



Les Documents de Travail de l'IREDU Working Papers

Institut de Recherche sur l'Education
Sociologie et Economie de l'Education
Institute for Research in the Sociology and Economics of Education

Jusqu'où peut-on comparer les apprentissages scolaires ?

Jean Bourdon

Juin 2007
DT 2007/7

Les Documents de travail de l'Irédud n'engagent que leurs auteurs. L'objet de leur diffusion est de stimuler le débat et d'appeler commentaires et critiques. Ils peuvent ensuite donner lieu à des publications dans des revues scientifiques. *Working Papers do not reflect the position of IREDU but only their author's views.*



Pôle AAFE – Esplanade Erasme – B.P. 26513 - F 21065 Dijon Cedex
Tél.+33 (0)3 80 39 54 50 - Fax +33 (0)3 80 39 54 79
iredu@u-bourgogne.fr- <http://www.u-bourgogne.fr/iredu>



Jusqu'où peut-on comparer les apprentissages scolaires ?

Jean Bourdon, IREDU, CNRS-Université de Bourgogne

juin 2007

Résumé

L'analyse des apprentissages scolaires et de leur qualité répond, dans un souci d'éducation comparée, à des processus complexes. La référence de plus en plus fréquente à un savoir de base universel, en rapport avec un socle de connaissances, pose la question suivant laquelle la mise en évidence comparative des facteurs d'efficacité de l'école est un exercice délicat. Dans cette note, l'objet est uniquement centré sur une question de la mesure comparée des résultats d'un test scolaire équivalent, ceci à des enfants situés dans des niveaux scolaires équivalents, où la langue d'enseignement est la même, mais vivant dans des milieux sociaux économiques largement différenciés. Cette comparaison montre, malgré les différences de contextes, l'existence de similitudes dans la « production d'école » au niveau des apprentissages initiaux et ceci permettrait de définir ces savoirs comme un bien collectif dont il resterait évidemment à mieux décrire les sources de qualité.

Abstract

The pertinent review for the early knowledge learning in school and their links with quality matters is a complex issue, expressly in a comparative approach when crossing the variety on the range for economic development. But the increasingly frequent reference for a universal basic knowledge, in a connection with a claim concerning a global convergence, as a necessity in self determination for everyone to apply for a minimum level in literacy, raises the question whereby the comparative description of the factors of effectiveness of the school. In this note, the objective is only centred on the concerns for a comparison in measurement relating to results of a common school test led to pupils living in countries largely different in the economic development scale (Cameroon and France), are equivalent, this with children located in equivalent school grades, where the language of teaching is the same one. This comparison shows, in spite of the differences in contexts, similarities in the "education production function" and thus would allow defining this basic knowledge as a near homogenous collective good.

Mots clés : apprentissages initiaux, tests scolaires, France, Cameroun.

Plan de la note

1. Un étalon : le socle de compétences ?	1
2. Le socle d'un savoir de base comme norme d'un bien public universel.....	2
3. Contexte de mesure	4
4. Application.....	6
5. Retour sur le concept de valeur ajoutée scolaire	8
6. Exploitation des items communs entre les deux tests.....	12
Bibliographie	15
Annexe A Comparaison sur les scores aux tests.....	16
Annexe B : Items et compétences.....	17
Annexe C : Tests de cohérence sur les items.....	18
Annexe D : Détails des items de chaque test et champs de compétences couverts	26
Annexe E : Adaptation des instruments à la pratique et au programme en France	30
Annexe F : Vue d'ensemble des programmes en France du primaire en 2002, par année, pour les thèmes des nombres et du calcul.....	32

1. Un étalon : le socle de compétences ?

Aux Etats-Unis, la commission sur l'excellence en éducation dans le rapport intitulé *A Nation at Risk* avançait, en 1983, la nécessité de retourner aux fondamentaux (lire, écrire, compter)

dans les savoirs de base. Cet investissement éducatif initial doit servir à la fois de socle commun pour ceux qui n'iront pas plus loin et de base pour ceux qui tendront vers l'excellence. Cette idée est reprise près de vingt ans plus tard dans le *No child Left Behind Act of 2001*. Le *Livre blanc* sur l'éducation et la formation (Commission européenne, 1996), propose la création d'un processus permettant de confronter et de diffuser largement les définitions des « compétences clés » et de trouver les meilleurs moyens de les acquérir, de les évaluer et de les certifier dans l'espace européen. Dans le plan d'action de la Commission européenne (2002) en matière de compétences et de mobilité, la recommandation insiste sur le fait que tous les citoyens aient la possibilité d'acquérir gratuitement certaines compétences clés ; en dehors des aspects technologiques et d'aptitude favorisant l'apprentissage, l'accent est mis sur les bases en lecture, écriture et calcul. Ces réflexions s'appuient sur la question des standards de formation qui sont l'un des impacts principaux des enquêtes d'évaluation comme celles de l'IEA ou de PISA/OCDE, pour les élèves, ou IALS/OCDE pour les adultes. Dans le cas français, surtout depuis la mise en place du collège unique, l'idéal républicain d'un socle commun de connaissances et de compétences a été maintes fois réaffirmé. Avec le débat toujours latent que pour arriver à un socle commun, il faut tenir compte de la territorialité des savoirs et favoriser des politiques locales et la discrimination positive, comme lors de la création des ZEP, Charlot (1994). La commission Fauroux (1998) propose une définition des savoirs primordiaux des enfants, à la fin de la scolarité obligatoire, dont l'acquisition doit être garantie à tous avec en particulier une liste de compétences de base au-delà des compétences techniques et de citoyenneté. Parmi ces savoirs primordiaux certains se rattachent directement aux apprentissages fondamentaux :

- lire, écrire, parler correctement et aisément le français ;
- calculer, connaître les figures et les volumes, être familier de la proportionnalité et savoir apprécier les ordres de grandeur ;
- se situer dans l'espace et le temps, et d'abord dans son environnement immédiat.

Le rapport Thélot (2002) précise ce socle commun : il ne constitue pas la totalité de ce qui est enseigné à l'école et au collège, mais doit contenir ce qui est indispensable pour réussir sa vie d'adulte. Le rapport distancie ce socle par rapport aux programmes : « Le socle commun des indispensables ne s'identifie pas aux programmes tels qu'ils sont actuellement en vigueur à l'école ou au collège. D'abord, parce que, contrairement à eux, il ne délimite pas un périmètre idéal de ce que le bon élève devrait théoriquement savoir ; ensuite parce que dans les faits, les programmes constituent un assemblage parfois lourd, manquant de cohérence, peu motivant et dont les évaluations montrent que trop d'élèves échouent à les maîtriser » Thélot (2002, p.38-39). Toutefois le rapport se refuse de délimiter le socle, puisqu'il doit émaner d'un choix politique. De fait, le Haut Conseil de l'éducation a été institué par la loi du 23 avril 2005 d'orientation et de programme pour l'avenir de l'école. C'est cette loi, issue des conclusions du rapport Thélot qui a créé la notion de "socle commun de connaissances et de compétences" à acquérir par les élèves et le Haut Conseil est chargé de définir le contenu du socle commun de connaissances et de compétences que tous les élèves devront maîtriser à 16 ans.

2. Le socle d'un savoir de base comme norme d'un bien public universel

Cette question de socle s'apparente donc à un processus de validation des apprentissages. Pour cette validation, deux moments particuliers paraissent se présenter, en premier, celui de la fin de la scolarisation obligatoire et celui de l'entrée dans la vie active. Ce socle particulièrement à la seconde étape correspond à un mélange de compétences abstraites essentielles et de savoir-faire dont les essais autour des résultats et des exploitations secondaires des enquêtes PISA ont montré la complexité d'analyse. Une seconde vision serait celle des apprentissages fondamentaux et répond aux compétences de bases fondamentales

telles qu'elles étaient identifiées par le rapport Fauroux. Ici le socle devrait donc concrétiser l'acquisition des savoirs de base de manière homogène pour l'ensemble d'une génération. On retrouverait un objectif de qualité universelle prenant corps autour de quatre concepts relatifs aux procédures et aux effets de l'éducation : les résultats, le rendement, l'efficience, l'efficacité. En suivant Bailly et Chatel (2004), la qualité est une mesure de processus comme la performance à une enquête d'évaluation des élèves ; mais, aussi, tout autant une vision extensive des effets qu'elle a produits sans intention, comme les attitudes sociales et les externalités des effets de pairs. Cette question pose alors l'articulation de la compétence comme bien universel. Dans la mesure où existe un objectif d'éducation de base universelle, dans les objectifs du millénaire, l'éducation peut être perçue comme un bien public impur parce qu'elle présente, comme la plupart des biens publics mondiaux, des degrés variables de non rivalité ou de non exclusion. Pour les pays en voie de développement, l'éducation serait moins accessible comme bien public, selon Tubiana et Sévérino (2002), étant la partie "bien public" d'un bien privé. Ceci se réfère aux défaillances des marchés et aux recours aux arrangements collectifs pour y remédier. Ainsi, la coopération internationale peut être améliorée sans remettre en cause les équilibres de pouvoirs actuels. A l'évidence l'accès universel à l'instruction ainsi décliné dans les ODM correspondrait à un actif essentiel de l'humanité ; ceci n'est pas sans rappeler le second principe de la justice au sens de Rawls. Si les inégalités sociales et économiques peuvent exister, elles doivent satisfaire à deux conditions : 1 / elles doivent d'abord être attachées à des fonctions et à des positions ouvertes à tous, dans des conditions de juste (*fair*) égalité des chances, et 2 / elles doivent procurer le plus grand bénéfice aux membres les plus désavantagés de la société. De fait, pour l'éducation, la théorie de la justice stipule que ce second principe est inhérent à l'existence du premier : chaque personne possède un accès égal à un système pleinement adéquat de libertés de base égales pour tous. Ceci nous renvoie bien vite à un dilemme proche de celui de la « poule et de l'œuf » : l'accès au bien éducation, second principe, peut être conditionné, pour les plus désavantagés, par la défaillance au respect du premier principe. Ceci est l'illustration parfaite de la limite de la demande de l'éducation dans les pays les plus pauvres par les coûts d'opportunité : les jeunes enfants ne peuvent fréquenter l'école car devant contribuer, par les revenus de leur travail, à la subsistance d'eux-mêmes et de leurs proches. Ceci a donc conduit à une nécessité d'action vers les plus pauvres et ainsi évité que l'éducation y soit accaparée pour des intérêts personnels. A l'évidence, ceci induit d'un point de vue économique l'importance de l'aide pour combler les défaillances qui contraignent l'objectif : la scolarisation universelle ; mais tend surtout à privilégier le résultat, par rapport aux modes d'organisation. Ainsi l'important est l'acquisition de ces apprentissages, obligations de résultats, indépendamment de l'organisation de l'école. Ainsi une école du projet BRAC au Bangladesh est efficace puisqu'elle s'adapte à un contexte hostile à la fréquentation scolaire et ceci en s'affranchissant du cadre d'organisation usuel de l'école. Cette nécessité d'action justifie l'aide mais aussi et surtout un certain contrôle de la qualité et de l'universalité du bien éducatif reçu par chacun. Contrôle nécessaire, si l'on juge indispensable que l'éducation possède une finalité en restant une activité utile au plein sens économique dans les PED, Pritchett (2001).

Si cette articulation peut paraître convaincante dans l'absolu, peut-on lui donner un sens pratique permettant de conduire à une mesure de la qualité des apprentissages fondamentaux qui ne soit pas trop ancrée à un contexte local ? Une partie des opinions a critiqué l'utilisation de tests d'acquisition scolaire comme PISA, car prétendue trop réductrice de l'adéquation des systèmes éducatifs à la richesse culturelle d'un pays ou d'une société. Ici dans un contexte à la fois de similitude d'outils de mesure et de communautés de langage, nous souhaitons tester si ces apprentissages de base peuvent être représentés sur une même échelle lorsqu'ils

s'attachent à décrire le contexte socio-économique distant et aussi sans passer par la prise en compte de ces contextes pour conserver de ce point de vue une échelle neutre.

3. Contexte de mesure

Le programme d'analyse des systèmes éducatifs de la CONFEMEN (PASEC) réalise depuis dix ans des évaluations des systèmes éducatifs pour les niveaux de deuxième et cinquième année de l'enseignement primaire. Les tests administrés aux élèves ont été conçus au Sénégal par une équipe conjointe du CIEP¹ et de l'INEADE², ceci au moyen de tests conçus au Sénégal par des inspecteurs et conseillers pédagogiques sénégalais. Ces tests sont cadrés sur les compétences de base (lire, écrire, compter), avec une approche plus centrée « résolution de problèmes » qu'une démarche de stricte validation du programme. Utilisés depuis 10 ans par les enquêtes du PASEC, plus de 1000 écoles et de 35 000 élèves ont été enquêtés au moyen de ces tests dans 11 pays. Même si le français est dans ces analyses majoritairement la langue d'enseignement, certains facteurs locaux, comme la pratique dans le milieu familial d'une langue nationale peuvent introduire des biais dans toute comparaison. Aussi l'analyse menée sur le Cameroun, lors de l'année scolaire 1995-1996 peut retenir l'attention. Dans les pays de l'Afrique subsaharienne francophone, et ceci évidemment pour les provinces non anglophones, le Cameroun est le pays où la langue française est la plus largement répandue comme vecteur d'expression quotidienne. Cette situation s'explique par la très forte variété des langues locales recensées suivant le dénombrement des atlas linguistiques³ ; cette dernière situation du choix pour une langue d'échange s'explique puisque l'inventaire linguistique a mis en évidence près de 270 langues locales, se répartissant entre plus de 20 groupes de langues africaines⁴. De ce fait, les premiers échanges sur de plus vastes territoires on conduit à adapter la langue du pays colonisateur comme langue d'échange⁵.

Les enquêtes du PASEC sont des enquêtes de valeur ajoutée scolaire, c'est-à-dire qu'elles se déroulent en deux temps : un test initial de début d'année complété par un test en fin d'année. La comparaison de ces deux tests permettrait ainsi d'apprécier les progrès de l'élève ou valeur ajoutée scolaire. Ces tests, partagés en deux groupes, autorisent respectivement d'évaluer les compétences en expression française et en mathématiques.

	Pré-test	Post-test
Français	32	41
Mathématiques	34	41
Ensemble	66	82

Tableau 1 : *nombre d'items aux tests PASEC Cameroun*

Les tests ont été administrés sur 122 classes du 5^e grade, ainsi plus de 2000 élèves ont passé pré et post-test. On note que le PASEC de manière générale se contente de limiter

¹ Centre international d'études pédagogiques (France).

² Institut national d'études et d'actions pour le développement de l'éducation (Sénégal).

³ D'ailleurs l'enquête a assez mal couvert les provinces du Nord Cameroun, provinces où justement la place du français comme vecteur de communication fait place parfois à des langues locales.

⁴ Voir Breton R. et , Fohtung B. (1991), Atlas administratif des langues nationales du Cameroun, Paris, ACCT, et Yaoundé, Cerdotola, 143 p.

⁵ Période de colonisation elle même complexe, puisqu'à l'origine la puissance colonisatrice était l'Allemagne ; avec un partage, dans un second temps, entre l'Angleterre (les deux provinces de l'Oust proche du Nigéria) et la France. Les provinces anglophones ne sont pas prises en compte dans l'enquête scolaire du PASEC.

l'échantillon d'élèves retenus dans une classe à 20⁶. Les écoles sont échantillonnées en fonction d'un sondage par grade reprenant la diversité de l'habitat (régions, urbain/rural) et les principaux types d'écoles.

Dans le contexte d'une recherche sur la nutrition, il est apparu souhaitable de faire passer un test de valeur ajoutée scolaire à des élèves vivant en France dans l'agglomération de Dijon, enfants suivant un cours de CM1, et ceci, au cours de l'année scolaire 2005-2006. De fait, l'expérience devait porter sur 8 classes et là tous les élèves de la classe passaient le test. Ici le choix d'école ne s'est pas réalisé par tirage aléatoire, mais par la sélection de classes qui présentaient la composition de la CSP du chef de ménage des familles des élèves les plus proches de celle de la moyenne nationale⁷. A la différence du PASEC, tous les élèves de chaque classe ont été testés, ce qui conduit à ce que 198 élèves, représentant l'effectif total de 8 classes issus de 8 écoles passent les deux tests, pour des tailles de classe entre 19 et 27 élèves⁸. On peut noter que le CM1 est en France le 4^e grade de l'enseignement primaire. Toutefois avec la généralisation du cycle préélémentaire, le niveau y est comparable suivant les pédagogues, en termes d'objectifs des acquisitions, avec le CM1 du Cameroun de grade 5 là où la découverte de l'école se fait lors d'un cycle préparatoire de deux années. La graduation des acquisitions pour les mathématiques est donnée, pour les six années de l'enseignement initial, dans le cas français, par l'annexe F.

Le test PASEC prévoit des durées très précises de passation des tests. Le reproduire à l'identique à Dijon s'est heurté à l'objection de la tutelle de l'inspection qui ne souhaitait pas que la passation du test accapare la classe plus de 75 minutes compte tenu des temps additionnels de préparation et de mise en condition. Il a fallu réduire la taille du test de 7 et 8 items, pour les deux disciplines, respectivement pour le pré-test et le post-test. Le choix d'élimination s'est fait en tentant de ne pas biaiser les grands domaines de compétence testés. Puis par la référence aux points bisériaux⁹ en éliminant les items les plus longs en temps de passation et dont les valeurs des points bisériaux et les variations de l'alpha de Cronbach¹⁰ en cas de retrait de ces items montraient qu'ils se situaient en relativement faible adéquation avec le résultat moyen sur l'ensemble des items, ceci pour chaque élève. Une seconde difficulté est apparue par rapport aux programmes actuels, et surtout à leur pratique, en France. Des items du test PASEC sont à la frange du programme de CM1 à l'exemple des calculs de la surface et de la circonférence du cercle. Dans certains cas, ceci a conduit à renforcer le ciblage de l'élimination nécessaire de certains items par la contrainte temps. Dans d'autres cas, ceci a conduit à reformuler, lors de la passation du test, l'item dans un mode plus en adéquation à la pratique pédagogique, ce point est précisé dans l'annexe E : « adaptation des instruments à la

⁶ Si tous les élèves sont pris en compte pour les classes inférieures en taille, un tirage au sort pour les plus grandes classes limite l'échantillon à 15 élèves.

⁷ Cet exercice s'est trouvé simplifié dans la mesure où justement les différences inter-écoles sur ce critère sont faibles dans l'agglomération de Dijon et que par ailleurs ce district possède une composition sociale proche de celle de la moyenne nationale. Ceci a pu permettre d'apporter facilement un substitut aux quelques classes qui ont décliné la participation à cette recherche.

⁸ Les huit écoles de la sélection présentent aussi des structures très proches, chacune étant organisée avec 2 classes pour chaque grade ; cette organisation étant par ailleurs la plus courante en France en milieu urbain.

⁹ La notion de point bi-sériel, s'attache à décrire la cohérence d'un item par rapport à l'objectif global du test dans lequel il s'inscrit. Dans ce sens, le point bi-sériel se rapporte à un item et estime la probabilité de réussir cet item et le test global. Il se définit plus concrètement par un coefficient de corrélation entre la note obtenue à l'item spécifié et le score global obtenu au test.

¹⁰ Le coefficient alpha de Cronbach est un indicateur de la fiabilité d'un test. Une valeur élevée de l'Alpha de Cronbach – dont le maximum est 1 – garantit que les tests mesurent effectivement le niveau d'acquisition des élèves. Dans le cas du Niger, l'alpha de Cronbach est supérieur à 0,70 à tous les tests.

pratique et au programme en France ». Ceci entraîne, dans quelques cas, à modifier le sens de la question, ainsi les questions sur le calcul autour du cercle n'ont pas fait appel à une connaissance acquise, mais à une présentation d'un formulaire basé sur son application pratique immédiate par l'exemple. Toutefois, ces items ont été clairement identifiés, ceci pour 5 items supplémentaires au post test et autorise éventuellement à retenir un traitement différencié. Enfin, signalons que la correction des épreuves a été centralisée et effectuée sur la base du barème commun, à l'exemple de ce qui est pratiqué par PASEC.

4. Application

Un premier point a pour but de rappeler les résultats pour le Cameroun et de les recalculer sur le seul résultat des items « comparable Dijon », soit que les items n'aient pas été retenus pour la contrainte de temps ou pour la question d'adéquation stricte au programme en usage.

	Note/100	Erreur type	Plage incertitude à 5% de risque	
Initial français	53,44	0,38	52,70	54,18
Final français	55,18	0,40	54,40	55,96
Initial mathématiques	56,67	0,39	55,90	57,44
Final mathématiques	49,80	0,37	49,07	50,54
Initial français et mathématiques	55,06	0,35	54,37	55,74
Final français et mathématiques	52,49	0,34	51,82	53,16
Initial français – comparable Dijon (- 2 items)	53,14	0,37	52,30	54,48
Final français – comparable Dijon (-4 items)	55,46	0,38	54,71	56,22
Initial mathématiques– comparable Dijon (-5 items)	58,29	0,40	57,50	59,08
Final mathématiques– comparable Dijon (-9 items)	55,68	0,43	54,84	56,53
Initial français et mathématiques– comparable Dijon	55,87	0,35	55,18	56,55
Final français et mathématiques – comparable Dijon	55,57	0,36	54,87	56,28

Tableau 2 : Résultats sur 100 aux *items PASEC Cameroun*

L'essentiel de l'écart vient donc de la nouvelle définition des items du test final en mathématiques. A ce niveau, on peut calculer les scores pour Dijon, ceux-ci sont représentés dans le tableau suivant. On remarque en tout premier lieu la question des erreurs types. Suivant la théorie statistique, cette marge d'incertitude est liée à la taille des échantillons. Aussi ces écarts sont à apprécier par rapport à la racine carrée du rapport des tailles sur les populations testées (soit $\sqrt{(N1/N2)} = \sqrt{(2030/198)} \approx 3,1$); ceci nous donne une valeur acceptable de comparaison des grandeurs par rapport aux plages d'écarts enregistrés, entre les deux échantillons, et peut être avancé comme un indicateur de vraisemblance pour la distribution propre à chacun des deux échantillons.

	Moy/100	Erreur type	Plage incertitude	
Initial français	71,68	0,87	69,87	73,40
Initial mathématiques	84,04	0,94	82,18	85,86
Final français	72,78	0,98	70,84	74,72
Final mathématiques	60,79	1,05	58,73	62,86
Final français items adéquation « programme »	75,77	1,00	73,81	77,74
Final math.items adéquation « programme »	62,21	1,07	60,09	64,33
Final tous les items	65,78	0,87	64,07	67,50
Final tous les items adéquation « programme »	68,32	0,90	66,50	70,09

Tableau 3 : Résultats sur 100 aux *items PASEC comparables strictement Dijon*

De cette comparaison, on pourrait retenir un avantage de 20 points des élèves de Dijon aux tests initiaux de français et seulement de 7 points au test initial de mathématiques. À peu de choses près, ces écarts se conserveraient, du moins dans la plage d'incertitude lors du test final. La réintroduction des items « aux limites du programme » pour Dijon, ne conduit à aucun résultat significatif en différences pour le français, mais conduirait à réduire le score de 1,5 point en moyenne au test final de mathématiques.

Dans un exercice proche, Jarousse et Mingat (1993) en complément d'une évaluation sur le système éducatif du Togo ont souhaité mener une comparaison prenant en compte des élèves de CM1 du département de Côte-d'Or. Ces élèves ont passé les mêmes tests de fin d'année scolaire que leurs camarades togolais. Ceci a été réalisé en appliquant les consignes et les conditions de passation des épreuves utilisées au Togo. Dans les deux matières, les scores des élèves togolais s'avèrent aussi inférieurs à ceux des élèves des écoles françaises. La différence entre les acquis des écoles françaises et togolaises est là aussi importante en français puisque les jeunes élèves de l'échantillon français de Côte-d'Or obtiennent un résultat post-test de 12% supérieurs en mathématiques et de 28% en français. Dans notre comparaison avec le Cameroun, les écarts respectifs sont de 19% et de 32%

Ces différences supérieures à l'encontre du Cameroun seraient toutefois contraires aux résultats d'un test du PASEC (voir tableau A1, annexe A) où les résultats pour le Cameroun se situent, tant en français qu'en mathématiques, dans des valeurs plus hautes que les résultats enregistrés pour le Togo. Toutefois, cet exercice comparatif est très délicat à mener dans la mesure où l'on fait référence à deux évaluations sur le Togo menées à dix années d'intervalle.

Une comparaison plus fine peut se réaliser en ne s'arrêtant pas aux écarts sur la performance globale, mais en appréciant cette performance sur quelques domaines de compétences ; ces derniers étant organisés autour du cumul des résultats pour un ensemble d'items qui tendent à tester les capacités de l'élève dans ce domaine de savoir. Jarousse et Mingat (1993) réalisent ceci, p.176 et suivantes, dans leur comparaison. Sur les résultats en mathématiques, il apparaît nettement que les élèves de Côte-d'Or surperforment relativement en géométrie et conversion et mesure alors que les résultats en résolution de problèmes seraient en relatif moins à leur avantage. Pour l'expression française, les avantages comparatifs des élèves de Côte-d'Or seraient la grammaire et, à l'inverse, ils en posséderaient moins en compréhension de textes.

Comme le soulignent Jarousse et Mingat (1993), tester des équivalences de niveau entre systèmes reste très délicat : ils soulignent explicitement : « il serait excessif de tirer des résultats présentés, des indications négatives quant à l'équivalence réelle des niveaux d'acquisitions des élèves dans les systèmes togolais et français » (p 185). Leur argumentaire se rattache à deux raisons :

-Les élèves togolais, ne sont pas majoritairement élevés dans un environnement familial francophone et subissent un handicap initial important qui ne peut être comblé que progressivement;

-La question de l'équivalence des savoirs se pose surtout en fonction de la comparabilité des bases, au Togo, en 1989 seulement 40% de la génération était scolarisée en CM1 et moins de 20% seulement de ces scolarisés a eu accès au second cycle secondaire. Pour le Cameroun, en 1995 la probabilité d'une génération d'accéder en fin de secondaire, certes plus élevée qu'au Togo, n'atteignait pas les 10%. Cette régulation des flux d'élèves par l'offre entraîne évidemment des conséquences en termes de sélection de ceux qui subsistent dans le système. Plus cette frange se réduira, plus les attentes envers l'école seront fortes pour parvenir à cette équivalence des savoirs.

5. Retour sur le concept de valeur ajoutée scolaire

La double comparaison entre les acquisitions de nos échantillons d'élèves en France et dans l'Afrique francophone souligne l'importance du rôle de la langue d'enseignement. Pour nous en convaincre, le tableau A1 de l'annexe A présente une collation des résultats réalisés avec des tests homogènes. On peut remarquer que le résultat du test en mathématiques pour Madagascar est le plus élevé des pays du Sud, s'il reste significativement différent de celui enregistré sur Dijon au seuil de risque de 5% ; il ne l'est plus si on limite le risque d'indifférence à 1%. Ceci montrerait que l'acquis dans les compétences d'expression courante serait un facteur d'efficacité pour l'apprentissage des mathématiques. Cette complémentarité des savoirs a été mise en évidence par les comparaisons internationales menées par les enquêtes de l'IEA, puis plus tard par PISA. Dans le contexte des pays du Sud, ce constat n'est jamais établi sereinement, car souvent la question de la langue d'enseignement y est abordée au travers d'a priori ethno-politiques. Ainsi, Naumann et Wolf (2001) mettent en cause la priorité accordée aux manuels scolaires en matière de politique éducative. Pour eux, ceci ne se révèle efficace que si l'on cherche en même temps des moyens didactiques pour que ces manuels prennent en compte la problématique de la langue du milieu de l'élève par rapport à la langue d'enseignement, et là il faut tenir compte des forces d'inertie.¹¹ Les conséquences en sont la présence de facteur de renforcement des inégalités, puisque les familles les plus aisées ont un accès plus facile à la langue d'enseignement ce qui relativement pénalise les plus pauvres. L'analyse réalisée par Naumann et Wolf a été contestée par sa méthode, CONFEMEN (2002) ; il reste néanmoins l'idée qu'une langue d'enseignement non majoritairement répandue dans le cadre local créerait un déficit cumulatif de maîtrise des savoirs. Ceci montrerait pour notre sujet le rôle central de la déficience des compétences d'expression qui par ricochet pénaliserait la valeur ajoutée scolaire.

Une tentative pour préciser ces derniers éléments peut venir d'une approche par grandes compétences, pour ceci nous pouvons reprendre le groupement proposé par le PASEC, dans la présentation des résultats, des items utilisés dans les présents tests.

Ces compétences sont pour l'expression française :

- Compréhension de phrase ;
- Compréhension de texte ;
- Grammaire ;
- Conjugaison ;
- Orthographe.

Que l'on pourrait dans un second temps rassembler en deux macrocompétences en regroupant les tests de compréhension, puis les tests grammaticaux au sens large. Du côté des mathématiques, les compétences seraient :

- Opération ;
- Numération ;
- Mesure ;
- Géométrie ;
- Problème.

Là aussi, dans un second temps, deux macrocompétences pourraient être distinguées, les trois premières donnant le calcul et les deux autres la résolution de problèmes.

¹¹ Volonté justifiée par le fait que la langue d'enseignement « cimente » le secteur éducatif et que si des langues locales étaient en usage, elles feraient éclater ce secteur qui serait ainsi plus faible et plus à la merci de pressions locales.

Le tableau 4 donne la comparaison des résultats. Pour le français, les plus nets avantages comparatifs des élèves de Dijon se remarquent pour la compréhension de phrase, la grammaire et l'orthographe. Ceci s'accorde, en partie, avec les conclusions de Jarousse et Mingat précitées, puisque dans le cas présent aussi on remarque un plus fort avantage comparatif des élèves camerounais pour la compétence de compréhension de texte. Pour les compétences de mathématiques, les deux cas montreraient un moindre avantage comparatif des élèves de Dijon et de Côte d'or pour la résolution de problèmes.

Compétences	Cameroun		Dijon	
	Part réussie	Erreur type	Part réussie	Erreur type
Compréhension de phrase	0,72	0,28	0,93	0,15
Compréhension de texte	0,46	0,23	0,71	0,16
Grammaire	0,52	0,20	0,62	0,18
Conjugaison	0,59	0,33	0,88	0,23
Orthographe	0,58	0,30	0,75	0,28
Opérations	0,48	0,25	0,65	0,30
Numération/conversion	0,47	0,20	0,75	0,25
Géométrie	0,41	0,32	0,61	0,30
Problème	0,38	0,28	0,51	0,25
Mesure	0,47	0,20	0,61	0,16

Tableau 4 : Scores comparés par champs de compétences

De manière générale dans les deux populations, il apparaît que les corrélations, pour un même champ de compétences, sont relativement faibles entre pré et post-test. Au-delà d'un fait sur le faible maintien des compétences acquises, on peut tout autant donner une interprétation basée sur le caractère quelque peu arbitraire de cette classification des items élémentaires en champs de compétences.

Tableau 5 Tests sur le Cameroun

N=2023

Score au test final	Score au test initial										Impacts autres
	Texte	Phrase	grammaire	conjugaison	orthographe	texte	Phrase	grammaire	conjugaison	orthographe	
Texte	0,1597	0,2053	0,2257	0,1971	0,1445	0,2092	0,2092	0,3819	0,2934	0,2617	géométrie
Phrase	0,0783	0,2236	0,1551	0,1922	0,0966	0,2092		0,5239	0,2743	0,2286	toutes
Grammaire	0,1372	0,2975	0,3185	0,3357	0,2449	0,2743			0,5179	0,4698	opération
Conjugaison	0,1168	0,2350	0,2095	0,2354	0,2005	0,2934		0,5179		0,4246	opération
Orthographe	0,0837	0,1971	0,2018	0,2400	0,2568	0,2617		0,4698	0,4246		opération
Numération	0,1924	0,2395	0,1441	0,1644	0,1960		Numération	Géométrie	Problèmes	Mesure	
Opération	0,2435	0,1617	0,1751	0,2244	0,3962	0,3936		0,2723	0,3794	0,4174	phrase
Géométrie	0,0961	0,1640	0,1341	0,1220	0,0873	0,2723		0,2464	0,3858	0,4552	phrase
Problèmes	0,1060	0,1395	0,0890	0,1698	0,1705	0,3794		0,2684	0,2684	0,3062	texte
Mesure	0,1980	0,2575	0,1448	0,1579	0,1931	0,4174		0,4552		0,3892	texte
								0,3062	0,3892		grammaire, ortho

Note : Ce tableau s'interprète comme une part de variance de l'item en ligne expliqué par l'item en colonnes.

Dans le cas du Cameroun, tableau 5, les compétences de grammaire et conjugaison apparaîtraient assez largement impactées au niveau du score final par le score initial ; a contrario les items d'expression paraissent comme des compétences plus autonomes. Les compétences de grammaire, au sens large, paraîtraient plus liées à la compétence d'opération pour les mathématiques (apprentissage des règles). Sur le plan des compétences en mathématiques, celle en « opération » paraît fortement influencée par les scores initiaux de « problèmes et mesure ». Alors que « géométrie et problèmes » paraissent plus autonomes, donc marquées par de modestes corrélations, tout en étant liées aux compétences initiales en compréhension de textes. Sur l'inertie des apprentissages représentée ici, partie droite du tableau, par la corrélation entre le score initial et le final pour un même groupe de compétences, de manière générale, ce phénomène est plus fréquent en français qu'il ne l'est pour les mathématiques. En dehors de la complémentarité déjà signalée entre grammaire et conjugaison, ces corrélations témoignent surtout de la relative indépendance entre ces groupes de compétences ; une vision rapide permettrait de conclure que les compétences sont en grande partie liées à l'ensemble des compétences acquises lors du test précédent.

Tableau 6 : Tests sur agglomération de Dijon

N=198

Note : du fait de la modification du questionnaire, la compétence initiale « mesure » est trop dissimblable ici, il n'y est pas fait référence.

Score au test final	Score au test initial										Impacts autres
	texte	Phrase	grammaire	conjugaison	orthographe	texte	Phrase	grammaire	conjugaison	orthographe	
Texte	0,2512	0,2693	0,2145	0,1563	0,1851		0,1888	0,3171	0,1679	0,1688	géométrie
Phrase	0,0606	0,0183	0,0046	0,0350	0,0696	0,1888		0,2939	0,1808	0,0492	géométrie, mesure
Grammaire	0,2316	0,2274	0,2020	0,2527	0,2864	0,3171	0,2939		0,2999	0,2790	mesure, numération
Conjugaison	0,1289	0,1120	0,1695	0,0996	0,2930	0,1679	0,1808	0,2999		0,1834	géométrie, mesure
Orthographe	0,2245	0,0745	0,2575	0,1925	0,5198	0,1688	0,0492	0,2790	0,1834		opération, problème
Numération	0,1173	0,1954	0,0291	0,2590			0,1372	0,2092	0,2691		texte
Opération	0,0231	0,1359	0,0161	0,1563		0,1372		0,0488	0,1228		phrase, texte; conjugaison
Géométrie	0,0861	0,0642	0,0305	0,2092		0,2092	0,0488		0,2016		conjugaison
Problèmes	0,1654	0,0625	0,0104	0,2351		0,2691	0,1228	0,2016			conjugaison
Mesure	0,2053	0,2426	0,0303	0,2691							grammaire, texte

Même s'il est évident que la comparaison est ici délicate avec le cas camerounais, du fait des différences de taille et de structure dans l'échantillon, on trouve ici certains points communs dont le premier est de montrer que la compétence en orthographe se confirme comme une compétence de base. A l'inverse, l'impact ici de la compréhension initiale de phrase paraît moins stratégique que dans le cas camerounais ; une explication rapide peut évidemment venir de la pratique différente du français dans la vie extrascolaire.

Par ailleurs et de manière générale, les corrélations entre test initial et test final sont largement moins importantes dans le cas dijonnais ; ceci peut évidemment venir de la faible taille de l'échantillon, mais aussi de la variété entre groupes, l'échantillon dijonnais présentant par sa construction une moins forte variété.

6. Exploitation des items communs entre les deux tests

Une autre possibilité de comparaison est offerte par les items communs aux deux étapes du test. Dans la structure du test appliqué à Dijon, 20 items sont communs au post et au pré-test et par ailleurs ces items sont aussi communs aux tests sur le Cameroun¹². Un problème vient toutefois du déséquilibre dans la répartition de ces items communs entre le français et les mathématiques, puisque seulement 15% concernent les mathématiques ; aussi seule une exploitation globale peut en être réalisée.

Dijon	Elèves/groupes	moyenne	ect	Min	Max
DEBUT	198/8	0,734	0,146	0,263	0,947
FIN	198/8	0,788	0,139	0,421	1,000
Cameroun					
DEBUT	2023/108	0,493	0,174	0,000	0,947
FIN	2023/108	0,568	0,192	0,053	1,000

Tableau 6 : Comparaison sur items communs

Dans les deux cas, il existe bien un meilleur progrès de réponse au post-test de 7,3% sur Dijon et 15,21% au Cameroun, ceci surtout si l'on tient compte de la performance relative, la création de valeur ajoutée scolaire y est plus prononcée. La comparaison est poursuivie en établissant un modèle de panel avec effets aléatoires estimé par maximum de vraisemblance. Si ce modèle paraît estimé de manière plus fiable dans le cas dijonnais (pseudo R2 construit de 0,28 contre 0,21), ceci s'estompe en partie si l'on prend en compte les degrés de liberté.

La corrélation interclasse apparaît beaucoup plus forte sur Dijon ($\rho=0,32$) qu'elle ne l'est sur les écoles camerounaises ($\rho=0,11$), ceci découle en grande part des modes de sélection des classes dans l'échantillon.

		Coef.	Std.	z .	P> z 95%	Confiance	
	Pré test	,4646979	,0602938	7,71	0,000	,346	,582
Cameroun.	_cons	,4479823	,047339	9,46	0,000	,355	,540
N=2023	/sigma_u	,0407368	,013689	,021084	,0787084		
Classes=108	/sigma_e	,1169709	,0062302	,1053756	,1298421		
	rho	,1081688	,0661589	,0265385	,2951653		
Likelihood-ratio test sigma_u=0:chibar2(01)=11.17 Prob>=chibar2=0.000							
	Pré test	,3234621	,0234482	13,79	0,000	,277	,369
Dijon	_cons	,4083023	,0155901	26,19	0,000	,377	,439
N=198	/sigma_u	,1025316	,0078533	,0882391	,1191391		
Classes=8	/sigma_e	,1468709	,0023737	,1422915	,1515976		
	rho	,3276651	,0346814	,2628431	,3981515		
Likelihood-ratio test of sigma_u=0: chibar2(01)= 542.22 Prob>=chibar2 = 0.000							

Tableau 7 : Comparaison au moyen d'un ajustement de panel à effets aléatoire

Une seconde illustration de cette comparaison des progressions en CM1 peut se réaliser au niveau des classes. Pour y parvenir dans le cas des écoles en France, nous les illustrons au

¹² Quelques modifications mineures ont été faites sur des noms propres ou d'objet afin d'en faciliter la compréhension dans chaque contexte local.

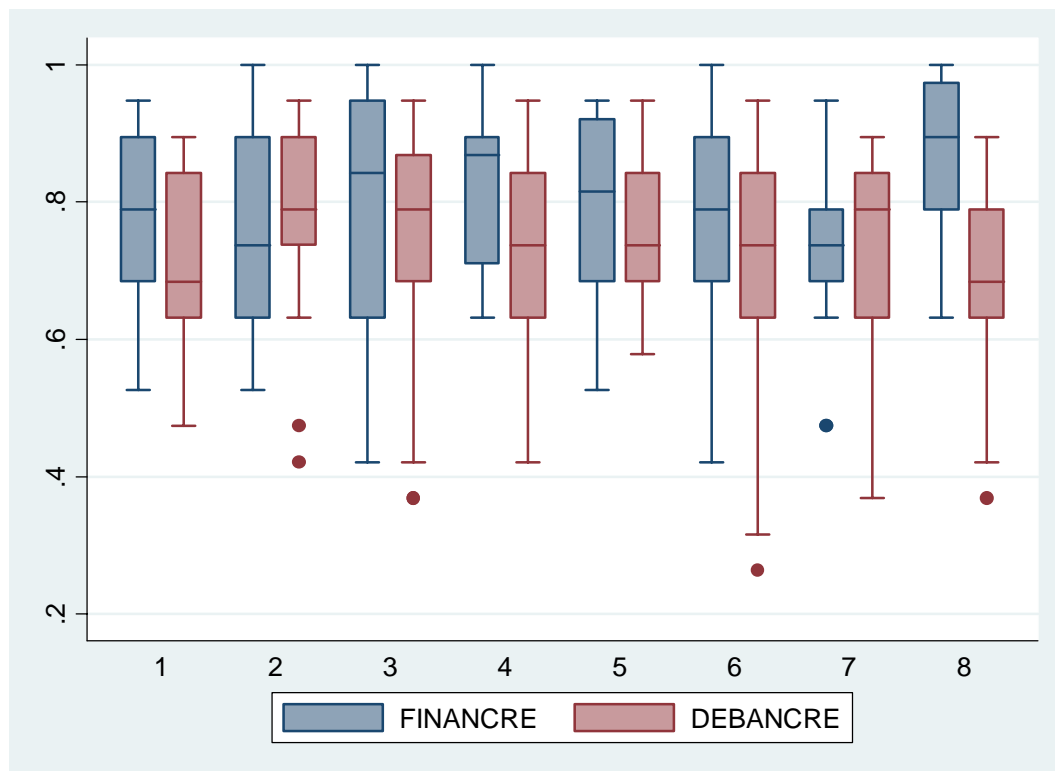
moyen de graphes de Tuckey ou plus prosaïquement « boîtes à moustaches ». Ce graphique donne un résumé du pré et du post-test pour les huit classes considérées.

Le test est donc représenté comme la somme de 20 items strictement identiques au pré (DEBANCRE) et au post test (FINANCRE) afin d'avoir un pur effet de valeur ajoutée.

Ce sont des boîtes à « moustache » classiques avec la médiane de distribution de classe comme barre dans la boîte, les 50% centraux comme taille de boîte, celle à 2 écarts-types pour les moustaches et éventuellement les outliers représentés par les points extrêmes.

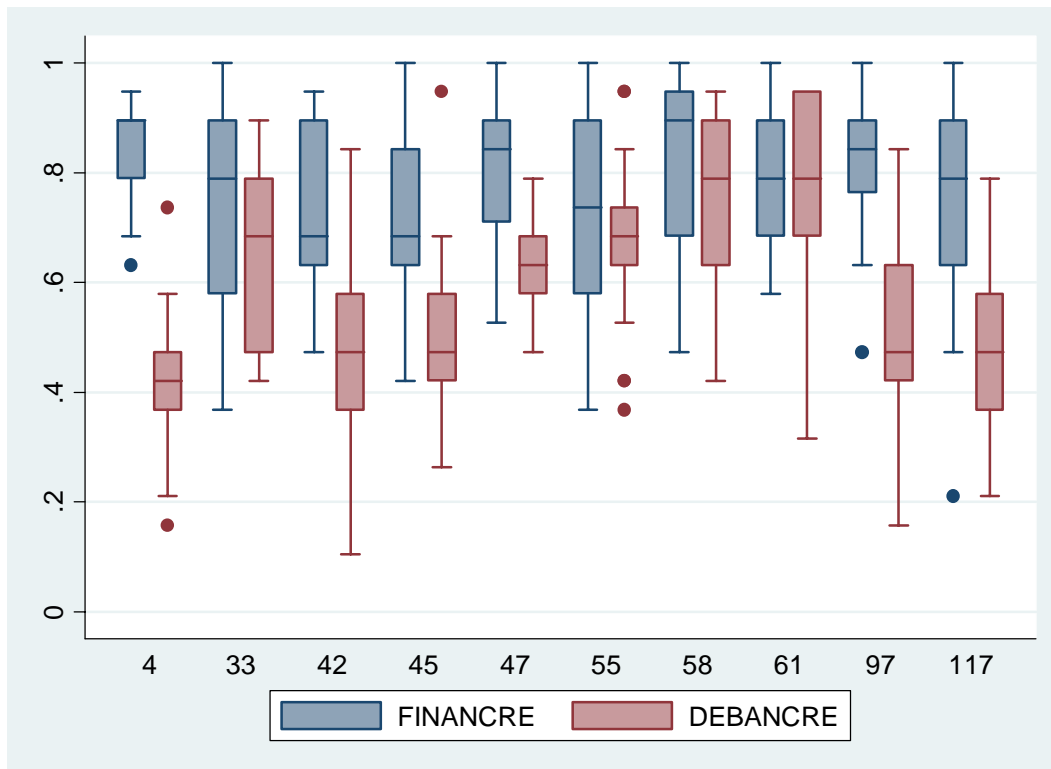
On enregistre un peu tous les cas :

- deux classes « régressent », mais une crée de l'homogénéité de manière nette,
- des classes « progressent », mais en créant de l'hétérogénéité ou en la maintenant,
- enfin deux, si ce n'est trois classes, « vertueuses » progresseraient tout en réduisant la dispersion.



Graphique 8 : La progression sur le cas des écoles dijonnaises.

Dans le cas du Cameroun, des choix doivent être faits pour rendre la comparaison plus directe tant sur la taille des classes que sur le niveau de performance. Comme cela a été mentionné, le PASEC constitue au maximum des échantillons de classe de taille 20¹³. Un certain nombre d'écoles très performantes au Cameroun sont caractérisées par des tailles de classe très faibles. Aussi a-t-on retenu de ne conserver que les classes dont au moins 15 élèves ont été pris en compte et pour lesquelles le niveau de performance au post test est comparable à celui enregistré en moyenne par les écoles dijonnaises. Ceci conduit, par appariement, à conserver 10 classes, celles-ci sont représentées dans le graphique 9. La comparaison entre ces deux derniers graphiques montrerait des hétérogénéités en grande partie comparables entre les deux références géographiques ; toutefois, les progrès sur ces items identiques paraissent avoir été plus significatifs dans le cas camerounais, mais ces progrès plus rapides se réaliseraient dans un contexte de plus grande variété de la performance au sein de la majorité des classes.



Graphique 9 : La progression sur le cas des écoles camerounaises sélectionnées.

¹³ Si la classe compte plus de vingt élèves, un tirage aléatoire est réalisé pour en obtenir vingt.

Bibliographie

Bailly F. et E. Chatel (2004), « Comment analyser la construction de la qualité éducative ? » Colloque Conventions et Institutions, Paris-X-Nanterre, repris in Problèmes économiques n° 2850, avril 2004, pp. 24-27.

CONFEMEN (2002), « Recherche empirique de la performance des systèmes africains d'enseignement primaire ». *CONFEMEN au Quotidien*, n° 45, février/mars 2001, pp. 10-11

Jarousse J.-P. et A. Mingat (1993), *L'école primaire en Afrique : Analyse pédagogique et économique (le cas du Togo)*. Paris, L'Harmattan.

Naumann, J. et P. Wolf (2001). *La performance de systèmes africains d'enseignement primaire - Critique et ré-analyse de données PASEC pour le Sénégal*. Perspectives, vol. XXI, n° 3, S, pp. 443-463

Commission européenne, Direction générale XXII - Éducation, Formation et Jeunesse, Direction générale V - Emploi, Relations industrielles et Affaires sociales (1996), *Livre blanc sur l'éducation et la formation. Enseigner et apprendre. Vers la société cognitive*. Luxembourg : Office des publications officielles des Communautés européennes.

Commission européenne, Direction générale de l'Emploi et des Affaires sociales, (2002). Plan d'action de la Commission en matière de compétences et de mobilité, Communication de la Commission au Conseil, au Parlement européen, au Comité économique et social et au Comité des Régions, COM (2002) 72, Commission européenne, Bruxelles.

Charlot B., dir. (1994), *L'école et le territoire : nouveaux aspects, nouveaux enjeux*, Paris : A. Colin.

Fauroux R. (1996). *Pour l'école : Rapport de la commission dirigée par R. Fauroux*, Paris : Calmann-Lévy - La Documentation française.

Thélot C. (2004), *Pour la réussite de tous les élèves : rapport de la Commission du débat national sur l'avenir de l'Ecole*, Paris : La Documentation française - SCEREN, 162 p.

Tubiana L., Severino J.M. (2002) "Biens publics mondiaux, gouvernance mondiale et aide au développement", in Jacquet P., Pisani-Ferry J., Tubiana L. (eds) "Gouvernance mondiale : Rapport de synthèse", *Conseil d'analyse économique*, Paris : la Documentation française.

Pritchett, L. (2001), "Where has all the education gone?", *World Bank Economic Review*, vol. 15, pp.367-391.

Annexe A Comparaison sur les scores aux tests

Tableau A1 : Résultats en français et mathématiques Post-test dans diverses études¹⁴

Année	Pays	Score de français (sur 100)	Score de mathématiques (sur 100)
1995-1996	Burkina Faso	44,0	46,1
1995-1996	Cameroun ¹⁵	55,1 (55,5)	50,0 (55,7)
1995-1996	Côte d'Ivoire	50,0	40,5
1997-1998	Madagascar	42,3	58,3 (*)
1995-1996	Sénégal	34,9	37,6
2000-2001	Togo	43,6	44,0
2003-2004	Tchad	28,8	32,6
2005-2006	RCA	23,8	27,8
2005-2006	Dijon¹⁶	74,3 (76,2)	60,8 (62,2)

Tableau A1 : Comparaison de scores en CMI suivant la même batterie de tests,

Source : collation des rapports PASEC

(*) Le test de mathématiques est administré en langue malgache,

¹⁴ Le résultat du Togo est à relativiser car la procédure d'échantillonnage pour ces enquêtes a été basée sur les enseignants et non sur la représentativité des élèves. Malgré tout, dans ces pays, plus de 1000 élèves ont été enquêtés à travers le territoire, ce qui rend confiant quand à l'idée que, bien que n'étant pas très précis, les scores moyens reflètent la qualité de l'éducation.

¹⁵ Le score entre parenthèses indique le calcul où sont éliminés les items qui peuvent susciter des interrogations quant à leur adéquation aux actuels programmes de CM1 en France. Il s'agit donc d'une mesure identique à celle des écoles dijonnaises, ce redressement n'a pas été réalisé dans les autres cas.

¹⁶ Le score, ramené sur une échelle sur cent, indiqué pour Dijon est celui enregistré sur l'ensemble des items administrés au test, le nombre entre parenthèses indiquant le calcul où sont éliminés les items qui peuvent susciter des interrogations quant à leur adéquation aux actuels programmes de CM1 en France.

Annexe B : Items et compétences

Table B : Tests PASEC, nombre d'items dans chaque macro compétence

Compétences	Français		Compétences	Mathématiques	
	Initial	Final		Initial	Final
Compréhension de phrase	6	6	Opération	12	8
Compréhension de texte	5	11	Numération	12	13
Grammaire	11	18	Mesure	4	4
Conjugaison	6	3	Géométrie	3	9
Orthographe	4	3	Problème	3	7
Total	32	41		34	41

Annexe C : Tests de cohérence sur les items

La construction d'échelle de performance scolaire relève un peu d'un dilemme. Comment tester un ensemble de savoirs multiples et variés tout en assurant la cohérence sur une échelle de performance, à caractère monodimensionnel, simple d'interprétation. La boîte à outil de l'évaluation scolaire tente de mesurer ces cohérences par divers moyens dont les plus répandus sont le calcul des points bisériaux et de l'alpha de Crombach.

Différences par le point bi-sériel

La notion de point bi-sériel, s'attache à décrire la cohérence d'un item par rapport à l'objectif global du test dans lequel il s'inscrit. Dans ce sens, le point bi-sériel se rapporte à un item et estime la probabilité de réussir cet item et le test global. Il se définit plus concrètement par un coefficient de corrélation entre la note obtenue à l'item spécifié et le score global obtenu au test.

Plus le point bi-sériel d'un item est proche de +1, plus la réussite à cet item s'accorde à la réussite du test tout entier. En d'autres termes, la mesure de la maîtrise par l'élève de l'objectif propre associé à cet item est une bonne mesure de la maîtrise générale, par cet élève, de tous les objectifs du test.

Un point bi-sériel de 0 révèle un manque de lien entre la réussite à un item particulier et la réussite au test. Il est donc dans ce cas, impossible de savoir, à partir de la réussite à cet item, si l'élève se situe parmi les bons ou parmi les mauvais.

Enfin, un point bi-sériel négatif indique que la réussite à l'item est corrélée négativement à la réussite à l'ensemble du test : c'est probablement, alors, que l'item est mal conçu, puisque les meilleurs élèves ont tendance à répondre faux à cet item, et les mauvais élèves à répondre de manière exacte.

Formule de calcul du point bi-sériel

Notons X_i la note que l'élève i a obtenue en répondant à l'item X et soit Y_i le score que ce même élève à obtenu pour tout le test auquel appartient l'item X .

- $X_i = 1$ si l'élève i trouve la bonne réponse que requiert l'item X ;
- $X_i = 0$ si l'élève ne trouve pas la bonne réponse.

Nous avons défini le point bi-sériel comme un coefficient de corrélation linéaire entre la variable qui mesure la réponse de l'élève à l'item spécifié (X) et le score global obtenu par l'élève (Y).

Notons pbi le point bi-sériel entre l'item X et le score Y de l'élève.

En tant que coefficient de corrélation linéaire entre X et Y , pbi est tel que :

$$pbi = \frac{\sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2} \times \sqrt{\sum_{i=1}^N (X_i - \bar{X})^2}}$$

On peut condenser cette formule en utilisant quelques propriétés particulières de la variable X et en utilisant l'écart type des scores obtenus par l'ensemble des élèves.

En effet, $\sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2} = \sqrt{\frac{N}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2} = \sqrt{N} \times \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2} = \sqrt{N} \times S$ où

S représente l'écart type des scores obtenus par l'ensemble des élèves ; et N représente le nombre total d'élèves ayant été soumis au test.

D'un autre côté on a :

$$\sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X}) = \sum_{i=1}^N X_i \times (Y_i - \bar{Y}) - \bar{X} \sum_{i=1}^N (Y_i - \bar{Y})$$

Or,

$$\sum_{i=1}^N (Y_i - \bar{Y}) = \sum_{i=1}^N Y_i - \sum_{i=1}^N \bar{Y} = \sum_{i=1}^N Y_i - N \times \bar{Y} = \sum_{i=1}^N Y_i - \sum_{i=1}^N Y_i = \mathbf{0}$$

Donc,

$$\begin{aligned} \sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X}) &= \sum_{i=1}^N X_i \times (Y_i - \bar{Y}) - \bar{X} \times \mathbf{0} \\ &= \sum_{i=1}^N X_i \times (Y_i - \bar{Y}) \\ &= \sum_{i=1}^N X_i \times Y_i - \bar{Y} \sum_{i=1}^N X_i \\ &= \sum_{i=1}^N X_i \times Y_i - N \bar{X} \bar{Y} \end{aligned}$$

$$\text{En effet, } \sum_{i=1}^N X_i = N \bar{X}$$

Mais, puisque $X_i = \mathbf{1}$ si l'élève i découvre la bonne réponse que requiert l'item X et que $X_i = \mathbf{0}$ si l'élève i donne une mauvaise réponse, alors on peut aussi écrire que :

$$\sum_{i=1}^N X_i = \sum_{i \in D} X_i + \sum_{i \notin D} X_i$$

$D = \{i / X_i = \mathbf{1}\}$ est l'ensemble des élèves qui trouvent la bonne réponse à l'item X . On considèrera entre autres que N_1 représente l'ensemble des élèves qui trouvent la bonne réponse à l'item X et N_0 représente l'ensemble des élèves qui proposent une mauvaise réponse à l'item X .

$$\text{On a donc } \sum_{i=1}^N X_i = \sum_{i \in D} X_i + \sum_{i \notin D} X_i = \sum_{i \in D} \mathbf{1} + \sum_{i \notin D} \mathbf{0} = N_1 \times \mathbf{1} + N_0 \times \mathbf{0} = N_1$$

Et alors on a $\sum_{i=1}^N X_i = N \bar{X} = N_1$ et $\bar{X} = \frac{N_1}{N} = p$ où p est la proportion d'élèves qui trouvent la bonne réponse à l'item X .

Ainsi,

$$\begin{aligned}
\sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X}) &= \sum_{i=1}^N X_i \times Y_i - N \times p\bar{Y} \\
&= \sum_{i \in D} X_i \times Y_i + \sum_{i \notin D} X_i \times Y_i - N \times p\bar{Y} \\
&= \sum_{i \in D} \mathbf{1} \times Y_i + \sum_{i \notin D} \mathbf{0} \times Y_i - N \times p\bar{Y} \\
&= \sum_{i \in D} Y_i - N \times p\bar{Y} \\
\sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X}) &= N_1 \bar{Y}_1 - N \times p\bar{Y}
\end{aligned}$$

Dans cette dernière égalité, \bar{Y}_1 représente le score moyen des élèves qui ont bien répondu à l'item X , et

$$\sum_{i \in D} Y_i = N_1 \times \bar{Y}_1.$$

On peut aussi décomposer le score moyen global \bar{Y} en fonction des scores moyens \bar{Y}_1 et \bar{Y}_0 , \bar{Y}_0 représente le score moyen des élèves qui n'ont pas bien répondu à l'item X .

$$\bar{Y} = \frac{N_1}{N} \bar{Y}_1 + \frac{N_0}{N} \bar{Y}_0$$

En utilisant cette dernière expression et sachant que $p = \frac{N_1}{N}$, on peut aussi écrire que :

$$\begin{aligned}
\sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X}) &= N_1 \bar{Y}_1 - N \times p\bar{Y} = N_1 \bar{Y}_1 - N \times \frac{N_1}{N} \times \left(\frac{N_1}{N} \bar{Y}_1 + \frac{N_0}{N} \bar{Y}_0 \right) \\
&= N_1 \bar{Y}_1 - N_1 \left(\frac{N_1}{N} \bar{Y}_1 + \frac{N_0}{N} \bar{Y}_0 \right) \\
&= N_1 \bar{Y}_1 - \frac{N_1}{N} N_1 \bar{Y}_1 - N_1 \frac{N_0}{N} \bar{Y}_0 \\
&= \left(\mathbf{1} - \frac{N_1}{N} \right) N_1 \bar{Y}_1 - N_1 \frac{N_0}{N} \bar{Y}_0 \\
&= \left(\frac{N - N_1}{N} \right) N_1 \bar{Y}_1 - N_1 \frac{N_0}{N} \bar{Y}_0 \\
&= \frac{N_0}{N} N_1 \bar{Y}_1 - N_1 \frac{N_0}{N} \bar{Y}_0 \\
\sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X}) &= \frac{N_1 \times N_0}{N} (\bar{Y}_1 - \bar{Y}_0)
\end{aligned}$$

Sachant aussi que $\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} = \sqrt{N} \times \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}$ et que

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2 &= \frac{1}{N} \sum_{i=1}^N (X_i - p)^2 = \frac{1}{N} \left(\sum_{i \in D} (X_i - p)^2 + \sum_{i \notin D} (X_i - p)^2 \right) \\
&= \frac{1}{N} \left(\sum_{i \in D} (1-p)^2 + \sum_{i \notin D} (0-p)^2 \right) \\
&= \frac{1}{N} \left(N_1 \times (1-p)^2 + N_0 \times p^2 \right) \\
&= \frac{N_1}{N} \times (1-p)^2 + \frac{N_0}{N} \times p^2 \\
\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2 &= p \times (1-p)^2 + \frac{N_0}{N} \times p^2
\end{aligned}$$

Posons $1-p = q$, on montre aussi que $q = \frac{N_0}{N}$

$$\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2 = p \times q^2 + q \times p^2 = pq(q+p) = pq \text{ car } p+q = p+1-p = 1$$

Donc

$$\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2 = pq = \frac{N_0 \times N_1}{N^2}$$

Ainsi,

$$\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} = \sqrt{N} \times \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2} = \sqrt{N} \times \sqrt{pq}$$

On montre ainsi que $pbi = \frac{\sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2} \times \sqrt{\sum_{i=1}^N (X_i - \bar{X})^2}}$ peut encore s'écrire

$$pbi = \frac{N_1 \times N_0}{N} \frac{(\bar{Y}_1 - \bar{Y}_0)}{\sqrt{N} \times S \times \sqrt{N} \times \sqrt{pq}} = \frac{N_1 \times N_0}{N^2} \frac{(\bar{Y}_1 - \bar{Y}_0)}{S \times \sqrt{pq}}$$

Puisque $\frac{N_1 \times N_0}{N^2} = pq$,

$$pbi = pq \times \frac{\bar{Y}_1 - \bar{Y}_0}{S \times \sqrt{pq}} = \frac{\bar{Y}_1 - \bar{Y}_0}{S} \sqrt{pq}$$

d'où la formule simplifiée suivante du point bi-sériel :

$$pbi = \frac{\bar{Y}_1 - \bar{Y}_0}{S} \sqrt{pq}$$

Cette formule a été appliquée à nos deux échantillons sur le Cameroun et les écoles de l'agglomération dijonnaise. Dans le premier cas les résultats sont sur le tableau C1 et concernent les 2006 élèves qui ont répondu à tous les items et ont participé au test initial et au test final. On le comprendra aisément le point bi sérial pose un dilemme dans ses valeurs de référence. On l'on recherche l'adéquation à une vision homogène de la performance

scolaire où l'on estime que diverses compétences peuvent se substituer dans la connaissance révélée par le test. Dans un essai de réponse, Cohen (1988), souligne cette difficulté qui doit combiner des règles autour des tests statistiques et le pur empirisme. Dans ce sens est recommandée une fourchette assez large de valeur du point bi sérial, une valeur inférieure à 0,1 serait le signalement d'un item à éliminer qui apporterait surtout de l'entropie au test. A l'inverse, la cohérence relative pourrait s'apprécier pour des valeurs au dessus de 0,4.

Cette règle de décision conduirait à ne rejeter aucun item des tests initiaux pour le Cameroun et à rejeter l'item Y pour le post-test de français, les items AA et AB pour le test de mathématiques. Dans le cas dijonnais la situation est plus contestable puisqu'au niveau du pré-test de français seraient rejetés les items T, X et Y. Pour le post-test de français l'item Y serait, comme dans le cas du Cameroun, concerné par le rejet, mais ici accompagné des items A et C. Le post-test de mathématiques serait le plus impacté avec cinq items P, T, V, W et AA. Un élément assez flagrant vient du rejet des items Y du post test de français et AA du post test de mathématiques pour les deux échantillons. Pour l'item Y, il s'agit de repérer un complément de lieu, et dans l'autre item AA de valider un ordonnancement de diverses fractions (comme pour AB, lui non repris sur Dijon vu la contrainte de temps). Dans le cas dijonnais se remarque cette question des items A et C post test de français sur les prépositions ; de fait, ces items figurent à l'identique pour le pré test (respectivement items D et F) et se caractérisaient par des valeurs faibles (tableau C2) ; le fait de les avoir placés en début de questionnaire aurait-il contribué à accroître l'aléa de réponse par effet d'angoisse ? Une seconde explication, complémentaire, peut venir d'un facteur d'oubli, ces prépositions ayant un aspect de situation spatio-temporelle ; comme cet aspect dépend d'apprentissages des années antérieures, il y aurait un facteur d'oubli net¹⁷. Pour l'item P en mathématiques post test, il s'agit d'une addition avec chiffres décimaux, l'item T est une simple reconnaissance de forme géométrique qui reprend les items AC et AD du pré test qui eux ne posaient pas ce problème. L'explication pour les items V et W est plus immédiate avec leur rapport à la géométrie du cercle, connaissance qui peut être, comme ceci a déjà été évoqué, considérée comme hors programme et dont l'approche a été reformulée dans l'application dijonnaise où l'on transite de l'application d'une connaissance à un exercice de mise en œuvre d'une formule. Comme de toute manière, il restera un risque quant à l'interprétation, il semble plus prudent d'éliminer ces deux items, dans le cas dijonnais, pour les tests qui vont suivre.

Définition de l'alpha de Cronbach

Lorsqu'un test est composé d'items dont les points bi-sériaux sont proches de +1, il présente une grande cohérence interne. L'alpha de Cronbach représente l'indicateur de synthèse de cette cohérence interne. Il varie entre zéro (test incohérent), et +1 (test totalement cohérent).

- Pour évaluer la réussite à un ensemble d'objectifs opérationnels, un test cohérent indique l'expression d'une compétence générale homogène. En d'autres termes, dans le cas du test de mathématiques, il indique qu'il y a bien de bons élèves en mathématiques, et pas seulement de bons élèves dans les divers champs de compétences, calcul, géométrie et algèbre, de la connaissance mathématique.
- 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.

Les critiques de l'alpha de Cronbach sont les mêmes que celles adressées au point bi-sériel : 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 alpha de Cronbach de niveau élevé. Si l'on ne voulait tester qu'un seul objectif opérationnel, il serait facile d'obtenir pour chaque item un point bi-sériel proche de +1, et pour le test en général un alpha de Cronbach également proche de +1. Aussi dans la pratique, l'alpha de Cronbach, sert surtout à opérer un arbitrage entre la longueur finale du test qui se traduit par le nombre d'items que l'on soumettra aux élèves, sans pour autant compromettre la fiabilité de la mesure. Si la variation de l'alpha qui suit la modification en nombre des items est limitée, alors la cohérence reste assurée. Dans le cas contraire, on sera allé trop loin, ou bien on aura exclu des items qui avaient un bon point bi-sériel et qui donc contribuaient à la bonne tenue de l'alpha ; ou alors on ajoute des items qui ne montreront pas les mêmes champs de compétences que ceux inclus à présent.

Calcul de l'alpha de Cronbach

L'alpha de Cronbach, se calcule suivant la formule ci-dessous.

¹⁷ La meilleure illustration vient de la baisse de performance de 10% enregistrée en moyenne sur ces items de prépositions formulés en termes identiques entre le pré et le post-test.

$$\alpha = \frac{K}{K-1} \times \left(1 - \frac{\sum_{k=1}^K p_k \times q_k}{S^2} \right)$$

La formule donnée ici s'applique seulement au cas où le test serait composé d'items à réponse binaire (vraie/fausse, 1 ou 0). Dans ce cas particulier, l'alpha de Cronbach s'appelle aussi coefficient de fidélité de Kuder-Richardson ou test KR20.

K Représente le nombre d'items du test

p_k est le pourcentage de bonnes réponses, ou index de facilité, qui a été obtenu pour l'item k du test et

q_k est le pourcentage de mauvaises réponses, ou index de difficulté, qui a été obtenu pour le l'item k du test.

Le tableau suivant donne les index moyens, calculés sur l'ensemble des items, et le coefficient de fidélité de Kuder-Richardson (KR20)

	Nombre items (k)	Index de facilité moyenne	Index de difficulté moyenne	KR20	Test de Feldt
Cameroun					
INIF	34	0,5305	0,2998	0,8068	
FINF	34	0,5513	0,3238	0,8337	
INIM	45	0,5599	0,3117	0,8221	
FINM	41	0,4989	0,3040	0,8366	
Dijon					
INIF	34	0,7168	0,2194	0,7175	0,9996
FINF	25	0,7326	0,2533	0,7413	0,9998
INIM	31	0,8404	0,3155	0,7677	0,9924
FINM	29	0,5879	0,2277	0,7131	1.0

Une question restante est celle de la comparaison de la difficulté entre nos deux références construites sur des échantillons indépendants comme par exemple ici l'échantillon du Cameroun et les écoles dijonnaises. La statistique W de Feldt (1969) permet de comparer les coefficients de fidélité, alpha de Cronbach ou KR20, pour deux échantillons indépendants. Cette commande nécessite, outre la valeur des deux coefficients de fidélité, la taille des échantillons et le nombre d'items dans chacun des tests. Ce test montre donc ici la quasi-certitude de l'hypothèse du rejet d'égalité des valeurs du KR20.

Tableau C1 Points bisériaux : tests Cameroun

Les items non repris dans le test sur Dijon sont indiqués en italique sur fond grisé

INI5F__A	0,4551	INI5M__A	0,3391	FIN5F__A	0,3135	<i>FIN5M__A***</i>	0,4330
INI5F__B	0,3713	INI5M__B	0,4719	FIN5F__B	0,3348	<i>FIN5M__B</i>	0,4128
INI5F__C	0,3454	INI5M__C	0,4599	FIN5F__C	0,3178	<i>FIN5M__C</i>	0,4114
INI5F__D	0,3247	INI5M__D	0,3843	FIN5F__D	0,3811	<i>FIN5M__D</i>	0,3818
INI5F__E	0,3021	INI5M__E	0,2764	FIN5F__E	0,4113	FIN5M__E	0,3731
INI5F__F	0,3900	INI5M__F	0,4402	FIN5F__F	0,4108	FIN5M__F	0,3996
INI5F__G	0,3185	<i>INI5M__G *</i>	<i>0,2864</i>	FIN5F__G	0,4112	FIN5M__G	0,3446
INI5F__H	0,3295	<i>INI5M__H</i>	<i>0,3349</i>	FIN5F__H	0,2505	FIN5M__H	0,3029
INI5F__I	0,2674	<i>INI5M__I</i>	<i>0,4257</i>	FIN5F__I	0,2364	FIN5M__I	0,3085
INI5F__J	0,4448	<i>INI5M__J</i>	<i>0,4554</i>	<i>FIN5F__J*</i>	<i>0,3864</i>	FIN5M__J	0,2401
INI5F__K	0,3985	INI5M__K	0,2805	<i>FIN5F__K</i>	<i>0,4355</i>	FIN5M__K	0,3324
INI5F__L	0,3589	INI5M__L	0,3415	<i>FIN5F__L</i>	<i>0,2783</i>	FIN5M__L	0,3660
INI5F__M	0,3845	INI5M__M	0,4576	FIN5F__M	0,5229	FIN5M__M	0,3478
INI5F__N	0,3622	INI5M__N	0,4675	FIN5F__N	0,5437	FIN5M__N	0,4087
INI5F__O	0,4258	INI5M__O	0,5108	FIN5F__O	0,5560	FIN5M__O	0,4035
INI5F__P	0,4651	INI5M__P	0,5298	FIN5F__P	0,4215	<i>FIN5M__P</i>	<i>0,2870</i>
INI5F__Q	0,4705	INI5M__Q	0,4744	FIN5F__Q	0,3806	FIN5M__Q	0,4394
INI5F__R	0,4818	INI5M__R	0,3226	FIN5F__R	0,4475	<i>FIN5M__R</i>	<i>0,3740</i>
INI5F__S	0,3051	INI5M__S	0,3524	FIN5F__S	0,3862	FIN5M__S	0,4103
INI5F__T	0,4877	INI5M__T	0,3203	FIN5F__T	0,4612	FIN5M__T	0,3012
INI5F__U	0,3007	INI5M__U	0,3225	FIN5F__U	0,4844	FIN5M__U	0,4741
INI5F__V	0,3682	INI5M__V	0,3783	FIN5F__V	0,4217	FIN5M__V	0,3808
INI5F__W	0,3500	INI5M__W	0,3686	FIN5F__W	0,4693	FIN5M__W	0,4169
INI5F__X	0,3666	<i>INI5M__X**</i>	<i>0,3501</i>	FIN5F__X	0,4147	FIN5M__X	0,3205
INI5F__Y	0,3602	<i>INI5M__Y</i>	<i>0,3239</i>	FIN5F__Y	0,1179	FIN5M__Y	0,3922
INI5F__Z	0,4235	<i>INI5M__Z</i>	<i>0,4351</i>	FIN5F__Z	0,2352	FIN5M__Z	0,3470
INI5F__AA	0,3570	<i>INI5M__AA</i>	<i>0,3982</i>	FIN5F__AA	0,3793	FIN5M__AA	0,0940
INI5F__AB	0,3732	INI5M__AB	0,2790	FIN5F__AB	0,3255	<i>FIN5M__AB</i>	<i>0,0720</i>
INI5F__AC	0,3500	INI5M__AC	0,3789	FIN5F__AC	0,3728	FIN5M__AC	0,3153
INI5F__AD	0,2686	INI5M__AD	0,3538	FIN5F__AD	0,3871	FIN5M__AD	0,3921
INI5F__AE	0,2987	INI5M__AE	0,3456	FIN5F__AE	0,1837	FIN5M__AE	0,3679
INI5F__AF	0,3577	INI5M__AF	0,3772	FIN5F__AF	0,3763	<i>FIN5M__AF</i>	<i>0,4333</i>
INI5F__AG	0,3673	INI5M__AG	0,3630	FIN5F__AG	0,5247	<i>FIN5M__AG</i>	<i>0,3066</i>
INI5F__AH	0,3931	INI5M__AH	0,3059	<i>FIN5F__AH*</i>	<i>0,5549</i>	FIN5M__AH	0,3883
				<i>FIN5F__AI</i>	<i>0,4407</i>	FIN5M__AI	0,2606
						FIN5M__AJ	0,3846
						FIN5M__AK	0,5103
						FIN5M__AL	0,5091
						FIN5M__AM	0,3782
						FIN5M__AN	0,4464
						FIN5M__AO	0,4332

Items non repris dans les tests sur Dijon

pré

* opérations, forte hétérogénéité de valeur des points bisériaux

** changement d'unités, mal assurée en début de CM1, vu les pratiques en France,

post

* compréhension de texte, items longs en temps de passation,

** phrase avec une proposition principale et une proposition subordonnée relative, difficultés adaptation,

*** opérations, temps disponible

Pour les autres items il s'agit soit de suppression d'items dans une série répétée ou liée à la valeur du point bisérial.

Tableau C2 Points biséaux : tests Dijon

INI5F__A	0,2479	INI5M__A	0,3600	FIN5F__A	0,1656	FIN5M__E	0,2930
INI5F__B	0,3473	INI5M__B	0,4581	FIN5F__B	0,2350	FIN5M__F	0,3402
INI5F__C	0,2250	INI5M__C	0,5081	FIN5F__C	0,0684	FIN5M__G	0,3830
INI5F__D	0,2194	INI5M__D	0,4975	FIN5F__D	0,2772	FIN5M__H	0,3560
INI5F__E	0,2220	INI5M__E	0,3013	FIN5F__E	0,2722	FIN5M__I	0,5057
INI5F__F	0,2486	INI5M__F	0,4493	FIN5F__F	0,3187	FIN5M__J	0,2168
INI5F__G	0,2297	INI5M__K	0,2609	FIN5F__G	0,4822	FIN5M__K	0,3678
INI5F__H	0,4154	INI5M__L	0,2640	FIN5F__H	0,2819	FIN5M__L	0,3392
INI5F__I	0,3041	INI5M__M	0,2746	FIN5F__I	0,3378	FIN5M__M	0,3621
INI5F__J	0,2253	INI5M__N	0,3621	FIN5F__M	0,5482	FIN5M__N	0,3300
INI5F__K	0,3912	INI5M__O	0,4651	FIN5F__N	0,5808	FIN5M__O	0,3540
INI5F__L	0,3003	INI5M__P	0,4487	FIN5F__O	0,5416	FIN5M__P	0,1853
INI5F__M	0,2797	INI5M__S	0,5394	FIN5F__P	0,5138	FIN5M__S	0,2079
INI5F__O	0,3011	INI5M__T	0,5393	FIN5F__Q	0,3925	FIN5M__T	0,1466
INI5F__P	0,3376	INI5M__U	0,4856	FIN5F__R	0,3180	FIN5M__U	0,4600
INI5F__Q	0,3376	INI5M__V	0,4654	FIN5F__S	0,2999	FIN5M__V	0,0750
INI5F__R	0,4037	INI5M__W	0,3864	FIN5F__T	0,3214	FIN5M__W	0,1109
INI5F__S	0,2069	INI5M__X	0,4235	FIN5F__U	0,3893	FIN5M__X	0,2250
INI5F__T	0,1284	INI5M__AC	0,3785	FIN5F__V	0,3333	FIN5M__Y	0,2459
INI5F__U	0,3627	INI5M__AD	0,3416	FIN5F__W	0,4160	FIN5M__Z	0,2551
INI5F__V	0,2398	INI5M__AE	0,2640	FIN5F__X	0,4078	FIN5M__AA	0,0534
INI5F__W	0,2450	INI5M__AF	0,3258	FIN5F__Y	0,1716	FIN5M__AD	0,2277
INI5F__X	0,0823	INI5M__AG	0,5444	FIN5F__Z	0,2413	FIN5M__AE	0,4061
INI5F__Y	0,0425	INI5M__AH	0,3689	FIN5F__AA	0,2069	FIN5M__AH	0,2640
INI5F__Z	0,4827			FIN5F__AB	0,3423	FIN5M__AJ	0,4645
INI5F__AA	0,3546			FIN5F__AC	0,3424	FIN5M__AK	0,4895
INI5F__AB	0,3481			FIN5F__AD	0,3842	FIN5M__AL	0,4874
INI5F__AC	0,5634			FIN5F__AE	0,4000	FIN5M__AM	0,3944
INI5F__AD	0,3670			FIN5F__AF	0,2340	FIN5M__AN	0,5304
INI5F__AE	0,3333			FIN5F__AG	0,3347	FIN5M__AO	0,5394
INI5F__AF	0,4297						
INI5F__AG	0,3843						
INI5F__AH	0,5282						

Bibliographie de l'annexe C :

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum.

Feldt L.S (1969) "A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests", *Psychometrika*, vol. 34, n° 3, pp. 363-373,

Cronbach, L.J. (1971). "Test validation". In *Educational measurement, 2e éd.*, sous la dir. de R.L. Thorndike, p. 443-507. Washington, D.C.: ACE.

Laveault, D. et J. Grégoire (1997). *Introduction aux théories des tests en sciences humaines*, Bruxelles : De Boeck Université, (Méthodes en sciences humaines)

Seheuneman J.D. (1987), "An Experimental, Exploratory Study of Causes of Bias in Test Items", *Journal of Educational Measurement*, Vol. 24, n° 2, pp. 97-118.

Annexe D : Détails des items de chaque test et champs de compétences couverts

Les items pouvant poser des problèmes, par rapport aux pratiques liées au programme, en France sont signalés par une case Code en grisé.

Tableau D1 : Pré-test de français 5ème année				
Domaines	Exercices	Items	Code dans la base	Objectifs
Compréhension de phrase	1	A	INI5F_A	identifier la phrase qui a le même sens que la précédente
		B	INI5F_B	
Compréhension de phrase	2	C	INI5F_C	identifier la préposition qui donne du sens à la phrase
		D	INI5F_D	
Compréhension de texte	15	E	INI5F_E	Répondre à une question dont la réponse est explicite dans le texte
		F	INI5F_F	Choisir la bonne réponse selon le texte
Compréhension de texte	15	AD	INI5F_AD	Choisir la bonne réponse selon le texte
		AE	INI5F_AE	
Compréhension de texte	15	AF	INI5F_AF	Choisir la bonne réponse selon le texte
		AG	INI5F_AG	
Compréhension de texte	15	AH	INI5F_AH	Choisir la bonne réponse selon le texte
Grammaire	3	G	INI5F_G	Accord d'un adjectif et d'un participe passé
		H	INI5F_H	
	8	P	INI5F_P	Transformer une phrase affirmative en une phrase interrogative
		Q	INI5F_Q	
		R	INI5F_R	
	9	S	INI5F_S	identifier le COI dans la phrase
	10	T	INI5F_T	Remplacer un groupe nominal par un pronom personnel qui convient
U		INI5F_U		
11	V	INI5F_V	Remplacer un groupe nominal par un adjectif possessif qui convient	
11	W	INI5F_W	Remplacer un groupe nominal par un adjectif possessif qui convient	
12	X	INI5F_X	Accord d'un mot au pluriel	
13	Y	INI5F_Y	identifier le groupe sujet	
Conjugaison	4	I	INI5F_I	Choisir le verbe accordé correctement
		J	INI5F_J	
		K	INI5F_K	
	5	L	INI5F_L	choisir le verbe conjugué à l'imparfait de l'indicatif
M		INI5F_M	choisir le verbe conjugué au passé composé de l'indicatif	
6	N	INI5F_N	Choisir la phrase dans laquelle le verbe est accordé correctement	
7	O	INI5F_O	Choisir le temps et la mode du verbe conjugué dans la phrase	
Orthographe	14	Z	INI5F_Z	Orthographier correctement des homophones grammaticaux
		AA	INI5F_AA	
		AB	INI5F_AB	
		AC	INI5F_AC	

Tableau D2 : Post-test de français 5ème année				
Domaines	Exercices	Items	Code dans la base	Objectifs
Compréhension de phrase	1	A B C	FIN5F_A FIN5F_B FIN5F_C	Identifier la préposition qui donne du sens à la phrase
Compréhension de texte	9	Z AA AB AC AD	FIN5F_Z FIN5F_AA FIN5F_AB FIN5F_AC FIN5F_AD	Répondre à des questions dont la réponse se trouve explicitement dans le texte
Grammaire	2	D E F	FIN5F_D FIN5F_E FIN5F_F	Accorder le verbe avec son sujet
	3	G H I	FIN5F_G FIN5F_H FIN5F_I	Remplacer un groupe nominal par un pronom personnel
	5	M N O	FIN5F_M FIN5F_N FIN5F_O	Transformer une phrase affirmative en une phrase interrogative
	8	W X Y	FIN5F_W FIN5F_X FIN5F_Y	Identifier la nature d'un complément de verbe (COD - CCL - CCT)
	10	AE AF AG	FIN5F_AE FIN5F_AF FIN5F_AG	Identifier la fonction d'un adjectif qualificatif
Conjugaison	7	T U V	FIN5F_T FIN5F_U FIN5F_V	Identifier des verbes conjugués au présent du subjonctif
Orthographe	6	P Q R S	FIN5F_P FIN5F_Q FIN5F_R FIN5F_S	Orthographier correctement des homographes grammaticaux

Note : les items repris en caractère gras italique ont nécessité une légère adaptation au cas français,

Tableau D3 : Pré-test mathématiques 5ème année				
Domaines	Exercices	Items	Code dans la base	Objectifs
Opération	1	A	INI5M_A	Choisir la réponse qui convient à la multiplication
		B	INI5M_B	
		C	INI5M_C	
		D	INI5M_D	
Opération	2	E	INI5M_E	Arrondir un nombre
		F	INI5M_F	
		W	INI5M_W	
		X	INI5M_X	
Opération	3	W	INI5M_W	Choisir le tableau où les résultats de l'opération sont corrects
		X	INI5M_X	
		K	INI5M_K	
		L	INI5M_L	
Opération	9	M	INI5M_M	Choisir le tableau où les résultats de l'opération sont corrects
		N	INI5M_N	
		O	INI5M_O	
		P	INI5M_P	
Numération	5	Q	INI5M_Q	écrire un nombre en chiffres
		R	INI5M_R	
		S	INI5M_S	
		T	INI5M_T	
Numération	6	U	INI5M_U	déterminer la position du chiffre dans le nombre
		V	INI5M_V	
		R	INI5M_R	
		S	INI5M_S	
Numération	7	T	INI5M_T	Décomposition d'un nombre
		U	INI5M_U	
		V	INI5M_V	
		S	INI5M_S	
Numération	8	T	INI5M_T	Comparaison des nombres
		U	INI5M_U	
		V	INI5M_V	
		S	INI5M_S	
Géométrie	11	AC	INI5M_AC	Identifier un rectangle parmi plusieurs figures géométriques
		AD	INI5M_AD	
Géométrie	12	AE	INI5M_AE	Identifier un point dans un repère à partir de ses coordonnées
		AH	INI5M_AH	
Problème	13	AF	INI5M_AF	Résoudre un problème concret faisant intervenir des prix Résoudre un problème concret faisant intervenir des longueurs
		AG	INI5M_AG	
		AH	INI5M_AH	

Note : l'item AH faisant appel à une suite d'opérations peut-être considéré comme « difficile » en début de CM1,

Tableau D4 : Post-test mathématiques 5ème année				
Domaines	Exercices	Items	Code dans la base	Objectifs
Opération	6	P S	FIN5M_P FIN5M_S	Effectuer des additions et des soustractions de nombres entiers et décimaux
Numération	2	E F	FIN5M_E FIN5M_F	Ranger des nombres entiers dans l'ordre croissant
	3	G H I	FIN5M_G FIN5M_H FIN5M_I	Ranger des nombres décimaux dans l'ordre décroissant
	10	Z AA AB AC	FIN5M_Z FIN5M_AA FIN5M_AB FIN5M_AC	Comparer des fractions de même numérateur Comparer des fractions de même dénominateur
	11	AD AE AF AG	FIN5M_AD FIN5M_AE FIN5M_AF FIN5M_AG	Simplifier des fractions
Problème	9	X Y	FIN5M_X FIN5M_Y	Résoudre un problème concret faisant intervenir des prix Résoudre un problème concret faisant intervenir des longueurs
	5	N O	FIN5M_N FIN5M_O	Résoudre un problème faisant intervenir des mesures agraires Résoudre un problème faisant intervenir des mesures de capacités
Mesure	4	J K L M	FIN5M_J FIN5M_K FIN5M_L FIN5M_M	Convertir des mesures de grandeurs (masse, capacité, longueur)
	8	U V W	FIN5M_U FIN5M_V FIN5M_W	Dans un problème concret - Calculer le côté d'un carré connaissant son périmètre - Calculer la surface d'un disque connaissant son rayon - Calculer le diamètre d'un cercle connaissant sa circonférence
	12	AH AI	FIN5M_AH FIN5M_AI	Estimer une quantité sur un dessin
Géométrie	7	T	FIN5M_T	Identifier un rectangle parmi plusieurs figures géométriques
	13	AJ AK AL AM AN AO	FIN5M_AJ FIN5M_AK FIN5M_AL FIN5M_AM FIN5M_AN FIN5M_AO	Tracer les médianes d'un losange Tracer les médianes d'un rectangle et d'un carré Tracer les diagonales d'un losange Tracer les diagonales d'un rectangle et d'un carré

Note : les items V et W peuvent être considérés comme hors programme, aussi une adaptation de la formulation a été introduite, voir annexe « adaptation des instruments ».

Annexe E : Adaptation des instruments à la pratique et au programme en France

Les programmes nationaux fixés en 2002 recherchent évidemment à assurer par la deuxième année du **cycle des approfondissements**, le CM1, la consolidation des apprentissages de CE2. L'enseignant doit affirmer les savoirs fondamentaux par plus d'autonomie de l'élève, l'apprentissage du sens logique, et de la précision, notamment au travers de travaux pratiques orientés vers la résolution de problèmes. Au total, le programme est basé sur une charge de 26 heures de cours par semaine dont 2 heures d'études dirigées pendant lesquelles le professeur des écoles reprend les notions importantes de la semaine et assure une aide individualisée à chaque enfant.

Dans ces aspects principaux, le programme repose sur l'acquisition des savoirs et des pratiques suivantes:

Français :

- Lire des textes de plus en plus longs (5-6 pages),
- Ecrire des textes de son propre cru,
- Apprendre des poèmes, proses, etc, par cœur, Jeux théâtraux,
- Approfondir et maîtriser les règles de grammaire, conjugaison et orthographe,

Mathématiques :

- Calcul sur les nombres naturels et décimaux : multiplication, soustraction, début division,
- Calcul mental et tables de multiplication,
- Résolution de fractions simples (avec même dénominateur),
- Construire des volumes en papier, tracer des figures géométriques à l'aide d'instruments... pour apprendre les notions de médianes, de segments, d'angles, d'arêtes, d'aires, de symétrie...
- Connaître les mesures d'aires, de périmètres, Savoir faire des conversions

Il est évident que ces détails ne sont donnés que pour le français et les mathématiques, sujets de savoirs sur lesquels portent les tests utilisés ici ; le programme officiel comporte aussi des prescriptions pour l'histoire-géographie, les sciences et technologie, éducation civique, éducation artistique, éducation physique et sportive,

Adaptations nécessaires ou accessoires

Toutefois, même si le programme reste une référence, il est l'objet de diverses adaptations et pratiques de la part de l'enseignant et de l'équipe pédagogique. Pour l'ensemble des 32 items présents au pré test de français, aucun ne pose un problème d'adaptation à ce programme de CM1. Sur le pré-test de mathématiques, l'item AH, constitué d'une résolution de programme pratique conduisant à la soustraction d'un nombre lui-même total de deux chiffres intermédiaire, est apparu « difficile » à des acteurs du système éducatif pour un test de début d'année. Cette difficulté peut aussi paraître correcte pour la dernière question d'un test dont le but reste de discriminer les savoirs. Le post test de français ne posait pas de difficultés en rapport au programme sur le fond, toutefois l'énoncé des items AE à AG utilisait les termes « attribut » et « épithète » pour les fonctions de l'adjectif. Comme ces termes peuvent être

considérés comme tombés en désuétude, nous avons retenu lors de la passation du test de donné le renseignement suivant :

La fonction attribut d'un adjectif précise la qualité, la nature ou l'état, il se place devant, ou après, des verbes d'état ;

L'épithète caractérise un adjectif qui se joint à un nom ou à un pronom pour le qualifier, le décrire.

Au post-test de mathématiques, une difficulté a porté sur le calcul des surfaces, Ainsi les deux items V et W sont des questions sur le calcul autour du cercle. Il nous a été objecté que les calculs sur la base de π sont pratiqués de fait en CM2. Aussi l'item a été conservé, mais avec la présentation préalable de la formule à appliquer. L'exercice ne faisait plus appel à la mémoire d'une connaissance supposée acquise, mais à une présentation d'un formulaire basé sur son application pratique immédiate par l'exemple.

Ces cinq items expliquent la différence entre le test dans son ensemble et la notion d'items de stricte adéquation programme utilisée dans le corps du texte.,

Annexe F : Vue d'ensemble des programmes en France du primaire en 2002, par année, pour les thèmes des nombres et du calcul

GS*	CP	CE1	CE2	CM1	CM2
<p>Connaître la comptine jusqu'à 31</p> <p>Produire une collection identique à une collection donnée jusqu'à 15 en utilisant l'écrit ou/et la comptine</p> <p>Savoir écrire les premiers nombres (de 1 à 9),</p>	<p>- De l'écriture "naïve" à l'écriture structurée :</p> <ul style="list-style-type: none"> - Construire l'écriture des nombres à deux chiffres (jusqu'à 99) <p>Numération écrite et orale, Les nombres de 0 à 99...</p> <p>Calcul automatisé : tables d'addition</p> <p>Calcul réfléchi</p>	<p>- Numération écrite et orale, Les nombres de 0 à 999</p> <p>Calcul automatisé : tables de multiplication par 2 ; 5 et 10</p> <p>Calcul instrumenté et réfléchi</p>	<p>- Les nombres de 0 à 9999</p> <p>Calcul automatisé : tables de multiplication</p> <p>Calcul instrumenté et réfléchi</p>	<p>- Les classes de mille : les grands nombres (million), Numération orale et écrite jusqu'à la classe des millions,</p> <ul style="list-style-type: none"> - Désignation écrite et orale des nombres <p>Entretien et extension aux nombres du CM1</p> <p>Entretien et extension aux nombres du CM1</p> <p>Introduction de fractions simples et de fractions décimales</p> <p>Les nombres décimaux : désignation orale et écriture chiffrée, Comparaison de nombres décimaux</p> <p>Addition, soustraction de nombres à virgule, (addition dans D posée),</p> <p>Entretien et extension aux nombres du CM1</p> <p>Entretien et extension aux nombres du CM1</p>	<p>- Les grands nombres,</p> <ul style="list-style-type: none"> - Désignation écrite et orale des nombres, <p>Entretien et extension aux nombres du CM2</p> <p>Entretien et extension aux nombres du CM2</p> <p>Multiplication d'un nombre décimal par un nombre entier, (multiplication par un entier posée, soustraction posée),</p> <p>Entretien et extension aux nombres du CM2</p> <p>Entretien et extension aux nombres du CM2</p> <p>Entretien et extension aux nombres du CM2</p> <p>Problèmes relevant des quatre opérations dans les nombres entiers,</p>
<p>...En liaison avec les écritures additives,</p> <p>L'addition comme procédé de réduction d'une suite additive,</p> <p>Problèmes simples relevant de la soustraction,</p> <ul style="list-style-type: none"> - Utilisation de l'écriture a-b pour désigner le nombre cherché, 	<p>L'addition comme procédé de calcul automatique, (addition posée)</p> <p>Calculs soustractifs sur la droite numérique par retranchements successifs ou complément (addition à trou)</p> <p>Construction progressive de la multiplication et d'un algorithme dans du calcul réfléchi,</p>	<p>Entretien et extension aux nombres du CE2</p> <p>Technique opératoire de la soustraction, (Soustraction posée),</p>	<p>Entretien et extension aux nombres du CM1</p> <p>Calculer le produit de deux nombres entiers par un calcul posé (multiplication posée),</p>	<p>Calculer quotient et reste de la division d'un nombre entier (au plus 4 chiffres) par un entier (au plus 2 chiffres), (Division posée, soustractions apparentes)</p>	

* Grande section de maternelle.