

Chapitre 3 : La conception des instruments d'enquête

Les instruments d'enquête, dans une évaluation standard, sont les tests d'acquisition et les questionnaires. En quelque sorte, les tests d'acquisition nous permettent de construire un indicateur de qualité, qui sera la variable à expliquer du modèle, par opposition aux variables explicatives (les facteurs de la qualité), qui seront pour leur part constituées à partir des données recueillies par les questionnaires¹.

A. Les tests d'acquisition

Souvent, le mandat des évaluateurs ne comprend pas de définition très claire de la qualité qu'il s'agit d'évaluer. En effet, les missions assignées à l'école sont multiples : transmission de connaissance, de valeurs, socialisation, insertion, ... la liste ne peut être close.

Toutes ces dimensions de la qualité sont susceptibles d'être évaluées, et la conception d'instruments portant sur la transmission des valeurs, pour ne prendre que cet exemple, serait un exercice légitime.

Des réponses précises ne pouvant être apportées qu'à des questions précises, il faut bien hiérarchiser les objectifs d'une évaluation. Le PASEC, pour sa part, a choisi d'évaluer les acquis fondamentaux des élèves en français et en mathématiques, ce qui suppose déjà la conception de deux séries de tests correspondant à ces deux disciplines.

Aucun test n'est neutre. Une évaluation peut être entreprise dans le cadre d'une réforme des contenus d'enseignement, avec le sous-entendu que ces contenus sont à changer. Dans ce cas, l'évaluateur sera tenté de tester les nouvelles compétences que l'on souhaite introduire, et n'aura sans doute pas de mal à montrer que ces compétences ne sont pas acquises, puisqu'elles ne sont pas encore enseignées !

¹ Cette dichotomie variable expliquée (tests), variables explicatives (questionnaires) n'est pas tout à fait respectée, puisque le score de début d'année des élèves est inclus parmi les variables explicatives.

Dans le cas d'une évaluation sur les facteurs d'efficacité, le problème est assez facile à résoudre. Ce sont moins des contenus d'enseignement particuliers qu'il s'agit d'évaluer, que la capacité d'un système à transmettre des contenus. La question de la pertinence ou non des contenus actuels est ainsi écartée. L'attitude la plus neutre consiste à fonder les tests sur les programmes en vigueur.

Parmi les objections faites à un test trop fidèle aux programmes en vigueur, il y a, dans certains contextes, le constat d'un non respect, par de nombreux maîtres, de ces programmes, par ignorance ou par routine. On comprend bien que cette objection mène à une impasse, car quand bien même ces maîtres se révéleraient efficaces sur certaines compétences, qui ne sont pas au programme, l'absence de délimitation précise de la notion de qualité brouillerait les conclusions de l'enquête.

Dans la pratique, il est difficile de résister à la tentation d'introduire dans un test des éléments portant sur des compétences un peu particulières, qui n'existent pas en tant que tels dans les programmes (par exemple, la résolution de problèmes de la vie courante, qui font intervenir des compétences transversales à plusieurs disciplines). Dans ce cas, il convient de faire de ces particularités un objectif secondaire, qui ne doit pas être introduit au péril de la compatibilité entre les contenus effectifs d'enseignement, et les compétences mesurées.

D'une manière générale, ce n'est pas aux évaluateurs de définir les compétences pertinentes ou non. Il est souhaitable que les responsables pédagogiques prennent leurs responsabilités à ce niveau : chacun son métier !

Admettons donc que les tests d'acquisition à concevoir soient fondés sur les programmes en vigueur. Les choix ne sont pas pour autant terminés. Il est en effet matériellement impossible de tester toutes ces compétences : la plupart du temps, les évaluateurs ne disposent au mieux que d'une demi-journée pour tester les élèves dans chaque école, et de toute manière, l'attention des élèves faiblit au-delà d'une heure de test.

Face à ce problème, les spécialistes des tests élaborent parfois des procédures d'administration fractionnées. Selon ce principe, aucun élève n'est soumis à l'ensemble des items d'un test, mais le test complet est subdivisé en

deux ou trois parties, et chaque élève n'est testé que sur une de ces parties. Un système rigoureux de répartition aléatoire de ces tests entre les élèves, ainsi que d'items communs d'ancrage, permet de reconstituer le degré de maîtrise globale moyen.

Ce système est recommandé lorsque l'objectif prioritaire est l'évaluation du niveau des élèves, mais dans le cas des analyses sur les facteurs d'efficacité, cela introduit une complication inutile. Des choix sont possibles, à condition qu'ils soient explicites. Seul un nombre limité de compétences, jugées prioritaires, seront testées. Une fois de plus, ce ne sont pas les compétences que l'on teste dans une analyse sur les facteurs d'efficacité, mais le processus d'acquisition. L'évaluateur n'a pas à répondre ici à la question "qu'est-ce que les élèves savent en fin d'année ?" mais plutôt à la question "qu'est-ce qui explique que certains élèves ont appris plus que d'autres en cours d'année ?".

Cette particularité facilite également l'utilisation de tests standardisés, y compris à l'échelle internationale. Même si ces tests standardisés sont plus ou moins conformes aux programmes en vigueur dans différents pays, leur utilisation simultanée permet de comparer l'efficacité de l'apprentissage dans des contextes d'enseignement encore plus variés.

L'utilisation de tests standardisés déjà existants présente ainsi de multiples avantages : elle donne une idée du niveau d'acquisition relatif de l'échantillon testé par rapport à d'autres échantillons auxquels ces mêmes tests ont été administrés, elle permet de comparer directement l'efficacité relative des différents facteurs d'un contexte à l'autre, et enfin, sur un autre plan, elle permet de simplifier et d'économiser lors de la phase de mise en place de l'enquête.

Enfin, dans la mesure où les données d'analyse sont ouvertes à la communauté de chercheurs, l'utilisation d'instruments communs permet des analyses comparatives internationales qui multiplient l'intérêt analytique de chaque évaluation prise séparément, permet des conclusions plus robustes, et contribue ainsi à l'élaboration d'un savoir positif sur les grands facteurs d'efficacité à l'école primaire. La mise en valeur des dispositifs d'enseignement les plus efficaces dépendant du nombre et de la variété introduits dans l'analyse, on comprend aisément l'intérêt de doubler chaque étude portant sur un pays d'une contribution à un ensemble international de données mises à la

disposition de la communauté scientifique. Un champ est ainsi ouvert à la coopération entre institutions de recherche, dont le premier sillon devrait être une politique systématique de mise à disposition des données, ce que les nouvelles technologies (internet, CD-ROM) rendent aujourd'hui très simple sur le plan technique.

Dans cet esprit, la CONFEMEN met à disposition des chercheurs non seulement une collection complète de tests de français et de mathématiques pour le primaire¹, mais également les résultats à ces tests, et plus généralement, l'ensemble des données recueillies au cours de ses enquêtes.

Cela dit, un compromis doit être trouvé entre la possibilité d'une comparaison internationale, et l'adaptation des tests à des objectifs et des contextes particuliers. C'est pourquoi notre guide ne saurait se concevoir sans des indications sur le processus complet de conception de tests adaptés à une analyse sur les facteurs d'efficacité de l'enseignement.

1. Définitions et exemples :

Au cours des paragraphes qui suivent, nous allons utiliser quatre termes dont voici la définition. Ces définitions varient suivant les usages, et celle que nous adoptons (en particulier pour la notion d'item) se veut simplement adaptée à l'usage que nous en ferons.

- Test : Ensemble d'exercices administrés au cours d'une même séance aux élèves
- Compétence : A la fois objectif de programme et capacité à atteindre cet objectif
- Exercice : Partie d'un test possédant son propre énoncé
- Item : Plus petite partie d'un test, ou d'un exercice, correspondant à une notation de 0 (mauvaise réponse), ou 1 (bonne réponse).

Remarques :

- un test est composé de plusieurs exercices
- un exercice est composé d'un ou de plusieurs items
- un exercice n'est pas seulement défini par son énoncé, mais aussi par ses consignes d'administration (manière dont il est présenté aux élèves, y com-

¹ Voir en fin d'ouvrage

pris les consignes orales), ainsi que par ses consignes de correction et de codage.

Quelques exemples d'exercices et d'items tirés de la collection de tests de la CONFEMEN¹ permettent d'illustrer ces notions.

Exemple 1 : Il s'agit d'un exercice du test de français posé en fin de deuxième année, composé de 6 items. Ces items ont été empruntés aux enquêtes IEA (the IEA Reading Literacy Study, 1990-1991).

entoure l'image qui va avec le mot :

exemple :

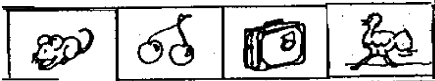
tasse



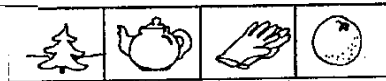
main



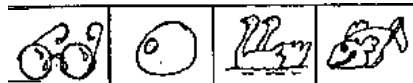
souris



orange



poisson



cinq



lit



A

67%

B

76%

C

60%

D

56%

E

70%

F

79%

¹ Tous les tests sont disponibles sur le site de la CONFEMEN : www.confemen.org. Ils figurent également dans le CD-ROM du PASEC.

Consignes d'administration

Note : Utiliser un cahier de l'élève pour la démonstration, sans aucune annotation, pour montrer la bonne page et pour aider à faire les exemples.

EXERCICE 1 : Reconnaître l'objet par son nom.

Faire mettre le doigt de l'élève sur le numéro 1.

Faire mettre le doigt de l'élève sur l'exemple.

Dire “ Regardez bien les images de l'exemple et lisez le mot à gauche des images ”.

Faire lire le mot “ tasse ” *successivement par trois élèves.*

Dire “ Dans vos cahiers, entourez l'image qui va avec le mot ‘tasse’ ”

Vérifiez que tous les élèves ont entouré l'image qui va avec le mot ‘tasse’ et aider ceux qui n'ont pas réussi.

Dire “ Maintenant, faites la même chose avec les autres images ”

Laisser les enfants travailler pendant deux minutes avant de passer à l'exercice suivant et ne pas donner plus d'explications.

Consignes de correction et de codage

Convention générale de codage :

Mauvaise réponse = 0
(y compris pas de réponse ou réponse non interprétable)

Bonne réponse = 1

Remarque générale sur l'exercice 1 : Tous les moyens d'identification de la bonne image seront acceptés, à partir du moment où il n'y a pas d'ambiguïté sur la réponse (par exemple, un trait au lieu d'un cercle, ou bien une flèche).

item A :

main la bonne image est celle **la plus à droite**

item B :

souris la bonne image est celle **la plus à gauche**



item C :

orange la bonne image est celle **la plus à droite**

item D :

poisson la bonne image est celle **la plus à droite**

item E :

cinq la bonne image est **la 2^{ème} partant de la gauche**

item F :

lit la bonne image est celle **la plus à gauche**

**

Exemple 2 : Il s'agit d'un exercice du test de mathématiques de fin de deuxième année, composé de 4 items.

pose les opérations suivantes:

exemple : $14 + 2$	<input type="text"/>
---------------------------	----------------------

a) $3 + 9$

b) $100 - 10$

c) $5 + 11 + 4$

d) $25 - 20$

A **56%**

B **31%**

C **30%**

D **50%**

Consignes d'administration

NB: Utiliser un cahier de l'élève de démonstration, sans aucune annotation, pour montrer la bonne page et pour aider à faire les exemples.

écrire au tableau l'exemple

Faire mettre le doigt sur le numéro 1

Dire: “ ***attention nous allons poser des opérations*** ”

Dire “ suivez bien au tableau , nous allons faire l'exemple $14 + 2 = ?$ ”

Faire trouver oralement la bonne réponse par les élèves

Ecrire la réponse au tableau, j'écris le nombre 14, ensuite je mets le signe + à gauche, puis j'écris le 2 sous le 4, je trace la barre et je mets le signe (=).

Dire: “ ***copiez la réponse dans vos cahiers*** ”

Vérifier

Dire: “ ***maintenant, continuons avec les autres exercices*** ”

Dire “ ***attention, posez seulement les opérations sans faire les calculs*** ”

Laisser les enfants travailler pendant une minute avant de passer à l'item suivant.

Consignes de correction et de codage

remarque : peu importe si l'élève inscrit le résultat ou pas (même s'il est faux) : l'important est que les chiffres soient bien alignés , unités sous les unités, et dizaines sous les dizaines. L'oubli de la barre et du signe = n'est pas grave, par contre le signe + ou - doit être présent.

item A :

Bonne présentation :

$$\begin{array}{r} 3 \\ + 9 \\ \hline = \end{array}$$

item B :

Bonne présentation :

$$\begin{array}{r} 100 \\ - 10 \\ \hline = \end{array}$$

(idem pour les items C et D)

**



Exemple 3 : Il s'agit d'un exercice d'un seul item extrait du test de mathématiques de 5^{ème} année.

Problème

Maman a 4,20 m de tissu. Elle utilise 2,75 m pour faire la robe de Fatou. Elle a besoin de 2,50 m pour coudre la robe de Sokhna. Quelle longueur de tissu lui manque-t-il ?

(Entoure la bonne réponse) :

Il reste pour faire la robe de Sokhna : $4,20 \text{ m} - 2,75 \text{ m} = 1,45 \text{ m}$

Il lui manque $2,50 \text{ m} + 1,45 \text{ m} = \mathbf{3,95 \text{ m}}$

Il reste pour faire la robe de Sokhna : $4,20 \text{ m} - 2,75 \text{ m} = 1,45 \text{ m}$

Il lui manque : $2,50 \text{ m} - 1,45 \text{ m} = \mathbf{1,05 \text{ m}}$

Il reste pour faire la robe de Sokhna : $2,75 \text{ m} - 2,50 \text{ m} = 0,25 \text{ m}$

Il lui manque : $4,20 - 0,25 \text{ m} = \mathbf{3,95 \text{ m}}$

Y

24%

Consignes d'administration

Dire : “ ***Lisez l'exercice n°9*** ”

Dire : “***Entourez la bonne réponse parmi les trois opérations proposées***”.

Laisser les enfants travailler trois minutes

Consignes de correction

item Y :

La bonne réponse est : $2,50 \text{ m} - 1,45 \text{ m} = \mathbf{1,05 \text{ m}}$ (réponse du milieu)

Exemple 4 : Cet exercice (le seul de ce genre) a mis les élèves face à une réalité très concrète (la lecture d'une notice de médicament), en leur faisant pourtant utiliser et exploiter les acquis scolaires. Il permet d'évaluer le caractère opératoire ou non des connaissances des élèves. Cet item a été emprunté aux instruments du projet Education pour tous UNESCO.

Dans une boîte de médicaments, on trouve la notice suivante : lisez-la attentivement et répondez aux propositions

PRIMALAN

INDICATIONS :

- *Dérangements intestinaux et plus spécialement:*
 - *diarrhées*
 - *vomissements.*

POSOLOGIE:

- *Adultes : 1 à 6 comprimés par jour.*
- *Enfants : de 3 à 5 ans ½ comprimé deux fois par jour;*
au dessus de 5 ans: ½ comprimé 2 à 4 fois par jour.

- *A prendre au début des repas en avalant avec un peu d'eau, sans croquer.*

PRECAUTIONS D'EMPLOI :

L'usage prolongé de ce médicament peut entrainer des maladies du rein.

PROPOSITIONS :

1. Est-ce que ce médicament guérit la diarrhée ?

(Mets une croix en face de la bonne réponse)

- oui
- non
- le texte ne le dit pas

☐
☐
☐

Y

61%

2. Un adulte peut prendre 8 comprimés par jour.

(Mets une croix en face de la bonne réponse)

- oui
- non
- le texte ne le dit pas

☐
☐
☐

Z

57%

tournez la page...

3. Un enfant qui a plus de 5 ans peut prendre :

(Mets une croix en face de la bonne réponse)

- 1 à 6 comprimés par jour ☐
- 1 comprimé deux fois par jour ☐
- ½ comprimé deux à quatre fois par jour ☐
- le texte ne le dit pas ☐

AA

32%

4. On doit prendre le comprimé :

(Mets une croix en face de la bonne réponse)

- en le croquant ☐
- en le suçant ☐
- en l'avalant ☐
- le texte ne le dit pas ☐

AB

64%

5. Voici le nom de certaines maladies. Si l'on prend le médicament trop longtemps, laquelle peut-on risquer ?

(Mets une croix en face de la bonne réponse)

- l'angine ☐
- les maux de tête ☐
- les maladies du rein ☐
- je ne sais pas ☐

AC

39%

Consignes d'administration

Vérifier que les enfants sont sur l'exercice numéro 9.

Lire le texte.

Laisser les enfants travailler pendant douze minutes.

(vérifier qu'ils n'oublient pas la dernière page de questions)

Consignes de correction

item Y :

Bonne réponse : **oui**

(mettre une croix dans la première case)

item Z :

Bonne réponse : **non**

(mettre une croix dans la deuxième case)

item AA :

Bonne réponse : **½ comprimé deux à quatre fois par jour**

(mettre une croix dans la troisième case)

item AB :

Bonne réponse : **en l'avalant**

(mettre une croix dans la troisième case)

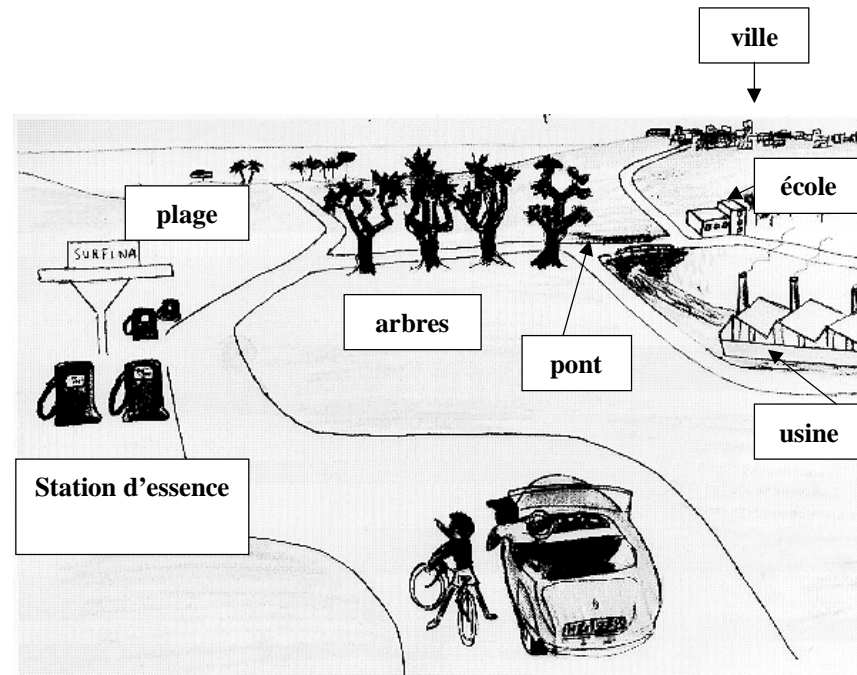
item AC :

Bonne réponse : **les maladies du rein**

(mettre une croix dans la troisième case)

**

Exemple 5 : Cet exercice, comme le précédent, met en jeu des compétences variées (de langue, de repérage dans l'espace, mais aussi de "savoir être"). Il est constitué de cinq items (même si une seule question est posée à l'élève), du fait que le barème de réponse est sur cinq points.



Explique au chauffeur de la voiture le chemin pour aller en ville

.....

.....

.....

.....

.....

.....

Consignes d'administration

Vérifier que les enfants sont sur l'exercice numéro 11.

Lire l'énoncé de l'exercice.

Laisser les enfants travailler pendant **douze minutes**.

Consignes de correction

Item AI : La question a-t-elle été comprise ? Ecrire 1 si la question a été comprise (l'élève décrit un itinéraire pour aller en ville, aussi erroné soit-il), 0 si la question n'a pas été comprise (il n'y a pas de tentative de donner un itinéraire)

Item AJ : L'itinéraire permet-il à l'automobiliste d'arriver en ville ? Ecrire 1 si oui, 0 si non.

Rem. C'est tout à fait indépendant de la qualité de l'expression. Il s'agit de noter si, même en style télégraphique, les éléments pour trouver le chemin sont là. Par exemple : «à droite aux arbres, tout droit puis à l'école vous voyez la ville à gauche» sera noté 1.

Item AK : Nombre d'indications exactes : donner le nombre.

Une «indication» est un élément concernant l'itinéraire ; il ne suffit pas de citer un lieu, mais il faut dire comment se positionner par rapport à ce lieu pour qu'on considère qu'il y a «indication».

Par exemple, «il y a un arbre» n'est pas une indication, ni vraie, ni fausse.

Par exemple, «tourner à droite aux arbres (ou : devant les arbres, après les arbres)» est une indication exacte.

Ou encore, «continuer jusqu'à la plage», «tourner à droite après le pont» est une indication fausse.

Item AL : Nombre d'indications fausses : donner le nombre (voir item AK)

Item AM : Expression et orthographe.

Donner une note entre 1 et 4 :

1 : texte incompréhensible à l'écrit comme à l'oral.

2 : texte correct à l'oral, mais l'écrit est complètement déficient (syntaxe, ponctuation, orthographe très lacunaires).

3 : texte écrit correct, mais imparfaitement maîtrisé.

4 : texte écrit maîtrisé.

2. De l'objectif à l'item

Quelques règles simples permettent d'organiser la conception des tests.

La première est de savoir sur quels programmes baser les tests : ceux de l'année en cours, ou également ceux des années précédentes ?

Dans le cadre analytique proposé en exemple, deux tests sont nécessaires, un test de début d'année scolaire, et un test de fin d'année scolaire. Une alternative se pose de manière simple :

- Utiliser le même test en début et en fin d'année, fondé sur un compromis entre les compétences au programme des niveaux inférieurs, et les compétences à acquérir du niveau considéré
- Concevoir un test de début d'année sur les compétences au programme des niveaux antérieurs, et un test de fin d'année centré sur les compétences du niveau en cours

La première option a souvent eu les faveurs des évaluateurs, en permettant de calculer pour chaque élève une différence de score entre le début et la fin d'année, et en faisant de cette différence, nommée “ progrès de l'élève ”, la variable à expliquer.

La deuxième option ne permet pas de calculer une telle différence, mais a l'avantage de concentrer la mesure de fin d'année sur les contenus de programme et donc sur l'enseignement de l'année en cours.

Même si le débat méthodologique est ouvert, nous recommandons de ne pas recourir au même test pour le début et la fin d'année. Car un élève qui aurait parfaitement réussi le test initial ne disposerait d'aucune marge de progression, et paraîtrait vraisemblablement régresser, alors qu'il se peut bien qu'il ait davantage appris qu'un élève qui est passé d'un score faible à un score moyen. Mieux vaut donc deux tests qui permettent de différencier à chaque fois l'élève faible de l'élève fort. Le deuxième test doit donc inclure des compétences acquises dans l'année, qui permettent de constater la progression de l'élève initialement fort (s'il y en a une).