	53.79	25.45	63.26	22.57	62.39	21.06	62.72	17.75

3.1.5.3.1. Premiers constats

Comme nous l'avons développé dans la partie méthodologique, notre échantillon a été divisé en 3 sous-groupes qui ont subi trois modes de passation différents. Un sous-groupe composé de 10 classes issues de 5 écoles ont été supervisées dans les trois phases par les chercheurs. Un deuxième sous-groupe composé de 20 classes issues de 10 écoles a été supervisé en phase 1 par les chercheurs et en phase 2 et 3 par les enseignants. Un troisième sous-groupe composé de 15 classes provenant de 8 écoles différentes a été supervisé uniquement par les enseignants.

L'objectif que nous avons en réalisant cette répartition était d'essayer de déterminer si un effet enseignant réel existait lors de la passation. En d'autres termes, nous avons comme idée que l'enseignant, par son attitude, pouvait avoir un effet positif et important sur les résultats de ses élèves. Pour tenter d'interpréter nos épreuves au regard de ce facteur, nous avons essayé de mesurer cette différence présupposée.

Tableau 16 :

Proportion d'élèves ayant obtenu au moins	Chercheur tout (A)		
	Phase 1	Phase 2	Phase 3
90	1.9	6.79	4.29
80	4.76	23.53	29.52
70	17.62	44.34	56.19
60	26.67	58.82	80
50	49.52	69.68	90.95
40	65.71	79.19	95.24
30	78.57	85.97	98.1
20	90.95	94.12	100
10	96.67	99.55	100

Proportion d'élèves ayant obtenu au moins	Chercheur Phase 1 (B)		
	Phase 1	Phase 2	Phase 3
90	0.29	2.68	4.72
80	2.31	18.23	23.89
70	12.1	37	43.89
60	20.75	51.74	62.78
50	39.19	64.08	78.61
40	58.21	72.12	89.72
30	73.2	80.16	93.33
20	86.74	90.08	97.5
10	93.66	96.51	99.44

Proportion d'élèves ayant obtenu au moins	Enseignant (C)		
	Phase 1	Phase 2	Phase 3
90	1.61	7.58	2.43
80	8.06	17.54	16.99
70	17.74	31.75	42.23
60	29.44	45.97	59.22
50	52.42	58.29	74.76
40	68.15	67.77	87.86
30	79.44	76.3	95.15
20	93.55	89.57	99.51
10	96.37	95.73	100

Les résultats présentés dans le tableau 14 et 15 nous ont à vrai dire surpris. Ainsi en phase 1, nous avons mesuré une différence de 5 % dans les résultats en faveur du sous-groupe

dont les épreuves ont été administrées par les enseignants. (Ces 5 % provenant de la comparaison entre ce groupe et les deux autres sous-groupes (« chercheur tout » et « chercheur phase 1 »)) Mais, en phase 2 et 3, nous observons le contraire car ce sont les élèves surveillés par les chercheurs qui sont les plus performants. D'une manière très significative, en phase 2, 69.68 % des élèves surveillés par les chercheurs obtiennent un résultat égal ou supérieur à 50 % alors qu'ils ne sont que 58.29 % d'élèves surveillés par les enseignants à obtenir le même résultat. En phase 3, la différence est encore plus significative car ils sont plus de 90 % à obtenir un résultat égal ou supérieur à 50 % lors de la passation exercée par les chercheurs alors qu'ils ne sont qu'environ 76 % à obtenir un résultat identique lorsqu'ils sont surveillés par les enseignants. Même s'il faut rester prudent compte tenu des aléas de notre échantillon, nous pouvons émettre quelques hypothèses sur base de ce constat.

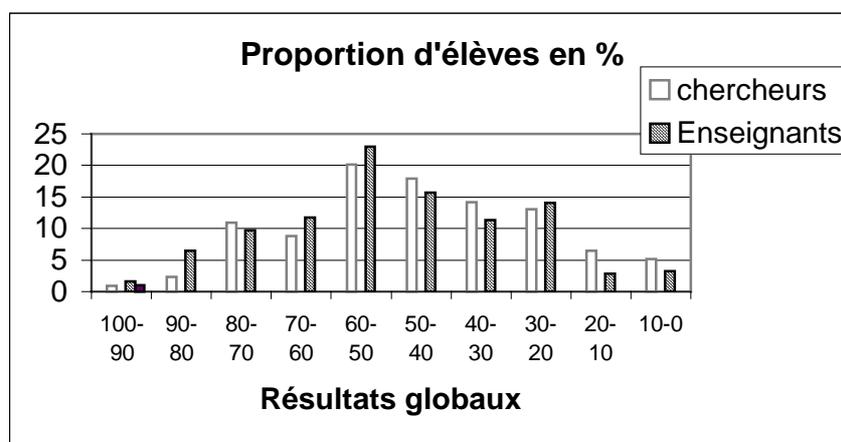
Avant de développer celles-ci, nous avons réalisé une analyse plus fine de ces résultats en croisant nos différents facteurs. Cette analyse étayera ces hypothèses.

3.1.5.3.2. Analyse statistique des phases 1 et 2

En phase 1

Si on examine les résultats obtenus par tous les élèves surveillés par un chercheur (« chercheur tout » et « chercheur phase 1 ») ou par leur enseignant sur un graphique, on s'aperçoit qu'il est impossible de vérifier dans quel sens va cette différence rapportée.

Tableau 17 :



Cependant, comme nous l'avons déjà énoncé, on note une différence significative entre les moyennes obtenues par les élèves qui ont été surveillés par leur instituteur et par ceux qui l'ont été par un chercheur. (5% de + chez enseignants)

Parmi les différents facteurs considérés dans cette recherche, y en a-t-il qui pourraient expliquer cette différence de résultats?

- En croisant les facteurs cycle et passation, on ne constate pas de différence significative, il ne semble donc pas y avoir d'influence de l'âge sur l'effet de la passation.
- Le sexe a relativement peu d'influence, les filles réussissent mieux que les garçons surtout si elles sont surveillées par leur instituteur.
- Les élèves issus d'un milieu plus favorisé ont de meilleurs résultats surtout s'ils sont surveillés par leur instituteur.

- Le projet pédagogique par contre, a une influence différente: les résultats obtenus par les élèves sont bien moins bons lorsque ce sont les chercheurs qui font passer l'épreuve plutôt que les enseignants, mais dans l'ensemble, avec ou sans projet, les moyennes diffèrent peu. Cependant, avec projet ce sont les élèves surveillés par leur instituteur qui réussissent mieux et sans projet, ce sont ceux qui passent la phase 1 surveillés par les chercheurs.

Pour illustrer et vérifier ces constatations, nous avons considéré le sous-groupe particulier des filles issues d'un milieu favorisé, dont les enseignants participent à un projet pédagogique, et dont l'épreuve a été surveillée par les enseignants : la moyenne obtenue est éloquente : 57.10 sur 100 (écart type = 20.91) contre 44.42 sur 100 (écart type = 20.43) pour celles qui ont les mêmes particularités mais dont l'épreuve a été surveillée par un chercheur.

En phase 2

- Chez les élèves du cycle 5-8, les résultats ne varient guère selon les modes de passation, ils sont légèrement supérieurs lorsque les chercheurs effectuent eux-mêmes la passation. Par contre, la différence est nettement plus marquée chez les élèves plus âgés. Ils obtiennent en moyenne 10% de plus si les chercheurs font passer eux-mêmes les épreuves.
- Quel que soit le mode de passation, les filles obtiennent des résultats semblables, par contre les garçons réussissent mieux quand l'épreuve est surveillée par les chercheurs (très significatif), et de plus, ils obtiennent de meilleurs résultats que les filles dans ce cas.
- Le fait d'appartenir à un milieu favorisé ou non n'influence pas la passation (non significatif).
- Quand le chercheur effectue lui-même la passation, la différence est pratiquement inexistante entre les résultats des élèves qui ont un projet pédagogique considéré comme faible et les autres : légèrement plus élevés quand le projet pédagogique est considéré comme faible. Elle est beaucoup plus sensible quand l'enseignant le fait. La différence est très forte quand c'est l'enseignant qui a surveillé les trois phases: avec un projet pédagogique considéré comme faible, la moyenne des points obtenus par les élèves est de 61 sur 100, contre 49 sur 100 dans le cas d'un projet pédagogique positif.

3.1.5.3.3. Hypothèses et conclusions

En phase 1, nous constatons une différence significative entre les résultats obtenus par les élèves surveillés par les chercheurs et ceux surveillés par leur enseignant. La différence est en faveur de ce dernier sous-groupe. L'analyse nous montre que cette différence a été observée auprès des élèves du cycle 5-8 et ceux du cycle 10-12. Elle nous montre également que ce seraient les enfants socialement favorisés, les filles et les élèves fréquentant des établissements développant un projet pédagogique qui seraient les plus influencés par la passation de leur enseignant. Ces constats, au delà de leur caractère anecdotique, nous conduisent à deux hypothèses. La première est que tous les enseignants n'exercent pas la même influence positive sur leur classe. La seconde est que chaque élève pris individuellement réagit différemment face au mode de passation proposé et que l'enseignant administrant une épreuve à sa classe influence plus certains élèves que d'autres.

Cependant, si l'on considère notre échantillon global et la répartition des différences, nous pouvons considérer que ces différences statistiquement démontrées sont toutefois limitées.

En phase 2, les résultats enregistrés indiquent qu'au cycle 10-12 les élèves – les garçons - obtiennent des meilleurs résultats lorsqu'ils sont surveillés par les chercheurs. Même si l'on

considère que certains enseignants exercent plus que d'autres une influence positive sur les résultats, nous devons avancer l'hypothèse qu'en phase 2 et également en phase 3, il y aurait un effet chercheur. Cette différence très significative observée entre les 2 sous-groupes laisserait supposer que les élèves, par la présence des chercheurs, sont restés mobilisés par les tâches proposées. Cela n'aurait pas été le cas de tous les élèves surveillés par les enseignants. On peut supposer, soit que ce sont les élèves qui se sont démotivés en considérant qu'il n'y avait plus d'enjeu, soit que ce sont certains enseignants qui par leur attitude ont montré un désintérêt par rapport à nos épreuves.

En conclusion, certains enseignants ont vis-à-vis de certains élèves une influence positive et importante sur leurs résultats en phase 1. Cette influence aurait permis à certains élèves d'obtenir des résultats supérieurs à ce qu'ils auraient pu prétendre s'ils avaient été surveillés par les chercheurs. Cependant, même si l'influence est importante pour ces élèves, la différence de résultats enregistrés pour l'échantillon global est limitée et ne remet pas en question les constats énoncés précédemment. En phase 2 et 3, même si certains enseignants exercent une influence, nous avons constaté que c'était la présence du chercheur qui serait la condition la plus favorable. Cette donnée intéressante pourrait nous amener à penser que les résultats globaux enregistrés en phases 2 et 3 auraient pu être meilleurs si nous avions administré les trois phases à tout l'échantillon.

3.1.6. Analyse des épreuves par classe

3.1.6.1. Introduction

L'analyse des résultats par classe nous éloigne d'un traitement statistique reposant sur un échantillon global et nous rapproche de la volonté, que nous avons entretenue durant ces deux années de recherche, de récolter toutes les informations possibles qui contribueraient à cerner les facteurs de réussite à des épreuves mesurant l'acquisition de compétences et à préciser les conditions d'apprentissage les plus adaptées à cette acquisition. Cette perspective nous a amené à poser un certain nombre de questions et à chercher à vérifier certaines hypothèses. Comme nous l'avons déjà annoncé, nous n'avons pas pu pour des raisons méthodologiques et matérielles réellement approfondir nos hypothèses. En effet, le questionnaire adressé aux enseignants de sixième année et celui adressé à leurs élèves ne nous ont pas apporté les informations escomptées. Nous espérons qu'une analyse des réponses au regard des résultats des élèves par classe auraient pu nous éclairer sur certaines conditions d'apprentissage à l'acquisition de compétences. Très court, trop court sans doute, ils n'ont permis ni l'un, ni l'autre d'apporter des éléments probants aux questions posées.

Pour le questionnaire enseignant, la désirabilité sociale ou l'interprétation de termes tels que projet ou situations – défis pourraient être la cause de ce constat. L'analyse des réponses ne laisse supposer aucun lien entre des types de pratique et les performances des élèves. Peut-être que ce lien n'existe pas, mais nous n'oserions pas affirmer, ni même suggérer cette hypothèse sur la base du questionnaire proposé.

En ce qui concerne le questionnaire élève, nous avons pu constater que dans de très nombreux cas, les réponses sur les acquisitions qui auraient du être communes à une classe divergeaient totalement. Cette non cohérence a rendu impossible un traitement des résultats par classe comme nous l'avions imaginé. Nous aurions pu réaliser un traitement item par item et élève par élève mais le manque de temps ne nous a pas permis de réaliser ce travail.

Enfin, de nouveau par manque de temps, nous n'avons pas poursuivi l'analyse systématique des copies telle que nous le suggérons et qui aurait pu également apporter des éléments de

réponses aux questions posées. Nous espérons pouvoir nous y atteler dans un prochain avenir.

Au terme de ces deux années de recherche, il nous apparaît clairement que la collecte d'informations qui apporteraient ou non des éléments de vérification à nos hypothèses demande un contact privilégié et personnel avec les élèves et les enseignants. Au-delà de l'analyse des copies, une méthodologie composée d'entretiens et d'observations nous semblerait efficace

Tout ceci ne nous empêche pas cependant de réaliser au regard des résultats quantitatifs enregistrés par les différentes classes différents constats intéressants.

3.1.6.2. deux constats

- des différences entre les classes

Les deux tableaux suivants proposent les résultats par école. Nous présentons, vu les différences enregistrées entre les épreuves, les résultats classés par épreuve. Très clairement, pour les différentes épreuves proposées, nous constatons une répartition très dispersée des résultats. Ainsi, pour l'épreuve « Le premier mai », les résultats à la phase 1 sont compris entre 32.2 % à 59.38 %, pour l'épreuve « Les murs de la classe », l'école la plus faible obtient 29.48 % en phase 1 et la plus forte 57.5 %. Pour les épreuves du cycle 5-8, les résultats pour l'épreuve « St Nicolas » se situent entre 26.47 % et 65.87 % et ceux de l'épreuve « Le jardinage » sont compris entre 20.45 % et 58.04 %.

Pourquoi constatons-nous des résultats aussi contrastés entre des classes ayant subi la même épreuve ? C'est sans doute la question essentielle que suscite les résultats présentés dans ces deux tableaux.

Sans apporter de réponses réelles à cette interrogation, une analyse des caractéristiques des différentes écoles ouvrent des pistes de réflexion. En effet, et ceci va dans le sens de nos propos précédents, ni le niveau social, ni l'implication pédagogique ne peuvent être, dans le cadre de notre échantillon, utilisés comme éléments déterminants d'explication aux différences de résultats constatées. Certes les résultats des écoles de milieu favorisé sont meilleurs, dans l'ensemble, que ceux des écoles de milieu défavorisé, mais il est important de souligner que ce critère ne détermine pas automatiquement le classement des écoles. Nous retrouvons dans les 4 épreuves et de manière récurrente des écoles considérées comme défavorisées obtenir des résultats supérieurs à des écoles fréquentées par un milieu social favorisé. Est-ce dû au projet pédagogique développé par l'école ? Ceci est possible, car une lecture concentrée sur ces deux critères laissent apparaître que la plupart des classes considérées comme défavorisées, ayant été surveillées par les chercheurs et obtenant des résultats supérieurs à la moyenne font partie de la catégorie « pédagogique positif ». Cependant, la définition subjective de ce critère nous demande de rester prudent. Cette prudence est de plus indispensable lorsqu'on observe les résultats fort dispersés des écoles favorisées et ayant été considérées comme développant un projet pédagogique positif.

- La hiérarchie entre les phases

24 classes sur les 45 respectent, au niveau de leurs résultats, la hiérarchie entre les trois phases. Cette différence s'explique en partie par les mauvais résultats en phase 3 de l'épreuve « Le premier mai ». Elles sont 39 en phase 2 à obtenir des résultats supérieurs à ceux enregistrés en phase 1. Des facteurs de passation peuvent être à la base de ce constat.

La comparaison des classes entre elles en tenant compte des trois phases montrent que ce ne sont pas automatiquement les classes les meilleures en phase 1 qui sont les meilleures dans les autres phases. Ceci n'est pas, à nos yeux, surprenant car, comme nous l'avons montré, le lien entre les trois phases n'est pas automatique. En d'autres mots, nous avons pu constater lors de l'analyse des résultats globaux que des élèves, par exemple, forts en phase 3 ne l'étaient pas automatiquement dans les autres phases. Il est important également d'insister sur le fait que ces classes « fortes » en phase 1, sans être les meilleures dans les autres phases, présentent dans celles-ci des résultats convaincants.

Ces commentaires nous amènent au constat que nous trouvons des classes présentant des profils différents ce qui laisse supposer d'une influence possible de certaines pratiques pédagogiques. Ainsi, nous trouvons des classes fortes en phase 3 et présentant des

résultats faibles dans les autres phases. Nous trouvons des classes faibles dans les trois phases...

3.1.6.3. L'effet enseignant

La diversité des résultats observés d'une classe à l'autre qui ne semble pas être déterminée par le niveau social des élèves nous amène à émettre une hypothèse pouvant expliquer ces différences. Deux possibilités, dans la limite de notre travail, s'offrent à nous comme tentative d'explication. La première est qu'il y aurait une influence de l'école fréquentée sur les performances des élèves. Sans pouvoir rejeter cette proposition, nous avons récolté quelques informations allant à l'encontre de cette possibilité. En effet, ni les résultats provenant du traitement du sous-groupe « projet pédagogique », ni l'analyse des résultats obtenus par les classes du même école (cycle-5-8 et cycle 10-12) ne nous permettent d'énoncer le fait que les performances des élèves seraient liées à l'école fréquentée. Ceci nous conduit à penser que les élèves seraient surtout influencés par l'enseignant ou le groupe d'enseignants qu'ils fréquentent ou qu'ils ont fréquenté.

Au delà des facteurs individuels propres à chaque élève, l'enseignant pourrait être le premier facteur explicatif des différences rencontrées. Les constats réalisés à la suite de nos nombreuses rencontres avec les enseignants durant la passation de nos épreuves vont dans ce sens. Ainsi, il est intéressant de signaler que trois des enseignants dans les classes desquelles nous avons administré les trois phases et qui avaient attiré notre attention sont ceux dont les élèves ont obtenu les performances les meilleures. Au delà de leur pratique pédagogique d'ailleurs fort différente, c'est leur empathie et la confiance témoignée aux élèves quant à leur capacité qui a suscité notre intérêt. Des élèves fréquentant d'autres enseignants manifestant des capacités méthodologiques de haut niveau mais ne présentant pas cette empathie ont obtenu des résultats beaucoup plus faibles. Hasard ? La question mérite d'être posée.

3.2. La réflexivité de l'enseignant

Dans la continuité des propos énoncés ci-dessus, s'intéresser à la réflexivité de l'enseignant c'est s'interroger sur ses capacités à parler et à réfléchir sur sa pratique. Terme très à la mode dans certains milieux pédagogiques, la réflexivité pour certains auteurs est une caractéristique de l'enseignant professionnel et par conséquent efficace.

Influencé par ce courant, nous nous sommes demandés si nous ne pouvions pas catégoriser les enseignants de notre échantillon sur base de ce critère. Pour ce faire, nous avons administré un questionnaire tentant d'évaluer leurs connaissances sur les réformes en cours et le degré d'adhésion à celles-ci. Avons-nous réellement évalué la réflexivité de ceux-ci. ? Nous ne le pensons pas, mais une connaissance des réformes en cours est à nos yeux un signe de professionnalisation. De ce fait, notre idée était de confronter les résultats du dépouillement du questionnaire de chaque enseignant aux résultats obtenus par leurs élèves.

3.2.1 Méthodologie

Chacune des écoles faisant partie de l'échantillon total (20) a été subdivisée en quatre strates (cycle 5/8, cycle 8/10, cycle 10/12 et direction). Elles ont eu un délai d'environ trois semaines pour remplir le questionnaire proposé. Théoriquement nous aurions dû compter sur le retour de 80 questionnaires. Mais la mortalité fut importante, 39 questionnaires seulement furent renvoyés. Il est clair qu'une mortalité de 50% biaise la représentativité de cette recherche.

L'objet de ce travail consistait à évaluer les degrés d'information et d'adhésion des enseignants et directeurs de l'échantillon considéré. Nous avons majoritairement opté pour des questions ouvertes afin d'éviter au maximum les biais liés à la désirabilité sociale ou à une évaluation qui se limiterait à la reconnaissance de concepts liés au travail en cycles.

Afin de mesurer le niveau d'information des sujets, nous avons dû préalablement définir les critères attendus. Pour ce faire, nous avons passé en revue l'ensemble des circulaires ministérielles et décrets relatifs au travail en cycles depuis 1971. La principale source retenue est la circulaire du Ministre J-M Nollet du 31 mars 2000.

Nous avons attribué un score pour chaque question. Score qui est fonction du nombre d'éléments attendus par question. Il s'agit donc bien d'évaluer le contenu des réponses.

Rappelons que nous avons à la fois des questions ouvertes et des questions fermées.

Pour les questions fermées (échelle de Likert), nous avons procédé à un recodage de manière à dichotomiser les résultats obtenus.

Concernant chaque question ouverte, nous avons préalablement déterminé les catégories de réponses qui devaient être abordées afin que l'information soit optimum. Nous avons donc opté pour une évaluation des réponses, en fonction de leurs parentés ou non aux catégories que nous avons prédéfinies et qui correspondent aux différents aspects qu'impliqueraient une bonne information. La note attendue n'est donc pas la même pour toutes les questions, puisque le nombre de réponses attendues diffère d'une question à l'autre.

Précisons qu'il s'agit bien de catégories, car plusieurs réponses pourtant sensiblement différentes, peuvent convenir pour autant qu'elles fassent bien évidemment référence à la catégorie.

Par exemple, si nous avons la catégorie « travail en équipe », il va de soi que les personnes qui ont répondu en termes de concertation, ou de coresponsabilité sont considérées pour cette catégorie comme étant bien informées.

Seulement, une difficulté est rapidement apparue. En effet, certaines réponses proposées par les enseignants ne sont pas équivalentes à la catégorie car, bien qu'elles s'y apparentent, elles sont une réduction des implications de la catégorie considérée.

Prenons par exemple la catégorie attendue « évaluation formative » et la réponse proposée par un enseignant : « auto-évaluation ». L'auto-évaluation est bien un processus métacognitif qui fait partie de l'évaluation formative ; mais cette dernière ne se limite bien évidemment pas à l'auto-évaluation. Que faire dans ce cas ? Nous avons tout de même opté pour créditer cet enseignant comme s'il nous avait répondu « évaluation formative ». D'abord, parce que nous considérons que nous ne pouvons nier une certaine information, mais aussi parce que dans le cadre du mémoire nous nous sommes attardés dans un second temps de manière beaucoup plus spécifique sur chaque réponse, ce qui nous a permis alors de nuancer la qualité de l'information et de préciser les éléments les plus fréquemment cités ou omis.

Signalons également que les enseignants qui ont répondu à ce questionnaire ont bénéficié d'un délai de trois semaines pour le remplir et qu'il n'y avait pas d'obligation ou de contrainte pour nous le restituer. Nous considérons donc qu'une non-réponse à une question est considérée comme résultant d'une ignorance (à l'exception de deux enseignants qui ont explicitement refusé de répondre).

Le score maximum pouvant être obtenu est de 20. :

Les très insuffisamment informés ayant obtenus un score inférieur à 4 / 20

Les insuffisamment informés ayant obtenus un score allant de 5 à 8 / 20

Les moyennement informés ayant obtenus un score allant de 9 à 11 / 20

Les biens informés ayant obtenus un score allant de 12 à 13 / 20

Les très bien informés ayant obtenus un score supérieur à 14 / 20

Au terme de toutes les questions traitant de l'information, et de manière globale, nous sommes arrivés à la conclusion que seules 8 personnes interrogées et ayant répondu au questionnaire (8 sur 39) peuvent être considérées comme étant informées c'est-à-dire 20%.

A propos de l'adhésion, nous avons opté pour une majorité de questions fermées. Le principe était également d'attribuer des scores mais bien sûr sans critères de références.

Ici, ce n'est pas moins de 24 sujets pour lesquels nous pouvons affirmer qu'ils adhèrent au travail en cycles (24/39) c'est-à-dire 61,5%.

Signalons parmi nos observations pour cet échantillon-ci, que toutes les personnes informées sont adhérentes au travail en cycles.

Mais que 16 sujets sont à la fois adhérent et mal informés (41%).

3.2.2 Résultats et conclusions

La méthodologie proposée par monsieur Schoeters nous a amené à classer les enseignants de notre échantillon en quatre catégories : la catégorie des enseignants bien informés et adhérent aux réformes ; la catégorie des enseignants bien informés et n'adhérant pas aux réformes ; la catégorie des enseignants mal informés et n'adhérant pas aux réformes et enfin la catégorie des enseignants mal informés mais adhérent aux réformes.

19 enseignants sur les 45 ayant participé à notre expérience ont complété le questionnaire. Deux enseignants ont explicitement refusé de le faire. Le taux de mortalité, normal dans une enquête par questionnaire, mais fort élevé dans le cadre d'une recherche où nous avons eu un contact direct avec tous les enseignants, peut s'expliquer par la longueur du questionnaire, mais également la méfiance possible de certains d'entre-eux vis-à-vis d'une recherche où l'inspection était partie prenante.

Sur les 19 enseignants interrogés, seulement 4 d'entre-eux sont considérés, d'après nos critères, comme bien informés. Ces 4 enseignants adhèrent tous à la réforme. Les 17 autres enseignants, bien qu'insuffisamment informés, adhèrent tous à la réforme.

Sur base de ces résultats, il serait vain de vouloir tirer des conclusions quant à l'impact de la connaissance des réformes sur les performances des élèves à nos épreuves. Les élèves des quatre enseignants considérés comme bien informés ne font d'ailleurs pas partie des classes les plus performantes.

Comme nous l'avions annoncé, nous savions que nous étions dans une phase exploratoire qui risquait d'être non pertinente. Les résultats décevants des différents questionnaires administrés renforcent notre idée que pour tenter de saisir les facteurs de réussite à des épreuves mesurant des compétences, une méthodologie reposant sur des entretiens directs, des analyses cliniques de production d'élèves et des observations est nécessaire.

3.3. Le travail d'appropriation du modèle par les enseignants

3.3.1. Introduction

Dans le cadre d'une formation à l'Institut Supérieur de Pédagogie de Namur portant sur l'évaluation des compétences et destinée à des instituteurs ou des directeurs d'école, nous avons demandé à ceux-ci de construire et de tester des épreuves suivant notre modèle.

Durant environ 18 heures de cours, les éléments théoriques furent présentés et les 15 heures suivantes furent consacrées à une analyse des socles de compétences et à la

réalisation du travail. En fin de module, huit nouvelles épreuves originales testées auprès d'élèves furent présentées.

Le premier constat, et non le moindre, est que la qualité des travaux montre qu'après une période de formation, des enseignants motivés s'approprient correctement le modèle. Sans se prononcer sur l'originalité des épreuves, nous pouvons affirmer que toutes respectaient les conditions minimum afférentes à notre modèle.

Suite à ces travaux demandés, nous avons mené une enquête auprès des participants afin de connaître leur opinion sur l'expérience vécue.

Nous leur avons donc soumis, lors de leur dernier cours, un questionnaire écrit abordant le temps qu'ils avaient dû consacrer à cette réalisation, la manière dont il s'y était pris et ce qu'ils en retiraient.

Aussi, nous nous proposons de développer, ci-dessous, les réponses que nous avons reçues. Pour ce faire, nous articulerons notre article en cinq points : le temps nécessaire (recherche dans les socles, réflexion, documentation, réalisation), l'ordre de réalisation, le rapport à ce type d'épreuves, la passation, la correction, et finalement les conclusions et pistes de réflexion.

3.3.2. Analyse du dépouillement du questionnaire

3.3.2.1 Si peu de temps et tant à faire... (Oscar Levant)

Si une notion revient constamment pour l'ensemble des enseignants, c'est bien celle du temps. Ce temps nécessaire pour écouter, comprendre et intégrer un modèle ; celui pour la réalisation d'une épreuve ou encore pour la passation de celle-ci ; et enfin celui nécessaire à l'analyse des résultats. Tout ceci représente de nombreuses heures qui ne peuvent être utilisées à l'enseignement des élèves. Où trouver ces heures ? Quel compromis adopter entre les apprentissages et l'évaluation ? Cette caractéristique de « la consommation du temps didactique » est un élément présent et déterminant dans l'esprit des enseignants dont il faut tenir compte (Astolfi, 1992).

En regardant de plus près les données recueillies auprès du groupe d'enseignants, nous pouvons constater que les trois groupes qui ont détaillé leurs heures passées à la réalisation de ce travail ont respectivement passé 26,5 heures, 26 heures et 71 heures. Même si ces résultats ne présentent pas le caractère scientifique requis pour une étude digne de ce nom, nous pouvons toutefois admettre que réaliser une épreuve de ce type selon le modèle proposé engage véritablement l'enseignant dans un travail de longue haleine.

Il est intéressant, à ce propos, de constater que ces travaux s'effectuaient en équipe. Nous avons donc des enseignants qui s'étaient choisis et qui, par leur participation à cette formation, étaient volontaires et motivés. Cette précision ouvre le débat du travail en équipe chez les enseignants. Gérard De Vecchi (1992) a déjà pointé les difficultés de ce type de travaux. Aussi, nos enseignants ont pu gagner un peu de temps s'ils se sont réparti le travail et les tâches, mais ils en ont peut-être également perdu lors des mises en commun et des compromis. Mais de toute façon, aucun enseignant de notre échantillon n'est prêt à réaliser ce travail seul.

Une analyse plus fine des heures consacrées nous montre que quatre groupes ont consacré la plus grande partie de leur temps à la réflexion. Ce dernier arrive effectivement quatre fois premier dans le temps dépensé par les groupes et deux fois deuxième. Bien sûr, nous nous rendons compte que cette donnée est biaisée du fait même du public interrogé et de ces attentes vis-à-vis de la réalisation de cette épreuve. Ensuite, la réalisation est citée deux fois

comme étant la partie qui demande le plus de temps et trois fois comme deuxième activité. Suivent enfin, le temps de documentation et le temps de recherche dans les socles.

Pour expliquer cette hiérarchie, nous pouvons avancer comme hypothèse la connaissance préalable, pour ces enseignants déjà en place, des socles de compétences (ou plutôt la connaissance de ce qu'ils doivent évaluer chez leurs jeunes élèves) et la familiarité effective dans la recherche de documentation ou le réseau de connaissances qu'ils se sont construits pour trouver cette documentation. A cet égard, nous avons constaté que les enseignants se servaient souvent de nouveaux documents ou des documents déjà employés, mais que les emprunts à d'autres collègues restaient rares.

En ce qui concerne le temps de réflexion, cinq groupes ont passé la majorité du temps sur la phase1. La définition des tâches et la réflexion sur la grille de correction nécessite en effet un arrêt préalable utile à la construction de ces épreuves.

Enfin, pour le temps de réalisation, les résultats ne sont pas probants. Pour certains groupes la conception du portefeuille de documents occupe la plus grande partie du temps, pour d'autres ce sont les grilles d'évaluation. D'autres encore semblent consacrer pas mal de temps à la mise en page (ce qui n'est pas vain lorsqu'on connaît l'influence que celle-ci peut avoir sur l'entrée dans la tâche des élèves).

3.3.2.2. L'organisation du travail

Une deuxième entrée de notre étude s'intéressait à l'ordre suivant lequel les enseignants ont effectué leur travail.

En première analyse, nous pouvons constater que l'ensemble des groupes, excepté un, s'est d'abord attelé à la définition des tâches et au texte d'introduction de la phase1. Même si l'ordre de réalisation de cette création d'épreuve n'est pas linéaire, nous pouvons cependant observer une forte propension à réaliser la phase1 pour commencer et à terminer par la phase 3. Nous constatons aussi que la mise en page de l'épreuve arrive dans les dernières activités et, à l'inverse, les consignes des phases1, et dans une moindre mesure de la phase2, au début.

A ce sujet, l'élaboration de consignes mérite une attention particulière. Toute la difficulté, pour celui qui la conçoit, réside en effet dans la réalisation d'une consigne claire, courte et comprise de tous ! Ce qui ne va pas nécessairement de soi (Rey, 1998).

Enfin, la création d'une relation étroite entre les trois phrases n'est pas évidente et demande du temps. De plus, au vu de cette étude, nous ne pouvons nous permettre d'établir un quelconque ordre de réalisation des tâches et des phases. Nous pouvons simplement relever une tendance à réaliser cette épreuve dans un ordre hiérarchique suivant le modèle présenté : phase1 – phase2 – phase3.

3.3.2.3. Le rapport à ce type d'épreuves

Lors de cette étude, un seul enseignant (qui en fait est directeur d'école) avait déjà réalisé ce type de travail. Malheureusement, nous ne possédons pas assez de données pour connaître le rôle qu'il a joué au sein de son groupe quant à l'ordre préconisé pour la réalisation, ni l'influence qu'il a eue sur le temps nécessaire à la réalisation de cette épreuve.

Nous avons pu également constater que nos enseignants, alors qu'il s'agit d'un échantillon privilégié au niveau pédagogique (car déjà en recherche pédagogique), avouaient pratiquer une évaluation principalement centrée sur les phases 3, correspondant à des tâches

automatisées. Il n'est pas dans nos intentions de leur jeter la pierre, mais d'observer qu'il reste du chemin à parcourir pour arriver à l'évaluation de tâches complexes !

D'ailleurs, 14 enseignants sur 16 souhaitent continuer ce type d'épreuves et le quinzième travaille en maternelle où l'évaluation diffère quelque peu. A ce sujet, il y a là aussi matière à réflexion et une recherche à mener sur l'adaptabilité de ce modèle à l'enseignement maternel.

Enfin, 13 enseignants sur 18 déclarent que ce type d'épreuves leur a permis de réaliser un diagnostic de leurs élèves. « Cela permet d'observer les compétences à retravailler avec les élèves. »

3.3.2.4. Passation des épreuves

Jusqu'ici, nous avons peu parlé de la passation de l'épreuve et de l'analyse des résultats. Notre objectif, au départ de cette étude, étant de constater la faisabilité par les enseignants de ce type d'épreuves, nous n'avions pas prévu de véritable canevas dans notre enquête pour ces deux points. Cependant, à la lecture des réponses des enseignants, nous souhaitons tout de même apporter quelques remarques.

Lors de la passation, certains enseignants ont observé des élèves déroutés par la présentation inhabituelle de l'épreuve, confirmant ainsi le peu d'évaluations réelles des compétences complexes dans notre enseignement actuel. Ils ont aussi relevé l'importance du sens et du contexte pour l'enfant, et ceci plus particulièrement dans le cycle (5-8) où ils ont constaté une transformation de la situation en vécu chez certains élèves. A ceci vient se greffer deux autres aspects : le rapport actuel des élèves à la lecture (quel type de lecture et quel type de rapport ?) et la problématique du portefeuille de renseignements (ni trop, ni trop peu et accessible !). Quelques-uns ont aussi constaté un effet de lassitude lors de la phase 2 (répétition de la phase1). Enfin, d'autres s'interrogent sur le délai nécessaire entre le passage des différentes phases.

A ce sujet, et afin d'affiner notre étude, il pourrait être intéressant, comme certains enseignants l'ont fait remarquer, de demander leurs impressions aux élèves (Face à de telles tâches, je (donner un sentiment) ; Avis sur l'activité réalisée (émettre un avis) ; Les textes que j'ai entendus aujourd'hui (donner son appréciation, son avis)).

3.3.2.5. Correction des épreuves

En ce qui concerne la correction des épreuves, certains enseignants trouvent celle-ci beaucoup trop longue et ardue. Nombreux sont ceux qui se posent des questions concernant la validité de leur grille de correction (accord entre évaluateurs) et la construction de celle-ci. Il faut penser à tout ! ... Pas sûr ! De nombreux enseignants pensent encore actuellement que leur grille de correction doit reprendre tout ce qui est évaluable dans cette épreuve. Or, nous pensons qu'il serait plus intéressant pour l'enseignant de réfléchir à ce qu'il veut effectivement évaluer chez l'élève et se centrer sur cette option. A force de s'éparpiller, on oublie ce pourquoi on évalue. Mais il reste vrai qu'une fois le choix effectué, il faut du temps et de la patience à l'enseignant pour bien cerner tous les éléments rentrant en compte dans l'évaluation de cette compétence. La répétition de cette pratique pourrait amener ainsi à terme l'enseignant à clarifier sa pensée en ce qui concerne la connaissance des démarches qu'il (et que l'élève) utilise lors de la mise en application de cette compétence. Il n'est dès lors pas étonnant que les enseignants éprouvent également des difficultés à valider les résultats de leur élèves. Ainsi, face à des épreuves concernant des tâches complexes, très peu savent si ce sont leurs élèves qui sont bons ou si c'était l'épreuve, voire la grille de correction, qui était trop facile ! Il y a là tout un travail à réaliser sur les seuils de réussite attendus. Qu'est-ce qu'une compétence maîtrisée ? Quelles sont

les compétences que les enseignants doivent faire acquérir à leurs élèves ? Quelles sont les attentes de la Communauté française, des inspecteurs et de la société ? C'est sans doute cela qu'il faudrait expliquer ou réexpliquer aux enseignants.

Enfin, il apparaît que ces évaluations demandent un temps important de correction mais, à l'image de ce qu'elles évaluent, elles sont riches d'enseignement. Dès lors, il ne faut plus les concevoir comme de petites interrogations que l'on effectue toutes les semaines pour voir si l'élève a bien étudié mais comme des évaluations plus ponctuelles qui permettent de vérifier où l'élève en est et le travail qu'il reste à effectuer pour l'amener à maîtriser ces compétences.

3.3.2.6. Conclusions et pistes de réflexion

Faire créer par les enseignants des épreuves étalonnées plonge ceux-ci dans un travail long et exigeant. Les nombreuses heures qu'ils y ont consacrées ne les incitent pas à se lancer seul dans l'aventure. Pour eux, le travail en équipe semble indispensable pour atteindre l'objectif. En ce qui concerne l'organisation de leur travail, nous avons pu constater une tendance à réaliser celui-ci en suivant la présentation du modèle proposé. De même, il apparaît que les enseignants ainsi que les élèves n'ont pas encore l'habitude de rencontrer des tâches complexes. Enfin, la conception de grilles de correction semble poser question, les enseignants ne parvenant que rarement à valider celles-ci.

Pour clore ce travail, nous proposerons quelques pistes de réflexion. Tout d'abord, il serait intéressant de vérifier l'hypothèse que la répétition de la réalisation de ce type d'épreuves par les enseignants diminuera le temps nécessaire à sa conception. Théoriquement, l'intégration du modèle ne sera plus à faire et une série d'automatismes se mettra en place au fur et à mesure. En ce qui nous concerne, nous pensons que malgré cela, le temps nécessaire à la réalisation ultérieure d'autres épreuves restera conséquent ! Tenant compte qu'« il faut bien économiser quelque part ce qu'on consomme ailleurs » (Astolfi, 1992, p182), nous préconisons la réalisation d'une épreuve par l'enseignant (bénéfique pour l'intégration du modèle utilisé et la compréhension des outils qui seraient mis à sa disposition) mais aussi, et surtout, une diffusion des outils conçus (ce qui permettrait un gain de temps considérable au niveau de la recherche de documents et de la réalisation des épreuves). Bien sûr, cela peut se faire par un travail en équipe pédagogique, mais voyons plus large... Nous proposons que ce type d'épreuves soit réalisé en Ecole Normale ce qui permettrait aux futurs enseignants de se familiariser avec ces épreuves, aux Ecoles Normales de jouer un rôle de diffusion auprès des enseignants, et aux enseignants de gagner du temps qu'ils pourraient investir dans la compréhension de ces outils et dans l'enseignement. Par cette interaction, nous pourrions recréer un lien qui semble se détériorer entre les Ecoles Normales, les futurs enseignants et les enseignants en fonction. Enfin, nous pensons qu'il est essentiel que les enseignants suivent une solide formation continuée si l'on souhaite réellement obtenir de bons résultats. Les réponses données au cours de cette enquête montrent l'importance et l'intérêt d'une formation décrivant le modèle théorique, la construction de ce type d'épreuves et l'analyse des résultats. Il est également essentiel pour les enseignants de clarifier les compétences attendues et la façon de les évaluer ainsi que d'apprendre à valider leurs grilles de correction. Mais nous avons aussi relevé des attentes au niveau de la gestion du travail en équipe, de l'élaboration de consignes et la constitution de portefeuilles adaptés. Sans cela, nous risquons de nous retrouver face à de nombreux problèmes, des déceptions et un désinvestissement des enseignants.

3.4. Conclusion de la troisième partie

L'analyse des résultats globaux nous a permis de conforter trois constats établis lors de la première année de recherche : la confirmation de notre hypothèse qui supputait que la majorité des élèves éprouvent des difficultés lorsqu'ils doivent résoudre des tâches

complexes, la hiérarchie des phases des épreuves, la cohérence de celles-ci et l'importance de la maîtrise des procédures automatisées lors de la résolution de problèmes.

L'analyse des résultats par épreuves et par tâches des phases 1 et 2 nous ont permis, sur la base des différences observées, d'enrichir notre méthodologie de constructions d'épreuves. Les analyses réalisées pourront aider les personnes désirant construire d'autres épreuves suivant notre modèle.

Nos résultats montrent l'influence réduite du milieu social sur les performances des élèves et soulèvent des questions quant à une influence des projets pédagogiques développés par les écoles. Les trois modes de passation que nous avons testé semblent indiquer que certains enseignants exercent une influence « positive » sur les résultats de certains de leurs élèves en phase 1 et qu'un effet chercheur en phase 2 et 3 serait à l'origine de résultats inattendus

Enfin, les résultats par classe mettent en évidence des différences importantes de résultats entre celles-ci. Ni le niveau social, ni le projet pédagogique développé par l'école ne peuvent réellement expliquer ces différences. C'est pourquoi, l'analyse de ces résultats nous ont conduit à émettre l'hypothèse qu'un effet enseignant influencerait de manière significative les performances des élèves. Au centre de cet effet enseignant, à côté des qualités professionnelles de l'enseignant, on soupçonne que l'empathie de ce dernier jouerait un rôle non négligeable.

Tous les constats énoncés peuvent aider les personnes désireuses de construire et d'analyser des épreuves suivant notre modèle. Notre expérience de formation à l'Institut Supérieur de Pédagogie de Namur nous a montré que notre modèle pouvait être approprié par des enseignants.

4. Conclusion générale

Au terme de ces deux années de recherche, qu'avons-nous produit ?

Nous avons élaboré un modèle d'épreuves d'évaluation des compétences des élèves à l'école primaire. Car par delà les six épreuves construites (trois pour la fin de la deuxième année primaire et trois pour la fin de la sixième primaire), ce qui importe c'est le modèle, c'est-à-dire la forme générale en trois phases à laquelle les six exemples sont conformes.

4.1. Une conception des compétences

Pour cela, il a bien fallu trancher dans l'encombrante polysémie du mot "compétence". Nous l'avons fait en restant le plus scrupuleusement fidèle au décret du 24 juillet 1997 sur les missions de l'école et aux Socles de compétences qui en sont la concrétisation.

La décision de définir ce qui doit être acquis à l'école en termes de compétences correspond certainement à de multiples intentions. L'une d'elles est qu'on peut juger plus éducatif et plus formatif pour le jeune humain de lui faire acquérir des compétences intellectuelles plutôt qu'une accumulation de connaissances spécifiques. La "tête bien faite" est préférée à la "tête bien pleine".

Mais une autre intention importante est certainement de déterminer un équipement minimal que tous les élèves doivent avoir acquis à la fin d'un cycle. Il y a là une volonté unificatrice du système d'enseignement ; mais aussi et surtout la volonté de réduire l'effet sélectif de l'école et notamment de l'école primaire et du début du secondaire. Cette volonté de faire atteindre par tous les élèves sans exception un niveau minimal conduit à le formuler en termes de compétences plutôt qu'en termes de connaissances. Car les connaissances, surtout si elles sont conçues comme des accumulations d'énoncés, de données, d'informations, etc., sont susceptibles d'une inflation indéfinie : on trouvera toujours pour un champ de savoir donné, de nouveaux détails à ajouter, des précisions de plus en plus fines dont la connaissance ou l'ignorance permettront de discriminer toujours plus finement les élèves. Les savoirs, dans leur accumulation scolaire traditionnelle, permettent d'établir des procédés d'évaluation de plus en plus discriminatoires.

Une compétence, définie comme la possibilité d'accomplir une tâche déterminée, permet moins une telle dérive : l'élève sait ou ne sait pas l'accomplir. Le risque est moindre d'engendrer sur cette base les raffinements sélectifs qui sont les instruments de cette culture de l'échec dont parle Crahay.

Dans cette intention, il faut à la fois que les compétences puissent être définies d'une manière univoque et qu'elles ne soient pas trop attachées à l'accumulation inflationniste des connaissances. Or le respect de ces deux exigences peut entraîner deux dérives et il est important de montrer en quoi les Socles de compétences les évitent.

La première dérive est, pour éviter la collusion avec les connaissances, de définir les compétences à acquérir à l'école sous la forme des grandes dimensions de l'activité mentale humaine : savoir observer, savoir analyser, savoir identifier, savoir déduire, savoir émettre une hypothèse, savoir comparer, etc. Or rien ne prouve que des compétences ainsi définies existent véritablement : est-ce que "savoir observer" met en jeu les mêmes opérations mentales quand il s'agit d'observer des expressions algébriques et quand on observe les organes d'un insecte ? "Observer" est un mot qui regroupe, pour la commodité de la communication, différentes actions humaines ; rien n'assure que celles-ci mettent en œuvre une structure mentale unique. Nous avons montré ailleurs (Rey, 1996) en quoi les expérimentations psychologiques contemporaines rendaient pour le moins très incertaine l'existence de compétences ainsi définies.

L'autre dérive, en quelque sorte symétrique de la première, est de définir les compétences en des termes précis et univoques : savoir effectuer une soustraction entre deux nombres inférieurs à 100 ; savoir, dans une phrase, mettre le verbe à l'imparfait ; savoir calculer l'aire d'un triangle, savoir repérer une métaphore, etc. L'intérêt de telles compétences est que leur maîtrise par un élève peut être constatée d'une manière objective et non discutable, ce qui rend l'évaluation précise et juste ; en outre il est toujours possible d'entraîner des élèves à exécuter des opérations standardisées en réponse à des consignes préétablies. Pourtant, nous parlons encore de dérive pour qualifier cette manière de définir les compétences, car elle réduit singulièrement l'ambition éducative de l'école : elle conduit à ce que les élèves sachent finalement accomplir une série nécessairement finie d'actes stéréotypés en réponse à des stimuli. Or ce qu'on peut attendre d'une véritable éducation c'est qu'elle conduise les enfants et les adolescents à déterminer par eux-mêmes dans quel cas il y a lieu de mettre en œuvre une opération qu'ils ont apprise. C'est cette capacité de jugement et de repérage du bon escient qui permettra au sujet, devenu adulte, de s'adapter à des situations professionnelles et personnelles très mobiles. C'est cela qui garantira également qu'il pourra prendre pleinement sa place de citoyen dans une société démocratique.

En outre, comme nous l'avons vu, le sens courant du mot "compétence" correspond bien à cette conception : un professionnel compétent, c'est quelqu'un qui sans doute possède un large répertoire de comportements automatisés, mais qui surtout est capable de sélectionner et de combiner, parmi ceux-ci, ceux qui conviennent à une situation originale et singulière. Or c'est bien cette conception des compétences qui est prévalante dans les Socles de compétences. L'analyse que nous en avons faite durant la première année de recherche nous en avait convaincus. Depuis lors cette conviction a été renforcée par la comparaison que nous avons menée entre les Socles et des référentiels scolaires étrangers (Québec et France). Une telle comparaison fait ressortir que dans le référentiel belge, les compétences sont très fréquemment définies comme impliquant l'adéquation à une situation nouvelle (d'où l'abondance d'expressions comme : "adapté", "convenable", "adéquatement", "à bon escient", "en fonction de la situation", etc).

4.2. Une épreuve en trois phases

Cette conception de la compétence, une fois rendue opérationnelle, nous amène à distinguer trois degrés : la procédure automatisée, la capacité à choisir la procédure qui convient à une situation nouvelle (et qui exige donc l'interprétation de cette situation), la capacité à choisir et à combiner plusieurs procédures pour répondre à une situation non seulement originale, mais complexe.

Or même si c'est le troisième degré qui nous paraît représenter une compétence au sens fort du terme, il nous a paru essentiel de créer un modèle d'épreuves qui puisse évaluer pour chaque élève la maîtrise de chacun des trois degrés.

Selon la commande qui nous était faite, les épreuves à construire pouvaient avoir un usage formatif ou un usage certificatif. Mais il nous a semblé que dans les deux cas, il fallait que l'épreuve permette de toute façon une évaluation diagnostique, c'est-à-dire qu'elle permette de savoir, aussi précisément que possible, ce que l'élève sait faire et ce qu'il ne sait pas faire, et dans ce deuxième cas les raisons, les obstacles qui l'empêchent de réussir.

D'où l'idée d'épreuves à trois phases dont la succession représente un dégradé dans la complexité :

1ère phase : une situation complexe et originale, toujours pluridisciplinaire, et aussi semblable qu'il est possible à une situation de la vie réelle.

2ème phase : la même situation, mais, cette fois, découpée en petites unités de démarche dont chacune ne requiert de l'élève, en principe, que la sélection d'une procédure de base.

3ème phase : enfin, le contrôle de la maîtrise que l'élève peut avoir des procédures de base quand on lui en demande l'exécution par une consigne explicite.

Ce choix d'épreuves relativement lourdes, mais à usage résolument diagnostique, répond, dans notre esprit, à quatre préoccupations :

a) D'abord, cette forme d'évaluation diagnostique permet d'échapper à un dilemme. En effet, la construction d'épreuves étalonnées d'une manière classique, c'est-à-dire normative (au sens statistique du terme, c'est-à-dire par référence à ce que la majorité des élèves savent faire aujourd'hui) risquait de produire des épreuves qui ne correspondraient pas au niveau d'exigence impliqué par le décret "Missions". A l'inverse la construction d'épreuves critériées, sans autre précaution, conduisant à situer chaque élève par rapport aux critères tirés du décret et des socles, risquait, dans le contexte d'une réforme qui ne peut pas avoir encore produit tous ses effets sur le terrain, entraîner des résultats mauvais, voire désastreux pour beaucoup d'élèves et engendrer le découragement parmi eux et parmi les instituteurs.

Nous avons choisi une forme d'épreuve qui renvoie bien aux critères que nous avons tirés de l'analyse du décret et des socles, mais qui permet, grâce au dispositif des trois phases, de faire ressortir les points positifs de l'élève et, sur les points négatifs, de cerner à quoi tiennent les difficultés.

b) La forme d'évaluation ainsi présentée est plus tournée vers l'investigation fine du fonctionnement cognitif de l'élève que vers le souci d'établir d'une manière sommative des différences entre les élèves. Il s'agit d'opérer les repérages nécessaires en vue de faire progresser chacun au maximum de ses possibilités et non pas de chercher des indices de discrimination entre élèves. On s'éloigne ainsi des pratiques d'une école sélective.

c) Ainsi conçues, les épreuves sont un instrument d'évaluation formative : elles permettent aux enseignants et, à travers eux, aux élèves, d'identifier le positionnement de chacun dans le cheminement des apprentissages.

d) Enfin, le modèle d'épreuves que nous proposons peut constituer un outil didactique pour l'enseignant. C'est un point sur lequel nous n'avons guère insisté jusqu'à présent. L'analyse fine des réponses d'élèves aux phases 1 et 2 apporte des indices très riches sur les préconceptions des élèves et sur les raisons qui sont à la base de leurs erreurs éventuelles. Or dans une conception constructiviste de l'apprentissage, il est important que l'enseignant élabore ses dispositifs didactiques à partir de ce qui, dans l'esprit des élèves, fait obstacle aux notions à acquérir.

4.3. Les résultats de l'essai des épreuves

Une fois le modèle d'épreuve mis au point et testé sur deux exemples (un pour les élèves du cycle 5-8 et un pour le cycle 10-12) durant l'année 1999-2000, la tâche de la présente année 2000-2001 était de créer sur le même modèle quatre autres épreuves et de les tester sur un échantillon plus large. Les moyens dont nous disposions ne nous ont pas permis de constituer un échantillon strictement représentatif, mais l'échantillon choisi est déjà important et, grâce aux précautions que nous avons prises dans sa construction, offre certaines présomptions de représentativité.

Rappelons le sens principal des résultats obtenus et de leur traitement :

- D'abord se confirme l'hypothèse que nous faisons dès le début de la présente recherche : la majorité des élèves (à 12 ans comme à 8 ans) ont des difficultés à aborder des tâches complexes et inédites. Cette situation n'est pas alarmiste en elle-même puisque c'est précisément un des enjeux de la réforme en cours que d'inciter les enseignants à habituer leurs élèves à de telles tâches. Notons que la diffusion de nos épreuves peut jouer un rôle dans cette incitation.

- Les résultats font apparaître nettement une hiérarchie dans les résultats aux trois phases : la phase 3 est toujours mieux réussie que la phase 2, laquelle à son tour est mieux réussie que la première. Les petites variations qu'on peut constater dans le détail des résultats sur ce point (entre cycles, entre garçons et filles ou entre épreuves) peuvent trouver à chaque fois des éléments d'explication et ne remettent pas en cause le fait massif d'écart entre les trois phases.

Ainsi se trouve confirmé, non seulement la pertinence d'une épreuve à trois phases, mais le modèle de compétences à trois degrés : l'écart est si constant et si général entre les trois phases qu'il semble bien désigner une différence de nature entre les types de compétences que chacune de ces phases exige.

- Mais en même temps les calculs corrélationnels font apparaître un lien de dépendance très fort (quelle que soit l'épreuve et le cycle) entre les résultats aux différentes phases. Cela signifie qu'il faut nécessairement être performant à la phase 3 pour réussir à la phase 2, sans toutefois que cela soit suffisant ; et de même qu'il faut être performant à la phase 2 pour réussir la phase 1, sans que cela soit là aussi suffisant.

Il y aurait là, si elle était confirmée par une étude sur un échantillon plus large, une indication d'un intérêt pédagogique majeur : pour pouvoir aborder des situations nouvelles et complexes, il faut nécessairement avoir automatisé (en arithmétique, dans le domaine de la langue, de l'écriture, etc.) un certain nombre de procédures de base. Si le modèle constructiviste de l'apprentissage ne l'a jamais nié, il a pu arriver que certaines pratiques de terrain qui en étaient issues oublient cette nécessité.

- Enfin, entre les résultats des différents exemples d'épreuves que nous avons construits sur la base de notre modèle, il y a quelques différences de détail. Mais dans chaque cas on peut concevoir les raisons de ces différences et en tenir compte dans la construction d'épreuves ultérieures. Et dans chaque cas, la hiérarchie des phases et la cohérence de l'ensemble est intact.

Au total, comme nous l'avons exposé, la volonté de construire des épreuves conformes aux exigences des Socles de compétences nous ont amené à ne pas pratiquer un étalonnage de type normatif. Mais l'ensemble des essais que nous avons faits l'an dernier et cette année, et les résultats que nous venons de rappeler sont des indices forts de la cohérence et de la stabilité de l'instrument. Nous pouvons dire que nous avons là un outil d'évaluation raisonnablement fiable en rapport avec les Socles de compétences.

Il reste toutefois des limites qu'il est de notre devoir scientifique de rappeler :

- D'une part, comme nous l'avons déjà plusieurs fois signalé, notre échantillon n'est pas strictement représentatif.
- D'autre part, si nous voulions que nos épreuves permettent une évaluation plus juste des compétences émergentes des élèves face à des situations complexes, il serait plus adéquat, comme nous l'avons déjà signalé dans un des rapports intermédiaires, que les élèves soient évalués dans le cadre d'un travail de groupes. Car c'est souvent en petits groupes que sont menées, dans les classes, les activités de recherche qui ont quelques similitudes avec les situations de nos phases 1.

Mais ce qui est également décisif, c'est que les enseignants puissent s'approprier, non pas seulement les six épreuves que nous avons construites, mais le modèle même et soient capables de construire eux-mêmes des épreuves de ce type qui correspondent au niveau réel de leurs élèves et au contexte particulier de leur classe.

Or de cela, nous avons quelques indices positifs, puisque dans le cadre d'une opération de formation continue, des instituteurs en fonction ont créé des épreuves. Il faut toutefois rappeler qu'ils l'ont fait en équipe et qu'il s'agissait d'un groupe d'enseignants particulièrement motivés.

Ajoutons que, d'une manière plus général, grâce à la présence de la phase 3 et pour une part de la phase 2, les enseignants en exercice retrouvent dans notre modèle d'épreuve des formes d'exercices qu'ils connaissent et qu'ils ont l'habitude de faire pratiquer à leurs élèves. Ainsi, même si les phases 1 leur apparaissent nouvelles et ambitieuses, ils ne sont pas déconcertés par la totalité des épreuves et ne se sentent pas désavoués dans ce qu'ils savent faire. Nous l'avons constaté dans les échanges informels que nous avons eu sur le terrain, à l'occasion des passations des épreuves.

4.4. Facteurs et conditions d'apprentissage

Les considérations qui précèdent constituent la réponse à la commande qui nous était faite de construire des épreuves étalonnées pour l'école primaire en rapport avec les Socles de compétences.

Mais en tant que chercheurs intéressés à repérer les facteurs de réussite et d'échec, nous avons cherché, à l'occasion de l'essai de nos épreuves, à réunir certaines indications intéressantes. Les unes sont d'ordre quantitatif et les autres d'ordre qualitatif.

Du premier point de vue, la construction de notre échantillon nous permettait de collecter quelques indices d'un effet possible, sur les performances des élèves, du milieu d'origine, de la dynamique pédagogique propre à l'école et du mode de passation.

En ce qui concerne l'influence du milieu d'origine, on retrouve les résultats obtenus classiquement par les sociologues, c'est-à-dire des performances sensiblement meilleures chez les élèves venant de milieu plutôt favorisé. Mais les écarts ne sont pas aussi importants qu'on pourrait s'y attendre. Il faut dire que notre repérage du milieu d'origine se fait seulement à partir de l'appartenance ou non de l'école à une ZEP, ce qui évidemment ne permet pas de cerner finement le positionnement social particulier d'une classes et encore moins d'un élève.

C'est sur le plan de la dynamique pédagogique propre à l'école que nous avons eu des résultats surprenants, voire paradoxaux. Nous nous attendions à ce que soient plus performants les élèves des écoles qui nous avaient été signalées par les inspecteurs comme engagées dans des projets pédagogiques novateurs. Non seulement nous n'avons pas constaté cette tendance, mais dans certains cas les résultats provenant d'écoles réputées dynamiques ont été particulièrement mauvais. Dans le corps du présent rapport, nous présentons diverses hypothèses qui pourraient rendre compte de ces résultats, mais qui devraient bien entendu être mises à l'épreuve.

Nous nous attendions de même à ce que la présence de l'instituteur pour faire passer les épreuves (et notamment la phase 1) ait un effet sur les résultats. Or la répartition de ceux-ci de ce point de vue manque de netteté. On rencontre de bonnes performances de classes avec une passation assurée par les chercheurs, et des performances médiocres dans des classes où la passation a été assurée entièrement par l'instituteur ; mais aussi des cas exactement inverses à ces deux-là. Tout se passe comme si la présence de l'instituteur avait un effet, mais un effet qui pouvait être positif ou négatif selon la configuration relationnelle entre le maître et ses élèves, et peut-être aussi selon son attitude face à nos épreuves.

Nous avons voulu également voir si la connaissance et la réflexion des enseignants sur la réforme en cours pouvaient avoir des effets sur les performances à nos épreuves. L'étude menée sur ce point, dans les classes où nous avons fait passer nos épreuves, par un étudiant, Monsieur Schoeters, dans le cadre de son mémoire de licence, n'a pas apporté de résultats probants, peut-être par suite de forts effets de désidérabilité sociales.

Les résultats des trois derniers points sont clairement incertains et il n'est pas question d'en tirer des résultats fermes. Mais il nous semble que tout se passe comme si ce n'était pas l'engagement de l'instituteur dans telle démarche pédagogique, ni sa plus ou moins grande adhésion à la réforme qui avaient un effet déterminant sur la performance des élèves à une tâche complexe. S'il y a un effet de l'enseignant, les constats empiriques que nous avons fait sur le terrain nous porteraient à supposer qu'il tient plutôt à une attitude relationnelle de celui-ci par laquelle il marque à la fois ses attentes et sa confiance envers ses élèves. Il va de soi qu'une telle hypothèse exigerait une étude à elle seule.

Sur le plan qualitatif maintenant, la diversité des réponses des élèves en phase 1 offre un matériau d'une grande richesse pour comprendre les différents types de démarches qu'ils adoptent. L'analyse fine de ces réponses nous confirme qu'il y a dans la mise en œuvre d'une compétence une opération cognitive spécifique qui est l'interprétation de la situation, ou encore le cadrage de celle-ci. Ce qui apparaît également c'est que cette interprétation, les éléments qu'il y relève et qui lui paraissent pertinents, dépendent du sens qu'il donne à la tâche et, au-delà, au travail scolaire, à l'utilité qu'il y voit, à la manière dont il assume son statut d'enfant et d'élève, et à ce que certains auteurs appellent aujourd'hui le rapport au savoir.

Le temps et les moyens dont nous disposions ne nous ont pas permis d'exploiter systématiquement ce matériau et de mener complètement ces analyses.

Mais nous savons désormais clairement que notre modèle d'épreuve est à la fois un outil d'évaluation fiable en rapport avec les socles de compétences, mais aussi un outil didactique pour les enseignants, et un outil pour la recherche ouvert sur des investigations tant quantitatives que qualitatives.

Bibliographie

- ASTOLFI, J.P. (1992). *L'École pour apprendre*. Paris : ESF, collection pédagogies.
- BRUNER (J.), 1983, *Le développement de l'enfant : savoir faire, savoir dire*, (rad. M. Deleau), Paris, PUF.
- BRUNER (J.), 1987, *Comment les enfants apprennent à parler*, (Trad. Piveteau et Chambert), Paris, Retz.
- CRAHAY (M), 1996, *Peut-on lutter contre l'échec scolaire ?* Bruxelles, Pédagogies en développement, De Boeck Université.
- "Décret définissant les missions prioritaires de l'enseignement fondamental et de l'enseignement secondaire et organisant les structures propres à les atteindre", 24 juillet 1997, *Moniteur Belge*, 23.09.1997, pp. 24653-24674.
- DEPOVER (C), NOEL (B) (eds); 1999, *L'évaluation des compétences et des processus cognitifs ; Modèles, pratiques et contextes*, Bruxelles, Pédagogies en développement, De Boeck Université.
- DE VECCHI, G. (1992). *Aider les élèves à apprendre*. Hachette Education.
- DOISE (W.) et MUGNY (G.), 1981, *Le développement social de l'intelligence*, Paris, InterEditions.
- FABRE (M.), 1999, *Situations-problèmes et savoir scolaire*, Paris, PUF.
- GREGOIRE (J) (Ed.), 1996, *Evaluer les apprentissages. Les apports de la psychologie cognitive*, Bruxelles, Perspectives en éducation, De Boeck Université
- Groupe EVA, INRP, 1991, *Évaluer les écrits à l'école primaire*, Paris, Hachette Éducation, collection pédagogie pour demain, didactiques.
- LE BOTERF (G.), 1994, *De la compétence. Essai sur un attracteur étrange*, Paris, Les éditions d'organisation.
- LE BOTERF (G.), 1997, *Compétence et navigation professionnelle*, Paris, Les éditions d'organisation.
- LEVY-LEBOYER (C.), 1996, *La gestion des compétences*, Paris, Les éditions d'organisation.
- MEIRIEU (Ph.) et DEVELAY (M.), 1992, *Emile, reviens vite... Ils sont devenus fous*, Paris, ESF.
- MEIRIEU (Ph.), 1987, *Apprendre, oui... mais comment ?* Paris, ESF.
- MENDELSON(P.), 1990, "*La notion de transfert d'apprentissage en psychologie cognitive*", Cahiers Pédagogiques, n° 281, février 90, pp. 23-25.
- MINISTÈRE DE LA COMMUNAUTÉ FRANCAISE, 1999, *Socles de compétences, Enseignement fondamental et premier degré de l'enseignement secondaire*. Bruxelles.
- NOIRFALISE (R.), 1991, *Connaissances ou capacités*, Repères, Revue des IREM, n° 5, oct. 91, pp. 5-22.

- PERRENOUD (P.), 1994, *Métier d'élève et sens du travail scolaire*, Paris, ESF.
- PERRENOUD (Ph.), 1997, *Construire des compétences dès l'école*, Paris, ESF.
- PERRET-CLERMONT (A.-N.) et NICOLET (M.), 1988, *Interagir et connaître*, Fribourg, Delval.
- REY (B.), 1996, *Les compétences transversales en question*, Paris, ESF.
- REY (B.), 1998, "Compétences et connaissances", *Deux points, ouvrez les guillemets*, Lausanne, n°4, oct. 98, pp. 2-5.
- ROPE (F.) et TANGUY (L.) (sous la direction de), 1994, *Savoirs et compétences*, Paris, L'Harmattan.
- SCHNEUWLY (B.), 1990, *Diversifier l'enseignement du français écrit*, Neuchâtel et Paris, Delachaux et Niestlé.
- TARDIF (J.), 1992, *Pour un enseignement stratégique*, Montréal, Les éditions LOGIQUES.
- TURCO (G.), 1988, *Écrire et réécrire*, CRDP de Rennes.
- VIGOTSKI (L.), 1985, *Le problème de l'enseignement et du développement mental à l'âge scolaire*, in Schneuwly (B.) et Bronckart (J.-P.), *Vygotsky aujourd'hui*, Neuchâtel-Paris, Delachaux et Niestlé.
- VANDEVELDE (L.), 1982, *Aider à devenir*, Bruxelles, Editions Labor .

INTRODUCTION.....	1
CALENDRIER DE LA RECHERCHE POUR L'ANNÉE 2000 - 2001.....	3
PARTIE 1 : LE CADRE CONCEPTUEL.....	4
1.1. CONTEXTE DE LA RECHERCHE	4
1.1.1. <i>Introduction</i>	4
1.1.2. <i>L'organisation et les réformes de l'enseignement fondamental en Communauté française de Belgique</i>	5
1.2. L'HYPOTHÈSE DE TRAVAIL	6
1.3. LE CADRE CONCEPTUEL.....	6
1.3.1. <i>le concept d'évaluation</i>	6
1.3.1.1. le cadre des enjeux de l'évaluation	6
1.3.1.3. Construisons des épreuves diagnostiques	10
1.3.2. <i>Le concept de compétence</i>	11
1.3.2.1. L'élément essentiel de la notion de compétence : l'accomplissement efficace d'une tâche.....	11
1.3.2.2. Deux intérêts de la définition de la compétence comme aptitude à accomplir une tâche	11
1.3.2.3. Trois types de compétences.....	13
1.3.3. <i>Une épreuve à 3 phases</i>	14
1.4. CONCLUSIONS.....	15
2.1. LA MÉTHODOLOGIE ET LES RÉSULTATS DE LA PREMIÈRE ANNÉE	17
2.1.1. <i>La construction des épreuves</i>	17
2.1.2. <i>Passation des épreuves</i>	18
2.1.3. <i>Les résultats et constats</i>	18
2.1.3.1. La hiérarchie entre les phases des épreuves	19
2.1.3.2. La confirmation de notre hypothèse	21
2.1.3.3. La cohérence de l'épreuve et l'importance des procédures automatisées.	21
2.1.4. <i>Conclusions de la première année de recherche</i>	22
2.2. OBJECTIFS ET MÉTHODOLOGIE DE LA DEUXIÈME ANNÉE DE RECHERCHE	23
2.2.1. <i>Objectifs de la deuxième année de recherche</i>	23
2.2.2. <i>Méthodologie</i>	24
2.2.2.1. Une collaboration efficace avec les inspecteurs, des directeurs et des enseignants des cantons scolaires de la Province de Namur.....	24
2.2.2.2.1. les critères	25
2.2.2.2.2. L'élaboration de l'échantillon.....	26
2.2.2.2.3. L'intérêt de cet échantillon.....	27
2.2.2.3. Les épreuves : élaboration, passation et traitement statistique.....	28
2.2.2.3.1. Présentation des épreuves.....	28
2.2.2.3.2. Les modalités et critères de correction.....	28
2.2.2.3.3. L'analyse statistique	29
2.2.4. <i>Les questionnaires</i>	31
2.2.4.1. Analyse de « la réflexivité » de l'enseignant.....	31
2.2.4.2. Le travail d'appropriation du modèle par des enseignants	32

2.2.4.3. Les facteurs de réussite à des épreuves mesurant l'acquisition de compétences et les conditions d'apprentissage les plus adaptées à cette acquisition.	32
2.4.3.1. Les opérations cognitives qui entrent dans la mise en œuvre d'une compétence	33
2.2.4.3.2. Première étape de réflexion : procédures de base, habitude d'exercices spécifiques, familiarité.....	34
2.2.4.3.3. Deuxième étape de réflexion : une certaine visée sur la réalité....	38
TROISIÈME PARTIE : PRÉSENTATION ET ANALYSE DES RÉSULTATS.....	43
3.1. RÉSULTATS ET ANALYSE DES ÉPREUVES	43
3.1.1. <i>Introduction</i>	43
3.1.2. <i>Avant toute chose</i>	43
3.1.3. <i>Analyse des résultats globaux : la confirmation des différents constats</i> ...	44
3.1.3.1. La confirmation de l'hypothèse	44
3.1.3.2. La hiérarchie des phases	45
3.1.3.3. La cohérence de l'épreuve	46
PHASE 1 (B) / PHASE 2 (N).....	47
PHASE 1 (B) / PHASE 3 (N).....	47
PHASE 2 (B) / PHASE 3 (N).....	48
3.1.4. <i>Les résultats par épreuve : observation de différences</i>	49
3.1.4.1. Constats	49
3.1.4.3. Analyse des phases 1.....	51
3.1.4.4. Analyse des phases 2	54
SUR BASE DE CE GROUPEMENT (TABLEAU 9), NOUS POUVONS CONSTATER DE MANIÈRE ÉVIDENTE QUE LE TYPE DE CONSTRUCTION DES TÂCHES PROPOSÉES EN PHASES 2 A UNE INFLUENCE SUR LES GAINS OBTENUS. AINSI LES TÂCHES PROPOSANT UNE DÉCOMPOSITION ITEM PAR ITEM SONT CELLES QUI OBTIENNENT LE GAIN MOYEN LE PLUS ÉLEVÉ. LES DEUX TÂCHES UTILISANT UN SYSTÈME DE COTATION « MIXTE » OBTIENNENT UN GAIN MOYEN QUASIMENT DEUX FOIS MOINS ÉLEVÉ QUE DANS NOTRE PREMIER GROUPEMENT. ENFIN, LES TÂCHES OÙ L'ÉLÈVE REÇOIT UNE TÂCHE EN PHASE 2 TOTALEMENT IDENTIQUE À CELLE PROPOSÉE EN PHASE 1 ET POUR LAQUELLE ON UTILISE LA MÊME GRILLE DE CORRECTION OBTIENNENT UN GAIN MINIME.	56
3.1.4.5. Analyse des phases 3	56
3.1.4.6. Conclusions des analyses par épreuve	57
3.1.5. <i>Les influences</i>	58
3.1.5.1. Existe-t-il réellement une influence du milieu social sur les compétences des élèves à résoudre des tâches complexes ?	58
3.1.5.1.1. Premiers constats	58
3.1.5.1.2. Analyse statistique des phases 1 et 2	59
3.1.5.1.3. Conclusions.....	60
3.1.5.2. Y a-t-il une influence de la dynamique de l'équipe pédagogique d'une école sur les compétences des élèves à résoudre des tâches complexes ? ..	60
3.1.5.3. Les modalités d'administration des épreuves ont-elles une influence sur les résultats des élèves ?.....	61

3.1.5.3.1. Premiers constats	62
3.1.5.3.2. Analyse statistique des phases 1 et 2	63
3.1.5.3.3. Hypothèses et conclusions	64
3.1.6. <i>Analyse des épreuves par classe</i>	65
3.1.6.1. Introduction	65
3.1.6.3. L'effet enseignant	70
3.2. LA RÉFLEXIVITÉ DE L'ENSEIGNANT	70
3.2.1 <i>Méthodologie</i>	70
3.2.2 <i>Résultats et conclusions</i>	72
3.3. LE TRAVAIL D'APPROPRIATION DU MODÈLE PAR LES ENSEIGNANTS	72
3.3.1. <i>Introduction</i>	72
3.3.2. <i>Analyse du dépouillement du questionnaire</i>	73
3.3.2.1 Si peu de temps et tant à faire... (Oscar Levant)	73
3.3.2.2. L'organisation du travail	74
3.3.2.3. Le rapport à ce type d'épreuves	74
3.3.2.4. Passation des épreuves	75
3.3.2.5. Correction des épreuves	75
3.3.2.6. Conclusions et pistes de réflexion	76
3.4. CONCLUSION DE LA TROISIÈME PARTIE	76
4. CONCLUSION GÉNÉRALE.....	78
4.1. UNE CONCEPTION DES COMPÉTENCES	78
4.2. UNE ÉPREUVE EN TROIS PHASES.....	79
4.4. FACTEURS ET CONDITIONS D'APPRENTISSAGE.....	82
BIBLIOGRAPHIE	84