



Journée d'étude en hommage à Jean Cardinet

Lausanne, 10 novembre 2016

Comment définir la validité d'une évaluation ?

Quelques réflexions autour de futures épreuves romandes

Jean-François de Pietro & Murielle Roth

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

(J. Cardinet, 1991, 19)

Plan de la présentation

1. Le contexte de nos travaux
2. Le « mandat » de l'IRDP : du projet EpRoCom au *Masterplan 2016*
3. Des « items » à « valider »...
4. Quoi évaluer ?...
5. Compétences, connaissances et références culturelles
6. Vers une validité didactique...

1. Le contexte de nos travaux

Une longue tradition de recherche, dans 3 directions principales

- Les fondements de l'évaluation (questions de finalités, de validité, etc.)

- La fiabilité de la mesure

- Des évaluations ponctuelles ciblées (enseignement rénové du français, mathématiques, moyens d'enseignement...)

1. Le contexte de nos travaux

Un changement de perspective...

Un changement aux causes multiples

- climat politique (*cf.* polémiques à propos des notes...),
- impact du néo-libéralisme (ou : de la recherche de la validité à la mesure de l'efficacité et de l'efficacité)
- (re)prise en main politique... → Concordat HarmoS, Convention scolaire romande, PER...

Convention scolaire romande – ch.2, sect.2, art. 15, al.1 et 2 (<http://www.ciip.ch/documents/showFile.asp?ID=2518>)

*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.

1. Le contexte de nos travaux / Un changement de perspective...

→ Repositionnement de la place et du rôle de la recherche

Cf. Mandat de prestations attribué pour les années 2016 – 2019 à l'Institut de recherche et de documentation pédagogique (IRDP)

B : Evaluation du système, épreuves de référence

Art. 6 Epreuves romandes communes pour la scolarité obligatoire

1 L'IRDP 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. 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.

Un changement de perspective... problématique

Quelques observations

- Une multiplication des dispositifs de mesure : épreuves internationales (PISA), tests de référence nationaux (HarmoS), régionaux (EpRoCom), cantonales (épreuves communes), travaux en classe pour l'attribution des notes et des moyennes...
- Une évaluation aux finalités floues:
 - Fournir des informations utiles pour le pilotage du système? **Enjeu : faire évoluer le PER sur la base de résultats globaux (échantillon) concernant l'atteinte des objectifs du PER**
 - Fournir des informations utiles à chaque canton (remplacer les épreuves cantonales)? **Enjeu : permettre de situer chaque élève (orientation, sélection) sur la base du PER**
- ➔ Quelle place pour les EpRoCom? ➔ **Pas de réponse claire, divergences politiques...**

Intermède 1

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

(J. Cardinet, 1991)

2. Le « mandat » de l'IRDP : du projet EpRoCom au *Masterplan* 2016 **Planification et... réajustement du projet**

- En 2016 : modalités de réalisation du projet précisées dans le cadre d'un *Masterplan* → Réajustement du projet:

1. 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

→ **De nombreuses questions... non résolues, en lien notamment avec la recherche d'une *validité* didactique.**

3. Des « items » à « valider »... Mais comment définir la validité d'un item, d'une évaluation?

Deux aspects du critère de validité :

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

→ Qu'est-ce qu'un « item » dans les épreuves et autres évaluations que nous avons examinées ?

Quelques exemples d' « items »

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.

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.

... réponse(s)
correcte(s)/10

A1 .../5 pts

Quelques exemples d' « items »

Marco Polo, *Le devisement du monde* ou *Le livre des merveilles*.



Comprenons le texte ensemble

- 1 Où se passe ce récit ?
À quelle époque ?
- 2 Qui est le narrateur ?
- 3 Qui est le Grand Khan ?
Quel autre nom lui donne-t-on ?
- 4 Combien de palais le Grand Khan a-t-il
à Ciandu ? En a-t-il ailleurs ? Qu'est-ce qui
te permet de l'affirmer ?
- 5 Qu'est-ce qui montre la richesse
et la puissance du Grand Khan ?
- 6 Pourquoi ce récit de voyage a-t-il pour
titre *Le livre des merveilles* ? Retrouve le mot
« merveille » dans le texte. Que désigne-t-il ?
- 7 Fais la liste de ce qui étonne le narrateur.
- 8 Qu'est-ce qui te semble aujourd'hui
le plus surprenant dans ce récit ?

Quelques exemples d' « items »

- D1. Dans l'extrait suivant (lignes 19 à 24), souligne tous les groupes nominaux et tous les pronoms que l'auteur utilise pour désigner la danseuse.

La danseuse était vraiment belle dans sa position gracieuse : une jambe lancée en avant et le bras à peine arrondi vers le haut. Elle était très élégante dans son tutu de tulle un peu jauni par le temps.

Malheureusement, cette belle créature ne remarquait pas le nouveau venu. Elle regardait de l'autre côté et elle ne se retournait jamais, même pas la nuit, quand tous les autres jouets s'animaient. Le soldat avait essayé une fois de lui adresser la parole [...].

Quelques exemples d' « items »

B3. Réponds aux questions suivantes.

a. Quel temps verbal est le plus fréquemment utilisé aux lignes 1 à 18? → _____

Pourquoi l'auteur emploie-t-il ce temps verbal? → _____

b. Quel temps verbal est le plus fréquemment utilisé aux lignes 33 à 49? → _____

Pourquoi l'auteur emploie-t-il ce temps verbal? → _____

Et quelques questions récurrentes à propos de tels exemples

- Qu'est-ce qu'on évalue vraiment ? Comment ?
- Qu'est-ce, finalement, que « comprendre un texte » ? Est-ce l'addition de la compréhension des diverses informations qu'il comporte ?
- Est-ce que cette notion conserve le même sens lorsqu'il s'agit d'un récit, d'un texte documentaire, d'un poème, d'une consigne ?
- Sur quelles bases peut-on finalement porter un jugement sur les compétences et connaissances d'un élève dans le domaine de la compréhension écrite (CE) ?

Qu'est-ce qu'un « item » ?

Quelques constats:

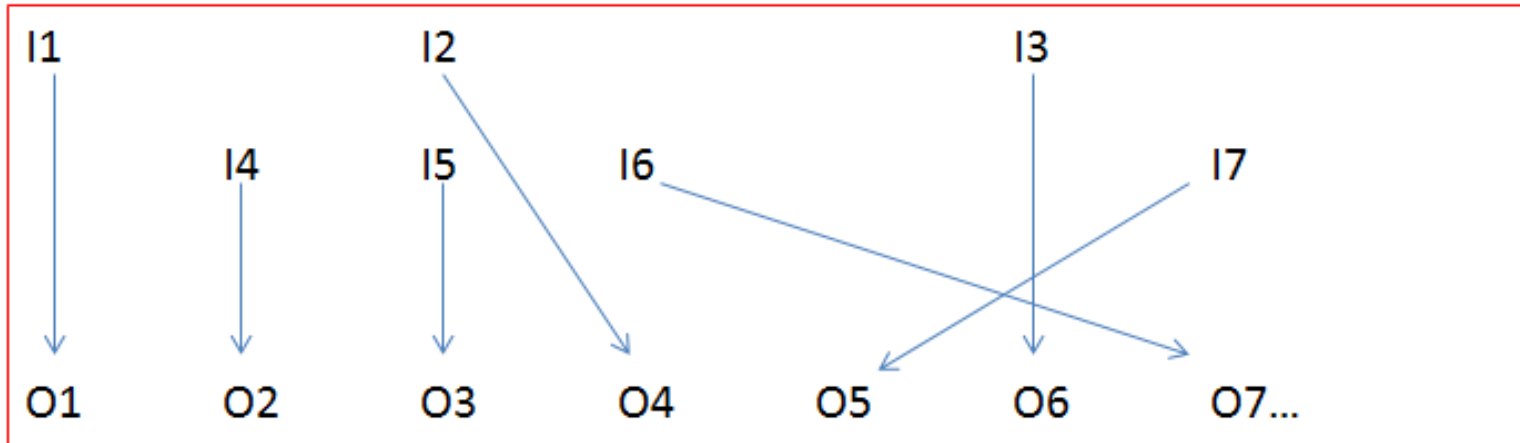
- Un item peut donner des informations de nature et de portée très différentes.
- La somme d'items disparates ne suffit pas à fournir une information globale, dépassant l'information ponctuelle fournie par chacun des items (conception « cumulative »).
- Nécessité d'une structuration cohérente des items entre eux (conception « intégrative »).

« La standardisation des épreuves privilégie l'objectivité de la correction, en négligeant le caractère arbitraire du choix des questions. Leur forme conditionne et limite les activités scolaires proposées aux élèves. »

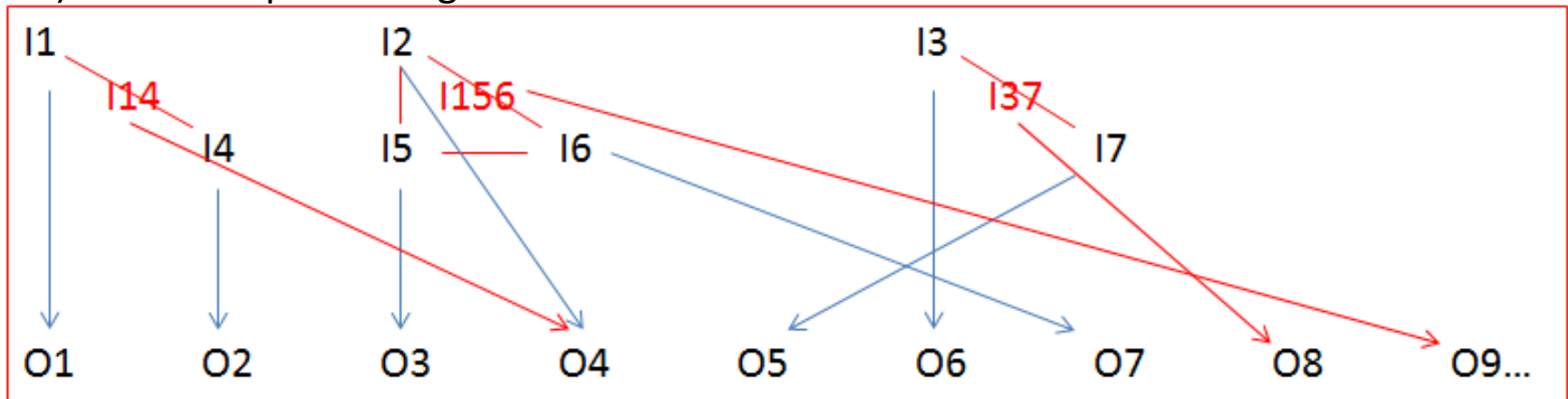
(Davaud & Cardinet, 1992, 7)

Qu'est-ce qu'un « item » ?

a) Une conception cumulative



b) Une conception intégrative



3. Des « items » à « valider »... Mais comment définir la validité d'un item, d'une évaluation?

Comme le dit J. Cardinet, à propos de la validité de contenu, «chaque épreuve [item ?] doit correspondre aux objectifs du curriculum enseigné.» (1990, 2)

- A quoi les items (voire les regroupements d'items) renvoient-ils dans une évaluation « fondée sur les objectifs et progressions du PER » ?**
- Qu'est-ce que cela nous dit des *connaissances* et *compétences* des élèves ?**

4. Quoi évaluer?... Une analyse du PER

Le PER est un curriculum qui définit ce que les élèves doivent apprendre, un référentiel pour l'enseignement-apprentissage – et non pour l'évaluation.

Il 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 des potentialités de manière optimale.» (CIIP, 2003 et 2010)

➔ Qu'est-ce qui doit, finalement, faire l'objet de l'évaluation? Des objectifs ponctuels? Les compétences et/ou les connaissances ? Plus généralement, quels « objets » trouve-t-on dans le PER?...

Quoi évaluer? Une analyse du PER

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*, 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. (*Lexique* du domaine Langues)

Quoi évaluer? Une analyse du PER

		Axes thématiques									
		Espace	Nombres	Opérations	Grandeurs et mesures	Modélisation	Phénomènes naturels et techniques	Corps humain	Diversité du vivant		
Cycle	Premier cycle	MSN 11 Explorer l'espace... Mathématiques	MSN 12 Poser et résoudre des problèmes pour construire et structurer des représentations des nombres naturels... Mathématiques	MSN 13 Résoudre des problèmes additifs. Mathématiques	MSN 14 Comprendre et saisir des grandeurs... Mathématiques	MSN 15 Représenter des phénomènes naturels, techniques ou des situations mathématiques. Mathématiques / Sciences de la nature	MSN 16 Explorer des phénomènes naturels et des technologies... Sciences de la nature	MSN 17 Construire son schéma corporel pour tenir compte de ses besoins... Sciences de la nature	MSN 18 Explorer l'univers et la diversité du vivant... Sciences de la nature		
	Deuxième cycle	MSN 21 Poser et résoudre des problèmes pour structurer le plan et l'espace... Mathématiques	MSN 22 Poser et résoudre des problèmes pour construire et structurer des représentations des nombres relatifs... Mathématiques	MSN 23 Résoudre des problèmes additifs et multiplicatifs... Mathématiques	MSN 24 Utiliser la mesure pour comparer des grandeurs... Mathématiques	MSN 25 Représenter des phénomènes naturels, techniques, sociaux ou des situations mathématiques. Mathématiques / Sciences de la nature	MSN 26 Explorer des phénomènes naturels et des technologies à l'aide de démarches caractéristiques des sciences expérimentales. Sciences de la nature	MSN 27 Identifier les différentes parties de son corps, en décrire le fonctionnement et en tirer des conséquences pour la pérennité de la vie... Sciences de la nature	MSN 28 Déterminer des caractéristiques du monde vivant et de divers milieux et en tirer des conséquences pour la pérennité de la vie... Sciences de la nature		
	Troisième cycle	MSN 31 Poser et résoudre des problèmes pour modéliser le plan et l'espace... Mathématiques	MSN 32 Poser et résoudre des problèmes pour construire et structurer des représentations des nombres réels... Mathématiques	MSN 33 Résoudre des problèmes numériques et algébriques... Mathématiques	MSN 34 Mobiliser la mesure pour comparer des grandeurs... Mathématiques	MSN 35 Modéliser des phénomènes naturels, techniques, sociaux ou des situations mathématiques. Mathématiques / Sciences de la nature	MSN 36 Analyser des phénomènes naturels et des technologies à l'aide de démarches caractéristiques des sciences expérimentales... Sciences de la nature	MSN 37 Analyser les mécanismes des fonctions du corps humain et en tirer des conséquences pour la pérennité de la vie... Sciences de la nature	MSN 38 Analyser l'organisation du vivant et en tirer des conséquences pour la pérennité de la vie... Sciences de la nature		

Objectif d'apprentissage

Composante d'un objectif

MSN 23 – Résoudre des problèmes additifs et multiplicatifs...

1. en traduisant les situations en écritures additives, soustractives, multiplicatives ou divisives

2. en sélectionnant les données numériques à utiliser

3. en choisissant l'outil de calcul le mieux adapté à la situation proposée

4. en anticipant un résultat et en exerçant un regard critique sur le résultat obtenu

5. en utilisant les propriétés des quatre opérations

6. en construisant, en exerçant et utilisant des procédures de calcul (calcul réfléchi, algorithmes, calculatrice, répe...)

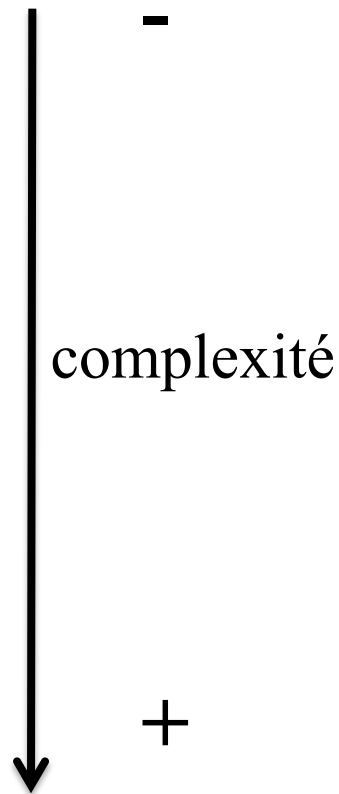
OPÉRATIONS	Progression des apprentissages		Objets	Attentes fondamentales
	5 ^e – 6 ^e années	7 ^e – 8 ^e années		
Multiples, diviseurs, suites de nombres	Liens MSN 22 – Nombres; FG 21 – MITIC		...détermine si un nombre est un diviseur ou un multiple d'un autre	Au cours, mais au plus tard à la fin du cycle, l'élève... ...établit la liste des premiers multiples et celle des diviseurs d'un nombre inférieur à 100 ...utilise les critères de divisibilité par 2, 5, 10, 100 ...reconnait et complète une suite arithmétique de nombres naturels inférieurs à 1 000 (6 ^e année) et de nombres inférieurs à 10 000 dont les termes ont au plus 1 décimale (8 ^e année)
	Recherche des multiples d'un nombre	Recherche des multiples et des diviseurs d'un nombre		
Recherche de quelques critères de divisibilité: 2, 5, 10, 100		Utilisation de quelques critères de divisibilité: 2, 3, 5, 9, 100		
Reconnaissance et établissement de suites arithmétiques		Reconnaissance, établissement de suites numériques et expression de leur loi de formation (progressions, multiples, puissances,...)		
Calculs	Utilisation d'outils de calculs usuels: calcul réfléchi, algorithmes, rétroaction mémorisée, calculatrice			...utilise des procédures de calcul réfléchi pour effectuer de manière efficace une addition, une soustraction

→ Pas de « compétences » ni de « connaissances », mais des *visées prioritaires*, des *objectifs d'apprentissage* (déclinés en composantes et organisés selon une progression) et des *attentes fondamentales*.

Où trouve-t-on des compétences dans le PER?

Cf. PER, attentes fondamentales:

- 1) « accorde le participe passé employé avec l'auxiliaire *être* avec le sujet du verbe »
(PER, L126, p. 65)
- 2) « repère les différentes parties d'un texte réglant des comportements » (PER, 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, L121, p. 37)



Prendre les compétences au sérieux ?

Qu'est-ce que cela signifie?

- La compétence langagière est indissociablement liée à l'idée d'activité ou de tâche complexe
- La compétence langagière est située et a un sens social / scolaire
- La compétence consiste en la mobilisation d'un ensemble intégré de ressources et non en l'addition de ressources.
- La compétence n'est pas observable en tant que telle; elle peut être approchée à partir de la performance de l'élève face à la tâche...

Intermède 2

« 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 & Cardinet, 1992, 5)

➔ Conséquences de ces changements : les situations d'évaluation doivent être en accord avec l'enseignement proposé (mais effet en retour, aussi : ce qui est évalué doit avoir été travaillé préalablement en classe).

5. Compétences, connaissances et références culturelles

De la difficulté à évaluer les compétences...

En analysant le PER, il apparaît en fait que :

- Il semble relativement aisé de lier des items (ou regroupements d'items) à des connaissances [ou praxéologies] et, par conséquent, de les évaluer.
 - Il s'avère en revanche bien plus délicat
 - a) d'y localiser les compétences,
- et, b) de les évaluer :

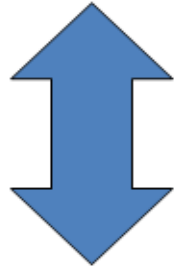
«...lit un texte de manière autonome, en dégage le sens général et le reformule oralement.»

(L1 21, p. 23 ; Objectif d'apprentissage, CE)

Mais... Quel texte ? Pour qui? Dans quel contexte? Dans quel but?...

Mais... faut-il évaluer les compétences ?

Enseigner en vue de favoriser le développement de compétences (cf. PER)

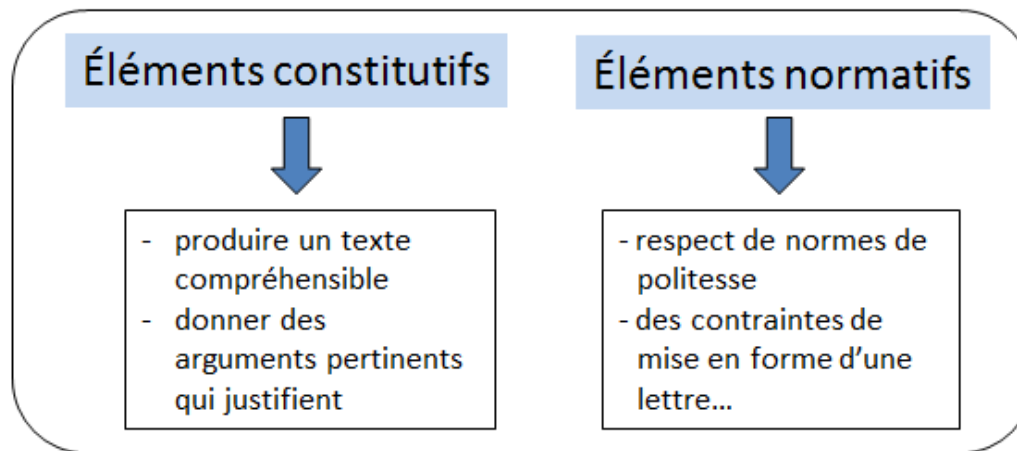


Évaluer de telles compétences ??? (cf. caractère situé, finalisé, inédit... vs forme scolaire...)

« 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. » (J. Cardinet, 1990)

→ pour être évaluables, les compétences doivent être réinterprétées pour la forme scolaire

- Exercer efficacement une tâche complexe, c'est satisfaire aux exigences constitutives et normatives de cette tâche:



Par ailleurs, on peut documenter les ressources – diverses et jamais totalement prévisibles – qui ont été mobilisées.

Par exemple:

- Savoir interpréter les reprises nominales dans un récit,
- Être capable de repérer le schéma narratif
- ...

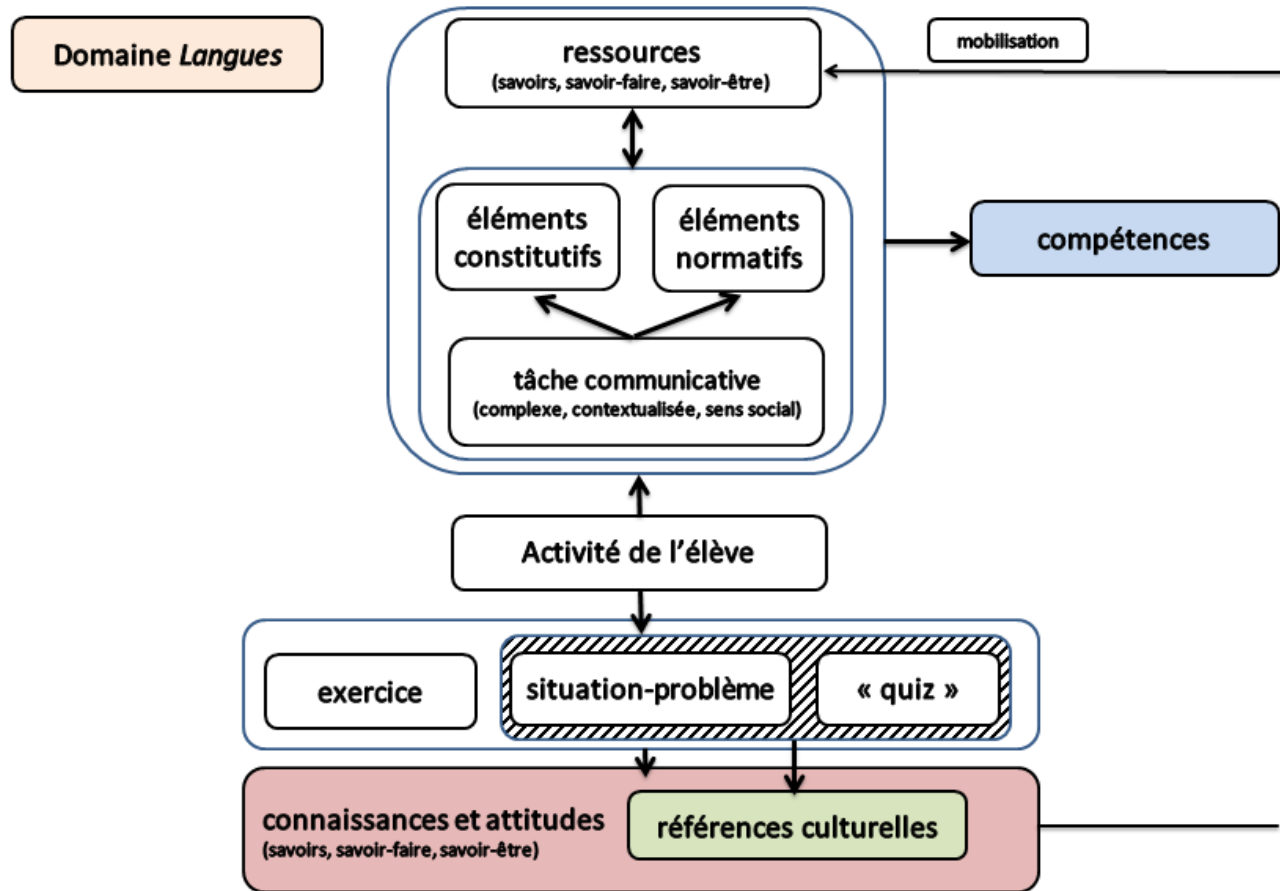
Mais alors, à part les compétences?...

→ Il importe par conséquent de clairement distinguer – pour l'enseignement / apprentissage *et* pour l'évaluation – les autres « objets » également présents dans le PER et de les articuler avec les compétences.

Nous proposons de distinguer également:

- des **connaissances** (savoirs et savoir-faire) et **attitudes** (savoir-être)
- des **références culturelles** socialement / scolairement légitimés

Une tentative de modélisation pour l'enseignement, l'apprentissage et l'évaluation



→ Lien entre « objet évalué » et tâche qui permet de le rendre visible

Intermède 3

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

(Davaud & Cardinet, 1992, 9)

6. Vers une validité didactique...

Quelques interrogations et ouvertures en guise de conclusion

- Nécessité d'un référentiel *pour* l'évaluation
- Nécessité de clarifier le statut des « objets » figurant dans le PER
- Lier statut des objets et format d'évaluation
- Lier enseignement et évaluation, sans les réduire l'un à l'autre !
- Validité?

« 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 & Cardinet, 1992, 9)

Retour à notre « banque d'items »...

Structure de la base : esquisse

