

On comprend qu'il faudra retenir en priorité les items pour lesquels le point-bisérial calculé à partir des résultats de la mise à l'essai est proche de +1. Un test composé exclusivement d'items aux valeurs point-bisérial proches de 1 est en effet particulièrement cohérent : sous des formes variées, il teste la même compétence avec constance. Si " par accident " un élève rate un item, tous les autres lui permettront de prouver sa vraie valeur, et réciproquement pour les mauvais élèves.

La nuance à apporter à un tel indicateur, c'est qu'à la limite, le meilleur test est fait de la répétition du même item (on est sûr dans ce cas là d'avoir des points bisériaux tous égaux à 1). Or dans la réalité, on veut tester des objectifs opérationnels distincts, et il est normal que certains élèves réussissent sur certains, et moins sur d'autres. Plus le test sera varié, et plus les points bisériaux associés à chaque item auront du mal à se rapprocher de 1. Néanmoins, si deux items testent le même objectif opérationnel, c'est celui dont le point bi-sérial est le plus proche de +1 qu'il faudra retenir.

L'alpha de Cronbach, ou la cohérence interne d'un test pris dans son ensemble

Lorsqu'un test est composé d'items dont les points-bisériaux sont proches de +1, il présente une grande cohérence interne. Un indicateur au niveau du test mesure cette cohérence interne, c'est l'alpha de Cronbach.

L'alpha de Cronbach varie entre zéro (test incohérent), et +1 (test totalement cohérent). Un test cohérent présente bien des avantages :

- d'abord il indique que ce que nous voulons mesurer (la réussite à un ensemble d'objectifs opérationnels) correspond bien à une compétence générale homogène. En d'autres termes, dans notre exemple, il indique qu'il y a bien de bons élèves en mathématiques, et pas seulement de bons élèves en calcul, d'autres en géométrie, d'autres en algèbre, etc.
- ensuite, un test cohérent sur le plan interne est fiable, c'est-à-dire qu'il y a de bonnes chances pour que le même élève, passant le test deux fois, obtienne à peu près la même note. Comme une évaluation a le défaut de ne tester chaque élève qu'une ou deux fois, à la différence du maître qui teste le même élève de nombreuses fois au cours de l'année, c'est un critère important¹.

¹ Cela explique le bon usage qui peut être fait des points bi-sériaux et de l'alpha de Cronbach dans un domaine connexe à celui de l'évaluation : la conception des épreuves d'examen et de concours.

Les critiques de l'alpha de Cronbach sont les mêmes que celles adressées au point bi-sérial : on veut parfois tester des compétences relativement variées au sein d'un même test, et cela va dans un sens contraire à l'obtention d'un bon alpha de Cronbach. Si l'on ne voulait tester qu'un seul objectif opérationnel, il serait facile d'obtenir pour chaque item un point bi-sérial proche de +1, et pour le test en général un alpha de Cronbach également proche de +1.

En fait, l'usage que l'on fait de l'alpha de Cronbach dans le cadre présent nous dispense de trop de considérations sur sa valeur absolue. En effet, l'alpha sera surtout pour nous un contrôle pour réduire le nombre d'items (et donc la longueur du test infligé aux élèves) sans compromettre la fiabilité de la mesure. En passant de 60 items à 30 items, comme dans notre cas, nous évitons un test trop long pour les élèves, mais disposons-nous toujours d'une mesure fiable des différences de compétence entre élèves? Si la diminution de l'alpha qui accompagne l'élimination d'items est limitée, tout va bien. Dans le cas contraire, on sera allé trop loin, ou bien on aura éliminé à tort des items qui avaient un bon point bi-sérial, et qui donc contribuaient à la bonne tenue de l'alpha.

D'une manière générale, on considère qu'un test est bien adapté à l'objectif d'une évaluation, qui est de hiérarchiser les élèves entre eux, si l'alpha de Cronbach est supérieur à 0,80. Par construction, plus un test comporte d'items, plus l'alpha a de chances d'être bon. Cependant, éliminer des items dont le point bi-sérial est faible ou négatif permet aussi d'améliorer l'alpha.

Le compromis idéal entre la longueur d'un test et sa validité se fait donc par aller-retour entre élimination des items dont les points bi-sériaux sont faibles, et vérification que le test ainsi raccourci conserve un bon alpha.

La procédure n'est pas entièrement automatique, puisque des considérations pédagogiques (équilibre et variété du test) pourront altérer le caractère un peu trop automatique de l'élimination progressive des items dont les points bi-sériaux sont les plus faibles.

En fait, c'est surtout lorsque un exercice est composé d'un grand nombre d'items (comme par exemple une série de verbes à conjuguer), que le point bi-sérial est un outil précieux de sélection des items à conserver.

Encadré : calcul du point bi-sériel de l'alpha de Cronbach

Calcul du point bi-sériel associé à chaque item après la mise à l'essai du prototype (60 items) :

Le calcul du point bi-sériel (comme celui de l'alpha de Cronbach), est souvent une fonction présente dans les logiciels statistiques (quelquefois, il n'est pas signalé en tant que tel, mais correspond à un usage particulier du calcul des corrélations). Dans notre cas, nous illustrons le calcul à partir d'une simple feuille de tableur Excel.

La formule de calcul du point bi-sériel pour un item est la suivante :

$$r_{pbi} = \frac{M_1 - M_0}{S} \times \sqrt{pq}$$

Avec :

M_1 = Score moyen des élèves qui ont juste à l'item

M_0 = Score moyen des élèves qui ont faux à l'item

p = Proportion de réponses justes à l'item (= $N_1/100$ sur le tableau)

q = Proportion de réponses fausses à l'item (= $N_0/100$ sur le tableau)

S = Ecart-type des notes de la population étudiée (calcul de l'écart-type de la population à partir de l'écart-type des notes de la colonne total du tableau)

Calcul de l'alpha de Cronbach pour le test prototype (60 items) :

La formule donnée ici s'applique seulement au cas où le test est composé d'items à réponse vrai/faux (1 ou 0). Dans ce cas particulier, l'alpha de Cronbach s'appelle aussi coefficient de fidélité de Kuder-Richardson.

$$KR20 = \frac{N}{N-1} \times \left[1 - \frac{\sum_{i=1}^N p_i q_i}{S^2} \right]$$

N est le nombre d'items (60 dans le cas du test prototype).

La somme des $p_i q_i$ (probabilité de la réussite multipliée par la probabilité de l'échec à l'item i , i variant ici de 1 à 60) s'obtient en additionnant les produits deux à deux des valeurs $N_1/100$ et $N_0/100$ portées sur la feuille (voir le tableau précédent).

Sélection des items :

Une fois le point bi-sérial calculé pour chaque item, la réduction du nombre d'exercices (et d'items), conformément au tableau donné plus haut peut commencer.

Pour reprendre l'exemple des trois exercices (C,D,E) du test prototype consacrés au chapitre II du Programme, nous pouvons suivre le raisonnement :

L'exercice C comprenait sept items (numéros 11 à 17)

L'exercice D comprenait six items (numéros 18 à 23)

L'exercice E comprenait sept items (numéros 24 à 30)

L'objectif est de passer de trois exercices et 20 items à deux exercices et 10 items.

Les points bi-sériaux associés aux items de l'exercice C sont relativement faibles par rapport aux points bi-sériaux associés aux items des exercices D et E. Nous décidons donc de l'éliminer entièrement.

Il nous reste alors deux exercices et 13 items pour le chapitre II. Il nous faut encore éliminer trois items. Ce sera l'item numéro 21 de l'exercice D (point bi-sérial de 0,36, le plus faible de tout l'exercice), et les items numéro 26 et 27 de l'exercice E (points bi-sériaux de 0,28 et de 0,45).

En procédant de façon similaire sur l'ensemble du test prototype, on arrive finalement à un test final de 30 items sur 12 exercices (voir tableau " test définitif ", page suivante).

Chapitre	I						II						III						IV						V						Total (sur 30)																						
	A						D						E						F						G							H						I						J						K			
N° d'exercice	1	2	3	4	6	18	19	20	22	23	24	25	28	29	30	32	33	34	35	36	46	47	48	49	50	53	54	58	59	60	Total																						
Élève\item	1	0	0	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	19																						
	0	0	1	0	1	0	1	1	1	1	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	1	1	1	1	15																							
	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	1	0	1	21																							
	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5																							
	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	1	0	0	1	0	0	0	1	1	1	16																							
	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	0	0	1	0	1	1	18																							
	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	0	0	1	0	0	0	1	1	19																							
	1	1	0	1	0	0	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	15																							
	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	6																							



96	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
97	1	0	0	1	0	1	1	1	1	1	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1	0	1	1	15
98	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	5	
99	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
100	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	

N _i	61	27	35	44	74	34	25	42	67	49	20	44	46	45	37	38	37	32	55	16	13	55	45	6	7	20	53	20	45	
N ₀	39	73	65	56	26	66	75	58	33	51	80	56	54	55	63	62	62	63	68	45	84	87	45	55	94	93	80	47	80	55
M _i	15	18	19	18	14	18	18	18	15	17	20	18	17	17	17	19	19	20	21	16	19	19	14	15	21	19	20	16	20	17
M ₀	6	9	7	6	4	8	9	6	4	6	9	6	7	6	8	7	7	6	7	6	10	10	7	8	11	11	9	6	9	6
rbpi	0,55	0,51	0,67	0,74	0,52	0,58	0,46	0,70	0,59	0,69	0,53	0,69	0,64	0,69	0,57	0,74	0,80	0,81	0,62	0,39	0,38	0,43	0,40	0,29	0,26	0,55	0,63	0,53	0,66	

N_i = Nombre de bonnes réponses pour l'item de la colonne
 N₀ = Nombre de mauvaises réponses pour l'item de la colonne
 M_i = score moyen des élèves donnant une bonne réponse à l'item de la colonne
 M₀ = score moyen des élèves donnant une mauvaise réponse à l'item de la colonne

Nombre d'élèves (N) : 100

Ecart-type des notes de la population étudiée (S) : 8,11

Coefficient de Kuder-Richardson (KR20) : **0,937**

En réduisant de moitié le nombre d'items nous ne perdons que 0,01 point sur l'alpha, qui reste à un niveau élevé nettement supérieur au seuil de 0,8.

Ainsi, ce processus nous a permis d'élaborer un test près de deux fois plus court, et presque aussi fiable que le test d'origine.

B. Les questionnaires

Dans notre protocole, les questionnaires distribués aux élèves, aux maîtres, et aux directeurs, ont pour fonction de nourrir la base de données qui servira à l'estimation statistique du modèle de prédiction des scores des élèves.

Dans le chapitre de ce guide sur l'analyse des données, nous verrons que la traduction statistique de ce modèle de prévision ne peut intégrer qu'une vingtaine de variables d'explication du score final¹. Au delà, en effet, les effets des différents facteurs risquent trop de se confondre les uns avec les autres.

Pourquoi, dès lors, s'embarrasser de trois questionnaires différents, si c'est pour ne conserver à la fin qu'un nombre limité de variables ?

C'est principalement la difficulté de prévoir exactement à l'avance quelles informations seront retenues. Nous avons vu que des facteurs réels (les causes agissantes), aux variables instrumentales (les mesures ou catégories) qui les représentent, il n'y a pas de correspondance directe. On aura souvent besoin de plusieurs questions, pour constituer une seule variable instrumentale, représentative d'un seul facteur explicatif ou de contrôle.

¹ Ce chiffre est à relativiser selon le nombre de variables "dichotomisées" dans le modèle, c'est-à-dire de variables qui détaillent une seule dimension explicative (comme le niveau académique du maître) mesuré par une série de catégories d'appartenance mutuellement exclusives (ce qui ont été recrutés au BEPC, au BAC, etc.). Du point de vue de l'algorithme d'estimation, en effet, ces catégories étant mutuellement exclusives, leur multiplication ne nuit pas, en général, à l'estimation des autres variables explicatives du modèle.

Il reste alors à calculer le nouvel Alpha de Cronbach (ou plutôt KR20) associé à ce test définitif. Le calcul donne 0,937, contre 0,947 pour le test prototype.

Par exemple, une hypothèse raisonnable lie le niveau de vie des élèves aux performances scolaires. Par quelle variable instrumentale faut-il mesurer ce niveau de vie ? Il nous faudra probablement synthétiser un certain nombre d'indicateurs, et donc autant de questions, pour constituer un indice fiable de ce niveau de vie. Il faudra même, la plupart du temps, tester différentes manières de constituer cet indice. Certaines informations du questionnaire se révéleront pertinentes, et d'autres non.

L'enjeu, lors de la constitution des questionnaires, sera précisément d'anticiper au mieux quelles informations sont pertinentes pour le modèle et quelles informations ne le sont pas. Car une deuxième contrainte pèse sur les questionnaires : ils doivent être aussi courts et clairs que possibles. Les questionnaires longs et complexes découragent les enquêtés. Mieux vaut se concentrer sur quelques informations, et veiller à la meilleure manière de les recueillir, que de ratisser trop large, en se disant que " ça pourrait bien servir "

Mieux vaut peu d'informations fiables qu'une information très large, mais lacunaire ou peu fiable. C'est d'autant plus vrai lorsque cette information doit faire l'objet d'un traitement statistique sophistiqué¹.

Organisation des informations à recueillir selon le questionnaire :

Dans le chapitre 2 (paragraphe sur le modèle à estimer), nous avons rassemblé les différents facteurs causaux à prendre en compte au sein d'un schéma général " Modèle temporel de production des acquis à l'école primaire "

Le tableau de la page suivante ordonne les grandes catégories qui avaient été retenues en fonction des questionnaires qui seront chargés de les documen-

¹ les anglo-saxons ont un proverbe très imagé pour qualifier les analyses statistiques sophistiquées faites à partir de données peu fiables : " garbage in, garbage out ", ce qui signifie peu ou prou " si on fait rentrer des détrit (dans le modèle), il en sortira (après traitement statistique) d'autres détrit ", ou encore : «Modèle poubelle, résultats poubelle».

ter : le questionnaire élève (QE), le questionnaire maître (QM) et le questionnaire directeur (QD).

Répartition de la collecte de données d'analyse entre le questionnaire élève (QE), le questionnaire maître (QM) et le questionnaire directeur (QD) :

Catégorie de facteurs	QE	QM	QD
Caractéristiques personnelles de l'élève	✓		
Milieu familial de l'élève	✓		
Milieu environnant			✓
Infrastructure et équipement de l'école			✓
Encadrement administratif de l'école			✓
Encadrement pédagogique de l'école			✓
Caractéristiques et style pédagogique du maître		✓	
Encadrement administratif du maître		✓	
Encadrement pédagogique du maître		✓	
Infrastructure et équipement de la classe		✓	
Équipement de l'élève durant l'année	✓		
Effectifs et organisation de la classe		✓	

Il est parfois bon, pour certaines variables d'approche délicate, de recouper les informations issues des questionnaires avec une observation directe effectuée par l'administrateur d'enquête lors de son passage dans l'école (voir ci-dessous l'exemple concernant les effectifs de classe).

C'est essentiellement une contrainte de coûts qui fait reposer autant la collecte de données sur la bonne volonté des enquêtés. Sans cette contrainte, les enquêteurs pourraient rester à loisir dans les écoles, et observer directement les phénomènes, en particulier pédagogiques.

Les élèves, les maîtres et les directeurs peuvent avoir des contraintes ou des volontés qui mettront à mal la sincérité de leurs réponses. Cela est vrai, en particulier, de toutes les questions portant sur la charge de travail, ou le style pédagogique. Certaines réponses sont faussées pour se conformer aux attentes que l'enquêté anticipe chez l'enquêteur, ou pour faire passer un message.

Il n'est pas vraiment productif de se lamenter de ces éventuels manquements à la neutralité. Il est parfois utile de poser plusieurs questions pour permettre des recoupements. La meilleure attitude cependant ne consiste pas à essayer de " piéger " l'enquêté par des questions détournées, mais à exposer le plus clairement et le plus simplement possible la volonté de mesurer les choses telles qu'elles sont, et non telle qu'on voudrait qu'elles soient, et à rappeler que l'anonymat de l'enquête sera respecté.

Comment formuler et mettre en page les questions :

Le principe est de réduire au minimum le travail d'interprétation au niveau de la saisie et du traitement des données.

L'erreur à éviter, est de classer ensemble des personnes qui ne doivent pas l'être, et réciproquement, de classer dans des catégories différentes des personnes qui devraient faire partie de la même catégorie.

Questions ouvertes et questions fermées :

Il existe principalement deux méthodes pour poser la même question, la méthode ouverte, dans laquelle aucune réponse n'est suggérée, et la méthode fermée, dans laquelle le destinataire du questionnaire doit choisir parmi un nombre limité de réponses toutes faites.

Par exemple, il y a deux façons d'interroger le maître sur les causes auxquelles il attribue l'échec scolaire :

Première approche : question ouverte :

Quelle est, selon vous, la principale cause de l'échec scolaire ?

.....

.....

.....

.....

Deuxième approche : question fermée :

A quelles causes attribuez-vous en priorité l'échec scolaire de certains de vos élèves? **(cochez au maximum cinq cases)**

- La santé des enfants
- L'absence d'aide du Directeur
- L'absence de devoir à faire à la maison

- La pauvreté culturelle du milieu
- La réforme
- Les emplois du temps
- Le manque de livres des maîtres pour chaque discipline
- L'inadaptation des programmes à la population de l'école
- Le manque de motivation des élèves
- L'absence de mobilier scolaire
- Le manque de motivation des maîtres
- L'absence d'aide pour les devoirs et les leçons à la maison
- Le multigrade
- La gestion de l'école
- L'absence de manuels
- Les capacités intellectuelles des élèves
- La formation des maîtres
- Les effectifs pléthoriques
- L'absence de matériel didactique
- Le désintérêt des parents pour l'école
- Le double flux (ou la double vacation)
- Vos absences
- La différence de culture entre l'école et la famille
- La sous-alimentation
- Votre compétence
- L'absence d'encadrement pédagogique par les inspecteurs
- Le manque de discipline de certains élèves
- Le travail de préparation trop lourd
- Les horaires trop lourds
- Le contenu inadapté de manuel pour les élèves
- L'environnement bruyant de l'école
- Le système de la journée continue

Les deux approches ont leurs inconvénients et leurs avantages. La première permet des réponses «plus riches» mais sera plus difficile à traiter. A contrario, la seconde est d'un traitement statistique aisé, mais restreint le choix (multiplier les possibilités n'est pas forcément la solution ; certaines risquent d'être trop proches, et des nuances de formulation risquent d'avoir des effets de suggestion sur l'enquêté).

Une illustration des pièges de l'enquête : combien y a-t-il d'élèves dans la classe ?

Quoi de plus simple que de donner la taille d'une classe ? L'expérience montre pourtant que les renseignements relatifs aux effectifs et à l'organisation de la classe, pourtant a priori évidents, sont propices à de nombreuses confusions.

La question “ combien y a-t-il d'élèves dans votre classe ? ”, adressée aux maîtres, revêt une importance stratégique : du point de vue pédagogique comme de celui des coûts, la question des effectifs par classe est tout à fait déterminante.

Or, si on n'y prend pas garde, la variété des situations de classe, comme la variété des interprétations des maîtres, peut rendre la réponse à cette question totalement inexploitable, au grand préjudice de l'analyse. Pire, ces différences et divergences peuvent passer inaperçues aux yeux de l'enquêteur, et conduire à de graves distorsions lors de la phase d'estimation statistique du modèle.

Cette question, a priori, est à poser dans le questionnaire maître, en fin d'année scolaire, (et non en début d'année scolaire, période où les affectations et les effectifs ne sont pas stabilisés) .

La première chose à se demander, pour bien poser cette question des effectifs, est la nature du phénomène dont on veut mesurer l'impact causal. D'un point de vue pédagogique, ce n'est pas la réalité administrative (combien d'inscrits) qui compte, mais la réalité pédagogique (combien de présents, en situation de cours, devant le maître, à un moment donné).

La question se complique du fait des nombreux modes d'organisation des classes que l'on rencontre sur le terrain. S'il s'agit d'une classe multigrade, comment savoir si l'enseignant a indiqué les seuls élèves du niveau (année) d'étude, où l'ensemble des élèves étudiant ensemble ? S'il s'agit d'une classe à double flux, comment savoir si le maître a indiqué l'effectif d'une cohorte (celle du matin, ou celle de l'après-midi), ou bien l'effectif des deux cohortes, qui bien souvent, du point de vue administratif, constituent une seule

classe ? Le maître est-il seul en classe, ou bénéficie-t-il de l'appui d'un collègue ou d'un assistant ?

Concernant la question des effectifs, dans le modèle, une seule variable (taille de classe) sera introduite. Mais pour être sûr de la validité de cette variable, il faudra tenir compte de toutes ces dimensions.

En fait, cette variable de la taille de la classe doit être déduite de toute une série de questions, dont l'extrait de questionnaire maître en page suivante donne une idée :

Extrait du questionnaire maître

19. Enseignez vous cette année dans une classe (cochez la case correspondante)

- à simple flux
- à double flux (les élèves qui viennent le matin ne sont pas les mêmes que les élèves qui viennent l'après-midi)
- multigrade (vous avez des élèves de cours différents dans la même classe)

20. Etes vous le seul enseignant auquel vos élèves ont affaire ? (cochez la case correspondante)

- oui, toujours
- oui, la plupart du temps
- non, je partage cette classe avec d'autres collègues ou adjoints

21. Si votre classe est à double flux (les élèves qui viennent le matin ne sont pas les mêmes que ceux qui viennent l'après-midi)

- vous enseignez un seul des deux flux (matin ou après-midi)
- Vous enseignez les deux flux (matin et après-midi)
- Votre classe n'est pas à double flux

22. A quel(s) cours vos élèves de cette année appartiennent-ils ? (Cochez la ou les cases correspondantes)

- CP1
- CP2
- CE1
- CE2
- CM1
- CM2

23. Quel est le nombre total d'élèves inscrits dans votre classe ?

(Si vous êtes en charge d'une seule cohorte d'une classe à double flux, donnez l'effectif de cette cohorte, si vous êtes en charge des deux cohortes d'une classe à double flux, donnez l'effectif des deux cohortes).

	nombre de garçons	nombre de filles	Total
CP1			
CP2			
CE1			
CE2			
CM1			
CM2			

24. Nombre d'élèves inscrits dans votre classe : _____

25. Nombre d'élèves de votre classe qui ont abandonné depuis le début de l'année : _____

26. Nombre d'élèves présents dans votre classe un jour ordinaire : _____

27. Nombre moyen d'élèves absents par jour (au cours du mois dernier) : _____

Enfin, d'autres recoupements sont souhaitables avec des questions adressées au directeur sur les effectifs des cours et leur organisation dans l'établissement.

Lorsque la marge d'incertitude concernant l'exactitude des données recueillies par questionnaire est trop grande, et seulement dans ce cas, il est alors préférable de former les administrateurs d'enquête au recueil direct de l'information. On veillera cependant à ne recourir à cette solution que dans les cas les plus délicats (celui de la taille des classes en est probablement un), afin de ne pas trop alourdir le travail de l'enquêteur.