

# Les dispositifs d'évaluation formative sont-ils efficaces ?

Gérard YVROUD

## Introduction

Peut-on juger de la réussite d'un apprentissage sans s'appuyer sur une évaluation ? La chose paraît difficilement concevable. C'est sans doute la raison pour laquelle l'école, dont la fonction est de permettre la réussite de nombreux apprentissages, est aussi le lieu où se déroulent tant d'évaluations. J.M. Monteil (Monteil, 1990, p. 34) fait observer que : *"Si dans les années 1970 il a paru illégitime, à certains, de procéder à des évaluations, on ne peut raisonnablement accepter une telle attitude."* De nos jours, la nécessité d'évaluer lors d'un enseignement, semble largement admise. C'est ce qu'avaient déjà indiqué G. Noizet et J.P. Caverni (Noizet et Caverni, 1978, p. 190) en ces termes : *"l'acte d'évaluer, qui permet de confronter un résultat à un objectif, est nécessairement inscrit dans tout processus d'apprentissage."* Actuellement, on ne semble plus contester la nécessité d'évaluer, à l'école. Au cours des vingt dernières années, de nombreuses recherches ont été conduites autour de l'évaluation. Elles ont mis en évidence la complexité de ce concept et son caractère multidimensionnel.

On admet aujourd'hui communément, dans le monde de l'enseignement, qu'une évaluation sommative se distingue d'une évaluation formative. C. Hadji (Hadji, 1989, p. 60) nous indique que : *"L'évaluation est dite sommative lorsqu'elle se propose de faire le bilan (la somme), après une ou plusieurs séquences ou, de façon plus générale, après un cycle de formation."* Cette fonction s'oppose à celle remplie par l'évaluation formative que P. Perrenoud (Perrenoud, 1998, p. 120) définit ainsi : *"Est formative toute évaluation qui aide l'élève à apprendre et à se développer, autrement dit, qui participe à la régulation des apprentissages et du développement dans le sens du projet éducatif."* Il convient cependant de ne pas oublier que cette distinction, pour commode qu'elle soit, ne demeure toutefois qu'une construction de modèles théoriques. J. Cardinet (Cardinet, 1986, p. 17) nous a fait observer que : *"La distinction entre les deux finalités, formative et sommative, représente une polarité importante, mais parmi beaucoup d'autres oppositions possibles."* Averti de ce parti pris, nous avons tout de même résolu d'utiliser cette opposition car elle nous permettait de bien circonscrire cette fonction pédagogique de l'évaluation qui va nous intéresser dans la suite de cet article.

Les nombreuses recherches conduites autour de l'évaluation formative se sont généralement attachées à clarifier ses intentions, et, plus rarement, à en exposer des modalités variées. Nous avons, pour notre part, décidé de tenter d'en observer ses effets. Il apparaît, aujourd'hui, que l'évaluation formative vise à informer pour, d'une part, faciliter les régulations indispensables du processus enseigner, et, d'autre part, optimiser les rendements du processus apprendre. Dans le contexte de la pratique, ces deux actions sont évidemment mêlées. Cependant, nous avons choisi de restreindre notre observation aux effets de cette évaluation sur l'apprentissage. N'est-ce pas là, d'ailleurs, sa finalité essentielle ?

Dans un premier temps, nous exposerons un des objectifs de l'évaluation formative qui vise à permettre à l'apprenant de mieux "voir" ce qu'on attend de lui, et nous examinerons certains des effets attendus. Ensuite, nous présenterons la méthode que nous avons retenue pour notre travail (1). Enfin, nous ferons état de quelques résultats obtenus que nous discuterons.

## 1. Quels effets peut-on attendre de l'évaluation formative ?

Suffirait-il de mettre en œuvre un dispositif d'évaluation formative pour obtenir une amélioration des apprentissages ? Les lectures que nous avons réalisées pour trouver une réponse à cette question ne nous ont pas apporté de réponse. Nous avons donc décidé d'essayer d'en fournir une. Nous faisons, alors, l'hypothèse que la mise en œuvre d'un dispositif d'évaluation formative devait amener l'élève et le maître à mieux "voir" ce qui se déroulait au cours des processus "apprendre" et "enseigner", donc à faciliter les apprentissages. Il semble, en effet, que pendant le déroulement de ces deux processus, les deux acteurs concernés soient placés dans une situation qui les rende, d'une certaine manière "aveugles". C'est ce que nous allons essayer de montrer maintenant.

### 1.1. Des acteurs aveugles

On admet communément la pertinence d'une modélisation de la situation d'enseignement / apprentissage sous la forme du triangle pédagogique (Houssaye, 1989). Or, dans une situation classique d'apprentissage, réalisée à l'école sous la conduite d'un enseignant, on peut montrer que les deux acteurs, l'élève et le maître, sont placés dans une position d'où ils ne peuvent pas "voir", c'est-à-dire qu'ils sont, normalement, privés d'indicateurs qui les renseignent sur la position de celui qui apprend, relativement aux objectifs d'apprentissage. C'est ce que fait apparaître la figure 1.

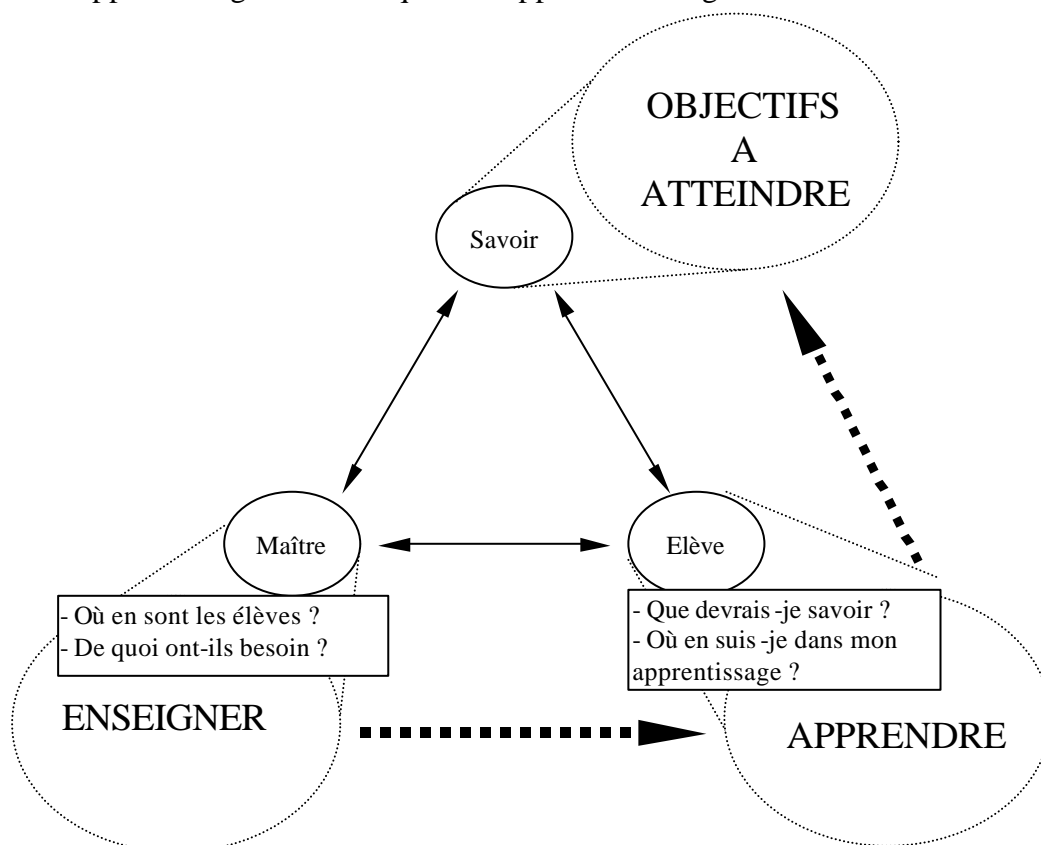


Figure 1 - Enseigner et apprendre : des processus réalisés "en aveugle"

Dans une situation d'apprentissage scolaire, c'est le maître qui sait ce que l'élève doit apprendre. Le processus "enseigner" consiste, pour lui, à amener l'élève à s'approprier un nouveau savoir. Or, l'enseignant ignore la situation exacte de ses élèves face aux

apprentissages considérés. On peut estimer que les réponses aux deux questions suivantes sont, pour lui, essentielles :

- Où en sont mes élèves en regard du savoir à enseigner ?
- Quels sont leurs besoins pour apprendre ce que je veux qu'ils sachent ?

Il ne suffit pas que le maître observe ses élèves pour trouver la réponse à ces deux questions. Seule une évaluation pourra apporter des réponses, réponses le plus souvent partielles dont il devra se contenter pour décider de ses stratégies. Sans l'information, toute relative, fournie par une telle évaluation, le maître est contraint de "piloter son enseignement sans visibilité réelle".

Du côté de l'élève, la situation ne semble pas plus "transparente". En effet, pour l'élève, placé en situation d'apprendre, le but ne peut guère apparaître nettement. Comment celui qui apprend pourrait-il se représenter clairement ce qu'il ne connaît pas encore ? Ce sera seulement quand il aura atteint l'objectif fixé que tout devrait s'éclairer, pour lui. Au début d'un apprentissage, alors qu'il n'a qu'une image forcément floue du but à atteindre, l'élève ne semble pas en mesure de savoir ce qui lui manque. On doit donc l'aider, d'une part, à mieux se représenter le but visé, d'autre part à mieux situer sa position par rapport à ce but. C'est bien là que peut intervenir l'évaluation formative. Si enseigner consiste à aider chaque élève à apprendre, l'information fournie par l'évaluation formative devrait concourir à faciliter les apprentissages. C'est donc autour de ce questionnement que nous avons construit notre problématique. Nous faisons l'hypothèse que la simple communication, à l'élève, de l'information obtenue au moyen d'une évaluation formative allait faciliter ses apprentissages.

Nous étions encouragé dans ce sens par les réflexions de certains chercheurs. Ainsi C. Hadji (Hadji, 1997, p. 16) nous indique que : "*Il nous paraît certain, de ce point de vue, que la bonne perception par l'élève, de la cible visée est l'une des conditions de sa réussite.*" De même, C. Delorme (Delorme, 1987, p. 23) précise : "*Or, c'est bien en ayant la possibilité de faire part de l'état de leur travail, en pouvant vérifier s'ils sont sur la bonne voie, ou analyser pourquoi ils sont dans une impasse, que les élèves peuvent le mieux apprendre et progresser.*" Il semble que ce soit, là, le bon sens même. Si l'on accepte l'idée que le processus "apprendre" est, pour une grande part, inobservable, on en conclut que sa réalisation nécessite une information que l'on doit trouver ailleurs que dans la simple observation de la situation d'enseignement / apprentissage. Expliciter, pour chaque élève, le but à atteindre, communiquer à chacun sa position, par rapport au but visé semble une démarche fructueuse. Encore conviendrait-il de s'interroger sur les raisons qui pourraient expliquer de telles améliorations.

## **1.2. Des effets positifs espérés**

On peut faire observer que, depuis que Bloom et ses collaborateurs ( Bloom, B.S., Hastings, J.T. et Madaus, G.F., 1971) ont tenté de mettre une évaluation devenue "formative" au service d'une amélioration des apprentissages, une hypothèse a guidé la mise en œuvre de ces dispositifs : l'idée que renseigner l'élève sur l'état de ses connaissances pouvait l'aider à mieux apprendre. Trois raisons, au moins, pourraient expliquer cet éventuel effet positif.

● Pour commencer, on peut dire qu'un dispositif d'évaluation formative, parce qu'il oblige le maître à expliciter ses attentes, va favoriser, chez l'élève, une meilleure connaissance de ce qu'on attend de lui. Pour permettre que l'apprentissage réussisse, l'enseignant organise des exercices où l'élève se trouve confronté à la réalisation d'une tâche. L'élève est alors amené à réaliser des actions pour réussir une production. Or, dans une classe, l'élève agit en orientant son action en fonction des attentes qu'il suppose chez l'enseignant. C'est ce que nous disent O. et J. Veslin (Veslin, 1992, p. 70) "*Quand il a une tâche à réaliser, l'élève se forme*

*une idée plus ou moins nette de ce qu'on lui demande. En général cette idée est au moins partiellement erronée : elle ne correspond pas à ce que l'enseignant attend.*" La mise en œuvre d'une évaluation formative va d'abord amener l'enseignant à expliciter ses attentes. Cependant, dans les premiers temps de l'apprentissage, le dialogue, pour essentiel qu'il soit, est tout de même une sorte de dialogue de sourds. En effet, l'enseignant nomme des éléments qui n'ont pas de sens pour l'élève qui ne sait pas encore. Que peut signifier, pour un élève qui apprend, qu'il devra être capable de trouver le quotient approché au dixième près par défaut, s'il ignore ce qu'est un quotient approché ? Mais, comment pourrait-il le savoir à ce moment, puisque c'est précisément ce qu'on veut qu'il apprenne !

L'apprentissage va consister à amener l'élève à se construire, progressivement, une représentation de plus en plus correcte des éléments nommés par l'enseignant, au travers d'essais successifs. C'est précisément parce qu'elle devrait faciliter cette élaboration progressive que l'évaluation formative peut apparaître comme une aide à l'apprentissage. Ainsi, l'explicitation des critères d'évaluation pourrait concourir à faciliter, chez l'élève, la construction d'une meilleure représentation des buts à atteindre.

- Un deuxième élément pourrait contribuer à faire apparaître l'évaluation formative comme une aide à l'apprentissage. Il s'agit d'une partie essentielle du dispositif, quand, après avoir obtenu l'information qu'il recherche, l'enseignant va la mettre à la disposition de l'élève. Cette information en retour, va permettre à celui qui apprend de corriger son action pour mieux l'adapter aux attentes. L'élève va ainsi pouvoir s'améliorer grâce aux informations qui feront apparaître ses points positifs, qu'il aura à cœur de mémoriser pour les reproduire au moment adéquat, et ses points négatifs, ceux-là seulement qu'il devra modifier pour améliorer ses performances. C'est par la communication d'une information, qui aide l'élève à mieux se situer par rapport aux attentes du maître, que l'évaluation formative peut apparaître également comme une aide à l'apprentissage. C'est ce qu'avait vu C. Hadji (Hadji, 1989, p. 124) qui affirmait que : *"Pour ajuster l'action réalisée au but poursuivi, il faut disposer d'informations sur sa situation par rapport au but."* Et, en utilisant la métaphore du pilotage d'un avion, l'auteur précisait, dans sa conclusion, (Hadji, 1989, p. 181) que : *"L'évaluateur est le navigateur qui fait le point pour permettre au pilote d'amener l'avion à la destination souhaitée."*

Ainsi, la mise en œuvre d'un dispositif d'évaluation formative, qui va organiser des rétroactions, dans l'espoir d'amener l'élève à analyser son action, en termes de réussite ou d'échec, pourrait apparaître comme une aide pour le guider vers un but qui ne peut que se préciser, au fur et à mesure du déroulement de l'apprentissage. Cependant, cette action bénéfique se réalise dans une logique de l'après coup. Cette information ne peut, en effet, apparaître que comme le diagnostic porté à partir d'une production achevée, ne serait-ce que provisoirement. Or, il semble que l'intérêt de l'évaluation formative puisse également être vu dans une logique d'action future.

- Un troisième élément peut, en effet, conduire à rechercher des avantages de l'évaluation formative dans une régulation proactive des actions des apprenants. On a vu que l'action du maître ne pouvait résider que dans un accompagnement, plus ou moins précis, des actions de l'élève. En dernier ressort, c'est toujours l'élève qui apprend ... s'il apprend ! Favoriser l'apprentissage ne peut que conduire à rendre l'élève encore plus capable d'autoréguler ses actions. L'évaluation formative se transforme ainsi quelque peu, pour devenir ce que G. Scallon a nommé l'évaluation formatrice. G. Nunziati, qui a été, avec J.J. Bonniol, à l'origine de cette évolution nous indique que (Nunziati, 1990, p. 48) : *"C'est G. Scallon qui suggéra, en 1982, le recours au vocable 'formatrice', pour bien spécifier les*

*différences entre ce que l'on mettait sous le mot 'formative' et les pratiques issues de la recherche de Marseilleveyre, notamment le fait que l'auto-évaluation correcte était considérée comme une compétence primordiale à construire." C'est ce qu'avait déjà fait observer C. Hadji (Hadji, 1989, p. 172) : "L'une des hypothèses sur lesquelles repose l'idée d'évaluation formatrice est que l'élève apprend d'autant mieux qu'il devient autonome."*

Dans cette logique, l'élève n'attend plus seulement le résultat de l'évaluation de son action pour tenter de la réguler. En l'invitant à participer à l'élaboration des critères d'évaluation, l'enseignant l'amène à mieux prendre conscience des attentes. Cette phase d'explicitation conduit ainsi l'apprenant à anticiper sur la conduite de sa propre action pour essayer de la rendre plus conforme aux buts qu'il conçoit. Ce ne sont donc plus seulement les résultats de l'évaluation qui vont influencer sur la conduite des apprenants, mais également une connaissance analytique des buts visés. Dans cette optique, P. Perrenoud (Perrenoud, 1998, p. 130) fait observer que : *"Il ne s'agit plus de multiplier les feed-back externes, mais de former l'élève à la régulation de ses propres processus de pensée et d'apprentissage, partant du principe que l'être humain est, dès sa prime enfance, capable de se représenter, au moins partiellement, ses propres mécanismes mentaux."* G. Nunziati, dans l'article cité (Nunziati, 1990, p. 51) distingue : *"deux hypothèses fondamentales de l'évaluation formatrice [...] à savoir que :*

- l'appropriation par les élèves des outils d'évaluation des enseignants,*
- et la maîtrise par l'apprenant des opérations d'anticipation et de planification,*

*sont les deux objectifs prioritaires d'une démarche d'évaluation qui se veut formatrice, c'est-à-dire d'une démarche de régulation conduite par celui qui apprend."*

De ce qui précède, il ressort qu'un troisième élément, constitutif d'une forme dérivée de l'évaluation formative, nommée évaluation formatrice, pourrait favoriser l'apprentissage. La "négociation" des critères d'évaluation conduirait l'élève à réguler, de manière proactive, sa propre action pour mieux l'adapter au but visé.

Ainsi, l'évaluation formative présenterait au moins trois avantages pour l'apprentissage.

- Elle faciliterait, chez l'élève, une meilleure connaissance des attentes du maître.
- Elle aiderait l'élève à profiter davantage du temps de l'apprentissage en utilisant les rétroactions pour mieux identifier ses propres processus, soit pour les conserver en mémoire s'ils s'avèrent corrects, soit pour les modifier s'ils apparaissent erronés.
- Elle permettrait également à l'élève de réguler, de manière proactive, ses processus pour les adapter, par anticipation, aux attentes qu'il découvrirait.

Il restait à montrer que l'élève, placé en situation d'apprentissage, progresse effectivement mieux, quand il reçoit une information en retour. Nous voulons faire deux observations qui nous amèneront à compléter notre investigation.

Si l'usage d'une évaluation formative améliore les apprentissages, on doit pouvoir observer une amélioration des performances en termes de progression globale. Il serait intéressant de pouvoir observer s'il existe des différences de progression qui seraient liées à l'usage d'un dispositif d'évaluation formative. Cela nous a conduit à émettre une première hypothèse.

**Hypothèse 1** : Au cours d'une séquence d'apprentissage, l'augmentation du niveau de connaissances des élèves, entre le début et la fin de la séquence, sera plus importante s'ils bénéficient d'une évaluation formative.

De plus, plusieurs auteurs, (Delorme, 1987 ; Amigues et Zerbato-Poudou, 1996 ; Hadji, 1997), mettent en avant l'importance qu'il y aurait à aider l'élève à se repérer dans ses apprentissages. Nous pouvions donc également nous interroger sur la possibilité que posséderait un dispositif d'évaluation formative d'aider l'élève à mieux se représenter le but visé, et à mieux estimer sa "position" par rapport à ce but. Cela nous a conduit à tester une seconde hypothèse.

**Hypothèse 2 :** L'usage d'un dispositif d'évaluation formative aide les élèves à mieux se repérer dans leurs apprentissages : ils connaissent mieux le but visé et ils estiment plus justement leur position par rapport à ce but.

Pour répondre à ces questions, nous avons décidé de construire un dispositif expérimental qui, tout en respectant le fonctionnement habituel d'une classe, permette néanmoins de répondre à ces interrogations.

## **2. Les attentes à l'épreuve du réel : l'expérimentation**

Nous avons donc dû imaginer un dispositif permettant d'observer s'il existait des différences, chez les élèves, selon que leurs apprentissages bénéficient, ou non, de "l'aide" d'un dispositif d'évaluation formative. Il fallait pouvoir comparer des classes qui travailleraient de manière, à la fois, différente quant à l'usage de dispositifs d'évaluation formative, mais suffisamment semblable, sur les autres points, pour pouvoir les comparer entre elles. Nous avons décidé de nous limiter à observer deux classes. En effet, s'agissant d'un travail exploratoire, il ne nous était pas possible de multiplier le nombre des classes soumises à observation et expérimentation.

### **2.1. Le dispositif expérimental**

Le principe du dispositif consiste à placer deux classes, jugées équivalentes, en situation d'enseignement / apprentissage. Les maîtres des deux classes organiseront la même séquence d'enseignement avec leurs élèves respectifs. Dans la classe 1, le maître dispensera son enseignement sans avoir recours à un dispositif d'évaluation formative. Dans la classe 2, en revanche, l'enseignant, pour la même séquence, mettra en œuvre un dispositif d'évaluation formative. Dans chaque classe, la séquence d'enseignement sera précédée d'un test initial et suivie d'un test final, qui permettra d'observer si les résultats des élèves évoluent de manière différente.

Notre problématique renvoie à la question de l'efficacité des dispositifs d'évaluation formative. Il convient donc de faire en sorte que le temps consacré à l'enseignement / apprentissage soit identique dans les deux situations. Nous avons donc imaginé que les élèves de la classe 1 - sans évaluation formative - réaliseront des exercices de renforcement, ce qui est classique à l'école élémentaire, pendant que leurs homologues de la classe 2 seront soumis à un dispositif d'évaluation formative. Dans cette dernière classe, l'information obtenue, grâce à ce dispositif d'évaluation, devrait mettre en évidence des différences entre les élèves, entre ceux qui réussissent bien et ceux qui éprouvent des difficultés. Il est donc très important de prévoir, alors, un temps de travail différencié, pour tenter de remédier aux difficultés constatées. Il conviendra de s'assurer que les deux activités, évaluation formative et exercices de remédiation, seront réalisées dans la même durée que les exercices de renforcement de la classe 1.

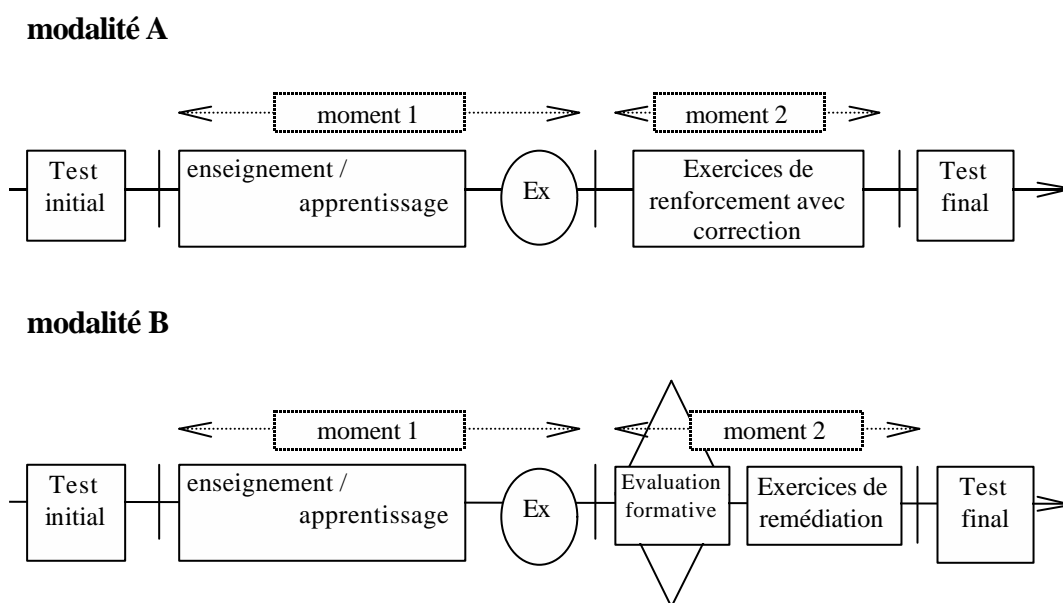
Cependant, un tel plan d'expérimentation ne permettrait pas de mettre en évidence un effet de l'usage du dispositif d'évaluation formative. Chaque classe fonctionnerait, ici, selon

une seule des deux modalités que nous avons retenues : avec une évaluation formative, ou sans elle. Si nous observions des différences entre les deux classes, la présence de la variable "évaluation formative" ne pourrait pas être considérée comme seule responsable de cette variation. Nous avons donc décidé de modifier notre plan d'expérimentation en faisant en sorte que chaque classe soit placée, tour à tour, en situation de bénéficier de l'usage d'une évaluation formative. Cela permet de comparer chaque classe à elle-même.

Nous avons donc fait réaliser deux séquences d'enseignement. Chaque séquence comporte deux modalités de mise en œuvre : une mise en œuvre sans dispositif d'évaluation formative - modalité A -, et une mise en œuvre avec dispositif d'évaluation formative - modalité B -. Ainsi, notre plan d'expérimentation est devenu le suivant.

- Mise en œuvre de deux séquences d'enseignement dans des classes de l'école élémentaire.
- Réalisation de chaque séquence selon deux modalités :
  - modalité A : sans dispositif d'évaluation formative
  - modalité B : avec dispositif d'évaluation formative.
- Pour chaque séquence d'enseignement, alternance des deux modalités, pour les deux classes retenues.

Le plan d'expérimentation peut alors être résumé à l'aide de la figure 2 et du tableau I suivants.



**Figure 2 - Deux modalités pour la réalisation des séquences d'enseignement / apprentissage**

*Remarque 1* : Les activités du moment 1 étant les mêmes pour les deux modalités, la durée qui leur sera consacrée sera la même. Les activités du moment 2 seront différentes selon les deux modalités, mais on veillera à leur consacrer la même durée dans les deux cas. Chaque classe réalisera, à son tour, un apprentissage en utilisant un dispositif d'évaluation formative, ainsi que le montre le tableau I.

*Remarque 2* : Le terme "Ex" qui apparaît dans chaque "moment 1" correspond à des phases d'activité des élèves quand ils sont appelés à réaliser des exercices où ils sont susceptibles de mettre en œuvre les procédures visées.

**Tableau I - Mise en œuvre alternée des séquences d'enseignement selon les deux modalités, dans les deux classes**

	séquence 1	séquence 2
classe 1	modalité A	modalité B
classe 2	modalité B	modalité A

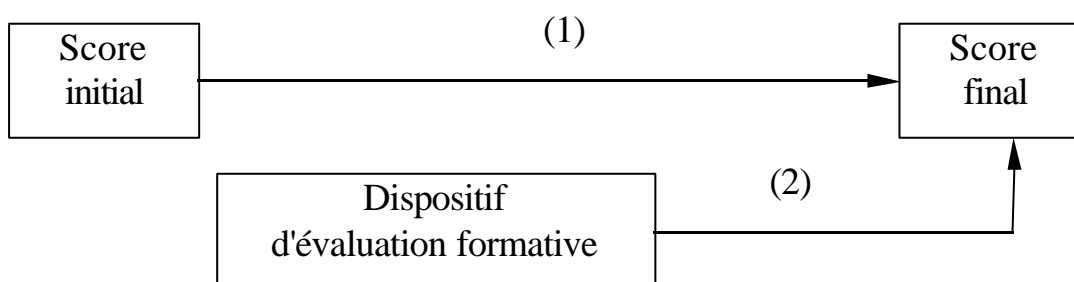
## 2.2. Le recueil des données

Le dispositif expérimental que nous avons imaginé nous a permis d'obtenir deux types de données : des données qualitatives et des données quantitatives.

### a. Les données quantitatives

Les différents tests mis en place nous ont permis de comparer les performances des élèves. Les tests initiaux ont fourni une image des compétences des élèves avant qu'ils commencent leur apprentissage. En décidant de noter ces tests, nous avons pu calculer des moyennes et observer la dispersion des notes grâce aux écarts types. Il était alors possible de comparer les classes à travers les performances de leurs élèves. Le même travail a pu être réalisé avec les résultats terminaux. De plus, en utilisant les mêmes séries de tests, initiaux et terminaux, nous avons pu avoir une image des évolutions des élèves entre le début et la fin de la séquence d'enseignement. Il apparaît, alors, possible de comparer les évolutions des deux classes, et également l'évolution de la même classe selon qu'elle a, ou non, bénéficié de l'usage d'un dispositif d'évaluation formative.

Pour estimer l'effet spécifique de l'évaluation formative sur les apprentissages, nous avons décidé d'insérer cette relation dans un modèle causal simple qui intègre deux variables : le niveau initial de l'élève et l'usage de l'évaluation formative. Ce modèle est représenté dans la figure 3.



**Figure 3 - Structure analytique adoptée pour cerner l'effet de l'évaluation formative**

La flèche (1) indique que le score final est fonction du score initial. En effet, un élève qui obtient une bonne note au test initial, est donc un élève qui avait déjà de bonnes connaissances dans le domaine avant de bénéficier de la séquence d'enseignement. On peut comprendre que cet élève pourra utiliser de nouveau ses connaissances lors du test final. De plus, elles l'aideront à construire son nouveau savoir. Dans les études qui mettent en relation score initial et score final, on observe, de manière classique, une forte corrélation entre les deux scores.

La flèche (2) indique l'effet du dispositif d'évaluation formative sur les performances au test final. Il est clair que ces deux variables, score initial et évaluation formative, ne sont pas les seules à influencer sur le score final. On peut penser que d'autres variables doivent jouer sur les performances finales de l'élève. On peut citer, par exemple, la capacité d'apprentissage propre à chaque élève, l'efficacité du maître à enseigner, la facilité relative de l'élève dans le domaine considéré. L'usage d'un modèle multivarié nous permettra de séparer les effets des variables évoquées et d'estimer l'effet net de l'évaluation formative.

## **b. Les données qualitatives**

Notre problématique nous a conduit à nous interroger sur le fait qu'un dispositif d'évaluation formative pourrait avoir un effet positif sur les capacités des élèves à mieux se repérer dans leurs apprentissages (Hypothèse 2). Afin de tester cette hypothèse, nous avons décidé de conduire des entretiens. Nous avons retenu l'idée de choisir, dans chaque classe, six élèves. Le choix a été réalisé selon la plus ou moins grande réussite des enfants au test initial. Les performances observées à ce test permettaient de classer les élèves en cinq groupes, rangés du groupe 1 - les meilleures performances - au groupe 5 - les moins bonnes -. Nous avons décidé de retenir deux enfants du groupe 1 - forts -, deux du groupe 3 - moyens - et deux du groupe 5 - faibles -. Les entretiens ont été réalisés à l'issue du moment 2 de la séquence (voir figure 2 ), c'est-à-dire avant le test final. Ils étaient de type semi-directif. Réalisés avec des enfants, ces entretiens ne devaient pas durer trop longtemps - dix à quinze minutes par élève -. Ces entretiens ont été enregistrés et retranscrits afin d'être analysés. Nous avons retenu le guide d'entretien ci-dessous. Les questions de 1 à 3 visent à observer comment l'élève se représente le but à atteindre. Les autres questions cherchent à voir comment l'enfant estime son propre niveau de performances.

### **Guide des entretiens avec les élèves**

1. Sur quoi as-tu travaillé au cours des deux séances ?
2. Que fallait-il connaître pour réussir ?
3. Est-ce que tu peux dire ce que le maître attend des élèves ?
4. Est-ce que c'était facile pour toi ?
5.
  - a. Qu'est-ce que tu penses avoir bien compris ?
  - b. Qu'est-ce que tu penses n'avoir pas compris ?
6. Est-ce que tu penses que tu vas bien réussir le prochain test qui portera sur ce sujet ?
7. Si on devait te noter, quelle note, sur vingt, penserais-tu avoir ?

Après avoir défini la méthodologie du recueil de données, il convenait de choisir nos classes, ce qui allait conditionner les séquences à réaliser.

### **2.3. Le choix des séquences**

Dès le départ, nous avons retenu le principe de travailler avec des classes du cycle 3 de l'école élémentaire, dans le domaine des mathématiques. Enseignant nous-même dans ce cycle de l'école primaire, il nous était, en effet, plus facile de concevoir des séquences d'enseignement pour les proposer à des maîtres qui accepteraient de travailler sur ce sujet. Le domaine des mathématiques nous apparaissait aussi présenter un avantage. Nous voulions que les séquences d'enseignement soient réduites dans leur durée afin de ne pas alourdir le travail des enseignants. Il semble plus simple de s'engager à collaborer pendant trois séances que pendant six. Et nous pensons qu'il est plus facile, dans cette discipline, d'isoler de courtes

séquences d'enseignement. Il nous paraissait bien plus compliqué de conduire le même travail à partir de séquences de production d'écrits.

Ce choix réalisé, nous avons dû trouver deux classes volontaires, constituées d'élèves suffisamment semblables. Il convenait également que les maîtres de ces classes possèdent des programmations assez proches afin qu'une séquence proposée dans une classe ne corresponde pas à un thème déjà traité dans l'autre classe. Nous nous sommes décidé pour retenir une grosse école, dans laquelle nous pourrions trouver deux classes du même niveau de scolarité, constituées d'élèves aux performances équivalentes. Nous avons retenu deux classes volontaires à l'école F. B. de Grenoble (classes de CM2 - Cours Moyen deuxième année). Nous avons réalisé la première séquence à la fin du mois de mars 1999. Il s'est avéré qu'une des séquences préparées n'a pas convenu à notre recherche. Les performances des élèves, au test initial, sont apparues trop élevées et ont conduit à ne pas engager la séquence d'apprentissage, car elle ne se justifiait que pour trois ou quatre élèves de chaque classe. Nous avons dû bâtir une autre séquence et notre expérimentation nous a donc conduit à retenir deux séquences, élaborées en mathématiques, pour des élèves du cycle 3.

La séquence 1 porte sur la division de deux entiers, avec quotient décimal. L'objectif est d'amener les élèves à donner du sens aux chiffres que l'on obtient quand on calcule un quotient décimal. La séquence 2 relève du calcul des durées. Elle vise à amener les élèves à effectuer des changements d'unités et à réaliser des calculs sur les mesures de durées.

Chacune des deux séquences s'inscrivait dans la programmation des classes. Pour chacune, et en accord avec les enseignants, nous avons élaboré les tests, proposé l'enchaînement des séances d'enseignement pour les maîtres, retenu les exercices pour les élèves et prévu le déroulement selon les deux modalités. Les thèmes choisis nous ont amené à construire des séquences de longueur légèrement différente, ainsi qu'on peut le voir dans le tableau II. Nous avons fait en sorte que le temps consacré à la réalisation d'une séquence soit le même pour les deux classes.

**Tableau II - La durée des deux séquences**

	moment 1	moment 2
séquence 1 : <i>la division</i>	1 séance	1 séance
séquence 2 : <i>les durées</i>	2 séances	1 séance

Pour la séquence sur la division, nous avons retenu l'idée que les exercices du test initial et du test final seraient exactement les mêmes, mais nous avons modifié les nombres proposés dans les calculs. Nous avons peur qu'apparaisse un effet d'apprentissage compte tenu du rapprochement, dans le temps, des deux tests. Après passation, nous regrettons ce choix. Il peut en effet apparaître qu'un élève ait réussi la division proposée au test initial parce qu'elle l'amenait à calculer  $42 : 6$  et qu'il connaît bien la table du 6. En revanche, le même élève aura échoué à la division proposée au test final parce qu'elle comportait  $56 : 8$  et que ce même élève connaît mal la table du 8. Après la réalisation de cette séquence et l'analyse de ce problème, nous avons décidé, pour la séquence sur les durées, que le test initial serait le même que le test final.

Chaque séquence a fait l'objet d'une préparation écrite qui en prévoit les grandes lignes. Ces documents ont été proposés aux enseignantes des deux classes. Un entretien a permis de préciser les points jugés obscurs. Après la passation du test initial, les enseignantes ont donc réalisé chaque séquence, en suivant le protocole prévu, mais en le traitant à leur

manière. Elles ont sûrement adapté ce qui était proposé selon leur propre conception de l'enseignement, et ce, pour chacun des thèmes retenus. Chaque protocole, aussi précis soit-il, ne peut jamais être qu'interprété à partir de sa transcription. Il y a, là, présence d'une variable "maître" qui a pu influencer sur les performances finales des élèves. Une façon de réduire ce biais consiste, comme nous l'avons prévu, à comparer chaque classe avec elle-même, pour chacune des deux modalités. Néanmoins, on peut penser que le changement de séquence a pu conduire chaque enseignante à être plus ou moins à l'aise avec le thème retenu, et, peut-être, plus ou moins efficace. La solution parfaite qui aurait consisté à observer, dans l'idéal, la même classe effectuer deux fois le même apprentissage, selon deux modalités, reste du domaine de l'impossible : les élèves ne peuvent découvrir deux fois les mêmes notions !

Si la mise en œuvre des séances du moment 1 a été confiée à chaque enseignante, nous avons voulu réduire les sources de différences pour le moment 2. A cette fin, nous avons décidé de confier la conduite de ce moment à un même enseignant qui a donc opéré dans les deux classes. Cela permettait de nous assurer que les dispositifs d'évaluation formative seraient bien mis en œuvre de la même manière. C'était aussi avoir l'assurance que les deux modalités seraient bien respectées, sans que l'une interfère dans l'autre.

#### **2.4. Le dispositif d'évaluation formative**

S'il est assez clairement établi que l'évaluation formative vise à informer le maître et l'élève, pour permettre d'effectuer les régulations dans les processus enseigner et apprendre, les modalités de mise en œuvre des dispositifs d'évaluation formative restent propres à chaque enseignant. Pour notre expérimentation, nous avons opéré de la manière suivante.

La mise en œuvre du dispositif d'évaluation dans la classe, et les exercices de remédiation réalisés de manière différenciée selon les besoins constatés, ont occupé une séance, soit environ une heure quinze. Cette séance intervenait après le moment d'enseignement proprement dit, au cours d'une journée différente. Nous considérons qu'elle devait contribuer, elle aussi, à l'apprentissage. Elle s'est toujours déroulée selon le scénario suivant.

Dans un premier temps, le maître organise un entretien avec l'ensemble des élèves de la classe. Il les amène à préciser les objectifs de la séquence, tels qu'ils peuvent les détailler. Il s'agit d'explicitier les sous objectifs visés. L'enseignant conduit les élèves à préciser ce qu'ils pensent qu'on attend d'eux. Que devront-ils montrer qu'ils savent faire ? Ce moment présente, pour le maître, deux avantages. D'abord, il lui permet de voir ce que ses élèves jugent important et il transcrit les éléments proposés au tableau noir. Ensuite, à partir des propositions de la classe, il en vient à expliciter ses propres attentes. Cette démarche, en deux temps, a pour objectif d'aider les élèves à mieux se représenter les buts à atteindre.

Dans un deuxième temps, les élèves découvrent une fiche d'évaluation où le maître a explicité ses attentes, par écrit. La phase précédente devrait faciliter, pour chaque élève, la compréhension de cette fiche. Chacun est alors appelé à s'auto-évaluer, par rapport à chaque point énoncé. Une colonne - *élève avant* - est prévue à cet effet. Quatre modalités sont proposées à chaque enfant :

- *oui*, qui signifie qu'il sait
- *non*, qui veut dire qu'il estime ne pas savoir
- *un peu*, qui rend compte que l'enfant pense ne pas bien savoir
- *et je ne sais pas*, qui signifie que l'enfant ne connaît pas sa position.

La figure 4 montre la fiche utilisée lors de la séquence 1, sur la division.

## Pour faire le point sur la division des nombres décimaux

Pour une division, je suis capable :

	<i>élève avant</i>	<i>élève après</i>
d'indiquer le quotient exact		
d'indiquer le quotient entier et le reste		
d'indiquer le quotient approché au dixième près par défaut		
d'indiquer le quotient approché au dixième près par défaut et le reste		
d'indiquer le quotient approché au centième près par défaut et le reste		
d'indiquer le quotient approché au dixième près par excès		

*symboles : + oui ; - non ; \* un peu ; / je ne sais pas*

**Figure 4 - La fiche d'évaluation formative utilisée pour la séquence sur la division**

Dans un troisième temps, le maître va proposer des exercices que les élèves vont réaliser. Chaque exercice sera suivi d'une correction collective. Chaque élève pourra alors constater s'il avait répondu de manière juste. Il devrait ainsi prendre conscience de ses réussites ou de ses manques. La fiche d'évaluation lui servira à s'évaluer dans la colonne prévue - *élève après* -. Les exercices prévus permettront, à chacun, d'analyser ses performances. L'élève devrait ainsi mieux comprendre ce qu'on attend précisément de lui. Les points qui restent des problèmes devraient être mis en évidence.

Le quatrième temps est une phase différenciée. Le maître va proposer aux élèves qui, au vu des résultats de l'évaluation, estiment qu'ils sont capables de bien réussir, de prolonger leur apprentissage. Des exercices d'une plus grande difficulté leur sont alors proposés. Pour les élèves qui ont découvert leurs insuffisances, l'enseignant proposera de reprendre ce qui leur pose encore des problèmes. Le groupe de ces enfants est ainsi réduit puisqu'il ne compte pas tous les élèves de la classe. Le maître est ainsi plus disponible pour chacun. Ce moment de la séance nécessite donc une gestion différenciée de la classe. La préparation des deux séquences de l'expérimentation en a tenu compte. Dans le fonctionnement habituel d'une classe, ce moment peut être allongé s'il s'avère nécessaire. Cependant, lors de l'expérimentation, nous avons décidé de le limiter à une seule séance afin de respecter l'égalité de durée des séquences selon les deux modalités.

### 3. Des résultats

Chaque séquence a fait l'objet d'un test initial et d'un test final, notés sur vingt. Les entretiens ont été réalisés auprès de six élèves pour chaque classe et à chaque séquence, soit au total vingt-quatre élèves.

Les tests, réalisés avant et après chaque séquence, rendent compte du niveau des performances des élèves. En rapprochant, pour chaque élève, et pour chacune des séquences, son niveau initial et son niveau final, nous pouvons obtenir une image de son évolution pour la séquence considérée. Nous avons calculé, pour chaque classe, la moyenne des élèves et l'écart type des distributions. De plus, pour chaque classe, et pour chaque séquence, les résultats ont été rangés dans l'ordre décroissant des performances obtenues au test initial. Ce classement nous a permis de regrouper les élèves de chaque classe en cinq groupes. Cela nous a conduit à voir comment évoluent les groupes selon leur niveau initial. C'est en utilisant ce

classement que nous avons choisi les élèves pour les entretiens, ainsi que nous l'avons indiqué ci-dessus.

### 3.1. Evolution des moyennes des classes

Pour observer l'évolution globale des élèves, nous allons utiliser les moyennes des classes. A cette fin, nous n'avons conservé que les performances des élèves qui avaient participé à l'ensemble des quatre tests. Les moyennes des résultats observés pour les deux classes sont reproduites dans les tableaux III et IV.

**Tableau III - Evolution globale des élèves de la classe 1**

	séquence 1 <i>(avec évaluation formative)</i>			séquence 2 <i>(sans évaluation formative)</i>		
	test initial	test final	différence	test initial	test final	différence
moyenne	11,5	13,17	1,67	9,62	13,85	4,23
<i>écart type</i>	4,36	5,60	4,36	4,26	3,75	3,86

**Tableau IV - Evolution globale des élèves de la classe 2**

	séquence 1 <i>(sans évaluation formative)</i>			séquence 2 <i>(avec évaluation formative)</i>		
	test initial	test final	différence	test initial	test final	différence
moyenne	9,96	12,44	2,48	9,68	15,24	5,56
<i>écart type</i>	3,51	4,81	3,44	3,96	4,32	3,61

Observons que les deux séquences nous placent dans deux situations différentes. Pour la séquence 2, les moyennes au test initial sont pratiquement égales (9,62 et 9,68). Les comparaisons de différences s'effectuent donc à partir du même niveau de départ. En revanche, pour la séquence 1, les moyennes initiales font apparaître un écart de près d'un point et demi ( $11,5 - 9,96 = 1,54$ ). Or, on sait qu'il est plus facile de progresser quand le niveau initial est plus faible. C'est d'ailleurs ce que l'on constate ici : le niveau le plus faible progresse davantage ( $2,48 > 1,67$ ).

Nous avons cherché à savoir si les différences des moyennes enregistrées étaient significatives, au seuil de significativité de 5 %. Pour cela, nous avons recherché la probabilité critique de l'hypothèse nulle en utilisant la distribution du  $t$  de Student (2). Pour chacun des quatre cas - deux séquences, dans deux classes différentes -, l'hypothèse nulle peut s'énoncer ainsi :

$$H_0 : \text{test final} - \text{test initial} = 0$$

Il n'y a pas de différence entre les moyennes observées aux tests initial et final.

L'hypothèse alternative est alors :

$$H_a : \text{test final} - \text{test initial} > 0$$

Les résultats au test final sont supérieurs à ceux du test initial.

Nous sommes dans la situation de recherche d'une probabilité unilatérale puisque nous préjugeons que les résultats au test final doivent être supérieurs à ceux du test initial. Les résultats obtenus apparaissent dans le tableau V.

**Tableau V - Significativité de l'évolution des moyennes**

		différ.	écart type	N	t calculé	probabilité critique	signif.
<b>classe 1</b>	séquence 1 avec EVF	+ 1,67	4,36	29	2,06	p < .025	oui
	séquence 2 sans EVF	+ 4,23	3,86	29	5,98	p < .0005	oui
<b>classe 2</b>	séquence 1 sans EVF	+ 2,48	3,44	25	3,60	p < .001	oui
	séquence 2 avec EVF	+ 5,56	3,61	25	7,70	p < .0005	oui

*Remarque :* pour N = 29, ddl = 28 ; pour N = 25, ddl = 24  
 "avec EVF " signifie "avec usage de l'évaluation formative"

Ainsi, au seuil de significativité de 5 %, nous pouvons rejeter l'hypothèse nulle dans les quatre cas. Nous pouvons donc dire, qu'au seuil de 5 %, la moyenne au test final est toujours significativement supérieure à la moyenne au test initial. Les moyennes des élèves progressent donc effectivement entre le début et la fin de chaque séquence.

### 3.2. L'évaluation formative : effet sur la moyenne des scores

Nous avons mis en évidence, pour les deux classes, que la progression moyenne des élèves étaient meilleure pour la séquence 2, la différence étant statistiquement significative. Or, notre plan d'expérimentation nous a conduit à alterner l'usage de l'évaluation formative selon les deux séquences. Ainsi, il s'avère que la plus grande progression de la classe 2, correspond à l'usage de l'évaluation formative. Mais cela n'est pas le cas dans l'autre classe où l'on observe le phénomène inverse ! Pour essayer de voir l'effet de l'usage du dispositif d'évaluation formative, nous avons utilisé la technique de la régression multiple. Nous avons construit le modèle à deux variables suivant :

$$\text{fin} = f(\text{ini}, \text{éva})$$

où :

- "fin" représente la variable dépendante, c'est-à-dire le score au test final ;
- "ini" représente une variable indépendante, c'est le score au test initial ;
- "éva" représente l'usage de l'évaluation formative ; c'est une variable qualitative dichotomique.

L'usage de la régression multiple nous conduit à obtenir des équations de la forme suivante :

$$Y_{\text{fin}} = a + b_1 \text{ ini} + b_2 \text{ éva}$$

où :

- " $Y_{fin}$ " représente le score au test final ;
- "a" représente la constante ;
- " $b_1$ " représente le coefficient affecté à la variable "ini" ;
- " $b_2$ " représente le coefficient affecté à la variable "éva" ;
- "ini" représente le score initial ;
- "éva" représente la variable qualitative "évaluation formative" qui peut prendre les valeurs 0 - absence d'usage de l'évaluation formative - et 1 - présence -

Pour le traitement, nous avons utilisé le logiciel "SAS". Nous avons travaillé à partir de l'ensemble des données des deux classes, en traitant séparément la séquence 1, puis la séquence 2. C'est dans cet ordre que nous présentons nos résultats.

#### a. La séquence 1 (séquence sur la division)

L'analyse des moyennes des classes a montré qu'au cours de cette séquence, les moyennes avaient progressé dans une moindre mesure. L'équation de la droite de régression obtenue est la suivante :

$$\text{fin1} = 4,50 + 0,82_{(0.131)} \text{ini1} - 0,67_{(1.089)} \text{éva} \quad R^2 = 43,09 \%$$

On observe que le coefficient de la variable "ini1" est positif et très significatif (p.c. < .0001). En revanche, le coefficient de la variable "éva" est négatif et non significatif (p.c. < .54). Cela nous conduit à dire que, à score initial égal, l'usage de l'évaluation formative fait baisser le score final de plus d'un demi point (0,67). Mais cet effet n'est pas significatif au seuil de 5 %.

La part de variance expliquée est de 43,09 %. Le modèle proposé ne rend compte que de moins de la moitié de la variance du score au test final. C'est presque essentiellement le score au test initial qui contient le pouvoir explicatif du modèle. Ce que l'on voit si l'on teste un modèle de régression simple, en ne prenant en compte que la variable "ini". Le tableau VI montre ce phénomène.

**Tableau VI - Pouvoir explicatif de deux modèles pour la séquence 1**

modèles	part de variance expliquée
(1) $\text{fin1} = f(\text{ini1})$	42,69 %
(2) $\text{fin1} = f(\text{ini1}, \text{éva})$	43,09 %

L'introduction de la variable "éva" n'augmente le pouvoir explicatif que de 0,4 %, ce qui est très faible. Quant à l'évaluation formative, l'absence de significativité statistique de son coefficient permet de dire qu'elle n'a pas eu d'effet discernable sur le score finale de cette séquence.

#### b. La séquence 2 (séquence sur les durées)

Nous avons vu que cette séquence avait permis, aux deux classes, de réaliser une meilleure progression. Cette plus grande évolution des élèves est observée en même temps qu'un moindre poids de la variable "ini2", qui représente le score initial de cette deuxième séquence. C'est ce que l'on peut voir dans le tableau VII.

**Tableau VII - Pouvoir explicatif de deux modèles pour la séquence 2**

modèles	part de variance expliquée
(3) $\text{fin2} = f(\text{ini2})$	33,99 %
(4) $\text{fin2} = f(\text{ini2}, \text{éva})$	35,88 %

On voit que l'introduction de la variable "éva" augmente le pouvoir explicatif du modèle de près de 2 % (1,89 %). Néanmoins, la part de variance expliquée du modèle (4) atteint à peine 36 % ; cela signifie que 64 % du phénomène restent à expliquer !

Les estimations de ce modèle (4) sont détaillées ci-dessous.

$$\text{fin2} = 8,53 + 0,57_{(.109)} \text{ini2} + 1,12_{(.903)} \text{éva} \quad R^2 = 35,88 \%$$

Le coefficient de "ini2" est positif et très significatif (p.c. < .0001). En revanche, le coefficient de la variable "éva" est non significatif (p.c. < .22). Il est cette fois positif, ce qui diffère du modèle (2). Ici encore, même si l'on constate une augmentation de la probabilité critique, par rapport au modèle (2), on voit que l'usage de l'évaluation formative n'a pas eu d'effet, statistiquement discernable au seuil de 5 %, sur le score moyen de la classe.

En résumé, on peut donc dire que les données de notre expérimentation nous conduisent à constater que **l'usage d'un dispositif d'évaluation formative, dans ces deux classes de l'école élémentaire et pour les deux séquences décrites, n'a pas eu d'effet, statistiquement significatif au seuil de 5 %, sur la moyenne des scores au test final.** Cependant, cette absence d'effet sur les performances globales ne nous empêchait pas d'imaginer qu'il puisse y avoir d'autres effets, plus qualitatifs. C'est ce que nous avons voulu observer à partir des entretiens que nous avons conduits.

### **3.3. L'évaluation formative : effet sur les capacités de repérage des élèves**

Nous avons donc vu que l'évaluation formative n'avait pas eu d'effet discernable sur l'évolution des moyennes des scores aux tests. Il semble curieux, et quelque peu frustrant pour ceux qui croient à l'intérêt de cette forme d'évaluation, de devoir constater une absence d'effet. Cependant, s'il n'est pas apparu d'effet sur les moyennes des classes, nous pouvons encore nous interroger. L'évaluation formative pourrait apporter une aide éventuelle dans la capacité des élèves à se repérer dans leurs apprentissages. Ces dispositifs d'évaluation facilitent-ils une meilleure représentation du but à atteindre ? Aident-ils l'élève à faire le point dans son cheminement vers le savoir ? C'est pour essayer de répondre à ces deux questions que nous avons interrogé des élèves.

Les entretiens conduits relèvent de la catégorie "*questionnaire ouvert*", telle que la définissent R. Ghiglione et B. Matalon (Ghiglione et Matalon, 1985, p. 58). Ils ont été enregistrés et retranscrits. Pour le traitement des données, nous nous sommes inspiré des techniques d'analyse de contenu. Nous avons utilisé une *procédure close* (Ghiglione et Matalon, 1985, pp. 173-184) pour la recherche d'un effet sur les capacités des élèves à se représenter le but à atteindre. Pour chaque séquence, six élèves ont été choisis dans chacune des deux classes. Nos vingt-quatre sujets appartiennent, pour moitié, à la catégorie d'élèves qui a pratiqué une évaluation formative. Pour chaque classe, ils se répartissent selon trois niveaux, bons, moyens et faibles, définis par leurs résultats au test initial. C'est ce que nous présentons dans le tableau VIII.

**Tableau VIII - Choix des élèves pour les entretiens**

		séquence 1		séquence 2		
		niveau	élèves	score tests	élèves	score tests
<b>avec</b> <b>évaluation</b> <b>formative</b>	<i>bon</i>	Léa Léo	19 - 19 17 - 15	Gilles Cécilia	16 - 20 15 - 14	
	<i>moyen</i>	Laurie Marie-Bé.	11 - 14 11 - 16	Boris Haï Dang	10 - 3 9 - 18	
	<i>faible</i>	Alexis Elsa	7 - 9 6 - 4	Jean-Bapti. Eve	6 - 14 5 - 10	
<b>sans</b> <b>évaluation</b> <b>formative</b>	<i>bon</i>	Geoffrey Emilie	20 - 20 15 - 19	Marion Guillaume	20 - 18 17 - 19	
	<i>moyen</i>	Anthony F. Didier	10 - 16,5 9,5 - 20	Isabelle Loïc	10 - 14,5 9 - 15	
	<i>faible</i>	Inès Anthony C.	7 - 11 6 - 19	Florent Sophie	5 - 12 2 - 9	

*Remarque* : pour les colonnes "score tests", les deux nombres correspondent respectivement au score au test initial et au score au test final.

Nous avons d'abord cherché à voir comment les élèves se représentaient le but à atteindre.

#### **a. Une meilleure connaissance des buts à atteindre**

Tout apprentissage repose sur ce paradoxe que l'on peut illustrer par une métaphore que l'on choisira dans le domaine de la montagne. La situation de l'apprenant est comparable à celle du randonneur qui cherche à atteindre, pour la première fois, un sommet. Ils savent tous deux que le but existe, mais ils ne le voient pas clairement et ils ne distinguent même pas le chemin qui y conduit. C'est au fur et à mesure de l'approche que les choses se précisent. Si une bonne représentation du but à atteindre permet au randonneur de guider, avec succès, ses pas vers le but visé, nous pensons qu'il en est de même pour l'élève. Il semble important que celui-ci connaisse le but, pour diriger ses actions. L'évaluation formative peut-elle améliorer les représentations du but chez les élèves ? Nous avons analysé les réponses des enfants à la question 2 (qui était exprimée en ces termes : "*Est-ce que tu peux dire ce que la maîtresse attend des élèves ?*") en comparant leurs réponses avec ce que les enseignantes considéraient comme le but à atteindre.

Pour la séquence 1, il s'agissait de :

- savoir calculer une division (quotient décimal) avec tout ce que cela comporte en termes de connaissances des tables de multiplication et de la technique du calcul écrit,
- pouvoir reporter les chiffres en colonnes de la partie gauche de l'opération (dizaines, unité et surtout dixièmes, centièmes, millièmes),
- retrouver les restes qui correspondent à un quotient tronqué.

Pour la séquence 2, il s'agissait de :

- connaître les correspondances entre les différentes unités de mesures (principalement : 1 h. = 60 min. et 1 min. = 60 s.),
- savoir effectuer des opérations avec des nombres entiers,
- appliquer les méthodes de calcul propres aux durées.

On observe que les enfants expriment, à leur manière, ces objectifs. Ainsi, pour la séquence 2 et à titre d'exemple, nous avons retenu les réponses de deux élèves.

*Remarque* : "q. 6 " signifie qu'il s'agit de la sixième question qui a été posée lors de l'entretien.

Voici ce que dit Guillaume, un élève qui exprime bien ce qu'on doit savoir faire :

*q. 6 : Est-ce que tu peux dire ce que la maîtresse attend de toi ?*

- Ben ... [silence prolongé] Ben, elle attend qu'on connaisse les règles pour pouvoir additionner et soustraire les nombres, enfin, les durées. Et heu ... ben pouvoir faire des conversions aussi, pouvoir, enfin, connaître les principales ... heu ... Quand on divise, ou on multiplie, savoir que cinq fois soixante, c'est trois cents....

*q. 7 : Donc, connaître les règles et savoir effectuer des calculs. Est-ce que tu penses qu'il y a autre chose ?*

- Heu ... Ben ... Non. J'ai à peu près tout dit.

Inès, l'élève suivante, répond d'abord de façon très générale, avant de pouvoir exprimer de façon vague ce qu'elle a perçu :

*q. 5 : Est-ce que tu peux dire ce que la maîtresse attend des élèves ?*

- Non. Ben, si, elle attend qu'on écoute bien et qu'on fasse un bon travail. Et qu'on sache répondre si elle pose une question, si on a bien écouté, bien sûr.

*q. 6 : D'accord. Et là, pour les mathématiques, est-ce que tu vois un peu ce qu'elle attend de toi ?*

- Non.

*q. 7 : Non. C'est un peu mystérieux ?*

- Oui.

*q. 8 : Tu ne vois pas ce qu'elle pourrait te poser comme question, dans un contrôle ?*

- Heu ... Si, des questions sur les durées ! De faire des additions et des soustractions, tout ça.

*q. 9 : Eh bien, ça peut être ça, ce que la maîtresse attend des élèves : qu'ils soient capables d'additionner des durées et de les soustraire ?*

- Oui.

La lecture de l'ensemble des réponses des élèves nous a permis de les classer en trois catégories :

- ceux qui ont exprimé, selon nous, les attentes en étant capables de les énoncer sans en oublier : ils expriment bien le but,
- ceux qui n'expriment que partiellement des buts à atteindre : ils expriment le but assez bien,
- ceux qui ne sont pas capables d'exprimer les attentes de l'enseignant : ils expriment mal le but.

Le tableau IX montre la répartition observée, en fonction du niveau initial des élèves. Pour chaque cas, nous avons fait apparaître l'effectif.

**Tableau IX - Répartition des élèves selon leur manière d'exprimer le but à atteindre**

	<i>niveau</i>	bonne expression		assez bonne expression		Mauvaise expression	
<b>avec évaluation formative</b>	bon	Léa Léo Emilie	3	Geoffrey	1		0
	moyen		0	Laurie Anthony F. Marie Bé.	3	Didier	1
	faible	Alexis	1	Elsa Inès Anthony C.	3		0
<b>sans évaluation formative</b>	bon	Gilles Guillaume	2	Marion Cécilia	2		0
	moyen	Isabelle Loïc	2		0	Boris Hai Dang	2
	faible	Jean-Baptis Florent Sophie	3		0	Eve	1

Si nous ne tenons pas compte des niveaux des enfants au test initial, nous obtenons le tableau d'effectifs suivants.

**Tableau X - Classement des élèves selon leur capacité à exprimer le but à atteindre**

	bonne expression	assez bonne expression	mauvaise expression	<i>ensemble</i>
<b>avec évaluation formative</b>	4 33,3 %	7 58,3 %	1 8,3 %	12 100 %
<b>sans évaluation formative</b>	7 58,3 %	2 16,7 %	3 25,0 %	12 100 %
<i>ensemble</i>	11 45,8 %	9 37,5 %	4 16,7 %	24 100 %

La faiblesse des effectifs nous empêche d'utiliser le test statistique du Khi Deux. On peut toutefois faire observer que les élèves qui ont bénéficié d'une évaluation formative sont plus nombreux à montrer une "bonne" ou une "assez bonne" expression du but : 11 contre 9 (soit 91,7 % contre 75 %). De plus, les élèves qui se représentent "mal" le but poursuivi apparaissent dans une proportion nettement plus faible parmi ceux qui ont bénéficié d'évaluation formative (8,3 % contre 25 %) Peut-on voir, ici, un effet bénéfique de cette forme d'évaluation ? En toute rigueur, notre expérimentation ne nous permet pas de répondre de

manière affirmative. Elle permet seulement de soupçonner un effet qui pourrait faire l'objet d'une future recherche.

Il nous reste à essayer de voir si le dispositif d'évaluation formative a eu un effet sur la capacité des élèves à mieux situer leur niveau.

### **b. Une meilleure connaissance de sa situation d'apprenant**

Notre grille d'entretien comportait des questions qui visaient à essayer de saisir comment l'élève évaluait sa situation face aux buts poursuivis. Avait-il trouvé cet apprentissage facile pour lui ? - question 4 : "*Est-ce que c'était facile pour toi ?*" - Comment estimait-il son niveau de connaissance ? - question 6 : "*Est-ce que tu penses que tu vas bien réussir le prochain test qui portera sur ce thème ?*", et question 7 : "*Si on devait te noter, quelle note sur vingt penserais-tu avoir ?*" - Examinons les réponses de quelques élèves.

Pour la question 4, l'analyse des entretiens fait apparaître que certains ont trouvé la séquence facile, comme Anthony F.

*q. 11 : Est-ce que c'était facile pour toi ?*

- Oui, surtout les additionner, les soustraire. Avant, c'était un petit peu, mais avec ce qu'on a fait tout à l'heure, j'en sais un peu plus et j'y arrive mieux.

Florent, parle de ses difficultés.

*q. 11 : Est-ce que c'était facile pour toi ?*

- Moi, je n'y arrivais pas trop. Les additions, j'ai compris. Les soustractions, je n'ai pas encore trop compris. Et les ... les conversions aussi.

*q. 12 : D'accord. Eh bien, justement, là, si on fait le point. Qu'est-ce que tu penses avoir bien compris ?*

- Les additions.

*q. 13 : Ca, c'est bon ?*

- Oui.

*q. 14 : Et il y aurait d'autres choses que tu aurais bien comprises ?*

- ... Non.

*q. 15 : Et alors, qu'est-ce que tu penses n'avoir pas bien compris ?*

- ... Ben, les soustractions.

Mais, la majorité des élèves explique que ce qui était difficile au début est devenu facile ensuite. Loïc l'exprime ainsi :

*q. 10 : Est-ce que c'était facile pour toi ?*

- Ben... La première évaluation, au début, ce n'était pas facile. Mais après, dès qu'on m'a expliqué les bases, en fait comment on pouvait... En fait, il ne faut pas mélanger les deux, ça allait beaucoup mieux. Et maintenant, j'arrive à faire des choses de tête sans... pas besoin de calculer. Rien qu'en sachant, je pouvais tout faire.

*q. 11 : Est-ce que tu saurais à quelle moment tu as commencé à comprendre ?*

- Heu... Dès la première explication, avec mademoiselle K. Là, c'est rentré dans la tête !

q. 12 : C'était quand il fallait calculer la durée de la journée, et la durée de la nuit ?  
 - Oui, voilà, voilà.

Cette question 4 visait à amener l'élève à donner son sentiment sur ce qu'il avait ressenti. Nous pensions le préparer ainsi à estimer son niveau afin de pouvoir répondre à la question 6- "Est-ce que tu penses que tu vas bien réussir le prochain test qui portera sur ce thème ?", et même à faire un pronostic sur sa note, avec la question 7.

Ici, on observe que, majoritairement, les élèves répondent par l'affirmative. C'est le cas de Guillaume, qui est sûr de lui :

q. 13 : Bon, est-ce que tu penses que tu vas bien réussir le prochain test qui portera sur ce thème-là ?

- Ben oui. Un peu ! Je suis pratiquement sûr d'avoir au moins la moyenne. Sinon, ouais, je pense que je vais bien réussir.

q. 14 : D'accord. Puisque tu parles de note, si on devait te noter, quelle note, sur vingt, penserais-tu avoir ?

- Moi ... [silence] Dix-huit ou dix-neuf.

Parfois, ils ont exprimé un doute. Un seul élève déclarera que le test final serait difficile pour lui. Son analyse apparaîtra d'ailleurs lucide puisqu'il aura douze sur vingt. Le tableau XI fait apparaître la situation des élèves telle que nous l'avons analysée.

**Tableau XI - Réussite envisagée au test final**

	Réussite envisagée au test final		
	<i>oui</i>	<i>peut-être</i>	<i>non</i>
<b>avec évaluation formative</b>	Léa Léo Laurie Inès Didier	Alexis Geoffrey Emilie Anthony F. Marie-Bé.	Elsa Anthony C.
<b>sans évaluation formative</b>	Cécilia Haï Dang Jean-Bapt Sophie	Marion Guillaume Isabelle Loïc	Gilles Boris Eve Florent

Ce classement des élèves permet de construire le tableau d'effectifs suivant.

**Tableau XII - Classement des élèves selon leur réussite envisagée au test final**

	Réussite envisagée au test final			<i>ensemble</i>
	<i>oui</i>	<i>peut-être</i>	<i>non</i>	
<b>avec évaluation formative</b>	10 83,3 %	2 16,7 %	0 0 %	12 100 %
<b>sans évaluation formative</b>	8 66,7 %	3 25,0 %	1 8,3 %	12 100 %
<i>ensemble</i>	18 75,0 %	5 20,8 %	1 4,2 %	24 100 %

La faiblesse des effectifs ne permet pas l'usage du test du Khi Deux qui nous aurait autorisé à trancher quant à une éventuelle relation entre la réussite estimée par les élèves et la variable "évaluation formative". En regardant les effectifs, on peut cependant avancer l'idée que les élèves semblent se répartir de façon similaire entre les deux modalités de la variable "évaluation formative".

La question suivante amenait les élèves à émettre un pronostic sur leur note. Nous avons analysé leurs réponses en comparant leur note estimée, et la note effectivement obtenue au test final. Le tableau XIII fait état de nos données. Nous avons ajouté une colonne "différence" qui rend compte de la capacité de l'élève à estimer ses résultats, donc son niveau par rapport aux attentes imaginées. Nous avons estimé que si l'écart entre l'estimation et la note réelle était inférieur ou égal à deux points, cela représentait une bonne capacité à situer sa future performance.

**Tableau XIII - Capacité à estimer sa note au test final**

	enfants	test initial	note estimée	test final	capacité à estimer
<b>avec évaluation formative</b>	Léa	19	15 - 16	19	non
	Léo	17	18	15	non
	Laurie	11	15	14	oui
	Marie Bérénice	11	17	16	oui
	Alexis	7	16	9	non
	Elsa	6	14	4	non
	Geoffrey	20	18	20	oui
	Emilie	15	> 14	19	non
	Anthony F.	10	15,5 - 16	16,5	oui
	Didier	9,5	16	20	non
	Inès	7	14	11	non
	Anthony C.	6	15	19	non
<b>sans évaluation formative</b>	Marion	20	> 14 - 18	18	oui
	Guillaume	17	18 - 19	19	oui
	Isabelle	10	16	14,5	oui
	Loïc	9	16 - 17	15	oui
	Florent	5	10	12	oui
	Sophie	2	13 - 14	9	non
	Gilles	16	16 - 17	20	non
	Cécilia	15	17 - 18	14	non
	Boris	10	16	3	non
	Haï Dang	9	> 14 - 15	18	non
	Jean-Baptiste	6	16	14	oui
	Eve	5	14	10	non

Le tableau XIV fait apparaître les effectifs selon que les élèves ont été, ou non, capables "d'estimer" leur note au test final, avec une marge d'erreur égale ou inférieure à deux points.

**Tableau XIV - Répartition des élèves selon leur capacité à estimer leur note au test final (écart inférieur ou égal à 2 points)**

	capable d'estimer		ensemble
	<i>oui</i>	<i>non</i>	
avec évaluation formative	4 33,3 %	8 66,7 %	12 100 %
sans évaluation formative	6 50,0 %	6 50,0 %	12 100 %
<i>ensemble</i>	10 41,7 %	14 58,3 %	24 100 %

Ici encore, en l'absence de test statistique, nous sommes réduits à émettre des commentaires prudents. Il ne semble pas y avoir un effet de l'évaluation formative. De plus, au cours des entretiens, aucun élève appelé à répondre aux questions relatives à sa réussite au prochain test n'a fait allusion à l'information qu'il avait obtenue grâce au dispositif

d'évaluation formative. La moitié des élèves interrogés possédait pourtant une information assez précise de son niveau de réussite. On peut avancer l'idée qu'ils n'associaient pas ce moment à une phase de repérage de leurs capacités.

#### **4. Discussion**

Notre projet était de parvenir à observer les effets de l'évaluation formative sur les apprentissages des élèves. Praticien régulier de l'évaluation formative, nous avons l'impression que cette forme d'évaluation favorisait les apprentissages. C'était pour ne pas rester à ce stade de "l'impression" que nous avons voulu "voir". Il nous semblait qu'une amélioration des apprentissages devait conduire à une amélioration des résultats des élèves. C'est donc pour observer des écarts que notre expérimentation nous a amené à tester les élèves avant et après les séquences. Or, nous n'avons pas observé d'effet significatif dans ce domaine. Alors, nous devons nous interroger. Faut-il renoncer à utiliser l'évaluation formative ? Celle-ci n'aurait-elle aucune efficacité propre ?

Nous savons que la mise en œuvre de tels dispositifs occupe le temps de classe. Elle absorbe également une part de l'énergie de l'enseignant. Une lecture simpliste de nos résultats conduirait à renoncer à l'usage de cette forme d'évaluation. Cependant, notre travail nous amène à ne pas faire cette lecture, et cela pour deux raisons. La première est liée aux limites de notre recherche. La seconde a trait aux fonctions de l'enseignant qui doit, entre autre, gérer les situations d'apprentissage.

##### **a. Des limites de notre étude**

Nous commencerons par évoquer certaines limites de notre étude relatives à trois choix que nous avons réalisés : le choix de notre échantillon, celui du dispositif d'évaluation formative mis en œuvre, et celui des séquences retenues.

En ce qui concerne les élèves de l'expérimentation, nous rappelons que nous n'avons travaillé qu'avec deux classes d'une même école. Nous avons pu voir que la relative faiblesse des effectifs nous avait parfois privé de l'usage du test statistique du Khi Deux (paragraphe 3.3). Une question se pose alors. Aurions-nous observé les mêmes résultats si nous avions travaillé avec un échantillon plus large ?

A propos de l'évaluation formative, notre étude a été construite à partir de l'usage, ou du non usage, d'un même dispositif. Rien ne nous empêche d'imaginer qu'un dispositif différent aurait pu produire des évolutions différentes chez les mêmes élèves. De plus, le dispositif n'existe que mis en œuvre par un enseignant. On voit qu'il y a là, également, une source de variation possible. Pour un dispositif donné, sa mise en œuvre doit sans doute varier assez fortement selon l'enseignant qui la réalise. Notre travail a cherché à supprimer cet "effet maître", en confiant cette phase à un même enseignant. Cependant, l'absence d'effet observé n'est-elle pas liée aux insuffisances du dispositif d'évaluation formative retenu ?

Quant au choix des séquences, l'observation de l'évolution des moyennes fait apparaître, pour une même classe, des différences de progression selon la séquence retenue (tableau V). Nous cherchions à observer s'il existait des différences de progression, pour des élèves d'une même classe, selon qu'ils auraient, ou non, bénéficié d'une évaluation formative. Notre travail a clairement mis en évidence que, selon la séquence, les progressions moyennes des élèves n'étaient pas les mêmes. La progression des élèves a été plus importante, en proportion, dans la séquence sur les durées, et ce, pour les deux classes de notre échantillon.

Or, nous ne pouvons comparer que si nous disposons d'éléments invariants. Dans notre situation, l'élément classe restait bien le même. Mais au vu des variations enregistrées d'une séquence à l'autre, il semble judicieux de poser la question suivante. Quand on engage successivement le même groupe d'élèves dans deux séquences d'enseignement différentes, s'agit-il effectivement de la même classe ? L'interrogation, dans notre situation, de ce qui reste semblable ne nous conduit-elle pas à réfléchir à la clause, parfois trop vite avancée, du "*toutes choses égales par ailleurs*" ?

Ainsi, certaines conditions de notre étude viennent relativiser le non effet que nous avons enregistré. Examinons maintenant, la nécessaire gestion des apprentissages par l'enseignant.

### **b. Evaluation formative et pédagogie différenciée**

Nous pouvons dire que l'apprentissage des élèves n'est pas seulement fonction de ce qui se déroule dans la classe, même si cela a beaucoup d'importance. Il est facile de le montrer en faisant remarquer que le temps de classe est le même pour tous les élèves d'une classe, mais que tous ne réussissent pas de la même manière. D'autres facteurs influent. L'action du maître n'est pas la seule à produire des apprentissages chez des élèves qui sont tous différents les uns des autres. C'est l'action réelle de l'élève qui semble essentielle dans sa réussite. Or, à l'école, c'est souvent le maître qui détermine les actions que les élèves doivent réaliser. Améliorer les performances des élèves consiste, sans doute, à optimiser les choix d'actions que chacun d'eux doit accomplir. Mais comment l'enseignant peut-il déterminer ces actions ? Comment les adapte-t-il aux besoins des élèves ? Nous pensons que la solution réside dans une meilleure connaissance de leurs besoins. L'évaluation n'apporte qu'une connaissance du niveau de l'élève. C'est l'analyse des réponses qui peut révéler les besoins. Il reste ensuite au maître à trouver les actions les plus adaptées aux besoins de chacun de ses élèves, puis à les mettre en œuvre, dans la classe. Il conviendrait donc de substituer une logique de l'analyse des besoins à une logique de l'appréciation des niveaux.

Nous croyons que l'évaluation formative est une condition nécessaire à l'amélioration des apprentissages, mais qu'elle n'est pas suffisante. Nous partageons l'avis de M. Grangeat (Grangeat, 1997, p. 23) qui affirme que : "*la question de l'amélioration des apprentissages revient à celle de l'articulation entre différenciation et évaluation.*" La mise en œuvre d'un dispositif d'évaluation formative ne pourrait donc, à elle seule, améliorer les performances des élèves. Mais cette forme d'évaluation apparaît indispensable à une gestion différenciée des apprentissages.

## **5. Conclusion**

En conclusion, nous pensons que l'évaluation formative n'est qu'un facteur de base de la différenciation pédagogique. Elle va guider les choix d'activités du maître, en lui permettant d'adapter au mieux ses propositions aux besoins de ses élèves. L'élève lui-même doit avoir connaissance des informations qui le concernent. Elles l'aident à faire le point, et elles éclairent les choix de l'enseignant. Elles contribuent à rendre l'élève davantage acteur de ses apprentissages. Nous pensons que c'est là un facteur qui devrait avoir un effet positif sur la réussite des apprentissages. Ces réflexions nous conduisent donc à une interrogation, à laquelle nous ne pourrions pas répondre ici, mais qui renvoie à une autre recherche. Un enseignement différencié améliore-t-il les apprentissages ? Faut-il croire à l'efficacité d'un tel enseignement différencié ?

## Notes

- (1) Cet article rend compte d'un travail de recherche conduit au sein du Laboratoire des Sciences de l'Education de l'Université Grenoble 2, équipe APEA, pour l'obtention d'un DEA (sous la direction du professeur Charles Hadji).
- (2) Pour le calcul du  $t$ , nous avons retenu la formule simplifiée suivante :

$$t = \frac{\overline{D}}{\frac{s}{\sqrt{N}}}$$

où :

- $\overline{D}$  est la moyenne des différences entre le test initial et le test final
- $s$  est l'écart type des différences observées
- $N$  est l'effectif de notre échantillon, c'est-à-dire le nombre d'élèves de la classe considérée.

L'usage de cette formule était rendue possible par notre situation d'expérimentation où nous avons des échantillons appariés.

## Bibliographie

- BLOOM, B.S., HASTINGS, J.T. et MADAUS, G.F. (1971) *Handbook on formative and summative evaluation of student learning*. New York : McGraw-Hill
- CARDINET, J. (1986) *Pour apprécier le travail des élèves*. Bruxelles: De Boeck.
- DELORME, C. (dir.) (1987) *L'évaluation en question*. Paris : E.S.F.
- GHIGLIONE, R. et MATALON, B., (1977) (4ème éd. 1985) *Les enquêtes sociologiques - Théories et pratique* . Paris : Armand Colin.
- GRANGEAT, M. (1997) *Différenciation, évaluation et métacognition dans l'activité pédagogique, à l'école et au collège*. Lyon : Université Lumière Lyon 2. (thèse)
- HADJI, C. (1989) *L'évaluation, règles du jeu*. Paris : ESF.
- HADJI, C. (1997) *L'évaluation démystifiée*. Paris : E.S.F.
- HOUSSAYE, J. (1989) *Le triangle pédagogique*. Berne : Peter Lang.
- MONTEIL, J.M. (2ème éd. 1990) *Eduquer et former : perspectives psychosociales*. Grenoble : P.U.G.
- NOIZET, G. et CAVERNI, J.P. (1978) *Psychologie de l'évaluation scolaire*. Paris : P.U.F.
- NUNZIATI, G. (1990) Pour construire un dispositif d'évaluation formatrice. *Cahiers pédagogiques* n° 280.
- PERRENOUD, P. (1998) *L'évaluation des apprentissages : de la fabrication de l'excellence à la régulation des apprentissages. Entre deux logiques*. Bruxelles : De Boeck, Paris : Larcier.
- VESLIN, O. et J. (1992) *Corriger des copies : évaluer pour former*. Paris : Hachette.