

Suivi pédagogique : grille critériée et dossier d'apprentissage

Marc Derycke

Volume 20, Number 3, 1998

L'évaluation par le portfolio

URI: <https://id.erudit.org/iderudit/1091376ar>

DOI: <https://doi.org/10.7202/1091376ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

Cite this article

Derycke, M. (1998). Suivi pédagogique : grille critériée et dossier d'apprentissage. *Mesure et évaluation en éducation*, 20(3), 33–53. <https://doi.org/10.7202/1091376ar>

Article abstract

A follow-up booklet contains assessment information the purpose of which is to enable the subsequent teacher to identify the pupil's level in order to follow a course of training consistently. Is this information a faithful and valid representation of the skills actually gained by the pupil? The article presents the results of a comparative experiment of two assessment and follow-up documents featuring different approaches learning : the criteria grid and the learning portfolio. The experiment concentrates on evaluating learning to read. The same protocol was suggested to teachers from three different educational systems. Results show that if the grid is best suited for the «good» pupil, the grid seems to be unsuitable method, or even detrimental, for the pupil said to be «experiencing difficulties». The learning portfolio is more faithful and does not depend so much on the pupil's level. Nevertheless, its optimum use requests, in addition to the records of the pupil's works, synthetical information giving a general overview of the pupil's skills.

Suivi pédagogique : grille critériée et dossier d'apprentissage

Marc Derycke
Université Jean Monnet

À la mémoire de Pierre Achard

Un livret de suivi pédagogique comporte des données évaluatives destinées à un enseignant extérieur afin de lui permettre de poursuivre les apprentissages de l'élève avec cohérence. Quels sont la fidélité et la validité de ces données, eu égard à l'état effectif des compétences qu'elles sont sensées représenter? L'article présente les résultats d'une expérimentation comparative de deux documents d'évaluation et de suivi de conception contrastée : la grille critériée et le dossier d'apprentissage; le support proposé était constitué de cas d'élèves en apprentissage de la lecture. Le même protocole a été soumis à des enseignants appartenant à trois systèmes éducatifs différents. Les résultats montrent que si la grille fournit des données acceptables pour un enfant en situation de réussite, elle est inefficace, voire nocive, pour un enfant en situation d'échec. Plus fidèle, le dossier d'apprentissage est nettement moins sensible à la situation de l'enfant. Néanmoins son usage optimum nécessite qu'outre les traces y soient consignée une information synthétique donnant un aperçu global des compétences de l'élève.

(apprentissage de la lecture, dossier d'apprentissage, évaluation formative, expérimentation de livrets évaluatifs, grille critériée, livret de suivi, métacognition, portfolio, suivi pédagogique)

A follow-up booklet contains assessment information the purpose of which is to enable the subsequent teacher to identify the pupil's level in order to follow a course of training consistently. Is this information a faithful and valid representation of the skills actually gained by the pupil? The article presents the results of a comparative experiment of two assessment and follow-up documents featuring different approaches learning : the criteria grid and the learning portfolio. The experiment concentrates on evaluating learning to read. The same protocol was suggested to teachers from three different educational systems. Results show that if the grid is best suited for the «good» pupil, the grid seems to be unsuitable method, or even detrimental, for the pupil said to be «experiencing difficulties». The learning portfolio is more faithful and does not depend so much on the pupil's level. Nevertheless, its optimum use requests, in addition to the

*records of the pupil's works, synthetical information giving a general overview of the pupil's skills.
(learning to read, learning portfolio, formative assessment, experiment of assessment booklets, criteria grid, follow-up booklet, metacognition, portfolio, pedagogical follow-up)*

Deux outils, deux variables pour le suivi

Une enquête réalisée en France montre qu'une grande majorité des livrets de suivi recueillis recourent à une grille critériée (cf. Gauch et al., 1995). Par ailleurs, on relève l'usage de nombreux outils proches du « dossier d'apprentissage », ces documents circulant en divers lieux surtout entre des classes maternelles et les parents.

L'étude spéculative du sens que construit une grille laisse présager une certaine inadéquation à la nouvelle fonction qui lui est dévolue : transmettre l'information pertinente pour la poursuite des activités pédagogiques (cf. Derycke, « Suivi pédagogique : quel sens construisent une grille évaluative, un dossier d'apprentissage? »). *A contrario*, on peut inférer de cette étude l'avantage que revêt l'usage d'extraits des travaux de l'élève. Alors que la grille résulte d'un codage supplémentaire qui occulte le référé brut et risque de le parasiter, ces extraits sont directement « parlants ». Le document qualitatif qui les consigne est proche du dossier d'apprentissage; n'est-il pas plus adapté au suivi pédagogique?

Pour répondre à cette interrogation, nous avons tenté de mettre en concurrence ces deux types d'outil pour juger de leur efficacité.

Tous les élèves font l'objet d'évaluations, tous font l'objet d'un suivi pédagogique, lequel est instrumenté notamment par divers types de livret scolaire. Toutefois, ces documents apparaissent dépourvus d'enjeu. Par exemple, en France, ils ne font pas l'objet d'une attention significative de la part de l'Administration dans leur mise en œuvre et dans la régulation de leurs effets. Enfin, la recherche néglige cette application de l'évaluation.

Or les variables du suivi sont multiples, et vu la vacance de l'investigation en ce domaine, on peut raisonnablement penser que toutes ne sont pas repérées dans leurs interrelations. Il convient alors de dégager ces variables en les soumettant à des épreuves majeures. Celles-ci se rencontrent auprès du public visé prioritairement par notre recherche, les enfants itinérants². Les déplacements occasionnent en effet des changements d'école et des ruptures

de scolarisation qui introduisent des discontinuités souvent néfastes à la poursuite normale des apprentissages. C'est pourquoi un document informant l'enseignant sur le niveau de compétence de l'enfant paraît utile pour

- 1° assurer la cohérence des activités pédagogiques,
- 2° gagner en efficacité vu leur moindre fréquentation scolaire,
- 3° favoriser l'accueil de ces enfants particulièrement insécurisés,
- 4° légitimer leur prise en charge par l'enseignant, leur présence passagère pouvant être négligée.

Ces contraintes fortes imposent donc des priorités et hiérarchisent des variables qui, dès lors, seront transposables dans leur principe aux autres publics.

Nous retiendrons ainsi, parmi les diverses fonctions du suivi, celle de fournir l'information permettant de déterminer le niveau de l'élève pour poursuivre avec cohérence ses apprentissages. Ainsi avons-nous cherché à établir en quoi un document favorise, chez l'enseignant-récepteur, la construction d'une image *fidèle* et *valide* des compétences acquises et de celles à acquérir par l'élève en priorité. À cette fin, nous nous sommes livré à une expérimentation comparative de ces deux documents de suivi de conception contrastée.

Leur fidélité a été évaluée conformément à deux seuils : au palier supérieur, le document fournit-il une information suffisante pour proposer des tâches adaptées à la poursuite des apprentissages? Au palier inférieur, ces données sont-elles suffisantes pour choisir un test qui mette l'enfant en situation de réussite raisonnable, afin de poursuivre sa progression? Si le document de suivi satisfait à ce critère, il remplit au minimum sa fonction puisqu'il favorise l'accueil de l'enfant dans le nouvel environnement pédagogique et relationnel où il est insécurisé.

La validité a été appréciée selon trois degrés de concordance entre la performance effective de l'élève, c'est-à-dire le référé du document reçu, et l'image des compétences construite par sa seule interprétation ; la question est donc : ce document me décrit-il bien les caractéristiques de tel élève en telle discipline?

À ces deux variables, il convient d'en associer une troisième qui, bien qu'absente de l'expérimentation, constitue, de notre point de vue, un critère de validation de tout document de suivi : quel rôle réserve le document à l'enfant quant à la prise en charge du suivi lui-même? L'enfant est-il passif

au sein de ce dispositif, est-il actif, et en quoi, enfin cet investissement s'intègre-t-il dans la mise en cohérence des apprentissages et dans leur activation?

Ces objectifs s'inscrivent dans une démarche d'opérationnalisation de l'accueil et du suivi cherchant à donner sens à l'école et à favoriser l'intégration des publics qui en sont socioculturellement éloignés.

Si l'usage de la grille aux fins du suivi est établi et partagé, celui du dossier ne l'est guère; son adaptation demande donc quelques précisions.

Soumis aux contraintes du suivi, l'usage et la composition du « dossier » ne repose plus sur un contrat *domestique* implicite liant l'élève, l'enseignant, les pairs; les extraits de travaux et les commentaires manifestent la nature de cette régulation qui s'arrête aux portes de la classe, aussi ce dossier échappe à la sphère publique.

Le suivi pédagogique impose donc de modifier ce contrat, ce qui renouvelle le sens de la responsabilité pédagogique : l'instruction de l'élève, et plus encore son éducation, ne sont plus affaire d'un seul, d'un maître, mais de plusieurs formant équipe, tenant conseil et davantage, tant dans le temps que dans l'espace... partenaires que le document devrait contribuer à lier (cf. Derycke, Colloque ADMEE). En conséquence, les traces et commentaires sont sélectionnés et rédigés non en fonction de la seule relation interne, domestique, mais aussi POUR l'enseignant suivant. Ceci se construit *au présent* avec le soutien de l'enseignant de manière que ce destinataire futur, maître de rencontre qu'élève je ne connais peut-être pas encore, je veux dès maintenant qu'il soit mien : mon dossier témoigne au jour le jour de ce lieu d'adresse. Il s'agit donc d'un contrat semi-domestique et double, *avec* le maître actuel et *pour* le suivant; il anticipe une appropriation par l'enfant, le projetant dans l'avenir proche. D'où la place centrale, stable, qu'il doit occuper au sein de la régulation.

Conformément à cet objectif, un dossier efficace doit être allégé pour ne comporter que les traces et commentaires significatifs actuellement pour leurs destinataires potentiels. L'aspect structurellement fragmentaire du dossier nécessite donc précisions et compléments oraux qu'il faut solliciter de l'enfant, puisqu'il l'a élaboré. Les autres traces, dont l'actualité a passé, peuvent être stockés ailleurs, dans un document à usage domestique.

On le voit, le suivi pédagogique des itinérants impose des contraintes méthodologiquement fécondes par leur degré élevé d'exigence et surtout

parce qu'elles émanent des principes du suivi. Transposés, les résultats intégrant ces contraintes concerneront le suivi en général, surtout celui des publics éloignés de l'attente des enseignants et de ceux qui subissent les ruptures plus sensibles (rupture entre école maternelle/école obligatoire, enseignement du premier degré/second degré, enseignement spécial/enseignement général, etc.). Par contre, la non-prise en compte de ces contraintes autorise à se satisfaire de solutions qui n'ont pas été vraiment éprouvées parce que privées d'enjeux. On pourra s'étonner, par exemple, que les outils de suivi actuels que sont les livrets scolaires n'ont guère été évalués quant aux variables que nous avons retenues³, comme si ces propriétés leur étaient acquises d'emblée.

Dispositif et protocole d'expérimentation

L'expérimentation *in situ* des variables du suivi que nous avons retenues est impossible avec les sédentaires, encore moins avec les itinérants : interférence de ou des observateurs dans la situation, impossibilité de confrontation entre les performances (le référé) passées, représentées par le document, et les performances présentes que pourrait observer le récepteur à l'aide d'un test identique : en effet le référé n'est pas reproductible en situation. De plus, le même élève ne peut être accueilli simultanément par plusieurs enseignants afin que le chercheur puisse saisir les variations d'interprétation, il en est de même en amont lors du renseignement du livret.

Afin de tourner cette difficulté, nous avons conçu une simulation proposée à des enseignants lors de regroupements. Les hypothèses, et le dispositif d'expérimentation destinés à isoler les deux variables, sont justifiés par une analyse sémiotique de la construction du sens par les deux types d'outils (cf. Derycke, Colloque CEPEC, 1997). Les conditions sont :

- discipline : apprentissage de la lecture, matière clé de la réussite de l'enfant,
- public : tous enseignants du primaire, en principe tous types d'élève; quatre enfants ont été retenus pour constituer des cas : un élève situé en dessous de l'attente des enseignants (Edmond est dit « en difficulté »), un autre dépasse ce niveau (Célimène) mais il n'a pas encore fini d'automatiser ses stratégies de lecteur, enfin deux enfants qui débutent par des stratégies opposées,
- supports : 1° cas vidéo : les enfants sont tour à tour mis dans la même situation complexe et socialement finalisée, soit lire oralement à un

adulte un texte inconnu d'un niveau correspondant à l'attente moyenne de l'enseignant à l'époque de la passation⁴, 2° traces écrites provenant d'exercices à trous, de dictée de mots.

L'expérimentation en simulation comporte trois phases, focalisées exclusivement sur le processus du suivi.

- I. *L'évaluation diagnostique* d'un cas vidéo en lecture : de manière à constituer deux sous-groupes équivalents, chaque enseignant reçoit, pour l'évaluation, l'un des deux types de document de suivi (grille/dossier) ; il lui est demandé de le renseigner après avoir observé la performance en lecture d'un élève sur support vidéo (v. *supra*) ; ensuite
- II. *La réception* en aveugle par un autre groupe de ces livrets associés à un questionnaire portant sur cette phase et la suivante; le questionnaire demande d'établir le niveau et de proposer une tâche, puis a lieu une interview de groupe; enfin
- III. *Le contrôle* : le même groupe observe le même cas vidéo dont le document présentait le diagnostic; chaque enseignant répond par écrit à une question ouverte⁵ ensuite a lieu une interview du groupe où il est tranché, par une question semi-fermée, sur la *concordance* ou la *discordance* entre l'activité observée (en III) et l'image que s'était faite l'enseignant à la lecture (en II) du document (rédigé en I).

Ce protocole a été proposé à des enseignants francophones de Suisse romande, de Belgique et de France afin de dégager une conception transnationale du suivi pédagogique⁶.

Signalons les insuffisances de cette simulation. 1° À la phase de réception, tous les enseignants ont connaissance des deux types d'outil puisqu'ils les ont déjà renseignés lors du diagnostic. Or un enfant peut arriver dans une école avec un document inconnu de l'enseignant; en conséquence, notre protocole sous-estime la difficulté de la réception. 2° L'évaluation, réalisée lors de l'expérimentation, est inhabituelle pour un enseignant puisqu'il s'agit d'une évaluation externe, unique, alors que celle que réalisent d'ordinaire les enseignants est interne et prend en compte le processus dans sa durée. 3° Enfin, les impératifs de l'expérimentation ont contraint à faire de l'enseignant le seul gestionnaire du dossier. Ces insuffisances sont ici négligeables si on s'en tient à l'examen des seules variables retenues; elles

doivent être prises en compte pour toute extrapolation des résultats destinés à la mise en œuvre du suivi.

Le « dossier d'apprentissage » a été prévu pour recueillir les traces de l'activité de l'élève ; les domaines et périodes d'acquisitions privilégiés sont donc ceux qui utilisent l'écrit. Or, centrée sur l'évaluation de l'apprentissage de la lecture, l'expérimentation sollicite bien au-delà son concept : ici les traces sont surtout orales, il faut donc passer par un codage intermédiaire réalisé par un tiers. Si l'efficacité du « dossier » se trouve vérifiée, on peut alors escompter qu'elle le sera d'autant plus dans les domaines qui lui sont originellement plus appropriés.

Les résultats présentés ici proviennent d'un double filtrage des données recueillies afin d'obtenir une meilleure homogénéité dans une perspective qualitative. Notre but en effet n'est pas d'établir des tendances statistiques qui gommeraient la diversité des démarches mais au contraire de préciser des variables et des stratégies *différenciées* chez des utilisateurs potentiels de ces documents.

Homogénéité du public. Sur les 182 personnes, nous avons retenu uniquement les enseignants en poste. Ceux-ci ont utilisé la grille et le dossier, tant au diagnostic de deux cas d'élève qu'à la réception des deux autres cas. Soit près de 200 documents testés sur trois phases.

Homogénéité des diagnostics. Sur quatre cas, les deux élèves qui « entrent » dans la lecture par des stratégies contrastées posent des problèmes de diagnostic : lors du contrôle (phase III), des enseignants récepteurs se trouvent en désaccord avec les collègues qui ont renseigné le document. Ces distorsions dans l'évaluation constituent une variable parasite qui empêche d'apprécier la fidélité et la validité des documents. Aussi avons-nous exclu de nos résultats ces deux cas afin de ne garder que ceux dont le niveau fait l'unanimité des enseignants, soit Célimène et Edmond.

Conséquence de ce double filtrage, il reste 82 enseignants qui tous ont renseigné et reçu les deux types de document. Nous nous intéresserons seulement à la réception (phases II et III), soit 41 grilles et autant de dossiers, 20 grilles décrivant les compétences de C., 21 celles d'E., 21 dossiers pour C., 20 pour E. L'objectif n'est pas de généraliser à partir de cet échantillon volontairement réduit mais de *déceler des variables et des stratégies spécifiques chez des professionnels chevronnés.*

Résultats quantitatifs

Dans un premier temps, nous présenterons les résultats issus du dépouillement des livrets et questionnaires écrits. En phase II (réception), les commentaires des enseignants ont été organisés selon deux rubriques : image/absence d'image. La première implique qu'il est possible d'assigner un niveau à l'élève selon le seuil supérieur (proposition d'une tâche appropriée) ou inférieur (proposition d'un test ajusté); l'absence d'image par manque d'information mène à l'impossibilité d'établir une quelconque concordance entre la représentation construite et le cas observé : le document ne remplit pas l'objectif minimum du suivi pédagogique, il est inefficace. Symétriquement en phase III (contrôle), la concordance comporte trois degrés ; la discordance indique que le document engendre une représentation nettement éloignée du niveau, laquelle hypothèque l'accueil. Eu égard à la seule validité, le document présente alors de sérieux risques (v. le tableau 1).

Grille. Cet outil convient le mieux au « bon » élève; néanmoins la majorité (12/20) manque d'information pour proposer une tâche. L'élève dit « en difficulté » est nettement desservi par cet outil puisque un quart le surévaluent nettement, la moitié manquant d'information pour établir son niveau. Avec cet élève, la grille apparaît généralement inefficace, voire nocive puisque un quart des enseignants lui proposeraient une tâche largement au-dessus de ses possibilités après l'accueil dans la nouvelle classe, ce qui risque de renforcer son sentiment d'échec.

Dossier. Edmond, ici encore, permet de trancher. Certes, l'image produite lors de l'interprétation est moins précise que pour Célimène, néanmoins elle est satisfaisante, bien que cet élève soit plus délicat à évaluer. Ceci doit s'apprécier comparativement avec la grille : il y a, avec le dossier, resserrement manifeste des résultats autour d'une image suffisante pour proposer un test ajusté, voire une tâche qui ne mettrait pas en difficulté cet élève fortement insécurisé. Le contrôle valide cette image, alors que l'on constate avec la grille une plus grande dispersion orientée vers un resserrement inverse du côté de l'absence d'image et même de la discordance quand les enseignants se sont trompés dans leur interprétation du diagnostic.

Enfin, l'analyse par localités montre, pour les deux élèves, une plus grande dispersion des résultats avec la *grille*, un plus grand resserrement avec le dossier, tous groupes d'enseignants confondus. À cet égard, un comportement apparaît significatif : méfiants avec un outil étranger, les enseignants suisses (de Fb et N) ne s'autorisent pas à porter un jugement sur

le niveau d'Edmond (7/8) alors que leurs homologues français le font pour les deux tiers (9/13), rassurés sans doute par un outil qui leur est familier, mais se trompent dans la même proportion (6/9) (cf. tableau comparatif des résultats quantitatifs en annexe). Par contre, bien que peu familiers du dossier, tous les enseignants s'autorisent à porter un jugement sur Edmond, alors même qu'ils déplorent l'insuffisance d'information, laquelle s'avérera suffisante malgré tout.

Tableau 1

Tableau comparatif des résultats quantitatifs ^a

		CÉLIMÈNE	EDMOND
G R	image		
	précise	4/20	1/21
	assez précise	4/20	2/21
	imprécise	12/20	7/21
I L	non-image	0/20	11/21

E	concordance		
	bonne	9/20	0/21
	assez bonne	6/20	3/21
	médiocre	3/20	1/21
	discordance	2/20	6/21

D O	image		
	précise	17/21	10/20
	assez précise	4/21	9/20
	imprécise	0/21	1/20
S S	non-image	0/21	0/20

I E R	bonne	16/21	13/20
	concordance	4/21	7/20
	assez bonne	0/21	0/20
	médiocre	0/21	0/20
	discordance	1/21	0/20

N.B. Les enseignants récepteurs de tel outil avec tel élève sont les mêmes qui auront l'autre outil avec l'autre élève. Cette stabilité des interprétants met en évidence la variation de l'outil en corrélation avec celle de l'élève.

^a La démarche expérimentale était trop complexe pour proposer un questionnaire fermé et univoque, aussi le comptage résulte-t-il de l'interprétation de commentaires ouverts. Sa précision est relative; l'essentiel est la polarisation par quoi s'opposent la grille et le dossier, d'une part, les deux élèves, d'autre part.

Résultats qualitatifs

Les résultats que nous allons présenter maintenant proviennent des interviews de groupe réalisés aux phases II et III après renseignement des questionnaires. Les détenteurs des grilles étaient interrogés en premier lieu.

Les inductions

Les commentaires des enseignants montrent que les formes du document de suivi, forme grille, forme dossier, favorisent des orientations interprétatives, voire induisent des jugements avant même la prise de connaissance détaillée du contenu; par ailleurs ces formes suscitent la mise en œuvre de stratégies différenciées de prise d'information et de construction de sens.

La forme grille

1) Certaines appréciations sont induites qui influencent l'interprétation :

- l'information fournie par les items est reçue comme étant précise et objective,
- plus il y a de croix, plus les enseignants disent être incités à considérer que le niveau de l'élève est élevé; inversement, moins il y a de croix, plus son niveau est faible,
- tendance à coder le plus d'items possible, mais puisqu'il n'y a pas de tests normalisés pour établir avec certitude le niveau, le codage « en voie d'acquisition » est largement utilisé; c'est le cas pour ne pas donner une image négative de l'enfant qui possède peu d'acquis (Edmond),
- inversement, réticence à attribuer « acquis » pour les compétences complexes (Célimène).

En conséquence, l'élève faible tend à être surévalué « en toute objectivité » alors que l'élève qui possède un niveau suffisant, mais non confirmé par l'étape ultérieure, se voit sous-évalué de bonne foi (v. le tableau comparatif, ci-dessus).

2) La forme grille impose un cadrage lâche et relativement distendu des compétences qui ne permet pas de distinguer des difficultés précises et, par conséquent, de proposer des remédiations ciblées, défaut qui devient un avantage lorsque les enseignants ont une conception behaviouriste de la remédiation.

La forme dossier

- 1) Induction des traces : considérée comme étant plus « subjective » que celle fournie par la grille, l'information provenant de documents comportant des traces est jugée néanmoins plus « objective » relativement à ceux qui offrent seulement les commentaires libres de l'enseignant, dépourvus d'extraits. De plus, et semblablement à la grille, cette absence de trace ne permet pas de se représenter la situation d'évaluation, à moins d'être contrebalancée par des commentaires particulièrement précis et judicieux... qui risquent de ne pas être pris en compte parce que tenus pour « subjectifs ».
- 2) Effets de l'absence de consigne :
 - 1° positivement : des enseignants informent sur les aspects spécifiques de la démarche de l'enfant selon une approche globale et personnalisée;
 - 2° focalisation sur ce qui tranche par rapport à l'ensemble de la tâche ou sur le détail. Il apparaîtra après coup (phase III) qu'avec un enfant mis en échec, les enseignants sont sollicités, non pas à souligner les manques comme avec Célimène (v. 3)), mais à repérer le peu d'acquisitions que possède Edmond tout en faisant état de ses difficultés;
 - 3° négativement : en l'absence de cadrage de l'observation par un recensement des points importants à prendre en compte dans la discipline et par souci de « remédiation », des enseignants oublient de se placer en informateur d'un destinataire absent et se focalisent sur les difficultés de l'élève, alignant des observations successives même lorsqu'elles sont contradictoires, comme si le dossier devenait une sorte de pense-bête privé destiné à orienter leur propre intervention pédagogique. Ils omettent, ce faisant, de signaler les acquis et de donner de l'élève une image globale; ceci dessert l'élève de bon niveau. Cette tendance risque de fournir un appui aux enseignants qui ont une conception behaviouriste de la remédiation. Conséquence de cette représentation négative, le niveau est sous-évalué et l'enseignant récepteur ne sait sur quelle compétence s'appuyer pour proposer des tâches appropriées. Néanmoins, la plupart de ceux qui ont observé ces contradictions ont fait les bons choix interprétatifs puisque les concordances sont acceptables (v. le tableau comparatif, ci-dessus).

Célimène

Grille⁷

Phase II, réception

L'interprétation, difficultés causées par :

- * la référence de certains items : des enseignants ne se représentent pas à quelle activité ou compétence précise ceux-ci renvoient. Cette imprécision provient de ce que 1° la situation d'évaluation est inconnue, 2° le récepteur ne partage pas avec l'évaluateur une culture commune des situations d'évaluation, bien qu'ils appartiennent au même groupe d'enseignants (région identique) lors de l'expérimentation⁸; inversement des enseignants possédant cette même culture (N) n'avaient pas cette difficulté,
- * le choix des items : des items antinomiques sont codés, lequel alors choisir⁹?
- * l'insuffisance de l'information apportée par l'évaluateur : des items de la grille ne sont pas renseignés alors qu'ils sont utiles pour établir le niveau; est-ce parce qu'ils ne sont pas observés ou parce que l'évaluateur n'a pas réalisé d'inférences à partir de la performance observée?
- * l'insuffisance d'information apportée par la grille : il n'existe pas 1° d'items intermédiaires, 2° d'items apportant des précisions nécessaires¹⁰, 3° d'items d'une autre nature que ceux existants permettant d'appréhender plus largement les acquis en lecture; ces trois points permettant une représentation plus ajustée à la performance de l'élève. La grille est jugée à la fois rigide et incomplète.

En conséquence, la « précision » et « l'objectivité » induites par la forme grille ne sont pas ici vérifiées.

Pas de convergence des jugements des enseignants lors de la phase de réception, alors qu'on aurait pu concevoir que des enseignants appartenant à une même région auraient un jugement convergent sur le niveau de l'enfant, à la lecture de documents rédigés par leurs confrères appartenant au même groupe.

Phase III, contrôle

Les divergences de la phase II sont confirmées lors du contrôle : 1° moins d'un quart ont un avis concordant sur le niveau de Célimène, 2° la moitié la sous-évaluent sensiblement, 3° un quart en ont une image déformée¹¹. Rappelons que Célimène est « bonne » en lecture; de l'avis des enseignants elle est une enfant particulièrement facile à évaluer.

Ces divergences sont imputables aux causes signalées ci-dessus, auxquelles il faut ajouter *la variation entre les enseignants*

- dans le choix des items : tous n'ont pas codé les mêmes compétences,
- dans l'évaluation du degré d'acquisition d'une compétence par le même élève : certains codent « acquis » ce que d'autres codent « en voie d'acquisition »,
- dans les inférences réalisées. Une tâche complexe et singulière ne se laisse jamais entièrement découper par des items préétablis ; l'évaluateur est donc contraint à faire des déductions. Or certaines, jugées légitimes, n'ont pas été faites et ont donc manqué aux données à interpréter, alors que d'autres, jugées injustifiées, ont été risquées et ont faussé l'interprétation.

Dossier d'apprentissage

Phase II, réception

L'interprétation.- Dans l'ensemble et pour tous les cas, si l'annotation du dossier d'apprentissage prend plus de temps que celui de la grille, par contre l'interprétation de ses données est aisée : « c'est parlant » disent de nombreux enseignants.

L'interprétation est facile et correcte lorsque

- 1° le codage de l'oralisation sur le texte est aisément compréhensible;
- 2° les traces sont complètes (extraits, observations ponctuelles et commentaires synthétiques) pour rendre compte de la globalité du processus et des stratégies. Les traces renvoient à la situation d'évaluation et sont interprétables avec précision. Celles-ci peuvent permettre de redresser un commentaire de l'évaluateur que le récepteur peut juger inapproprié, voire erroné. Les traces offrent ainsi une alternative au jugement de l'évaluateur;
- 3° les commentaires envisagent la globalité de la démarche de l'élève en lecture pour en faire ressortir les aspects saillants. Ainsi, des commentaires dépourvus de traces et jugés « subjectifs » à la réception se sont révélés concordants avec le niveau constaté en phase III.

L'interprétation risque d'être faussée quand le document ne transmet pas une image globale de l'activité-lecture mais seulement les difficultés, surtout lorsqu'elles ne sont pas désignées précisément. Ceci peut être aussi amplifié par un codage ambigu ou difficilement compréhensible du texte lu pour rendre compte de son oralisation.

L'interprétation est difficile 1° quand ne figurent que des traces partielles, 2° lorsque des annotations ponctuelles sont faites linéairement au long de l'observation et ne sont pas expurgées après coup de leurs incohérences, 3° quand l'évaluateur apporte des éléments d'information disparates ou d'ordre secondaire, voire contradictoires, comme si lui-même n'avait pas de compréhension globale de la démarche de l'élève et notait des bribes au coup par coup sans fil directeur¹².

Néanmoins, dans leur ensemble, les enseignants confrontés à ces points de vue unilatéraux ou contradictoires ont réalisé une interprétation adéquate, laquelle était vérifiée par une concordance acceptable (v. le tableau comparatif, ci-dessus).

Trois facteurs offrent des conditions positives pour l'interprétation :

- 1° l'attitude de bienveillance à l'égard de l'élève. Les compétences essentielles acquises sont recensées afin de constituer un appui pour celles à travailler,
- 2° le cadrage de l'observation évaluative grâce à des données et des formulations guidées par les objectifs pédagogiques de l'année,
- 3° l'ajustement à l'enfant pour apporter les compléments nécessaires à la compréhension de ses stratégies (attitude, aspect relationnel, entrée dans la tâche, etc.).

Convergence des jugements entre les enseignants : elle est beaucoup plus forte qu'avec la grille, les enseignants ont même été unanimes lors de certaines séances d'expérimentation.

Phase III, contrôle

Lors des interviews, bien que nuancée par une légère sous-estimation, la concordance apparaît bonne dans plus de huit cas sur dix entre les jugements portés à la réception (phase II) et la vérification (phase III).

À la phase III, les enseignants récepteurs se plaignent que les jugements les plus adéquats demeurent restrictifs. Cette légère sous-estimation du niveau de l'élève provient, selon eux, de la situation-test (brève évaluation externe). Ils souhaitent avoir d'autres supports pour apprécier la progression. Nous rencontrons ici une limite du dispositif d'expérimentation mais non pas du dossier d'apprentissage.

La fidélité et la validité du dossier d'apprentissage est donc démontrée pour ce qui concerne Célimène, tenant compte des restrictions que nous avons mentionnées ci-dessus.

Edmond

Grille

Phase II, réception

L'interprétation.- L'interprétation des grilles est difficile.

Il y a des contradictions entre des items codés, ceci même lorsque l'évaluation provient d'un évaluateur compétent : l'un d'eux par exemple avait fourni un dossier d'apprentissage concordant sur Célimène. Il manque des items sur le comportement devant la tâche et sur la situation d'évaluation.

Les grilles ne montrent ni les acquis, ni les difficultés.

En conséquence la référence (ici, le niveau de l'enfant), désignée par les grilles, est floue ou quasi inexistante. Les enseignants sont perplexes et se sentent souvent, non seulement dans l'impossibilité de proposer une tâche adéquate à l'enfant, mais même de lui proposer un test qui lui soit adapté.

Il n'y a pas de convergence des jugements entre les enseignants : les jugements des enseignants sont contrastés, certains le considèrent en échec, et d'autres, qu'il entre dans la lecture.

Phase III, contrôle

Le niveau d'Edmond est surestimé par un nombre important d'enseignants. Il y a majoritairement discordance entre le niveau constaté et l'interprétation du niveau à partir de la grille; l'outil de suivi ne remplit pas son objectif minimum, pire, il est nocif s'il favorise la mise en échec de l'élève.

Dossier d'apprentissage

Phase II réception

L'interprétation.- L'information est plus claire qu'avec la grille, plus précise. Des indications pertinentes sont données sur le comportement lors de l'exécution de la tâche, car celui-ci est particulièrement significatif. Cependant, les indications demeurent vagues. Néanmoins, les enseignants considèrent posséder suffisamment d'information pour proposer un test qui mette Edmond en situation de réussite dans sa progression lors de son accueil, ce qui constitue l'objectif minimum du livret de suivi pédagogique.

Cette situation d'évaluation est jugée insuffisante et il est demandé davantage de traces des travaux de l'élève. Comme pour la grille, ces

insuffisances proviennent de l'inadaptation de la tâche proposée, trop difficile pour Edmond.

Convergence des jugements entre les enseignants : forte convergence des jugements des enseignants dans la majorité des groupes.

Phase III, contrôle

Dans l'ensemble les jugements établis lors de la réception concordent avec le niveau constaté, mais avec une légère surévaluation. Seul un jugement est nettement surévalué, il provient d'une erreur d'appréciation commise par l'enseignante à la lecture du document (v. le tableau 2).

Conclusions

Les données traitées proviennent d'un échantillon particulièrement homogène d'enseignants; ceux-ci ont suivi deux cas extrêmes, Célimène et Edmond. Nous pouvons relever deux points principaux :

- 1° il apparaît nécessaire que le document de suivi comporte des traces des travaux de l'enfant, sélectionnées dans la durée afin de donner une représentation concrète des moments clés de son évolution, telle qu'elle donne des indications actuellement pertinentes;
- 2° l'observation évaluative qui guide la sélection de ces traces doit être cadrée par des consignes qui, partant d'une attitude « bienveillante », relève les acquis, les stratégies, l'entrée dans la tâche, dégage enfin les priorités pour fournir une représentation à la fois globale, personnalisée et pertinente qui permette la poursuite des apprentissages.

Si le premier point n'est pas une surprise, le second point constitue pour nous une véritable découverte. En effet, préalablement à l'enquête, la variable « cadrage de l'observation évaluative » n'avait pas été établie comme un facteur sensible du suivi : elle nous semblait aller de soi¹³.

En tout état de cause, le *dossier d'apprentissage* semble bien être plus efficace. Mais il faut que les conditions que l'on vient d'évoquer soient remplies. Reste qu'une simulation ne peut rendre compte de toutes les variables, en particulier les difficultés qui sont liées à la durée et à la richesse du document (cf. Simon & Forgette-Giroux, 1993). Néanmoins, l'argument le plus puissant qui milite en sa faveur pour assurer le suivi pédagogique est

Tableau 2

Tableau comparatif des résultats qualitatifs

	DOSSIER D'APPRENTISSAGE	GRILLE
	*homogénéité des jugements *concordance	*garantie d'une appréhension des compétences disciplinaires principales
P O S I T I F	*personnalisation par - centrage sur le spécifique, la difficulté - globalisation de la démarche de l'élève - notation sur l'attitude	*universalité *pas de jugement de valeur
	*indépendant du jugement précédant	
	*rapidité d'interprétation	*visualisation rapide *rapidité d'encodage
	*peu sensible au niveau de l'élève	*acceptable pour l'élève sans difficulté
N É G A T I F	*non-garantie de l'appréhension des compétences principales *focalisation négative (sur le «bon» élève), *manque de critères d'observation (cadrage)	*hétérogénéité des jugements *variation d'interprétations * <u>contradictions avec le niveau constaté</u>
	*lourdeur à l'encodage	*réduction des écarts entre niveau bon et médiocre *inefficace pour l'élève en difficulté
	*oublis	*dépendance du jugement du prédécesseur *réducteur + amalgames
	*codage non normalisé de l'oralisation en lecture	*imprécis, abstrait, ne permet pas de distinguer une difficulté précise
	*risque de jugement de valeur	*incomplet *manque de gradation *subjectif

la place active réservée à l'enfant, premier intéressé et seul médiateur stable entre les enseignants¹⁴.

Si on prend le parti de la grille, son caractère abstrait et imprécis peut-il être amendé par l'usage systématique de tests? Certes, ceux-ci sont utiles pour un bilan sommatif, annuel par exemple, et pour garantir l'investissement des enseignants vis-à-vis de chaque élève. Mais ceux-ci parcellisent les savoir-faire et sont administrés de l'extérieur; comment les associer à une pédagogie quotidienne, centrée sur une démarche globale à la fois sociale et culturelle qui donne sens à l'école et à laquelle participe l'enfant?

NOTES

1. Directeur de la revue *Langage et société*, Pierre Achard a contribué à la conception finale du questionnaire inclus dans le dispositif d'expérimentation dont nous proposons ici une partie des résultats. Il est décédé le 1^{er} octobre 1997.
2. Recherche financée par la Commission européenne dans le cadre de SOCRATES, Action II Chapitre 2 (Comenius, interculturel), pour le compte du projet A6 de l'EFECOT, Bruxelles. J'en assume la direction avec la collaboration de A. C. Silva (ISPA, Lisbonne), B. Heijs (Seminarium voor Orthopédagogie, Utrecht) pour l'axe I (suivi en lecture), F. Tollmien (Université de Essen) pour l'axe III (accompagnement et pilotage du suivi); j'assume l'axe II (fidélité, validité des outils) avec la coopération de E. Corroy, E. Grasset, L. Maéro, C. Segreto et T. Masson, enseignants.
3. Ceci notamment en France où existait un organisme affecté à des tâches de cette nature : la Direction à l'évaluation et à la prospective.
4. Il ne s'agit donc pas d'enfants mis devant des tests destinés à déterminer des compétences partielles hors de leur mobilisation dans une tâche sociale.
5. « Vous venez de voir l'activité de l'enfant dont le livret vous avait présenté le niveau. Qu'en pensez-vous? »
6. Les expérimentations ont eu lieu à l'occasion de modules de formation continue; ils ont duré chaque fois une journée complète. En tout, 182 personnes ont participé aux expérimentations, dont des enseignants stagiaires et des inspecteurs.
7. Remarque unanime à propos de tous les cas : des trois possibilités de codage des items sont insuffisantes (« acquis », « non-acquis », « en voie d'acquisition »), car une compétence non observée dans la situation d'évaluation n'est pas nécessairement non acquise. Pour parer à ce défaut, les enseignants inventent un codage particulier qui est parfois mal interprété par le récepteur.
8. En conséquence, on peut légitimement s'interroger sur la transférabilité de cette information pour les traduire en activités pédagogiques.
9. Par exemple, « déchiffrage - *acquis* » et « compétence combinatoire - *en voie d'acquisition* » ; « lecture expressive - *acquis* » et « ne pose pas de questions sur le texte », ce qui est un indicateur négatif de compréhension.

10. Par exemple, s'il existe un item signalant si l'enfant possède ou non un « capital mots », on ne sait rien sur l'étendue de celui-ci et sur ses caractéristiques, plus généralement, sur les différentes procédures d'entrée par le sens, les démarches et stratégies pour entrer dans la tâche.
11. L'évaluation approximative à partir des réactions orales fait apparaître des jugements plus sévères qu'au travers du questionnaire écrit.
12. En ce cas, il faut incriminer la compétence en évaluation de l'enseignant.
13. Ce résultat valide pour nous le choix d'une méthode qualitative de recueil et de traitement des données.
14. Les contraintes de l'expérimentation imposent de mettre entre parenthèses l'interaction puisqu'elle est impossible avec le support de la vidéo. Ces conclusions sont donc à intégrer à l'usage effectif du dossier d'apprentissage au sein d'une dynamique où l'enfant a de réels interlocuteurs, étant donné sa fonction médiatrice. Ainsi la relation avec l'élève, la négociation des critères, l'objectivation écrite de ses commentaires concernant sa production, essentiels pour favoriser chez lui une démarche d'appropriation de l'auto-évaluation, sont aussi rendues fonctionnelles.

RÉFÉRENCES

- Dericke, M. (à paraître). Évaluation et suivi pédagogique : la communication prise au piège. Actes du Colloque ADMEE 1996, Grenoble.
- Dericke, M. et al. (à paraître). Orientations pour le suivi pédagogique des enfants itinérants.
- Dericke, M. (à paraître). Suivi pédagogique : quel sens construisent une grille évaluative, un dossier d'apprentissage?. Actes du Colloque CEPEC 1997, Lyon.
- Gauch, A.-M., Chevalerias, F. & Ripoche, J.-L. (1995). Les outils de suivi : description comparative. In La scolarisation des enfants d'itinérants. Revue du C.R.E. n° 9, Saint-Étienne, Université Jean Monnet.
- Hadji, C., (1992). L'évaluation des actions éducatives. Paris : P.U.F.
- Simon, M. & Forgette-Giroux, R. (1993). Vers une utilisation rationnelle du dossier d'apprentissage. Mesure et évaluation en éducation, 16(3-4), 27-40.
- Weiss, J. (1994). Évaluer autrement. Mesure et évaluation en éducation, 17(1), 63-73.

ANNEXE
RÉSULTATS QUANTITATIFS

CÉLIMÈNE							
	Localités	Fb	N	F	CF	SE	Total
G R I L L E	précise					4/5	4/20
	image assez précise	1/5	2/3	1/4			4/20
	imprécise	4/5	1/3	3/4	3/3	1/5	12/20
	non-image						0/20

	bonne	4/5	2/3			3/5	9/20
	concordance assez						
	bonne	1/5		2/4	2/3	1/5	6/20
	médiocre		2/4	1/3			3/20
	discordance		1/3			1/5	2/20

		Fr	N	F	CF	SE	Total
D O S S I E R	précise	4/5	3/3	2/4	3/4	5/5	17/21
	image assez précise	1/5	2/4	1/4			4/21
	imprécise						0/21
	non-image						0/21

	bonne	4/5	3/3	2/4	2/4	5/5	16/21
	concordce assez bonne	1/5		2/4	1/4		4/21
	médiocre						0/21
	discordance					1/4	1/21

		EDMOND					
		Fr	N	F	CF	SE	Total
G R I L	précise	1/3	1/21				
	image assez précise			1/4		1/5	2/21
	imprécise		1/4	3/4		3/5	7/21
	non-image	5/5	2/3	2/4	1/4	1/5	11/21

L E	bonne						0/21
	concordce assez bonne		1/3		2/4		3/21
	médiocre				1/5		1/21
	discordance			2/4	1/4	3/5	6/21

		Fr	N	F	CF	SE	Total
D O S S	précise	3/5		2/4	2/3	3/5	10/20
	image assez précise	2/5	3/3	2/4	1/3	1/5	9/20
	imprécise				1/5		1/20
	non-image						0/20

I E R	bonne	4/5	3/3	1/4	3/3	2/5	13/20
	concordce assez bonne	1/5		3/4		3/5	7/20
	médiocre						0/20
	discordance						0/20