

A propos de la validité « didactique » d'une évaluation

Jean-François de Pietro

Institut de recherche et de documentation pédagogique

Jean-Francois.dePietro@irdp.ch

Murielle Roth

Institut de recherche et de documentation pédagogique

Murielle.Roth@irdp.ch

Résumé

En nous inscrivant dans la longue tradition des activités menées par l'Institut de recherche et de documentation pédagogique (IRDp) dans le domaine de l'évaluation, tradition illustrée en particulier par les travaux de Jean Cardinet, nous abordons ici – en vue d'une évaluation liée au *Plan d'études romand* (PER) introduit maintenant depuis quelques années – diverses questions qui touchent à l'élaboration d'une banque d'items et de matériaux langagiers à mettre à disposition des enseignants et des cantons francophones. Le but de notre contribution est de mettre en évidence les questionnements didactiques que pose une telle construction :

- Quels sont les « objets » du PER à prendre en compte pour l'évaluation ?

Ou, pour le dire autrement,

- Comment transposer des objets d'enseignement et d'apprentissage en objets d'évaluation ?
- Quels sont les (types de) tâches à même de rendre ces objets visibles ?

Nous abordons ces interrogations en nous appuyant sur des observations et analyses concernant en particulier le domaine du français. Plus généralement, l'enjeu de nos travaux est de contribuer à renforcer la *validité didactique* des évaluations conduites en lien avec le PER – autrement dit de définir certaines conditions afin qu'on évalue bien ce qu'on veut évaluer. Et, dans cette perspective, nous ferons résonner nos questionnements actuels par des citations de Jean Cardinet auxquelles ils nous semblent particulièrement faire écho.

Mots-clés

Validité didactique - modélisation - IRDP - banque d'items

Pour citer cet article : de Pietro, J-F., & Roth, M. (2017). A propos de la validité « didactique » d'une évaluation. *Evaluer. Journal international de Recherche en Education et Formation*, 3(3), 31-50.

1. Introduction

« L'évaluation est moins un problème de mesure que de dialogue. »

(Cardinet, 1991, p. 19)

Travaillant à l'*Institut de recherche et de documentation pédagogique* (IRDP), notamment dans le domaine de l'évaluation des apprentissages des élèves, nous nous inscrivons dans une longue tradition de travaux en ce domaine qui a forgé une part importante de la réputation de l'Institut et dont Jean Cardinet représente certainement la figure emblématique. Pourtant, héritiers – indirects (car nous n'avons nous-mêmes jamais eu l'occasion de travailler avec lui) – de ses travaux, et engagés aujourd'hui dans l'élaboration d'un dispositif d'évaluation destiné à la fois à être mis à disposition des enseignants et à servir de base pour d'éventuelles épreuves communes romandes, nous avons pu nous rendre compte à l'occasion de ce colloque d'hommage que nous ne percevions pas toujours l'immense importance de ses travaux¹.

Certainement impressionnés par la scientificité, et la complexité, de ses travaux d'orientation éducatrice, nous avons ainsi dû constater que nous avions un peu oublié ses autres écrits portant par exemple sur le sens et les fonctions de l'évaluation, illustrés notamment par son image de la fusée (Cardinet, 1977a) – travaux qui s'avèrent pourtant, ainsi que nous le verrons ci-après, d'une très grande pertinence *didactique* aujourd'hui encore. Bien sûr, les notions d'évaluation *formative* (Cardinet, 1979), *interactive* (Cardinet et Weiss, 1978 ; Cardinet, 1991), voire *informative* (Weiss, 1995, 2003) ou de *régulation* (Cardinet, 1977a ; Mottier Lopez et Tessaro, 2016) font aujourd'hui partie de notre « bagage notionnel ». Mais nous n'avions pas pris toute la mesure des apports de Jean Cardinet au développement de ces nouvelles perspectives pour l'évaluation.

Nous avons par conséquent souhaité, pour cette contribution, mettre en discussion quelques-uns de nos questionnements actuels en les faisant résonner par des citations de Jean Cardinet auxquelles ils nous ont semblé faire écho : dans quelle mesure ses apports, de divers ordres, sont-ils pris en compte – même sans que nous en soyons pleinement conscients – dans nos travaux actuels ? Que nous disent-ils de l'évolution actuelle de la réflexion et des actions mises en œuvre – au niveau romand – dans le domaine de l'évaluation ?... Cette réflexion concernera essentiellement des questions didactiques, à l'exemple de la discipline *français* et, plus particulièrement, de la compréhension écrite. Nous nous centrerons avant tout sur des questions dont l'enjeu est de renforcer la validité *didactique* des évaluations qui, dans l'école romande, devront désormais être conduites en lien avec le *Plan d'études romand* (CIIP, 2010 ; désormais PER) récemment introduit dans les différents cantons. Il s'agit, autrement dit, de définir par ces travaux certaines conditions nécessaires afin qu'on évalue bien ce qu'on veut évaluer car... « chaque épreuve doit correspondre aux objectifs du curriculum enseigné. » (Cardinet, 1990, p. 2)

Nous présenterons tout d'abord le contexte dans lequel nous travaillons (chapitre 1) et les mandats qui nous sont attribués dans ce contexte, du premier projet d'« EpRoCom » au *Masterplan* 2016 (chapitre 2). Nous en viendrons alors aux questions auxquelles nous essayons, actuellement, de répondre : à propos de la notion d'item et de la *validité* des items (chapitre 3), à propos des « objets » du Plan d'études romand (PER) à prendre en compte pour l'évaluation (*Quoi évaluer ?*, chapitre 4) et de leur statut (*Compétences, connaissances et références culturelles*, chapitre 5) et à propos des types de questionnement à mettre en place en

¹ Cette remarque n'engage bien sûr que les auteurs de cette contribution !

relation à ces objets (chapitre 6), pour conclure sur quelques réflexions générales sur l'importance et la place de l'évaluation dans le contexte scolaire (chapitre 7).

2. Le contexte de nos travaux

Ainsi que nous l'avons rappelé, l'IRDP a dès sa création en 1969 été particulièrement actif dans le domaine, au sens large, de l'évaluation. Les travaux réalisés s'inscrivent dans trois directions principales et concernent :

- les fondements de l'évaluation (questions de finalités, de validité, etc.) ;²
- la fiabilité de la mesure, notamment dans une perspective éducatrice ;³
- des évaluations ponctuelles ciblées portant surtout sur les enseignements rénovés du français et des mathématiques, couvrant des domaines aussi divers que les moyens d'enseignement élaborés dans le cadre de ces rénovations et leurs effets sur les performances des élèves, l'avis des parents quant à ces enseignements, etc.⁴

Dès le tournant du millénaire, une volonté politique nouvelle s'est manifestée – tant au niveau suisse qu'au niveau romand – en vue d'une plus grande harmonisation entre les systèmes scolaires cantonaux.⁵ Cela a conduit à diverses « Déclarations » et recommandations politiques, à l'élaboration au niveau fédéral d'un « Concordat » (« HarmoS »), plébiscité lors d'une votation populaire en mai 2006, qui définit certaines conditions minimales à respecter dans tous les cantons signataires (âge du début de la scolarité, compétences à atteindre à certaines étapes du cursus pour les disciplines considérées comme prioritaires⁶, etc.), à l'écriture de nouveaux plans d'études dans les diverses régions linguistiques du pays, etc.

En Suisse romande, un *Plan d'études romand*, le « PER », réalisé par des équipes disciplinaires d'enseignants, a ainsi été publié en 2010 et progressivement mis en œuvre dans les 7 cantons francophones. Il définit les *connaissances* et les *compétences* que l'école obligatoire doit transmettre aux élèves, « permettant à chacun et chacune de développer ses potentialités de manière optimale. » (CIIP, 2003, 2010). En lien avec ces développements et l'introduction du PER, les autorités éducatives ont dès lors engagé des travaux en vue de l'élaboration, d'une part, de moyens d'enseignement qui soient en adéquation avec les contenus et objectifs du PER, d'autre part d'épreuves de référence permettant de vérifier si ces objectifs sont atteints.

² Voir par exemple, en reprenant quelques jalons historiques de ces travaux : Cardinet, 1977b, 1981 ; Weiss, 1998 ; de Pietro, Roth et Sánchez Abchi, 2016 ; etc.

³ Cardinet, 1986 ; Cardinet et Tourneur, 1978 ; Eboulet et Matei, 2013 ; Sánchez Abchi, de Pietro et Roth, 2016 ; etc.

⁴ Quelques exemples parmi les très nombreuses publications à ce propos : Cardinet, 1977c ; Pochon, 1979 ; Wirthner et Weiss, 1981 ; Calame *et al.*, 1995 ; Genoud, 1999 ; Antonietti, 2005 ; de Pietro *et al.*, 2009 ; Elmiger et Singh, 2014 ; etc.

⁵ Rappelons ici que ce sont, en Suisse, les 26 cantons qui sont chacun responsables de la structure et du fonctionnement de leur système scolaire et qu'il en découlait une importante variabilité tant en ce qui concerne la structure même du système (âge d'entrée à l'école, types de filières, procédures de promotion et de sélection, etc.) que les moyens d'enseignement et l'enseignement lui-même.

⁶ Langue de scolarisation, langues étrangères, mathématiques et sciences naturelles. Voir à ce propos : <http://www.edk.ch/dyn/15415.php> (consulté le 28 août 2017).

Afin de bien comprendre où se situe le projet dans lequel s'inscrivent nos réflexions, voici, succinctement rappelées, quelques-unes des étapes par lesquelles il est passé :

- 2004 : l'IRD, en collaboration avec des groupes de spécialistes issus des cantons, mène une première réflexion sur les conditions de création d'épreuves romandes communes. Ce travail a tout d'abord l'intérêt de rendre visible la variété et le nombre importants d'épreuves évaluatives développées dans presque tous les cantons romands. Dans son rapport final, en 2007, le groupe propose dès lors, dans la perspective d'une harmonisation future, de réaliser une analyse plus fine de ces multiples épreuves afin de mieux en saisir la nature, la visée et les enjeux (apprentissage, sélection, certification...). Pour ce faire, il souligne la nécessité de créer un pôle romand de compétences en matière d'évaluation, afin de définir des buts communs et construire une méthodologie commune.
- 2007 : la *Conférence intercantonale de l'Instruction publique* (CIIP), organe politique responsable au niveau romand de la coordination de l'éducation entre les différents cantons, adopte une *Convention scolaire romande*, instituant ainsi un « espace romand de la formation ». Elle y définit des conditions-cadres à appliquer par les différents cantons. L'une d'entre elles concerne particulièrement l'évaluation :⁷

Convention scolaire romande – ch.2, sect.2, art. 15, al.1 et 2

1. La CIIP organise des épreuves romandes communes à l'Espace romand de la formation, en vue de vérifier l'atteinte des objectifs du plan d'études.
2. En fin de cycle ou à la fin du degré secondaire I, si la discipline choisie pour l'épreuve romande commune correspond à celle d'un test de référence vérifiant un standard national, le test de référence peut servir d'épreuve commune.

- 2008 : la CIIP confie à l'IRD le mandat d'étudier la possibilité de créer des épreuves romandes communes. Un *Consortium* romand, formé de délégués cantonaux, ainsi que des groupes de travail dans les disciplines *français* et *mathématiques* sont constitués en vue de l'élaboration de premières épreuves. Les travaux réalisés – dans le cadre du *Consortium* jusqu'en 2012 puis, suite à la dissolution de celui-ci, au sein de l'IRD – ont en particulier donné lieu à trois publications: une analyse des épreuves cantonales et des scénarios possibles pour une évaluation adaptée au PER (Marc et Wirthner, 2012), un modèle d'évaluation adapté au PER (Marc et Wirthner, 2013) et une analyse des éléments du PER et, subsidiairement, des moyens d'enseignement romands (MER) qu'il s'avérerait nécessaire de clarifier afin de mieux définir les « objets » à évaluer (Roth et de Pietro, 2013).⁸

C'est donc dans la suite de tous ces travaux que s'inscrivent ceux que nous menons actuellement et les quelques réflexions que nous souhaitons mettre en discussion dans cette contribution.

⁷ Convention scolaire romande – Texte adopté par la CIIP le 21 juin 2007, Art. 15, §1. <http://www.ciip.ch/documents/showFile.asp?ID=2518>

⁸ Durant cette période, l'IRD a également mené d'autres travaux en lien à cette problématique d'une évaluation romande. Ils portent sur le classement des items (Pochon, 2006), la perception des évaluations externes par les enseignant-e-s (Dierendonck, 2008), un inventaire des différentes réformes romandes dans le domaine de l'évaluation (Behrens, 2010), l'analyse des dispositifs de testing adaptatifs (Vermot, Behrens et Marc, 2011) et la question de la protection des données (Armi et Pagnossin, 2012).

3. Le travail actuel de l'IRD : du projet « EpRoCom » au *Masterplan* 2016

« (...) aucune technique ne peut estimer le niveau des élèves de façon généralisable. »

(Cardinet, 1988, p. 4)

2.1 Un projet prioritairement politique

Tout ce qui a été réalisé durant cette période, *grosso modo* entre 2003 et 2016, s'inscrit – parfois avec succès, parfois moins – dans le mouvement de reprise en main politique de la gestion du/des système(s) scolaire(s) : volonté d'aller vers une meilleure harmonisation – qui assure notamment une plus grande équité d'un canton à l'autre et facilite en même temps la mobilité intercantonale des travailleurs – et d'assurer un véritable *pilotage* du système via l'élaboration de conditions-cadres, d'un plan d'études commun et d'épreuves permettant d'en mesurer l'efficacité.⁹

Ce changement de perspective comporte de nombreux aspects positifs mais il pourrait en même temps conduire à un repositionnement ambigu, à divers égards problématique, de la place et du rôle de la recherche dans l'ensemble du dispositif – dont témoigne, par exemple, le point concernant l'évaluation du *Mandat de prestations* attribué par la CIIP, pour les années 2016 – 2019, à l'IRD :

B : Evaluation du système, épreuves de référence

Art. 6 Epreuves romandes communes pour la scolarité obligatoire

1. L'IRD assure la coordination générale de la préparation des épreuves romandes communes (EpRoCom) sur la base des objectifs du PER et du cadre de travail placé sous la responsabilité du secrétaire général et de la CLEO [Conférence latine de l'enseignement obligatoire]. Il collabore étroitement avec les responsables des épreuves cantonales dans les services d'enseignement ou de recherche des cantons membres et dirige, dans le cadre de groupes de travail intercantonaux, la mutualisation ou l'élaboration, la validation et le calibrage des items à partir desquels il construit les séries d'épreuves correspondant à la planification quadriennale adoptée par l'Assemblée plénière. Il rédige et soumet à l'aval de la CLEO les consignes d'application et de correction qui seront communiquées par la CIIP aux Départements cantonaux, responsables de la conduite et de la correction des épreuves. Il développe au besoin l'infrastructure technique et les réseaux permettant la mise en œuvre du dispositif.

Extrait du mandat de prestations de l'IRD pour années 2016-2019

Selon ce document, pris à la lettre, la recherche pourrait devenir un simple rouage de ce travail de pilotage, bien loin des recherches-actions ou autres menées dès la création de l'Institut et qui ont notamment soutenu le développement de nouvelles approches de

⁹ Soulignons tout de même que tout cela s'inscrit également dans un mouvement sociétal plus large encore, d'envergure mondiale, marqué par une influence de plus en plus forte du monde économique, qui conduit à une exigence de rentabilité et d'efficacité du système scolaire concrétisée notamment dans la volonté d'assurer un pilotage basé sur la mesure de performances : « Le développement de certaines formes modernes d'évaluation à l'école est issu de la culture de l'entreprise et de la technocratie, et (...) dans cet esprit, tout investissement doit être évalué. » (Frakowiak, 2005). Et cela conduit dès lors à élaborer de nombreux instruments de mesure : épreuves PISA, tests de référence nationaux portant sur les standards HarmoS, épreuves romandes communes visant à mesurer l'atteinte des objectifs du PER, etc.

l'évaluation, en particulier à travers les travaux de Jean Cardinet : évaluation formative (*cf.* Laveault, ici-même), puis interactive (Weiss, 1992) voire informative (Weiss, 2003), etc.¹⁰

Dans le cours de ces travaux, de plus, les divers acteurs impliqués – enseignant-e-s, didacticien-ne-s, mais aussi responsables politiques – ont également dû se rendre compte des difficultés considérables de cette tâche d'harmonisation en raison des différences d'approches qui subsistent entre les cantons. Il s'avère de fait très difficile de définir un cadre commun pour l'ensemble d'entre eux dans la mesure où, par exemple, ils n'attribuent pas nécessairement un même statut et une même fonction à de telles épreuves romandes communes.

Dans une lettre datée du 30 juin 2011, le Secrétaire général de la CIIP invitait ainsi l'IRDP à développer « sans retenue la production d'items pouvant être utilisés à terme dans de telles épreuves intercantionales » ... mais en précisant cependant que l'organisation de celles-ci ne pouvait être déterminée tant que l'organisation des tests nationaux de référence, fondés sur des standards *nationaux* de formation, n'aura pas été définitivement arrêtée...

Ces diverses péripéties et les difficultés rencontrées ne vont pas sans soulever certaines questions :

- L'école actuelle tend vers une multiplication des dispositifs de mesure : épreuves internationales (PISA), tests de référence nationaux (HarmoS), régionaux (EpRoCom), cantonaux (épreuves communes), travaux en classe pour l'attribution des notes et des moyennes... L'évaluation des apprentissages, certes nécessaire tant dans la perspective de la gestion du système que comme élément constitutif de tout apprentissage, ne finit-elle pas par prendre trop de place, par occuper une part trop importante des programmes ? Certaines réactions et résistances des enseignant.e.s peuvent le laisser penser, d'autant que ce nouveau pilotage est largement guidé selon un mode « top-down » dans lequel ils ne semblent guère avoir leur mot à dire.
- Le dispositif envisagé, les épreuves communes notamment, a encore, pour l'heure en tout cas, des finalités floues et qui ne semblent pas les mêmes pour l'ensemble des cantons : s'agit-il uniquement de fournir des informations utiles pour le pilotage du système, dans le but d'ajuster et faire évoluer le PER sur la base de résultats globaux – appréhendés via des évaluations partielles échantillonnées – concernant l'atteinte des objectifs qui y figurent ? Ou de fournir des informations utiles à chaque canton, dans le but de « situer » (orienter ? sélectionner ?...) chaque élève sur la base du PER ? Doivent-elles remplacer (par souci d'économie ?) les épreuves cantonales actuelles ou viennent-elles simplement s'y ajouter ?...
- Ne passe-t-on pas trop rapidement sur les nécessaires questionnements didactiques que soulève l'évaluation en décidant la production d'épreuves alors que de (trop) nombreuses questions restent ouvertes : qu'est-ce qui, dans le PER – qui est un référentiel conçu pour l'enseignement et l'apprentissage, non pour l'évaluation –, doit faire l'objet d'une évaluation ? Dans ce contexte d'une évaluation fondée sur le PER, qu'est-ce qu'un « item » ? Bref, que mesure-t-on vraiment et qu'est-ce que cela signifie ?

La citation placée en exergue à ce chapitre nous invite pour le moins à la prudence, et à cesser de voir l'évaluateur comme un *deus ex machina* gérant le système depuis son tableau de bord !

¹⁰ A propos de l'évolution de l'orientation des travaux au sein de l'IRDP, *cf.* Weiss *et al.*, 2001.

2.2 Une évolution significative et positive

« Il faut abandonner ce mythe de l'évaluateur à son tableau de bord. »

(Cardinet, 1991)

Le mandat reçu par l'IRDP s'inscrit dans ce contexte et comporte par conséquent une importante dimension politique. Afin de ne pas devenir qu'un simple rouage dans le système global de pilotage, les collaborateurs impliqués dans le projet se sont toutefois efforcés de poursuivre un véritable travail de recherche :

- en expérimentant des dispositifs d'évaluation (Marc et Wirthner, 2013),
- en menant des analyses approfondies, dans la perspective d'une potentielle évaluation commune, des éléments figurant dans le PER, des moyens d'enseignement et de quelques épreuves cantonales (Marc et Wirthner, 2012 ; Roth, de Pietro et Sánchez Abchi, 2014).
- en tentant d'élaborer une « modélisation » de l'évaluation mettant en relation les objets à évaluer et les types de tâches qui permettent de les rendre visibles (Roth, de Pietro et Sánchez Abchi, 2014); de Pietro, 2015).

Il s'agissait ainsi, dans la mesure du possible, de fournir les fondements scientifiques en vue de ce pilotage du système qui devrait être basé, notamment, sur des épreuves romandes communes. Autrement dit, il s'agissait pour nous d'y inclure une véritable *dimension didactique* dans la mesure où, comme nous l'avons vu, le PER constitue un référentiel pour l'enseignement / apprentissage et ne permet de définir clairement ni des « objets » à évaluer, ni des modalités d'évaluation. C'est pourquoi il nous a paru indispensable de mener cette réflexion didactique en amont, afin qu'une évaluation en lien au PER – quelle qu'en soit par ailleurs la finalité (évaluation-système ou évaluation des acquis des élèves) – ait une validité de contenu, autrement dit qu'elle évalue ce qu'on veut évaluer.

En 2016, les modalités de réalisation du projet ont été précisées dans le cadre d'un *Masterplan* – qui tient compte, par divers réajustements, des questions soulevées ci-avant et de ces travaux en cours, comme le montre l'extrait suivant de cette nouvelle planification :

2. Finalités et orientations retenues

FINALITÉS

Le projet EpRoCom vise en priorité à la mise à disposition d'une BANQUE D'ITEMS VALIDÉS, pertinents, fiables et fondés sur les objectifs et progressions du PER, dans laquelle les enseignants et les cantons pourront librement puiser et à partir de laquelle, sur décision de l'Assemblée plénière en fonction des besoins et opportunités, une épreuve commune romande pourra de temps à autre être mise sur pied et pilotée par l'IRDP et ses partenaires scientifiques dans le but de vérifier à plus large échelle certaines atteintes spécifiques du PER.

Extrait du Masterplan EpRoCom – vérification de l'atteinte des objectifs du plan d'études romand (PER)

Cette redéfinition du projet a le mérite de l'*ouvrir* à divers usages (par les cantons, mais aussi par les enseignant.e.s) et de redonner un véritable rôle à la recherche. Elle nous paraît très intéressante dans la mesure où elle suppose des expérimentations et des prises d'informations auprès des enseignant.e.s afin de déterminer ce qu'il est finalement possible / souhaitable de faire et dans la mesure où elle oriente désormais le travail vers l'élaboration d'une base de données fondée didactiquement, sans subordonner ce nécessaire travail scientifique à la

réalisation immédiate d'épreuves. Nécessaire, en effet, en raison des nombreuses questions qui restent pour l'heure non résolues, en lien notamment avec la recherche d'une *validité* didactique et avec la définition même de ce qu'on entend par *item*. C'est donc à une réflexion autour de ces questions que nous allons à présent nous intéresser.

3. Des « items » à « valider » ...

« La standardisation des épreuves privilégie l'objectivité de la correction, en négligeant le caractère arbitraire du choix des questions »

(Davaud et Cardinet, 1992, p. 7)

3.1 Quelle définition de l'item ?

Nous nous sommes donc d'abord demandé ce qu'il fallait entendre par la notion d'item. Dans une approche peu réfléchie – telle qu'on en observe trop souvent – la réponse pourrait sembler évidente : un item c'est une question de test et la réponse qui lui est associée. C'est par exemple ce qu'on observe dans les épreuves conçues sous la forme de QCM. Dans le meilleur des cas on pourra donner plus d'importance (plus de « points ») à certains items qu'on juge prioritaires – autrement dit pondérer l'épreuve – mais certaines ambiguïtés demeurent, rendant nécessaires diverses clarifications :

- Peut-on « mélanger » dans une même épreuve et les prendre en compte dans un calcul unique des questions portant sur des objets différents, tant en nature (orthographe, production orale...) qu'en statut (connaissances, compétences...) ?¹¹ Et qu'est-ce qui, au final, est alors évalué ?
- L'IRT (« Item Response Theory »), notamment, est fondée sur le principe que, pour être placés sur une échelle évaluative unique, les items doivent être *indépendants* les uns des autres, au risque sinon d'évaluer (et comptabiliser) plusieurs fois certains éléments ; cela est-il possible lorsque, par exemple, on veut évaluer, la compréhension d'un texte ? La compréhension, ou la non compréhension, d'un passage, d'un paragraphe, voire d'un mot, n'influence-t-elle pas la compréhension (ou non) du suivant et de l'ensemble du texte ?
- Comment, dans cette optique, appréhender ce qu'on pourrait considérer comme la compréhension globale du texte, nécessairement (mais pas seulement) fondée sur la compréhension de ses différents constituants ? Peut-on simplement inférer de la somme des points obtenus aux différentes questions à la compréhension du texte ?...

Pour réfléchir à ces questions, nous avons notamment examiné ce qui était considéré comme item dans diverses épreuves, cantonales ou autres. Cet examen nous a conduits à quelques constats qui confirment nos interrogations :

- un item peut fournir des informations de nature et de portée très différentes ;
- la création d'items qui soient véritablement indépendants les uns des autres s'avère très difficile et risque de conduire à privilégier des informations « locales » qui n'ont que peu d'impact sur le sens général du texte ;

¹¹ Le bon sens nous rappelle qu'il ne faut pas mélanger les pommes et les poires...

- la somme d'items disparates qui en découle alors ne suffit pas à fournir une information dépassant l'information ponctuelle fournie par chacun des items et ne permet dès lors pas – en tout cas en l'absence de procédures complémentaires – d'évaluer ce qu'on pourrait considérer comme la compréhension globale du texte¹², compréhension qui constitue selon le PER une finalité importante du travail à réaliser avec les élèves.

Nous avons par conséquent tenté de distinguer entre une conception « cumulative » et une conception « intégrative » de l'item et de réfléchir aux moyens de proposer une structuration cohérente et opératoire des items entre eux.

La première de ces deux conceptions pourrait être schématisée ainsi :

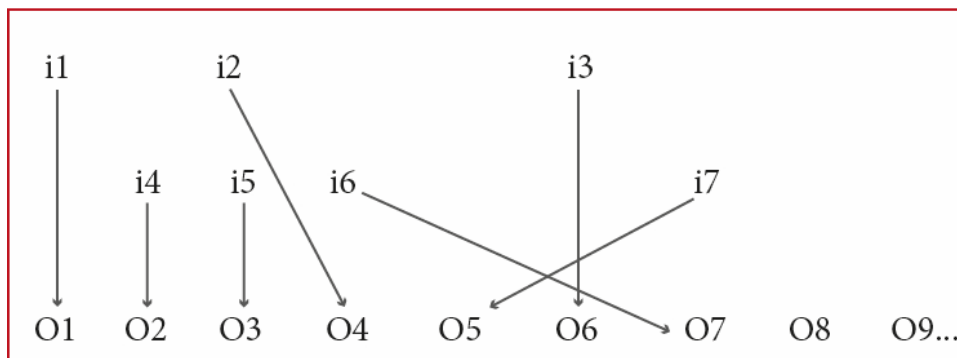


Figure 1. Conception cumulative de l'item (i = item ; O = objectif du PER)

Chaque item, ici, est considéré indépendamment des autres et relié à un seul et unique objectif du plan d'études. Selon la conception intégrative, en revanche, un item peut certes parfois être relié isolément à un objectif, mais il peut aussi être regroupé avec d'autres afin de fournir une nouvelle information – un nouvel item – également reliée à un objectif du plan d'études. « Pire », si on peut dire, un item peut à la fois renvoyer par lui-même à un objectif et entrer dans un regroupement qui renvoie à un autre objectif. On obtient ainsi un système en réseau, certes plus complexe à appréhender mais bien plus riche en termes d'informations fournies :

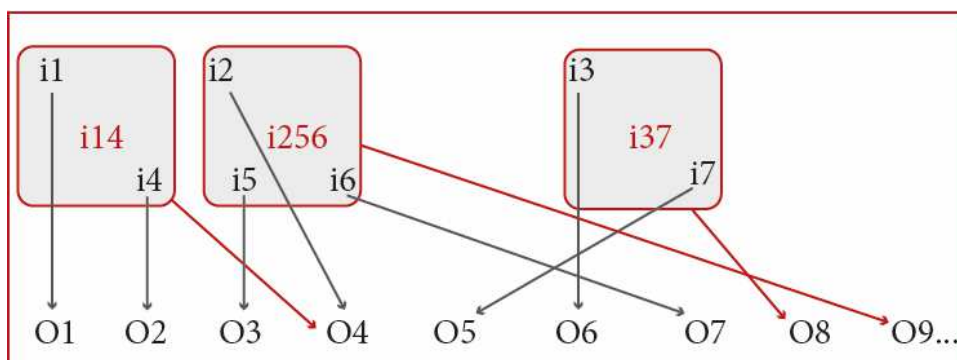


Figure 2. Conception intégrative de l'item

¹² Notons par ailleurs que cette notion de compréhension globale n'est pas non plus sans soulever diverses questions. Dans le PER, elle n'apparaît pas sous cette forme mais dans diverses formulations telles que : « lit un texte de manière autonome, en dégage le *sens général* » (cycle 2), « lit et *comprend un texte complet* de façon autonome », « dégage *les éléments essentiels du contenu* » (cycle 3) [nos italiques]. Selon nous, cette compréhension globale, du « *sens général* », est fondamentale et fait sens pour autant qu'on la réinterprète en fonction du genre de texte : le *sens général* d'un texte argumentatif n'est pas équivalent à celui d'un texte poétique, d'un texte narratif ou d'un texte qui transmet des savoirs.

Ceci peut paraître bien abstrait ! Prenons un exemple, emprunté à une (partie d')épreuve concernant la compréhension orale d'un (extrait !) de récit¹³ :

A. COMPRENDRE L'ORAL (SANS TEXTE) (5 points)		Réservé à l'enseignant-e
1. Réponds en notant dans la case V , F ou ?		
V = Vrai F = Faux ? = L'histoire ne le dit pas.		
a. Le récit se déroule dans une petite ville.	☐	... réponse(s) correcte(s) / 10 A1 .../5 pts
b. Tu as entendu le nom et le prénom des personnages.	☐	
c. Tous les habitants de la ville ont des enfants.	☐	
d. Les habitants repeignent leur maison à intervalles réguliers.	☐	
e. La maison blanche, tout en haut de la rue, a toujours été inhabitée.	☐	
f. Un tremblement de terre a provoqué de très gros dégâts dans la ville.	☐	
g. La première personne qui entre dans la maison blanche après le tremblement de terre est une petite fille.	☐	
h. Le piano mécanique s'est remis à fonctionner grâce au tremblement de terre.	☐	
i. Peu à peu, les gens attirés par les notes de musique entrent dans la maison blanche.	☐	
j. Avant le tremblement de terre, les gens manifestaient vivement leurs préférences et leurs émotions.	☐	
À la fin de l'activité, ce cahier est repris par l'enseignant-e.		

Figure 3. Un exemple d'épreuve de compréhension orale

¹³ Cet exemple est extrait d'une épreuve cantonale pour le cycle 2. Le propos, ici, ne consiste pas à analyser de manière approfondie ces exemples mais bien à faire ressortir quelques problèmes qu'on rencontre dans de nombreuses épreuves. On soulignera d'ailleurs que nous avons, comme on dit, trouvé de tout dans les épreuves examinées, de très bons « items », d'autres bien moins intéressants. Pour des analyses plus approfondies, voir Roth, de Pietro et Sánchez Abchi, 2014 ; de Pietro et Roth, 2016 ; de Pietro, Roth et Sánchez Abchi, 2016.

Dans cette partie d'épreuve, la question « tous les habitants de la ville ont un enfant » (c) permet certes de dire quelque chose en relation à la compréhension locale de certains éléments du texte, mais l'information apportée est peu pertinente pour la construction du sens global du texte. En revanche, le fait de savoir qu'« un gros tremblement de terre a causé d'énormes dégâts dans la ville » (f) non seulement renvoie à une information locale mais participe aussi de la compréhension globale du texte en fournissant une explication à ce qui se passe ensuite dans la ville. Autrement dit, le premier item peut sans problème être envisagé isolément... mais il ne dit (presque) rien de la compréhension globale du texte ; le second, à l'inverse, contribue à la compréhension globale mais ne peut véritablement être isolé, par exemple, de (j)... En outre, on peut facilement constater, à la simple lecture de ces diverses questions, qu'on ne peut guère dire quoi que ce soit à propos du sens global du texte – ce qui, autrement dit, suggère que ces questions ne nous disent à peu près rien de ce qui fait vraiment l'intérêt du récit qui sert de support à l'évaluation.

Notre conception intégrative de l'item soulève toutefois, comme nous l'avons relevé, la question de l'autonomie des items et pose dès lors un problème de validité éducatrice de la mesure. Conscients de cette difficulté à définir la notion d'item de manière à la fois rigoureuse et opératoire, nous réfléchissons à présent à une solution qui permette de dépasser ce dilemme entre les exigences de la mesure et celles de la pertinence didactique. Cardinet décrit ce dilemme ainsi : « la conception encyclopédique de l'enseignement s'accommodait fort bien d'une évaluation par objectifs atomisés : la conception plus récente d'un apprentissage significatif amène à développer une forme d'évaluation qui lui corresponde. Mais il faut veiller à ne pas croiser les modèles, en prétendant mesurer des compétences globales avec des questions à choix multiples, ou établir des niveaux d'aptitudes généralisables à partir de situations non standardisées. On ne ferait plus rien de valable. » (Davaud et Cardinet, 1992, p. 9)

3.2 Comment s'assurer de la validité d'un item et/ou d'une évaluation?

La validité d'une évaluation est le degré d'adéquation entre ce que l'on déclare faire (évaluer telle ou telle dimension) et ce que l'on fait réellement, entre ce que l'outil mesure et ce qu'il prétend mesurer (Laveault et Grégoire, 2002). Elle comporte deux aspects :

- La validité de la mesure : l'outil de mesure utilisé nous permet-il de dire ce qu'on dit ?
- La validité de contenu : évalue-t-on ce qu'on doit / ce que l'on cherche à évaluer ?

Dans nos travaux en cours, nous comptons sur les spécialistes de la mesure pour nous aider à résoudre, par exemple, le problème de l'indépendance des items. Dans notre perspective didactique, c'est avant tout la question de la validité de contenu qui nous intéresse.

A ce propos, comme déjà dit, Cardinet écrivait : « chaque épreuve¹⁴ doit correspondre aux objectifs du curriculum enseigné. » (1990, p. 2) Par conséquent, en nous référant au PER, nous nous sommes demandé quels en sont les éléments qui sont pris – ou qui devraient être pris – en considération dans les épreuves que nous avons examinées. La réponse, comme nous le verrons, n'est pas simple et devrait dépendre notamment, selon nous, de ce qui est visé à travers l'évaluation.

¹⁴ Dans notre contexte, « épreuve » peut être remplacé par « item ».

4. Quoi évaluer ?...

« Pourtant, un système scolaire qui se donne pour tâche de former le plus grand nombre d'élèves à maîtriser des compétences complexes ne peut pas se doter du même cadre d'évaluation qu'un système scolaire qui cherche à répartir les élèves en filières différenciées et à sélectionner les meilleurs pour les écoles qui suivent »

(Davaud et Cardinet, 1992, p. 5)

Là encore, le projet d'élaboration d'épreuves semble clair : ce qui doit être évalué ce sont « les objectifs et progressions » du PER. Mais... Le PER, nous l'avons dit, est un curriculum qui définit ce que les élèves doivent apprendre, un référentiel pour l'enseignement et l'apprentissage – et non pour l'évaluation. Ainsi, par exemple, il ne dit pas si tous les objectifs mentionnés sous la rubrique « Progression des apprentissages » doivent être « atteints », ni à quel degré, ni quand... Pour cela il faut se tourner vers les *attentes* dites *fondamentales* qui, dans le PER, « désignent les apprentissages que chaque élève devrait atteindre au cours, mais au plus tard à la fin d'un cycle. » (PER, *Lexique*, 2010) Toutefois, celles-ci ne sont pas très nombreuses et ne couvrent à l'évidence par l'ensemble des apprentissages « implicitement » attendus...

En outre, selon la Déclaration de 2003, texte fondateur de l'espace romand, le PER constitue l'un des éléments devant permettre à l'école publique d'assurer « la construction de connaissances et l'acquisition de compétences permettant à chacun et chacune de développer ses potentialités de manière optimale. » (CIIP, 2003, 2010) Il est donc question de *compétences* – comme dans nombre de référentiels aujourd'hui – mais aussi de *connaissances*. Dans les différents « Lexique » du PER, ces deux termes sont traités de la manière suivante :

« **Compétence** : Possibilité, pour un individu, de mobiliser un ensemble intégré de ressources en vue d'exercer efficacement une activité considérée généralement comme complexe. » (PER, *Lexique* de la présentation générale, 2010)

Connaissance : pas de définition dans le *Lexique* de la Présentation générale !

« **Connaissances linguistiques** - ensemble des savoirs et savoir-faire concernant le vocabulaire (...), qui permettent de mettre en œuvre les compétences langagières et communicatives. » (PER, *Lexique* du domaine *Langues*, 2010)¹⁵

La disparité est frappante. En fait, les connaissances semblent comme secondaires, là uniquement au service de la mise en œuvre des compétences... Ce seraient donc les compétences qu'il s'agit d'évaluer ? Mais où les trouve-t-on, concrètement, dans le PER ? En fait, excepté dans le *Lexique*, le terme n'est pas utilisé... On ne trouve ni « compétences » ni « connaissances », mais des *visées prioritaires*, des *objectifs d'apprentissage* (déclinés en composantes et organisés selon une progression) et des *attentes fondamentales*.

Il importe par conséquent de se demander ce qui doit, finalement, faire l'objet de l'évaluation ? Uniquement les attentes fondamentales (peu nombreuses) ? Ou également d'autres objectifs ponctuels, extraits des progressions d'apprentissage, mais dont on ne sait quand et à quel degré ils doivent être atteints ? Des compétences plus globales, proches des « visées prioritaires » qui semblent constituer le but ultime de l'enseignement, mais qui ne

¹⁵ On ne trouve pas de définition dans les *Lexiques* des autres domaines du PER (mathématiques et sciences, arts et mouvement, etc.).

sont jamais définies de manière précise ni déclinées en niveaux ? Et on se demandera aussi si les attentes fondamentales correspondent éventuellement à des compétences ?...

Mais on doit surtout se demander *pourquoi* évaluer telle ou telle chose : s'agit-il, ainsi que certains le souhaitent, d'évaluer l'efficacité du système ? Ou « de former le plus grand nombre d'élèves à maîtriser des compétences complexes » ? Ou encore « de répartir les élèves en filières différenciées et [...] sélectionner les meilleurs pour les écoles qui suivent » ? Comme le soulignent avec force Davaud et Cardinet, la forme et l'objet même de l'évaluation, son « cadre », ne peuvent être dissociés de sa visée.

5. Compétences, connaissances et références culturelles

« L'évaluation scolaire est ainsi prise en tenaille entre le désir de précision, qui lui ôte toute signification sociale, et un souci d'ouverture sur la vie, qui lui ôte toute valeur métrique. »

(Cardinet, 1990)

5.1 La compétence, une notion problématique...

Le développement des compétences des élèves semble tout de même constituer la visée principale de l'enseignement et des apprentissages. Toutefois, là encore, on le sait bien aujourd'hui¹⁶, cette notion soulève de nombreuses questions – en particulier dans la perspective d'une évaluation. Ainsi, outre les éléments figurant dans la définition du PER, les très nombreux travaux portant sur cette notion mettent notamment l'accent sur certaines caractéristiques des compétences qu'il apparaît bien difficile de prendre en compte :

- Une compétence est une disposition qui est activée en situation, contextualisée, face à une tâche finalisée. (Allal, 1999 ; Candelier, 2012 ; Pekarek Doehler, 2006 ; etc.)
- Une compétence consiste en la mobilisation d'un ensemble intégré de ressources et non en l'addition de ressources. Ces ressources peuvent être : d'autres compétences, des connaissances (savoirs, savoir-faire...), des attitudes (savoir-être) ... (Le Boterf, 1994 ; Perrenoud, 1999 ; Shavelson, 2012 ; etc.)
- La compétence n'est pas observable en tant que telle ; elle peut être approchée à partir de la performance d'accomplissement de la tâche (Candelier, 2012 ; Shavelson, 2012).
- La « complexité » d'une activité dépend de nombreux facteurs qu'il est difficile de cerner : familiarité avec la tâche, temps à disposition, etc. (Candelier, 2012 ; Rey *et al.*, 2006 ; Marc et Wirthner, 2013).

De fait, la définition du PER néglige plusieurs éléments qui apparaissent généralement comme centraux dans les travaux sur la notion de compétence et qui viennent encore compliquer la situation. En analysant les éléments qu'on trouve dans le *Plan d'études*, il apparaît en fait que, s'il semble relativement aisé d'en identifier certains éléments comme des connaissances et, peut-être, de les évaluer, il s'avère en revanche bien plus délicat d'y localiser des compétences, en particulier si – un peu au-delà de la définition du PER – on veut prendre appui sur ce que nous disent les théories actuelles.

¹⁶ Voir par exemple Crahay, 2005 ; Bronckart et Dolz, 1999 ; Pekarek Doehler, 2006, etc.

5.2. ... et difficile à repérer

Les attentes fondamentales, par exemple, qui semblent constituer des éléments particulièrement importants du PER, ne peuvent être assimilées à des compétences. On y trouve en effet des contenus très disparates, auxquels il serait incohérent d'attribuer un même statut. Voici par exemple trois de ces attentes pour le domaine français :

- 1) « *accorde le participe passé employé avec l'auxiliaire être avec le sujet du verbe* » (PER, attentes fondamentales, L126, p. 65)
- 2) « *repère les différentes parties d'un texte réglant des comportements* » (PER, attentes fondamentales, L121, p. 33)
- 3) « *écrit un texte à visée argumentative incluant l'exposition du sujet ou de la controverse, une opinion et au moins trois arguments* » (PER, attentes fondamentales, L122, p. 37).

(1) se rapproche d'un simple « savoir-faire ». Rien n'exclut de l'évaluer, mais... on n'est en tout cas pas alors en train d'évaluer des compétences : on recueille des informations sur la maîtrise d'un savoir-faire précis mais peu de choses sur ce qu'on pourrait considérer comme la compétence de production textuelle (« Écrire des textes de genres différents adaptés aux situations d'énonciation... », PER : visée prioritaire L132), voire sur « la compétence en français » des élèves.

On peut très bien décider d'évaluer des objets de statut différent mais il faut réfléchir à ce que signifie chaque prise d'information ainsi réalisée et, le cas échéant, à la manière d'articuler ces différentes prises d'informations si l'on souhaite aboutir à une information plus large sur les élèves. Autrement dit, il importe de ne pas tout mélanger en aplatissant les différences : il importe de savoir ce qu'on évalue, comment et dans quel but.

Pour l'attente (3), qui semble se rapprocher davantage de ce qu'on pourrait considérer comme une compétence, la situation est à l'évidence bien plus compliquée que pour les connaissances ou les savoir-faire. On constate en fait que :

- pour une telle activité « complexe », on peut généralement mobiliser des ressources très diverses et celles-ci ne sont donc guère prédictibles ;¹⁷
- les stratégies (diverses) et/ou motivations qui guident les choix dans le processus de mobilisation ne sont guère accessibles à l'observation ;
- une activité complexe est toujours, en partie au moins, nouvelle et elle ne correspond dès lors qu'en partie à des choses qui ont été « enseignées » – ce qui pourrait aller à l'encontre du « contrat didactique » qui, pour assurer l'équité de l'évaluation, veut qu'on n'évalue que des choses qui ont été enseignées...

Comme le relève Jean Cardinet, l'évaluation dans un cadre scolaire est confrontée à des injonctions contradictoires. En milieu professionnel, lorsqu'il est question de qualifications, ou encore lorsqu'il est question de donner un permis de conduire qui « certifie » la compétence à conduire, il semble – au moins partiellement – possible d'atténuer ces contradictions, sous deux conditions :

¹⁷ On peut en outre se demander, dans la formulation, pourquoi on parle de *trois* arguments... Et, pour illustrer la diversité potentielle des ressources mobilisables, on pourrait prendre ici quelques exemples « absurdes », du genre « demander à quelqu'un de plus compétent de faire l'activité » – ce qui s'appelle « déléguer » et est considéré comme une qualité dans l'entreprise: on dit bien de quelqu'un qu'il *sait déléguer*... D'une certaine manière, comme le dit la définition du PER, c'est l'efficacité qui compte !

- proposer au candidat une « famille de tâches », en contexte, qui *toutes* ont été exercées mais dans des contextes parfois différents ;
- accepter une certaine subjectivité du jugement évaluatif.

Cela est-il possible dans le cadre scolaire ? Ce n'est pas impossible pour autant qu'on ne se contente pas d'une seule tâche, qu'on diversifie et contextualise ces tâches et... qu'on accepte cette part de subjectivité qui va à l'encontre des exigences d'objectivité que, souvent, les autorités attendent de l'évaluation.

Cependant, dans le cadre d'une formation générale, il s'avère bien plus difficile de circonscrire de manière raisonnable la famille des tâches constitutives de la compétence visée – et à évaluer – chez les élèves ainsi que, *a fortiori*, l'ensemble des ressources qui devraient être mobilisables... Par exemple, si l'on considère l'objectif suivant qui, lui aussi, semble se rapprocher d'une compétence : « ...*lit un texte de manière autonome, en dégage le sens général et le reformule oralement* » (L1 21, p. 23 ; Objectif d'apprentissage pour la compréhension écrite), il paraît bien difficile de définir précisément ce qu'on attend : de quels textes parle-t-on ?¹⁸ Dans quel but le lit-on ? Dans quel contexte ? Avec quelles aides possibles (dictionnaires, experts...) ?...

5.3 Une conception « scolaire » de la compétence

Conscients que l'école ne peut pas se désintéresser des compétences, nous préconisons néanmoins une certaine prudence :

- en acceptant que, pour être évaluables, les « compétences » sont réinterprétées, adaptées, pour la forme scolaire et qu'elles se rapprochent autant de ressources – des praxéologies liant savoirs et savoir-faire (Chevallard, 1997) – de haut niveau que de véritables compétences dans le sens où le monde professionnel – responsable de la vogue de cette notion – les conçoit ;
- en acceptant la part de subjectivité qu'une telle évaluation implique nécessairement : l'évaluation, ici, ressemble plus à une inférence fondée sur un faisceau d'indices qu'à une mesure garantissant sa « valeur métrique ». Comme le soulignent Davaud et Cardinet : « On est bien loin d'une évaluation pointilliste, mesurant la maîtrise de connaissances ponctuelles. Cela conduit à renoncer à la mesure quantitative, qui véhicule une illusion de justesse, pour explorer des situations d'évaluation plus globales. » (1992, p. 7)

L'évaluation de compétences supposerait dès lors :

1. de s'appuyer sur une relative diversité de tâches « comparables » ;
2. d'identifier les ressources mobilisées par les élèves qui sont rendues visibles par l'accomplissement de la tâche – mais cela sans jamais pouvoir prédire totalement les ressources qui sont censées être mobilisées !
3. de fonder en premier lieu l'évaluation sur un jugement globalisant, autrement dit en répondant d'abord à la question « est-ce que la tâche a été effectuée avec succès ? », puis seulement sur une évaluation de la *manière* dont la tâche a été accomplie.¹⁹

¹⁸ Cette question recouvre déjà en elle-même de nombreuses autres questions : des textes relevant de quels genres ? De quels domaines thématiques ? De quel degré de complexité ?...

¹⁹ On rejoint en fait ici la distinction opérée par Searle (2009) dans sa théorie des actes de langage entre des règles *constitutives* – sans la satisfaction desquelles l'acte n'est tout simplement pas accompli – et des règles *normatives* qui portent sur la manière de les accomplir, mais ne sont pas déterminantes au même titre.

Par exemple, pour ce qui concerne la compétence de compréhension d'un texte, on devrait ainsi considérer que c'est la compréhension du « sens général » de textes diversifiés, relevant de genres différents, qui exprimerait en premier lieu la compétence – en se rappelant, de plus, que ce sens général n'existe qu'en lien au genre dont relève le texte et à une tâche qui oriente la lecture.²⁰ Une telle conception implique donc que l'évaluation inclue des questions qui permettent de dire quelque chose de la compréhension du sens général et qu'une compétence ne peut être considérée comme une simple addition de ressources.

5.4 Et qu'en est-il des connaissances ?

Suivant en cela la *Déclaration* de la CIIP (2003), nous considérons que l'évaluation ne porte pas uniquement sur des compétences mais également sur des connaissances. Il importe par conséquent de définir – dans la perspective de l'évaluation – les autres « objets » également présents dans le PER et qui doivent, d'une manière ou d'une autre, être pris en compte.

Les connaissances – qu'il s'agisse de savoirs, de savoir-faire, de praxéologies – peuvent dans une certaine mesure être évaluées de manière (partiellement) décontextualisée. On attend d'elles – celles qui ont été retenues comme nécessaires par le PER – qu'elles soient disponibles, mobilisables, en tout temps, dans n'importe quel contexte (ou presque...).

Plus précisément, les connaissances nous paraissent en fait de deux types :

- Certaines ont un statut en elles-mêmes dans la mesure où elles correspondent à des savoirs ou savoir-faire considérés – par la société, par l'école... – comme faisant partie d'une culture commune attendue. On parlera alors d'un statut de **référence culturelle**.²¹
- D'autres ne sont en quelque sorte que des **ressources** au service de compétences, qu'on peut exercer afin de les rendre mobilisables et dont on peut (parfois) observer la mobilisation dans l'accomplissement d'une tâche.

Dès lors, la fonction assignée à l'évaluation devrait influencer sur le choix des objets du PER évaluables / à évaluer. « Idéalement », une évaluation *sommative* (voire certificative) devrait porter sur des choses enseignées : ressources (dont les références culturelles) et compétences envisagées – avec prudence – dans une perspective scolaire ; enfin, pour ce qui est de l'évaluation *formative* – qui doit servir à soutenir les apprentissages –, elle peut porter sur n'importe quel objet, mais ne devrait pas donner lieu à une note. En tous les cas, il importe donc de bien définir pourquoi on évalue et distinguer clairement entre évaluation formative et évaluation sommative (ou certificative).

²⁰ On relèvera cependant que, dans certains cas particuliers (récit, poème...), la compréhension du texte peut constituer en elle-même la finalité de la tâche. Mais il importe de veiller au risque de réductionnisme scolaire dans ces cas où la tâche n'est envisagée que pour elle-même...

²¹ Dans la perspective d'une évaluation sommative, en particulier, cela constitue une tâche *socialement* délicate de définir quelles sont les connaissances qui constituent des références culturelles – par exemple pour le domaine du vocabulaire – et qui, dès lors, peuvent être évaluées pour elles-mêmes.

6. Des objets aux types de questionnement

« On voit bien là que le savoir ne dépend pas seulement de l'élève. (...) Le savoir est lié à la fois à l'élève et au mode de questionnement. Il n'y a pas de niveau de connaissance absolu. »

(Cardinet, 1991, p. 16)

Il importe donc, nous l'avons vu, de distinguer les objets du PER selon leur statut. Tous peuvent faire l'objet d'une évaluation mais... pas de la même manière, pas avec les mêmes types de tâches et formats de questionnement, pas avec les mêmes finalités. Les objets à évaluer et les tâches qui permettent de les rendre visibles sont liés. L'enjeu consiste dès lors à proposer des modalités d'évaluation qui permettent de prendre des informations pertinentes et valides en fonction des types d'objets du PER et de la fonction assignée à l'évaluation.

Nous considérons (au moins) quatre « types de tâches » qui permettent selon nous des prises d'informations portant sur des objets (partiellement) différents :

- des *tâches complexes* (finalisées, contextualisées...) à privilégier pour recueillir des informations concernant des compétences (et, pour autant qu'on puisse en observer la mobilisation, les ressources effectivement mobilisées) ;
- des *problèmes* qui correspondent à des tâches d'une certaine complexité, (a) peu (voire pas du tout) contextualisées et finalisées, (b) leur résolution dépend de la mise en œuvre de ressources prédictibles (savoir-faire) qui ont fait l'objet d'un apprentissage ;²²
- des *exercices* qui sont des tâches « simples », décontextualisées et non finalisées, portant sur la maîtrise de ressources (des savoir-faire) délimitées, ayant fait l'objet d'un apprentissage ;
- des *quiz* qui sont également des tâches « simples », décontextualisées et non finalisées, portant sur la maîtrise de ressources délimitées (des savoirs cette fois), ayant fait l'objet d'un apprentissage.

Pour les raisons que nous avons développées précédemment, les frontières, en contexte scolaire, ne sont pas toutes absolument claires, en particulier entre ce qu'on considèrera comme une tâche complexe et comme un problème : où commencera-t-on à parler d'évaluation de compétences ? Où parlera-t-on plutôt d'évaluation de ressources ?...

7. Vers une validité didactique...

« Se poser des questions sur la forme à donner aux épreuves communes, c'est donc beaucoup plus qu'un problème technique. »

(Davaud et Cardinet, 1992, p.9)

Malgré les nombreux problèmes qu'elle soulève – tant d'un point de vue didactique et éducatif qu'en lien à ses finalités sociales –, l'évaluation, sous ses différentes formes, représente un aspect constitutif de tout apprentissage et il est donc indispensable de l'aborder

²² Le problème, dans cette conception, représente en quelque sorte le pendant pour l'évaluation de la situation-problème lorsqu'il est question d'apprentissage.

avec tout le sérieux nécessaire. C'est pourquoi il ne nous paraît guère possible d'élaborer des épreuves communes romandes sans avoir, au préalable, sinon résolu les difficultés rencontrées du moins arrêté une position claire et scientifiquement étayée par rapport à ses finalités, par rapport aux objets sur lesquels elle doit porter et – ainsi que nous y invite la citation ci-dessus – à la forme à leur donner. La recherche, ainsi que l'a toujours soutenu J. Cardinet, a donc un rôle important à jouer dans l'élaboration de procédures évaluatives valides, et qui soient avant tout au service des élèves et de leurs apprentissages.

C'est dans cette perspective que – en adoptant une approche positive mais critique – nous travaillons actuellement à l'élaboration d'une modélisation de l'évaluation qui devra, au final, mettre en relation des fonctions évaluatives, des objets du PER évaluables / à évaluer, des types d'activités qui les rendent visibles et des formats de questionnement qui en assurent la fiabilité éducatrice.

Ce travail nous conduit à certains « compromis » raisonnables et conscients, par exemple en adoptant une conception réduite, « scolaire », de la notion de compétence. Il nous permet aussi d'affirmer avec force certaines conceptions concernant notamment la nécessité de clarifier le statut des « objets » figurant dans le PER et d'être au clair sur ce qu'on veut évaluer, les liens entre fonctions évaluatives et modalités d'évaluation ou la distinction entre connaissances, ressources et références culturelles. Nous sommes optimistes, au final, que ce travail nous permette de proposer, au niveau romand, une approche de l'évaluation qui lui garantisse une certaine validité didactique et d'apporter ainsi une contribution de la recherche à la création d'une banque d'items opératoire qui soit – en premier lieu – au service des enseignants et... des élèves.

En approfondissant cette réflexion dans le cadre de ce numéro d'hommage, et au vu de nos lectures reflétées dans les citations choisies, nous avons bien le sentiment d'avoir finalement suivi ainsi la voie magnifiquement tracée depuis – au moins – la naissance de l'IRDP par Jean Cardinet.

8. Références

- Allal, L. (1999). Acquisition et évaluation de compétences en situation. In J. Dolz, & E. Ollagnier (éds), *L'énigme de la compétence en éducation* (pp. 77-95). Bruxelles : De Boeck Université.
- Antonietti, J.-P. (éd.). (2005). Évaluation des compétences en mathématiques en fin de 4e année primaire : résultats de la seconde phase de l'enquête MATHÉVAL. Neuchâtel : IRDP.
- Armi, F., & Pagnossin, E. (éds). (2012). *École et protection des données personnelles*. Actes de la journée d'étude du 16 novembre 2011. Neuchâtel : IRDP.
- Behrens, M. (2010). Standards ou les escaliers mouvants de Poudlard. In P. Gilliéron Giroud, & L. Ntamakiliro (éds), *Réformer l'évaluation scolaire : mission impossible ?* (pp. 113-138). Berne : P. Lang.
- Bronckart, J.-P., & Dolz, J. (1999). La notion de compétence : quelle pertinence pour l'étude de l'apprentissage des actions langagières ? In J. Dolz, & E. Ollagnier (éds), *L'énigme de la compétence en éducation* (pp. 27-44). Bruxelles : De Boeck Université.
- Calame, J.-A., et al. (1995). *Math 5-6... pas si mal ! : évaluation des moyens romands d'enseignement de mathématique en 5ème et 6ème années*. Neuchâtel : IRDP.
- Candelier, M. (éd.). (2012). *Le CARAP : un cadre de référence pour les approches plurielles des langues et des cultures : compétences et ressources = FREP: a framework of reference for pluralistic approaches to languages and cultures: competences and resources*. Graz : Centre européen pour les langues vivantes (CELV).
- Cardinet, J. (1977a). *Objectifs éducatifs et évaluation individualisée*. Neuchâtel : IRDP. Repris dans Cardinet, 1986, 63-117.
- Cardinet, J. (1977b). *Objectifs pédagogiques et fonctions de l'évaluation*. Neuchâtel : IRDP.

- Cardinet, J. (1977c) *L'évaluation du nouvel enseignement de la mathématique en première année primaire : présentation de l'expérience romande*. Neuchâtel : IRDP.
- Cardinet, J. (1979). *Les deux visées de l'évaluation formative*. Neuchâtel : IRDP. Repris dans Cardinet, 1986, 197-223.
- Cardinet, J. (1981). *La cohérence nécessaire dans le choix des procédures d'évaluation scolaire*. Neuchâtel : IRDP.
- Cardinet, J. (1986). *Évaluation scolaire et mesure*. Bruxelles : De Boeck.
- Cardinet, J. (1988). *L'appréciation du travail des élèves face à la critique scientifique*. Neuchâtel : IRDP.
- Cardinet, J. (1990). *Remettre le quantitatif à sa place en évaluation scolaire*. Neuchâtel : IRDP.
- Cardinet, J. (1991). *L'histoire de l'évaluation scolaire des origines à demain*. Neuchâtel : IRDP.
- Cardinet, J., & Tourneur, Y. (1978). *Le calcul de marges d'erreurs dans la théorie de la généralisabilité*. Neuchâtel : IRDP.
- Cardinet, J., & Weiss, J. (1978). *L'observation interactive, au confluent de la formation et de la recherche*. Neuchâtel : IRDP.
- Cardinet, J., & Tourneur, Y. (1985). *Assurer la mesure : guide pour les études de généralisabilité*. Berne : P. Lang.
- Chevallard, Y. (1997). Les savoirs enseignés et leurs formes scolaires de transmission : un point de vue didactique. *Skholé*, 7, 45-64.
- Conférence intercantonale de l'Instruction publique de la Suisse romande et du Tessin (CIIP). (2003). *Déclaration de la CIIP relative à la politique de l'enseignement des langues en Suisse romande du 30 janvier 2003*. Neuchâtel : CIIP. Accès : <http://www.ciip.ch/FileDownload/Get/150> (consulté le 24.11.2017).
- Conférence intercantonale de l'Instruction publique de la Suisse romande et du Tessin (CIIP). (2010). *Plan d'études romand (PER)*. Neuchâtel : CIIP [Site Web]. Accès : <http://www.plandetudes.ch/> (consulté le 05.12.2017).
- Crahay, M. (2005). Dangers, incertitudes et incomplétude de la logique de la compétence en éducation. *Les cahiers du Service de pédagogie expérimentale*, 21/22, 5-40.
- Davaud, C., & Cardinet, J. (1992). *Quelles épreuves communes voulons-nous ?* Genève : CRPP.
- De Pietro, J.-F. (2015). Formats d'activités et statut des objets à enseigner/évaluer. *Formation et pratiques d'enseignement en questions*, 19, 43-53.
- De Pietro, J.-F., Pfeiffer Ryter, V., Wirthner, M., Béguin, M., Broi, A.-M., Clément, S., Matei, A., & Roos, E. (2009). *Évaluation du moyen d'enseignement "S'exprimer en français" : rapport final*. Neuchâtel : IRDP.
- De Pietro, J.-F., & Roth, M. (2016). Pour une évaluation sous contrôle didactique : qu'est-ce qui permet de dire qu'un élève a « compris » un texte oral ? *Babylonia*, 2, 19-23.
- De Pietro, J.-F., Roth, M., & Sánchez Abchi, V. (2016). Vers un référentiel pour l'évaluation : choix politiques et enjeux didactiques. In É. Falardeau, P. Lefrançois, J.-L. Dumortier, & J. Dolz (dirs), *L'évaluation en classe de français, outil didactique et politique* (pp. 111-132). Namur : Presses universitaires de Namur.
- Dierendonck, C. (2008). *Comment les évaluations externes des acquis des élèves sont-elles perçues par les enseignants du primaire dans les cantons de Neuchâtel, Vaud et Fribourg ? : enquête exploratoire*. Neuchâtel : IRDP.
- Eboulet, É., & Matei, A. (2013). *Application de la régression quantile et ses tests statistiques sur les données PISA en Suisse romande*. Neuchâtel : IRDP.
- Elmiger, D., & Singh, L. (2014). *Anglais dès la 7e : phase pilote More! 8e : évaluation des expériences de la deuxième année (2012-2013)*. Neuchâtel : IRDP.
- Frakowiak, P. (2005). *L'évaluationnisme, le malheur de l'école*. Accès : <https://www.charmeux.fr/evaluer.html> (consulté le 25.09.2017).
- Genoud, P. (1999). *Observation de la mise à l'épreuve des nouveaux moyens d'enseignement romands de mathématiques : évaluation collective de quelques connaissances et aptitudes des élèves en 4P*. Neuchâtel : IRDP.
- Laveault, D. (2017). Pour une évaluation scolaire source de progrès et d'innovation : contribution de Jean Cardinet. *Évaluer : journal international de recherche en éducation et formation*, 3(3).
- Laveault, D., & Grégoire, J. (2002, 2e éd.). *Introduction aux théories des tests en psychologie et en sciences de l'éducation*. Bruxelles : De Boeck.

- Le Boterf, G. (1994). *De la compétence : essai sur un attracteur étrange*. Paris : Les Éds d'Organisation.
- Marc, V., & Wirthner, M. (2012). *Épreuves romandes communes : de l'analyse des épreuves cantonales à un modèle d'évaluation adapté au PER : rapport final du projet EpRoCom*. Neuchâtel : IRDP.
- Marc, V., & Wirthner, M. (2013). *Développement d'un modèle d'évaluation adapté au PER : rapport scientifique du projet d'épreuves romandes communes*. Neuchâtel : IRDP.
- Mottier Lopez, L., & Tessaro, W. (éds). (2016). *Le jugement professionnel, au cœur de l'évaluation et de la régulation des apprentissages*. Berne : P. Lang.
- Pekarek Doehler, S. (2006). Compétence et langage en action. *Bulletin suisse de linguistique appliquée VALS/ASLA*, 84, 9-45.
- Perrenoud, P. (1999). D'une métaphore à l'autre : transférer ou mobiliser ses connaissances ? In J. Dolz & E. Ollagnier (éds), *L'énigme de la compétence en éducation* (pp. 45-60). Bruxelles : De Boeck Université.
- Pochon, L.-O. (1979). *Enquête romande auprès du corps enseignant de quatrième année primaire sur l'enseignement de la mathématique : présentation des résultats*. Neuchâtel : IRDP.
- Pochon, L.-O. (2006). *De la possibilité d'usage d'ontologies pour la gestion des contenus mathématiques*. Neuchâtel : IRDP.
- Rey, B., Carette, V., Defrance, A., & Kahn, S. (2006). *Les compétences à l'école : apprentissage et évaluation*. Bruxelles : De Boeck.
- Roth, M., & De Pietro, J.-F. (éds). (2013). *Quoi et comment évaluer en référence au Plan d'études romand (PER) : premières pistes : rapport scientifique intermédiaire*. Neuchâtel : IRDP (document interne, non publié).
- Roth, M., De Pietro, J.-F., & Sánchez Abchi, V. (2014). Comment évaluer la lecture en référence au Plan d'études romand ? *Forum lecture*, 3, 1-25.
- Sánchez Abchi, V., De Pietro, J.-F., & Roth, M. (2016). *Évaluer en français : comment prendre en compte la difficulté des items et des textes*. Neuchâtel : IRDP.
- Searle, J.R. (2009). *Les actes de langage : essai de philosophie du langage*. Paris : Hermann.
- Shavelson, R.J. (2012). An approach to testing & modeling competence. In S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn, & J. Fege (eds), *Modeling and measuring competencies in higher education: tasks and challenges* (pp. 19-33). Rotterdam : Sense Publishers.
- Vermot, B., Behrens, M., & Marc, V. (2011). Étude comparative de deux tests assistés par ordinateur en vue de l'élaboration d'épreuves communes en Suisse romande. In J.-G. Blais, & J.-L. Gilles (dirs), *Évaluation des apprentissages et technologies de l'information et de la communication : le futur à notre porte* (pp. 225-244). Québec : Presses de l'Université Laval.
- Weiss, J. (1992). *Vers une évaluation interactive à l'école*. Neuchâtel : IRDP.
- Weiss, J. (1995). *Une évaluation scolaire pour demain : une évaluation pour apprendre et se former*. Neuchâtel : IRDP.
- Weiss, J. (1998). *Une évaluation informative pour la régulation des apprentissages et des formations*. Neuchâtel : IRDP.
- Weiss, J. (2003). Une évaluation informative pour la régulation des apprentissages et des formations. In M. Laurier (dir.), *Évaluation et communication : de l'évaluation formative à l'évaluation informative* (pp. 239-266). Outremont : Quebecor.
- Weiss, J. (dir.). (1991). *L'évaluation : problème de communication : congrès de l'Association pour le développement des méthodologies d'évaluation en éducation (ADMEE)*, Neuchâtel, 24-26 septembre 1990. Cousset : DelVal ; Neuchâtel : IRDP.
- Weiss, J., De Pietro, J.-F., Jaquet, F., Pochon, L.-O., & Wirthner, M. (2001). L'Institut romand de recherche et de documentation pédagogique : trente ans de recherche en éducation. In J. Weiss (dir.) et al., *L'épopée des centres de recherche en éducation en Suisse 1960-2000* (pp. 31-52). Neuchâtel : IRDP.
- Wirthner, M. & Weiss, J. (1981). *Appréciation des moyens d'enseignement de français de 2e année primaire*. Neuchâtel : IRDP.