

Enquête nationale sur le niveau intellectuel des enfants d'âge scolaire

réalisée en 1965 par l'Institut National d'Études Démographiques (INED)
et l'Institut National d'Étude du Travail et d'Orientation Professionnelle (INOP)

CHEMINEMENT INFORMATIQUE DE L'INFORMATION (1965-1970) Fichier fossile 1970 : variables groupées en bagages généalogiques

[Rédaction au 6 octobre 2005]

Pour tirer parti des données conservées dans le fichier fossile 1970, il faudra tenir compte de la complexité voulue du plan expérimental, qui oblige à plusieurs systèmes de pondération (voir chapitre I). Une autre complexité, tout à fait involontaire, découle des contraintes des exploitations informatiques vers 1965-1970, et des imperfections qui en sont résultées. Toute utilisation des données conservées doit tenir compte du classement des variables en **bagages, découlant de leur généalogie**, autrement dit du cheminement informatique des informations.

Appelons "**bagage**" un sous-ensemble des variables dans le fichier fossile 1970, sous-ensemble ayant même origine. Ce classement en bagages (généalogiques) donne la structure logique du fichier fossile 1970. On doit distinguer 7 bagages (tableau 9 ou 10)..

L'utilité de distinguer les bagages est la suivante : les variables membres d'un même bagage ont certains traits communs. En particulier, dans chaque variable formant un bagage, on rencontre une lacune systématique de l'information sur un groupe particulier d'élèves : toutes les lignes du fichier ne sont pas dotées des quatre bagages importés.

II-1. Vue d'ensemble : sept "bagages" dans le fichier fossile 1970.

Le tableau 1 part de l'état du fichier fossile 1970. Ce fichier est segmenté en sept bagages, présentés dans l'ordre temporel du rassemblement des variables. Le tableau 1 bis résume aussi le cumul progressif des bagages.

Cette segmentation logique, en sept bagages, n'est pas directement visible dans l'état historique où le fichier fossile a été retrouvé (et transmis aux Archives nationales). Dans cet état historique, les variables de quatre bagages (F1 F3 F3x1 Fin) sont mêlées dans la dessin d'une ligne (enregistrement).

Tableau 1. Segmentation généalogique du fichier fossile 1970 (en 7 bagages)

(noter : en MAJUSCULES, les 4 BAGAGES importés des 4 fichiers spécialisés)

BAGAGE F1	= des variables importées du fichier Fic1 après son exploitation 1966 dont: notes et scores issus de la recherche psychotechnique .
BAGAGE F3	= des variables importées du fichier Fic3 après exploitation (1968): conditions de terrain (élève et milieu familial), et jugements.
Fichier UNION (FicU) créé: contient bagages F1 et F3	
bagage F3x1	= 10 variables créées en 1969 par intersection des bagages F1 et F3 dans l'exploitation autonome du fichier UNION.
L'importation des bagages F4 et F5 donne:	
BAGAGE F4	= 10 variables importées du fichier Fic4 après son exploitation 1969: compléments sur quelques aspects (notamment commune de résidence)
BAGAGE F5	= 93 variables importées du fichier Fic5 après son exploitation 1969: variables en chantier sur la composition de la fratrie de l'élève.
Opérations finales après toutes importations:	
bagage "Fin"	= 4 variables finales créées 1970 par intersection de tous les bagages réunis, après importation des bagages complémentaires F4 et F5
bagage F0	= Matricule d'un élève (M9 : 9 caractères en fichier fossile 1970) créé en fin 1970 pour remplacer la séquence initiale Sq10 de 1965

Tableau 1 bis. Cumul historique du rassemblement des bagages.

1969 (a)	$Sq10 + F1 + F3$	= bagages du fichier UNION à sa création en 1969
(b)	$Sq10 + F1 + F3 + F3x1$	= bagages du fichier UNION encore autonome.
1970 (a)	$Sq10 + F1 + F3 + F3x1 + F4 + F5$	=après apports des fichiers Fic4 et Fic5
(b)	$Sq10 + F1 + F3 + F3x1 + F4 + F5 + Fin$	=après création du bagage "Fin"
(c)	$F0 + F1 + F3 + F3x1 + F4 + F5 + Fin$	= en fichier fossile 1970.
Note: "Sq10" représente ici la séquence initiale à 10 caractères, créée en 1965		
"F0" ou "M9" est le matricule à 9 caractères, issu de Sq10, figurant dans le fichier fossile 1970		

Les quatre bagages importés constituent la partie la plus massive du fichier fossile 1970. Ces bagages (F1 F3 F4 F5) résultent de l'importation de variables venant des quatre fichiers spécialisés (Fic1 Fic3 Fic4 Fic5).

Mais il y a aussi trois autres bagages, qui n'ont pas été directement importés d'un unique fichier. Quelques variables ont été créées en combinant des informations prises dans plusieurs des bagages importés (bagage intersection "F3x1", et bagage "Fin"). Le matricule (ou bagage "F0") a un statut particulier.

II-2. Contraintes matérielles des exploitations informatiques 1966-1970.

Il n'existait pas encore de disquette ou de disque dur, ni d'écran associé à un clavier. Pour accéder aux ordinateurs, rares et grosses machines¹, les seuls canaux étaient les cartes perforées et les bandes magnétiques. Ces deux supports ont conduit à un partage des données initiales.

-a- Dans tous les cas, il fallait perforer d'abord des **cartes mécanographiques** pour la saisie des informations initiales². Sur une carte perforée (80 colonnes), on ne peut transmettre qu'une partie de toutes les informations à saisir sur un élève. La description d'un élève implique alors un fractionnement des données initiales en plusieurs cartes.

La correction des erreurs de frappe n'était pas simple. Dans un atelier de perforation, quand une opératrice percevait une erreur commise en cours de frappe, elle devait éliminer la carte erronée, et perforer entièrement une nouvelle carte. Ne pas éjecter la carte erronée revenait à introduire un élève fictif dans la cartothèque.

Pour permettre ultérieurement de reconstituer l'unité d'un élève, chacune de ses cartes portait des colonnes servant de matricule. Ce matricule a donc été perforé autant de fois que l'élève a de cartes. Les erreurs de perforation sur l'une ou l'autre des cartes vont contrarier le couplage ultérieur du contenu des cartes, et donc la reconstitution de l'unité de l'élève.

Sans un ordinateur, l'exploitation des cartes perforées était possible avec des machines mécanographiques ; l'appareil de base est la trieuse. C'est d'abord avec une trieuse de cartes que toute cartothèque était mise en ordre : les cartes sont rangées selon le contenu des colonnes formant le matricule. Ces mises en ordre des cartes étaient capitales, en vue du couplage ultérieur du contenu des fichiers spécialisés issus des cartothèques.

-b- Dans les procédures les plus modernes (1965-1970), c'est sur une **bande magnétique** qu'était d'abord transféré le contenu de toute cartothèque de données. Dans un fichier sur bande, on retrouve l'ordre de succession donné aux cartes perforées sources. C'est un stockage séquentiel. Une bande magnétique décrivant 120 000 élèves est comme un film cinématographique à 120 000 images. Admettons que l'image d'un élève occupe un centimètre, la bande est un "long métrage" (1,2 km) occupant matériellement plusieurs bobines. Au stockage séquentiel des élèves est associé un traitement séquentiel : les mêmes opérations (définies par des instructions sur des cartes perforées formant le programme) seront répétées successivement sur chaque élève dans l'ordre de la bande³. Un traitement aménageant les variables initiales, ou créant des variables dérivées, conduisait à créer

¹ Ont été utilisés les ordinateurs de la Société grenobloise d'applications hydrauliques (SOGREAH), par le canal d'une société prestataire de services : Institut international d'informatique.

² Présentation d'époque en : Grémy, Jean-Paul, "Le dépouillement d'enquêtes sur ordinateur : quelques notions de base", *Bulletin de psychologie*, XXI 9-11 (1968), p.648-667.

³ Présentation d'époque de la logique des traitements séquentiels : Grémy, Jean-Paul, *Initiation à l'informatique; introduction à la programmation*, [Paris] Hachette, 206p. 1973. L'exposé est fait en admettant un fichier de données sur cartes perforées, seul support à diffusion universelle à cette date.

un autre bande (but), image par image : finalement, ce fichier but contient, toujours dans l'ordre de la bande source, la suite des images nouvelles de tous les élèves.

Trop lent, le déroulement des bandes magnétiques imposait des temps d'attente à l'unité centrale de l'ordinateur, et accroissait donc le coût de chaque traitement. Pour réduire les temps morts, la solution était de porter sur la bande une image moins détaillée de chaque élève. Au départ, on choisit une partie de toutes les informations disponibles (fichiers spécialisés). Dans un traitement créant des variables nouvelles, on choisit d'abandonner certaines variables du fichier source, pour laisser la place aux variables nouvelles sur le fichier but. Ces pertes ne semblaient pas définitives, puisque les bandes sources étaient conservées. Mais l'évolution technique en a décidé autrement. Les premiers fichiers, créés en 1966, sont devenus illisibles dès 1969, suite à la disparition des premiers ordinateurs utilisés.

-c- Enfin, se sont ajoutés les limitations des logiciels en usage vers 1965. Un **"nouveau langage de tabulation" (NLT)** était mis au point par la société prestataire de service⁴. Ce logiciel prototype était spécialisé dans le dépouillement des enquêtes statistiques : recodage des données, traitement avec filtre et pondération, publication (listage papier) de tableaux d'effectifs et de pourcentages, ou de calculs simples (moyenne, écart type). Pour l'emploi du logiciel NLT, le fichier des données devait respecter des contraintes absolues. Tout caractère littéral est interdit (sauf dans une partie non exploitable, comme le matricule). Dans une variable destinée aux calculs (de type "quantité"), et qui ne pouvait entrer comme dimension d'un tableau, NLT n'admettait pas de stocker des nombres avec partie décimale, ni même des nombres entiers négatifs.

Dans une variable nominale (de type "code"), destinée à servir de dimension d'un tableau de fréquence, le seul codage admis par NLT était de désigner les modalités par la suite sans lacunes des nombres naturels (01 02 ...) ; le nombre des modalités ne peut dépasser 99; le code zéro '00' sert à repérer les sujets à valeur manquante. Le chiffrage manuel des informations avait été fait sans respecter ces règles, sur les cartes perforées. Après le transfert des chiffrages initiaux sur une bande magnétique, le premier traitement nécessaire a été de faire un recodage "numérisé" de chacune des variables, pour permettre leur emploi ultérieur avec le logiciel NLT. Certaines variables du fichier fossile 1970, "proches de la saisie", ont un codage numérisé qui diffère du codage écrit par le chiffeur sur les documents sources.

Une particularité du logiciel NLT explique la difficulté du décryptage, fait en 2001-2002, du dessin du fichier fossile 1970. Un fichier de type NLT conservait les variables dans l'ordre alphabétique des noms donnés à ces variables. La création d'une seule variable nouvelle introduisait un nom nouveau, ce qui donnait, sur la bande but, une nouvelle position physique aux variables qui existaient sur la bande source. Il n'était pas fourni de description du dessin du fichier, donnant la mise à jour de la position physique des variables sur la ligne (enregistrement). Mais d'autres logiciels ont été utilisés par la suite. Au total, l'ordre des variables dans le fichier fossile 1970 résulte du mélange de plusieurs conventions différentes, utilisés successivement.

II-3. Quatre fichiers spécialisés : sources de 4 bagages importés en fichier fossile 1970.

Le fichier fossile 1970 résulte du cheminement des informations de 1965 à 1970. Dès 1965, a été défini un partage raisonné des informations sur un élève.

Le partage raisonné de l'information répond à une suite d'objectifs, munis d'un ordre de priorité. Ont été créées quatre cartothèques, source de quatre fichiers spécialisés créés de façon échelonnée (tableau 2). Chaque fichier décrivait en théorie le même ensemble global d'élèves ; les élèves étaient décrits sous un angle particulier, répondant à un objectif partiel. Chaque fichier spécialisé devait pouvoir être exploité, d'abord, de façon autonome.

Les noms des fichiers spécialisés (Fic1, Fic3, Fic4, Fic5) évoquent directement les types de cartes perforées qui ont fourni le contenu initial de chaque fichier. Le souci de retenir au chiffrage de nombreux détails a conduit à perforer au moins 4 cartes par élève; parfois jusqu'à 8 cartes. Au total, on a eu une masse de 650 000 cartes pour 120 000 élèves environ.

⁴ Institut international d'informatique (ou "iii" ou "3i").

Tableau 2. Sommaire : les 4 fichiers spécialisés et leurs objectifs.

Extraits de ces fichiers conservés en fichier fossile 1970 : 4 bagages (F1 F3 F4 F5).

Quatre cartothèques	Quatre fichiers spécialisés (perdus)	Conservé en 1970 (*)
cartes n°0 ou (1+2)	<p>OBJECTIF A</p> <p>Recherche psychotechnique créant des scores unifiés en partant du détail des items des quatre cahiers de test</p> <p>>--> Fichier spécialisé Fic1 en 1966</p> <p>Exploitation autonome: variables disponibles. Extrait >--> F1</p>	BAGAGE
carte n°3	<p>OBJECTIF B</p> <p>Description sans considérer le test: faits et jugements tirés du Bulletin individuel (Bi) ou du bul. familial (Bf)</p> <p>>--> Fichier spécialisé Fic3 en 1968</p> <p>Exploitation autonome: variables disponibles. Extrait >--> F3</p>	BAGAGE
	OBJECTIF C : description compréhensive (après réunion des deux bagages F1 et F3)	
carte n°4	<p>OBJECTIF D</p> <p>Pour complément: nom et adresse en clair (source Bf)</p> <p>>--> Fichier spécialisé Fic4 en 1969 (les 10 variables déduites sont anonymes)</p> <p>Exploitation autonome: variables disponibles. Extrait >--> F4</p>	BAGAGE
cartes n°5 (+ 6 7 8)	<p>OBJECTIF E</p> <p>Pour complément: sur chaque enfant de la famille (source Bf) sexe, date de naissance, situation scolaire de chacun</p> <p>>--> Fichier spécialisé Fic5 en 1969</p> <p>Exploitation autonome: variables disponibles. Extrait >--> F5</p>	BAGAGE
	OBJECTIF F : description enrichie (après importation des bagages F4 et F5)	
(*) Le fichier fossile 1970 est formé de 7 bagages généalogiques. Les 4 bagages (F1 F3 F4 F5) ont été importés directement de chaque fichier spécialisé. Trois autres bagages contiennent des variables créées par combinaison de variables issues de plusieurs fichiers spécialisés (vue d'ensemble en tableau 9 ou 10).		

Lors de l'exploitation autonome de chaque fichier spécialisé, on a procédé à la mise en forme des variables et à la création de variables nouvelles. Certains résultats publiés résultent de tableaux produits dans l'exploitation autonome des deux fichiers de base (Fic1, puis Fic3).

Les **quatre fichiers spécialisés** ont été la source de **quatre bagages importés** (F1 F3 F4 F5), formant quatre groupes de variables dans le fichier fossile 1970. Chaque bagage importé provient directement d'un seul des 4 fichiers spécialisés (Fic1 Fic3 Fic4 Fic5) ; en général, les variables importées n'ont pas été modifiées dans la suite des travaux d'époque.

C'est en deux étapes que les variables, extraites de chaque fichier spécialisé, ont été progressivement rassemblées, suivant l'ordre de priorité des objectifs.

II-4. Trois objectifs essentiels en 1965 : vers la création du fichier UNION 1969.

Trois premiers objectifs (A B C) étaient essentiels. Pour y satisfaire, ont été créés deux fichiers spécialisés de base (Fic1 en 1966; Fic3 en 1968). Puis est intervenu, en 1969, la création d'un fichier UNION, issu des deux fichiers de base. À ce stade, on n'avait pas encore la totalité du fichier fossile 1970.

--A. *L'objectif préalable était la **recherche psychotechnique**, pour créer des "scores unifiés".* Pour satisfaire à cette recherche psychotechnique, a été créé un premier **fichier spécialisé (Fic1)**, dit aussi fichier "test". Certains élèves, présents dans ce fichier "test", n'avaient pas de test exploitable.

Au départ, le fichier Fic1 a été alimenté notamment par les réponses de l'élève à chaque item du test (jusqu'à 96 items au cahier 4), données perforées en général sur deux cartes (n°1 et 2) ; par exception, une seule carte (type n°0) suffisait pour certains élèves (ceux avec test fantôme, ou ceux avec cahier I).

Seule l'exploitation autonome de ce premier fichier spécialisé (Fic1) était disponible lors de la rédaction du Tome 1 (opus INED n°54, 1969). Cette première publication contient les tables pragmatiques établies selon la métrique INOP.

--B. *Le deuxième objectif était la mise en forme des informations sur les **conditions de terrain** :* caractéristiques de l'élève et de sa famille, jugements du maître et des parents. Pour satisfaire à un tel objectif, des informations ont été perforées sur une seule carte (n°3), source d'un second **fichier spécialisé de base (Fic3)**.

Dans l'exploitation autonome du fichier Fic3, des ventilations ont été produites, qui sont la base de certains tableaux publiés (opus INED n°64 et 83) ; les tableaux issus du fichier Fic3 décrivent l'élève ou sa famille sans mise en relation avec les résultats au test.

--C. *Le troisième objectif est la **description compréhensive** :* mettre en relation les conditions de terrain et un score unifié (le quotient intellectuel de la métrique INED). Dans ce but, le **fichier UNION (FicU)**, créé en 1969, a réuni des variables fournies par chacun des deux fichiers spécialisés de base (Fic1 et Fic3). On avait alors, lors de la création du fichier UNION, des variables importées de deux origines :

- **bagage F1** : des variables importées depuis le premier fichier spécialisé (Fic1), après son exploitation autonome : parmi ces variables, les scores unifiés ; l'un d'eux est le quotient intellectuel de la métrique INED.

- **bagage F3** : des variables importées depuis l'autre fichier de base (Fic3), après son exploitation autonome : pour décrire les conditions de terrain.

Ce fichier UNION (FicU) a donné lieu à une exploitation autonome, pour créer 10 variables, d'un autre type que les variables importées. Ces variables ont été définies par l'intersection logique de certaines variables des deux premiers bagages. Ce qui produit un troisième bagage :

- **bagage intersection (F3x1)** : variables construites en combinant le bagage F3 et le bagage F1, dès après la construction du fichier UNION. Six de ces créations répondaient au besoin de repérer la cohérence entre les deux bagages importés.

Les ventilations produites avec le fichier UNION, en 1969, croisaient le quotient intellectuel et des conditions détaillées de l'élève. C'est la source de la plupart des tableaux publiés, quand ils relient l'état de l'élève et son QI.

Les trois bagages (F1, F3, F3x1), déjà présents dans le fichier UNION, sont le cœur du fichier fossile 1970.

Les raccords entre cette vue historique et les publications sont les suivants. Le Tome 1 (opus INED n°54, 1969) repose sur l'exploitation autonome du premier fichier spécialisé (Fic1) en 1966-1967. Les deux derniers tomes (1973 et 1978; opus n°64 et 83) sont basés sur les exploitations de 1968-1969, avant l'importation des

bagages complémentaires (F4 F5), sauf exception⁵. Certains tableaux publiés ont été tirés de l'exploitation autonome du fichier Fic3 en 1968, d'autres sont issus de l'exploitation du fichier UNION en 1969.

II-5. Les objectifs complémentaires : état complet du fichier fossile 1970.

Le fichier UNION, en 1969, était la première forme du fichier récapitulatif que sera le fichier fossile 1970. Mais, outre les trois objectifs prioritaires (A B C), deux objectifs complémentaires (D E) avaient été fixés dès 1965, préliminaire à une description compréhensive enrichie, prévue comme un objectif final plus lointain (F).

-D. *Un objectif complémentaire* était de tirer parti du **patronyme de l'élève et de l'adresse** de la famille. Pour satisfaire à cet objectif complémentaire, a été perforée une carte (n°4), donnant en clair le nom et l'adresse, et qui a constitué l'état initial du **fichier spécialisé (Fic4)**. Son exploitation a permis de créer 10 variables, au contenu anonyme.

-E. *Le dernier objectif complémentaire* était de *décrire en détail la place de l'élève observé dans sa fratrie*, son "contexte fraternel". En vue de cet objectif, a été créé, puis exploité le **fichier spécialisé (Fic5)**. Le fichier Fic5 contenait initialement la description détaillée (10 caractères) de chacun des frères ou soeurs de l'élève observée. Ces détails ont été fournis par une carte (n°5), qui suffisait pour un élève ayant 1 à 5 frères ou soeurs; pour certains élèves, plus de 20 enfants étaient détaillés dans la fratrie : la carte n°5 était complétée par des cartes suites (n°6 à 8). L'exploitation autonome du fichier Fic5 a permis de créer 93 variables, orientées vers la description du contexte fraternel de l'élève observé dans l'échantillon de l'enquête.

⁵ Exception : le bagage F4 a fourni un aperçu sur la distance kilométrique entre la commune de résidence (famille de l'élève) et la commune de l'établissement scolaire fréquenté (opus INED n°64, p.102).

Tableau 3. Schéma d'ensemble simplifié du cheminement des informations.

Des fichiers spécialisés au fichier Union 1969; puis au fichier final (fossile 1970), lequel est composé de 7 bagages (groupes de variables ayant la même généalogie).

		Pour étude compréhensive: scores unifiés fonction des conditions de terrain	
Cartes perforées	Fichiers spécialisés et exploitations autonomes	Fichier UNION 1969 FicU (bagages F1 F3 F3x1)	Fichier fossile 1970
Deux fichiers de base (Fic1 et Fic3)		Séquence initiale Sq10(Sq10) --> (M9) définie en 1965	
n°0) n°1) n°2)->	Fic1 1966 Recherche psychotechnique = var. conservées	--> BAGAGE F1	BAGAGE F1
n°3 -->	Fic3 1968 Conditions de élève et milieu (sans test) = var. conservées	--> BAGAGE F3	BAGAGE F3
On crée 10 variables par intersection de F3 et F1		bagage F3x1	bagage F3x1
Cartes	Deux fichiers complémentaires (Fic4 et Fic5)		
n°4 -->	Fic4 1969 Nom et adresse = 10 var.créées	----- (importées) -->	+ BAGAGE F4
n°5) n°6) n°7) n°8)->	Fic5 1969 Frères et soeurs de l'élève = 93 var.créées	----- (importées) -->	+ BAGAGE F5
On crée 4 var. finales combinant F1 F3 F3x1 F4 F5			bagage Fin
- Sq10 = Séquence initiale à 10 caractères (incluant le matricule), créée en 1965 (M9) = Matricule hiérarchique à 9 caractères adopté à la fin des travaux.			

= Complément technique au tableau 9: **fichier "établissement" et fichier "communes"**
En dehors de la description individuelle des élèves, deux fichiers ont décrit des unités collectives : l'établissement scolaire, et la commune de résidence. Les apports de ces deux fichiers sont intervenus seulement lors de l'exploitation autonome du fichier spécialisé Fic4 (d'où a été importé le bagage F4).

- Issu de cartes n°9, un fichier décrivait chaque établissement scolaire. Ce fichier particulier a été utilisé pour créer 3 des 10 variables formant le bagage F4 (variables fossiles ALT PUP CLAS).

- Un autre fichier a été créé pour apporter des informations sur chaque commune de résidence rencontrée parmi les élèves d'un même établissement (élèves présents en Fic4); son apport est présent dans 4 variables du bagage F4 (variables fossiles NOCO KMK KMQ COM).

Sur les 10 variables du bagage F4, seulement 3 ont été alimentées exclusivement par les données initiales qui figuraient en carte n°4 (variables fossiles DERE HOMO PATR).

-F. *Objectif final visé : une recherche compréhensive enrichie.* Pour la préparer, a eu lieu en 1970 l'importation, sur les élèves du fichier Union, des variables complémentaires créées dans les deux derniers fichiers spécialisés (Fic4 et Fic5) ; ce qui a donné :

- **bagage F4** : importé du fichier Fic4 : 10 variables.
- **bagage F5** : importé du fichier Fic5 : environ 93 variables, qui sont restées dans un état semi-fini ; l'affinement envisagé n'était possible qu'en relation avec les variables du bagage F3, décrivant l'élève observé.

Enfin, deux opérations finales ont eu lieu :

- création du **bagage final (ou "Fin")** : 4 variables construites en combinant des variables prises dans les cinq bagages déjà rassemblés (F1, F3, F3x1, F4, F5) ; voir Notice générale -7H-
- adoption du **matricule** (ou bagage "F0") dans la forme à 9 caractères que présente le fichier fossile 1970. Cette forme (M9) a été adoptée à la fin des travaux, et diffère de la séquence à 10 caractères (Sq10) qui a identifié un élève au long de l'exploitation antérieure.

Peu après 1970, l'exploitation informatique a été interrompue brutalement. Les apports complémentaires du bagage F4 ont été peu exploités ; l'apport du fichier Fic5 (bagage F5) est resté un matériau semi-fini en attente de raffinement. L'objectif final n'a pas été atteint.

Au total, le **fichier fossile 1970** est un instrument de travail, dans l'état où les chercheurs l'ont abandonné peu après 1970. Il n'a pas été conçu pour être diffusé. Il inclut certaines variables qui sont la base des travaux publiés, et d'autres variables en attente de raffinement. Par ailleurs, le fichier fossile 1970 n'a pas conservé toutes les variables qui ont été exploitées dans les fichiers spécialisés.

Le tableau 3 donne un résumé de la généalogie des variables disponibles. On avait des cartes perforées, qui ont alimenté quatre fichiers spécialisés. Après l'exploitation autonome de chaque fichier, un choix de variables a été conservé d'abord dans le fichier UNION, puis dans le fichier fossile 1970 : ce sont les quatre bagages formés de variables importées. À quoi se sont ajoutés trois autres bagages.

II-6. Un bagage importé est un extrait sélectif d'un fichier spécialisé.

Un **bagage importé**, présent dans le fichier fossile 1970, **résulte d'une sélection** faite parmi les variables que contenait le fichier spécialisé source, à la fin de son exploitation autonome. Il ne s'agit jamais d'un choix parmi les élèves, mais d'un choix parmi les variables.

Le chercheur du XXI^e siècle qui voudrait refaire un traitement d'époque, ou reconstruire un tableau publié tiré de l'exploitation autonome d'un fichier spécialisé, pourra échouer à retrouver exactement les traitements faits autrefois. Refaire exactement le passé n'est possible que dans le cas où toutes les variables sources utilisées dans le traitement ancien ont été importées dans le fichier fossile 1970.

Le tableau 4 schématise le processus des importations sélectives, dans la situation théorique idéale (la réalité est plus complexe ; voir § II- 7).

Ce tableau 4 suggère la situation idéale où : tout élève se trouverait doté des quatre bagages importés (F1 F3 F4 F5), provenant des quatre fichiers spécialisés. Pour que la réalité reproduise un tel schéma théorique, il aurait fallu, d'abord, que les quatre fichiers spécialisés aient été formés exactement du même ensemble d'élèves (sans pertes, ni lignes excédentaires). Il aurait fallu aussi que chaque fichier source ait été sans erreur de saisie, dans les 9 caractères servant de clé de couplage (matricule $\mu 9$, partie de la séquence initiale Sq10).

*Tableau 4. Distinction entre bagage importé et fichier spécialisé source.
Schéma théorique : tout élève en fichier 1970 serait doté des 4 bagages.*

Légende: /1/ variables qui étaient présentes en fichier spécialisé Fic1 et qui n'ont pas été retenues pour former le bagage F1 /3/ variables du fichier Fic3 non retenues pour former le bagage F3 /4/ variables du fichier Fic4 non retenues pour former le bagage F4 /5/ variables du fichier Fic5 non retenues pour former le bagage F5			
Schéma théorique parfait: rassemblement des apports sélectif des 4 fichiers			
Fic1	= Sq10	/1/.....	<u> * </u>
Fic3	= Sq10	.. /3/.....	<u> * </u>
Fic4	= Sq10 /4/.....	<u> * </u>
Fic5	= Sq10 /5/.....	<u> * </u>
Théorie=	Sq10	<u>bag.F1</u> <u>bag.F3</u> <u>bag.F4</u> <u>bag.F5</u>
"Sq10" représente la séquence initiale à 10 caractères, créée en 1965, et contenant le matricule.			

Dans ce schéma idéal (tableau 4), prenons l'exemple du bagage (F1) importé du fichier Fic1. Parmi toutes les variables qui étaient disponibles dans le fichier spécialisé (Fic1) à la fin de son exploitation autonome, deux parts ont été faites :

- /1/ = des variables qui n'ont pas été retenues pour former le **bagage F1**,
- F1 = le bagage issu de Fic1, présent dans le fichier fossile 1970.

En disposant des variables du bagage F1, on ne peut refaire exactement toutes les opérations de la recherche psychotechnique de 1966-1967, qu'avait permis le fichier Fic1. On ne peut, non plus, retrouver, pas à pas, toutes les opérations qui ont conduit à la création des poids conservés dans le bagage F1. Le chercheur de 1969 était déjà dans cette situation.

Le cas du bagage F3 est particulier. Peu de variables ont été abandonnées (lot /3/), parmi les variables qui étaient disponibles en fichier Fic3. Cette particularité du bagage F3 le distingue des trois autres bagages importés (F1; F4; F5).

Il était inévitable d'opérer des sélections, pour s'adapter aux limitations matérielles de l'informatique de l'époque. Le choix raisonné des variables retenues a répondu à l'objectif de 1969 (objectif C) : il fallait disposer, dans le fichier UNION en 1969, des informations suffisantes pour la description compréhensive. À ce stade, la recherche psychotechnique était un objectif périmé : il n'était plus question d'être en mesure de refaire cette recherche, mais seulement d'en utiliser certains résultats.

II-7. Lignes à dotation incomplète : exemple du fichier "Union" (1969).

Dans le schéma idéal (tableau 4), chaque élève théorique était décrit par quatre bagages importés. Dans la réalité, lors de la réunion des apports des fichiers spécialisés, il a été impossible de reconstituer l'unité initiale de certains élèves. C'était l'effet des erreurs matérielles commises antérieurement, notamment dans la perforation des cartes mécanographiques.

-a) *Les trois profils de dotation apparus en fichier UNION.*

Examinons la création, en 1969, du **fichier "union"** ; son contenu est le coeur du fichier fossile 1970. Le fichier UNION résulte du couplage des élèves présents dans le fichier Fic1 ou dans le fichier Fic3. En théorie, ce sont les mêmes élèves ; en théorie, tout élève devait être doté de deux bagages (F1, F3), chacun issu des fichiers spécialisés Fic1 et Fic3.

En réalité, la création réelle du fichier UNION (FicU) a fait apparaître des lignes avec trois dotations différentes (tableau 5). Pour créer ce fichier UNION, la clé du couplage avait la même définition que le matricule à 9 caractères présent dans le fichier fossile 1970. Pour expliquer l'apparition de trois types de lignes, il faut distinguer, comme en tableau 5, deux formes pour ce matricule : la forme (μ9') qui était en Fic1, la forme (μ9'') présente en Fic3.

Tableau 5. Création 1969 du fichier UNION (FicU) des fichiers Fic1 et Fic3
Dotations des élèves: 3 profils selon les bagages F1 et F3

Création en 1969 de <u>FicU</u> fichier UNION des 2 fichiers de base (Fic1 Fic3)			
<u>Fic1</u>	=	$\mu 9'$.. /1/.....	<u>*</u>
<u>Fic3</u>	=	$\mu 9''$ /3/.....	<u>*</u>
<u>FicU</u>	=	-----	Lignes en fichier UNION selon 3 types de dotation :
116.699	M9= $\mu 9'$ = $\mu 9''$	<u>bagage F1</u> <u>bagage F3</u> avec double bagage F1+F3
5.375	M9= $\mu 9'$	<u>bagage F1</u> <u>/v/</u> avec bagage F1 sans F3
1.268	M9= $\mu 9''$	<u>/u/</u> <u>bagage F3</u> avec bagage F3 sans F1

123.342 = nombre total de lignes du fichier UNION <u>FicU</u> ;			
Légende: /1/ variables qui étaient présentes en fichier spécialisé Fic1 , non retenues en bagage F1 /3/ variables du fichier Fic3 non retenues pour former le bagage F3			
$\mu 9'$ représente le matricule présent en premier fichier spécialisé (Fic1) $\mu 9''$ représente le matricule présent dans l'autre fichier spécialisé (Fic3) M9 = matricule obtenu sur une ligne du fichier union (FicU)			
Variables fossiles aménagées en 2003 : la lettre 'u' ou la lettre 'v' est mise comme modalité pour les élèves à dotation tronquée: u = élève sans bagage F1 mais avec F3 ; marqué 'u' en toute variable du bagage F1. v = élève sans bagage F3 mais avec F1 ; marqué 'v' en toute variable du bagage F3.			

Suite aux lacunes et erreurs propres à chaque fichier source (Fic1, Fic3), la création du fichier UNION a fait apparaître trois types de lignes.

Cette structure fondamentale, apparue lors de la création du fichier UNION en 1969, est aussi celle du fichier fossile 1970. Les effectifs affichés sont ceux de ce fichier :

- 123 342 élèves ont un matricule (M9) explicite.
- 116 699 élèves ont un **double bagage** (F1+F3).
Ces lignes sont apparues quand il y avait égalité de matricule ($\mu 9'=\mu 9''$) en comparant les 2 fichiers sources (Fic1 et Fic3).
Ce contenu identique est le matricule retenu (M9).
- 1 268 lignes **sans bagage F1**, mais avec un bagage F3;
élèves identifiés en fichier Fic3 (M9= $\mu 9''$) mais non trouvés en Fic1 ;
c'est le contenu ($\mu 9''$) qui est retenu pour matricule (M9).
- 5 375 lignes **sans bagage F3**, mais avec un bagage F1 ;
élèves identifiés en fichier Fic1 (M9= $\mu 9'$) mais non trouvés en Fic3 ;
c'est le contenu ($\mu 9'$) qui est retenu pour matricule (M9).
-
- 123 342 élèves qui ont formé le fichier UNION en 1969 :
élèves avec matricule (M9) en fichier fossile 1970.

Lors de la création du fichier Union, il ne pouvait exister de lignes démunies à la fois des deux bagages de base (F1 et F3). C'est plus tard, lors de l'importation des bagages complémentaires (F4, F5), qu'apparaîtront 293 lignes sans bagage F1 ni F3, lignes dotées seulement des bagages complémentaires. Le matricule de ces lignes a été effacé en 1970 ; ce sont 293 lignes volantes dans le fichier fossile 1970.

-b) En 1969, utilisation des seuls élèves avec double bagage.

A partir de 1969, l'objectif visé était la recherche compréhensive : mettre en relation les résultats au test et la plupart des variables de terrain. Seul un élève doté du **double bagage de base (F1+F3)** pouvait répondre à cet objectif.

C'est seulement pour ces 116 699 élèves à double bagage qu'on pouvait avoir, à la fois : des scores tirés du test et importés en bagage F1, et les variables de terrain importées en bagage F3 (état des élèves, conditions familiales, jugement du maître, jugement des parents). Un élève à double bagage (F1+F3) est comme un organisme cohérent avec tous ses organes : une tête (bagage F1) et un corps (bagage F3) ; et son nom (matricule M9).

-c) Interprétation problématique des lignes à dotation partielle.

Les lignes avec dotation partielle ont été écartées des exploitations en 1969-1970, mais elles restent présentes dans le fichier fossile 1970. Le tableau 6 est une autre représentation de la répartition des élèves selon les trois profils de dotation apparus dans le fichier UNION.

*Tableau 6. Structure de base du fichier fossile 1970: trois profils de dotation apparus lors de l'UNION des deux fichiers de base (Fic1 et Fic3)
(En fichier fossile 1970: 123.342 lignes avec matricule).*

123.342 lignes du fichier UNION 1969						
		5.375	116.699	1.268		
Total avec BAGAGE F1		avec bagage F1 sans F3	avec double bagage F1+F3	avec bagage F3 sans F1	Total avec BAGAGE F3	
Fic1	--> 122.074	= 5.375	+ 116.699			
			116.699	+ 1.268	= 117.967	<-- Fic3
			116.699			

Le choix des élèves avec double bagage (F1+F3), en provoquant une réduction d'effectif assez faible, ne permet pas de retrouver intégralement l'effectif qui était présent en Fic1, ni en Fic3 :

- il y avait 122 074 élèves qui formaient le fichier Fic1 ; mais 5 375 élèves avec bagage F1 sont sans bagage F3 ; effectif réduit de 4%.

- 117 967 élèves formaient le fichier Fic3 ; mais 1 268 élèves avec bagage F3 sont sans bagage F1 ; l'amputation numérique est de moindre importance (1%).

Les lignes à dotation partielle (F1 sans F3 ; F3 sans F1) sont apparues pour des causes accidentelles diverses. Il était exclu de rechercher comment avaient pu apparaître, dans le fichier UNION, ces lignes à dotation incomplète : tête sans corps, ou corps sans tête.

L'effectif avec bagage F1 est supérieur à l'effectif avec bagage F3 (écart 4 107). Ce déséquilibre pose un problème technique. On pensait avoir, en théorie, deux effectifs égaux, dans les fichiers spécialisés Fic1 et Fic3 ; en comptant les élèves avec "test fantôme" en Fic1, et les élèves avec deux "bulletins fantômes" en Fic3. L'écart de 4 017 lignes peut résulter aussi bien d'un manque accidentel d'élèves dans le fichier Fic3, que d'un excès accidentel d'élèves dans le premier fichier (Fic1). Bien que paradoxal, l'excès de lignes dans le fichier Fic1 est vraisemblable, compte tenu des conditions délicates de la création initiale de ce premier fichier (Fic1). Pour avoir la liste complète des réponses aux items, il a fallu coupler les apports de deux cartes perforées (n°1 + n°2), en général. La clé de couplage des deux cartes a été le matricule de l'élève, inclus dans la séquence initiale (Sq10) répétée sur la carte initiale (n°1) et sur la carte suite (n°2). Les erreurs de perforation, dans la double saisie de ce matricule, ont produit dès 1966, dans le fichier spécialisé Fic1, des lignes incomplètes. Par exemple, on avait des tests incomplets : les 40 items de la carte n°1 sans les items de la carte suite. Très probablement, les tests incomplets de ce type ont été considérés comme décrivant des élèves réels démunis de test ; l'effectif initial

du fichier Fic1 aurait ainsi été gonflé artificiellement.

II-8. Raffinement 1969 du couplage : élèves avec double bagage bien couplé.

-a) Possibilité d'élèves "chimères" sur des lignes à double bagage.

L'apparence d'un élève complet sur une ligne à double bagage suggère que l'on a un élève réel. Mais la ligne à double bagage (F1+F3) peut aussi présenter un organisme "chimère", réunissant la tête de Pierre (bagage F1) et le corps de Paul (bagage F3). Un tel couplage erroné des organes de deux élèves distincts découle de certaines erreurs de matricule. On est parti de deux élèves réels (Pierre et Paul), chacun étant présent dans les deux fichiers de base (Fic1 et Fic3). Supposons des erreurs dans la double saisie du matricule de Pierre, ou dans la double saisie du matricule de Paul ; dans des cas particuliers, ces erreurs ont pu conduire, dans le fichier Union, à présenter sur une ligne un élève "chimère".

Ces lignes "chimères" ne sont pas identifiées individuellement. Avant les traitements compréhensifs, une précaution a été prise pour s'en protéger. Telle a été la fonction des variables de cohérence entre le contenu du bagage F1 et le contenu du bagage F3, dans le cas des lignes avec double bagage (F1+F3). Créés dès après la création du fichier "union", on a six variables de cohérence ; elles font partie des dix variables fossiles qui forment le bagage intersection (F3x1).

-b) Usage des variables en "tronc commun".

La cohérence ou l'incohérence des deux bagages (F1 F3) a été constatée sous l'angle de diverses variables en "tronc commun". Sous ce terme, on désigne des informations cruciales (le sexe de l'élève, sa date de naissance détaillée, et son niveau scolaire à 17 modalités) ; elles ont été perforées deux fois, pour figurer dans chacun des deux fichiers de base (Fic1, Fic3).

En un premier temps, ces variables en tronc commun ont été une source indispensable dans les deux exploitations autonomes :

- exploitation du fichier (Fic1) pour la recherche psychotechnique,
- exploitation du fichier (Fic3) pour l'étude des élèves sans apport du test.

Par la suite, les variables en tronc commun se sont trouvées importées, en parallèle, dans les deux bagages de base (F1 et F3). Lors de la création (1969) du fichier UNION, on avait donc sur une ligne avec double bagage (F1+F3), deux images de chaque information du tronc commun (tableau 7) : le sexe de l'élève, par exemple, était présent deux fois : dans la suite d'images (tc') en bagage F1, et la suite parallèle (tc") en bagage F3. On avait aussi deux images du premier caractère de la strate (alors que le second caractère, composante du matricule à 9 caractères, n'apparaît qu'une fois).

Tableau 7. Création 1969 des variables de cohérence entre bagages (F1 F3)
sur les lignes avec double bagage (F1+F3) du fichier UNION (FicU).

		bagage F1	bagage F3	
Fic1	= $\mu 9'$.. / 1 / ...	<u>A</u> <u>tc'</u>	
Fic3	= $\mu 9''$ / 3 /	<u>tc''</u> <u>B</u>	
		bagage F1	bagage F3	
FicU				3 types de dotation :
5.375	M9 = $\mu 9'$ <u>A</u> <u>tc'</u>	/v/___/v/	avec bagage F1 sans bag.F3
1.268	M9 = $\mu 9''$ <u>/u/</u> <u>/u/</u>	<u>tc''</u> <u>B</u>	avec bagage F3 sans bag.F1
116.699	M9 = $\mu 9'=\mu 9''$... <u>A</u> <u>tc'</u>	<u>tc''</u> <u>B</u>	avec double bagage F1+F3:
			'-----'	
	Comparaison	(tc' = tc'')	- - ->	variables de cohérence
				en bagage F3x1
				intersection de F3 et F1
Légende: $\mu 9'$ = matricule à 9 caract. présent dans le premier fichier spécialisé (Fic1) $\mu 9''$ = matricule à 9 caract. présent dans l'autre fichier spécialisé (Fic3) M9 = matricule obtenu sur une ligne du fichier union (FicU) A = variables propres au bagage F1 (exemple: résultat au test) B = variables propres au bagage F3 (exemple: catégorie socio-professionnelle)				
TC = variables en tronc commun (étaient présentes à la fois en Fic1 et Fic3) tc' = image des variables TC dans le fichier spécialisé Fic1 tc'' = image des variables TC dans l'autre fichier spécialisé (Fic3) Dans la création de variables de cohérence entre bagage F1 et F3, ont été utilisées en 1969 : date de naissance complète, niveau scolaire, 1 ^o caractère de la strate.				

-c) Cohérence entre les deux bagages (F1 F3).

Pour chaque information connue par deux images, on a cherché s'il y avait équivalence⁶ de contenu des deux images. La cohérence maximale entre les deux bagages (F1 et F3) est alors définie quand il y a équivalence pour chaque information, à la fois sous l'angle de la date de naissance complète, du niveau scolaire (à 17 modalités), et du premier caractère de la strate. Seule la double image du sexe n'a pas été utilisée pour définir la cohérence entre les deux bagages.

Sur les 116 699 élèves avec double bagage (F1+F3), le traitement fait en 1969 a montré 2 169 élèves sans cohérence maximale. Dans ce cas, le couplage des deux bagages a été jugé douteux, pouvant contenir un élève chimère. La cohérence maximale est apparue pour 114. 530 élèves. Seuls des élèves de ce type (double bagage bien couplé) ont été exploités en 1969-1970 dans la description compréhensive.

⁶ Dans le cas général, la cohérence se déduisait de l'identité stricte de contenu. Toutefois, ce n'est pas toujours le cas. Exemple pour la comparaison du jour de naissance, chiffré avec deux codages différents : avec redressement en bagage F1, mais sans redressement en bagage F3. Un jour de naissance inconnu en bagage F3 ne fait pas repousser l'équivalence avec le jour connu affiché en bagage F1 ; les deux bagages sont alors jugés cohérents, sous ce point de vue partiel..

Tableau 8. Double bagage bien couplé: effectifs dans le fichier fossile 1970.

123.342 lignes du fichier UNION 1969					
		5.375	116.699	1.268	
	Tous avec BAGAGE F1	avec bagage F1 sans F3	avec double bagage F1+F3	avec bagage F3 sans F1	Tous avec BAGAGE F3
<u>Fic1</u>	--> 122.074	= 5.375	+	116.699	
	(I)	(II)		116.699 + 1.268	= 117.967 <-- <u>Fic3</u>
Double bagage douteux 1969 (a)			116.699		
			- 2.169 (III)		
Double bagage admis en 1969			114.530 (IV)		

Pour définir les élèves admis pour l'exploitation (lot IV), la procédure historique a donc procédé en deux temps : après l'apparition des lignes avec double bagage, a eu lieu la sélection des seules lignes montrant la cohérence maximale entre les deux bagages rassemblés. Mais le contenu logique de la démarche aurait pu être obtenu d'un seul coup ; les élèves admis pour l'exploitation (lot IV) seraient apparus dès la création du fichier UNION, si le couplage des deux fichiers sources (Fic1 et Fic3) avait été fait avec une clé très large, équivalente à 17 caractères dans le dessin des cartes perforées initiales. Les 114 530 élèves du lot (IV) seraient apparus avec une clé comprenant, non seulement les 9 caractères du matricule (clé utilisée en 1969), mais aussi le premier caractère de la strate, et les 7 caractères qui donnaient initialement la date de naissance complète (jour, mois, année) et le niveau scolaire (17 modalités).

-d) Allègement final du tronc commun, initialement en deux images.

Dans le lot (IV) des 114 530 élèves avec double bagage bien couplé, les deux images des variables en tronc commun sont équivalentes ; et même, le plus souvent, elles ont un contenu identique (voir la note précédente). Rappelons toutefois que, pour le sexe, l'égalité des deux images, en F1 et en F3) n'a pas testée.

En 1969, le seul objectif était la description compréhensive, à réaliser avec des élèves à double bagage bien couplé ; dès lors, pour ces élèves, il était sans intérêt de conserver les deux images (équivalentes) des variables en tronc commun. Pour alléger le fichier sur bobine magnétique, une fois la cohérence maximale définie, la décision a été prise d'éliminer une des deux images de certains duos.

Une complexité du fichier fossile 1970 vient de ce que les élagages faits dans les duos d'images du tronc commun ont porté tantôt sur le bagage F1, tantôt sur l'autre (F3), et aussi que quelques duos ont été conservés.

Par exception à ces élagages, le fichier fossile 1970 a conservé le duo des deux images de 3 variables détaillant la date de naissance : le jour de naissance, le mois de naissance, et l'âge au 1/1/1965 en années, avant 20 ans (découlant de l'année de naissance, mais en excluant les élèves de plus de 20 ans, erreurs manifestes de perforation). Pour décrire les élèves avec double bagage bien couplé, ce sont les seules variables qui aient été conservées sous leur double forme. Vu le rôle essentiel de l'âge dans toute analyse, la conservation des deux images de la date de naissance pouvait permettre une analyse des lignes à dotation partielle (5 375 élèves avec F1 sans F3 ; 1 268 élèves avec F3 sans F1). On a ainsi les moyens de définir l'âge au 1/1/1965 pour tous les élèves formant le fichier UNION en 1969, en puisant dans l'un ou l'autre des deux bagages (F1 F3). Une telle information extensive n'a pas été conservée pour les autres dimensions : niveau scolaire, strate détaillée, sexe.

Les variables abandonnées sont surtout les images qui existaient dans l'état initial du bagage F1. Par exemple, c'est le bagage F3 qui a été préféré comme source exclusive des informations sur le sexe et sur le niveau scolaire. Ainsi, le souci d'alléger le fichier a contribué à rendre presque irréversible le choix, en 1969, des conditions dans lesquelles devait se faire la recherche compréhensive. Ces décisions vont retreindre aussi la liberté des chercheurs du XXI^e siècle, pour tirer parti des données du fichier fossile 1970. On le verra (chapitre III) dans les propositions des divers champs d'exploitation cohérents, sous forme de commandes "clés en main".

II-9. Choix des élèves fait pour la description compréhensive (1969-1970).

L'objectif, en 1969-1970, est la description compréhensive, reliant les résultats au test aux conditions hors du test. N'ont été exploités que des élèves avec double bagage, mais avec des conditions restrictives (tableau 17).

Au sein des 116 699 élèves avec double bagage (F1+F3), les conditions retenues sont au nombre de quatre :

- a)- exclure 2 169 élèves dont le couplage est douteux.
- b)- exclure les élèves sans test (et donc sans score interclassant).
- c)- exclure les élèves observés en CP avec le cahier II.
- d)- exclure les élèves sans score QI dans la métrique INED.

Tableau 9. Les élèves exploités dans la description comprehensive (1969-1970).
Effectifs retrouvés dans le fichier fossile 1970.

		123.342 lignes du fichier UNION 1969					
		Tous avec BAGAGE F1	avec bagage F1 sans F3	avec double bagage F1+F3	avec bagage F3 sans F1	Tous avec BAGAGE F3	
<u>Fic1</u>	--> 122.074	=	5.375	+	116.699		
	(I)		(II)		116.699	+	1.268
						=	117.967 <-- <u>Fic3</u>
				116.699			
				(a) - 2.169 (III)	Double bagage douteux 1969		
				114.530 (IV)	Double bagage admis en 1969		
Sans test (S) 7.066				(b)	En (IV) sans test exploitable		
				(c) - (c)	En (IV) avec cahier 2 en CP		
				- 9.962 (b+c)	En (IV) Elèves non exploités		
Avec test (T) 115.006				104.568 (V)	Exploitable avec score interclassant en champ agrégé prudent		
Sans QI (Th) 15.918				(d) - (14.568)	En (V) Elèves sans QI défini		
				(provisoire)			
Avec QI (Tq) 99.088				(90...:)(VI)	Exploités: 6 à <14 ans; avec QI en champ agrégé prudent.		

En se limitant aux trois premières conditions (a b c), on avait 104 568 élèves connus par le score interclassant dans le champ agrégé prudent. En ajoutant la quatrième condition (d), qui limite l'étude aux élèves munis d'un quotient intellectuel, et donc en âge d'obligation scolaire, l'effectif est de l'ordre de 90 000 (à préciser).

Toute la description comprehensive en 1969-1970 a pu être faite avec l'emploi exclusif d'un seul système de pondération, prévu pour les élèves avec test dans le champ agrégé prudent.

Hormis la condition (a) qui exclut les élèves avec un double bagage douteux, les trois conditions choisies (b c d) ont été définies à partir de variables du bagage F1 ; elles sont issues de l'exploitation autonome du fichier Fic1.

Aucune condition restrictive n'est intervenue provenant du bagage F3. Les lacunes dans les informations attendues de ce bagage sont très diversifiées. Les unes découlent de l'absence du bulletin individuel, les autres à l'absence du bulletin familial, mais certaines variables sont alimentées, pour certains élèves, par le bulletin individuel venant en substitut du bulletin familial fantôme.

= Les effectifs exploités selon les publications (opus INED n°64).

L'affirmation qu'aucune condition restrictive n'est intervenue provenant du bagage F3 semble contredite dans une publication (opus INED n°64, p. 26-27). On y lit les effectifs suivants :

-(a) 103 781 élèves avec test et au moins un des deux bulletins (Bi Bf)

-(b) 791 élèves avec test et démunis des deux bulletins

a+b 104 572.

Le total (a+b) retrouve l'effectif affiché (V) en tableau 17, presque parfaitement (104 568 ; l'écart de 6 unités n'est pas expliqué).

Mais la publication dit aussi que "l'exploitation qui fait l'objet de cette publication porte sur 103 781 élèves". C'est une rédaction cavalière, puisque certains des tableaux publiés sont issus du traitement autonome du fichier Fic3, et ne sont donc pas limités aux élèves avec test présent.

Si l'on part de la page citée (opus INED n°64), on ne retrouve pas les effectifs qui figuraient dans le fichier Fic3 :

-(a) 103 781 avec test et au moins un des deux bulletins (Bi Bf)

-(a') 6 218 sans test et au moins un des deux bulletins (Bi Bf)

- (A) 109 997 élèves (a+a') ayant au moins un des deux bulletins (Bi Bf)
- (b) 791 élèves avec test sans aucun des deux bulletins
- (C) 110 788 (A+b)

Ce total (C) est très inférieur au nombre des élèves avec bagage F3, donné ci-dessus (117 967 en tableau 9) ; c'est l'effet de trois exclusions :

- 1- exclusion des élèves qui n'ont pas le double bagage (F1+F3),
- 2- exclusion des élèves dont le double bagage est mal couplé,
- 3- exclusion des élèves de CP observés avec le cahier II.

Les deux premières conditions sont évoquées en ces termes "cohérence entre les trois sources de données", dans la note de bas de page 26 (opus INED n°64).

Au contraire, les ventilations faites du fichier Fic3 (en 1968) ont été faites sans ces trois exclusions. Certaines ventilations écartaient les élèves démunis des deux bulletins. Dans d'autres cas, on a écarté seulement les élèves sans bulletin individuel, ou seulement les élèves sans bulletin familial.

II-10. Dans le bagage F1: élèves sans test, système de pondération des élèves avec test

Quelques aspects de l'exploitation autonome du premier fichier spécialisé (Fic1) importent beaucoup pour préciser le contenu du bagage importé F1, et donc aussi pour tout emploi du fichier fossile 1970.

-a) Le fichier spécialisé Fic1 contenait aussi les élèves sans test.

Le fichier Fic1 contenait, en théorie, tous les élèves, y compris ceux démunis de test, qui étaient connus seulement par la présence d'un des bulletins (individuel ou familial). Pour ces élèves, au lieu du test réel dans le dossier papier, on avait un document "test fantôme". La distinction entre test fantôme et test réel est le constat fait par le chiffreur sur la présence des documents. De son côté, l'autre fichier de base (Fic3) devait contenir, lui aussi, tous les élèves : y compris ceux avec un "bulletin individuel fantôme" et un "bulletin familial fantôme" ; ces derniers étaient connus seulement par la présence d'un test réel.

Pour tout élève présent en fichier Fic1, on connaît la date de l'examen (passation du test) ; date prévue pour une classe entière, y compris pour tout élève sans test. Il a donc été possible de calculer pour tout élève membre du fichier Fic1, deux sortes d'âge : non seulement **son âge au 1/1/1965** (déduit seulement de son année de naissance), mais encore **son âge à la date du test**. Dans la recherche psychotechnique, c'est l'âge à la date du test qui intervient. Son calcul a été fait en utilisant la date de naissance détaillée (jour, mois, année) et la date du test (jour et mois en 1965) ; le résultat a été arrondi en classes d'âge en mois révolus le jour du test.

-b) Le numéro de cahier de test désigné (connu en Fic1) n'a pas été conservé.

En théorie, tout élève devait passer le **cahier de test désigné** par le plan de collecte. Le cahier **désigné** était connu, même en l'absence de test exploitable. Dans la présentation théorique du plan de recherche (chapitre I, tableau 4), il n'est pas prévu d'élèves sans test. C'est selon le cahier de test désigné qu'est défini ce plan. De même, la logique des différents systèmes de pondération est présentée en fonction du cahier **désigné**. La ventilation des élèves publiée en Tome 1 (p. 82-83) est faite selon le cahier **désigné**.

Mais la variable qui donnait le numéro du cahier désigné n'a pas été conservée pour former le bagage F1. Elle fait partie du lot de variables représenté par le symbole /1/ en tableau 4. Seul le numéro du cahier présent est disponible.

-c) Repérage des élèves avec un test présent et exploitable.

Pour la recherche psychotechnique, il a fallu repérer, dans le premier fichier (Fic1), les élèves avec un test exploitable. Ont donc été exclus, tout au long de la recherche psychotechnique, 7 068 élèves sans test exploitable : ils avaient seulement le document dit "test fantôme", ou un test réel non exploitable. Ont été exploités les 115 006 **élèves avec test présent exploitable** (selon le fichier fossile 1970).

-d) En bagage F1, sont disponibles des systèmes de pondération pour élèves avec test.

Antérieurement (chapitre I), la définition a priori de trois systèmes de pondération a été présentée en tenant compte du **cahier de test désigné**. Selon ces principes, on avait des poids simples (c'est à dire: sans redressement). En 1966, de tels poids sans redressement ont été d'abord calculés dans le fichier Fic1. Mais ils n'ont pas été conservés dans le bagage F1.

Seuls des poids avec redressement, définis seulement pour les élèves avec test, sont conservés en bagage F1. Ils ont été conçus pour des traitements limités aux élèves avec test exploitable. L'effet des redressements est, notamment, de majorer le poids des seuls élèves avec test, pour compenser l'absence des élèves sans test dans les distributions des notes. À la différence des exposés a priori (chapitre I, § 6), c'est seulement pour les élèves **avec test exploitable** que sont disponibles, en bagage F1, les trois systèmes de **pondération** :

- pondération, avec redressements, des élèves **avec test** dans le champ agrégé complet,
- pondération, avec redressements, des élèves **avec test** dans le champ agrégé prudent,
- pondération mono-cahier, avec redressements, des élèves **avec test**.

Il est donc absurde d'employer, sur tout ou partie des 7 068 élèves sans test, les systèmes de pondérations conservés en bagage F1. En d'autres termes, dans le fichier fossile 1970, il n'existe aucun système de pondération pour exploiter de façon représentative l'ensemble des élèves qui formaient le fichier Fic1 (y compris les sans test).

Mais on dispose, en bagage F3, d'un poids simple (= sans redressements) défini pour les élèves présents en Fic3. C'est le poids qui conviendra au champ agrégé complet incluant des élèves sans test.

II-11. Élèves avec ou sans bagage complémentaire F4.

L'apparition de lignes à dotation incomplète est un processus général, constaté ci-dessus dans la création du fichier UNION. D'autres imperfections vont conduire à l'absence du bagage F4 sur certaines lignes du fichier fossile 1970. En considérant les trois bagages importés (F1 F3 F4), une ligne peut être garnie, ou non, de chacun des trois. Le tableau 10 montre que les huit profils de dotation possibles se rencontrent, parmi les 123 635 lignes du fichier fossile 1970. Parmi les 8 profils, deux sont sans grand intérêt (lignes volantes sans bagage F1 ni F3). La fréquence inégale des 6 autres profils montre que les lacunes sont associées : l'absence du bagage F4 est forment lié à l'absence du bagage F3. Les lacunes se cumulent sur certains élèves.

Tableau 10. En tenant compte du complément possible en bagage F4, les 8 profils de dotation selon la présence des trois bagages (F1 F3 F4).

		123.342 lignes du fichier UNION 1969 (FicU) (3 profils de dotation)				Hors FicU	Total fichier fossile 1970
	En fichiers spécialisés	avec F1 sans F3	avec double bagage F1+F3	avec F3 sans F1	Total en fich. UNION FicU		
Fic1	-> 122.074	= 5.375	+ 116.699				
Fic3	-> 117.967	=	116.699	+ 1.268			
		5.375	+ 116.699	+ 1.268	= 123.342		
En gras (8 profils)							
Sans F4		4.885	2.925	132	7.942	145	8.087
Avec F4	Fic4 ->	490	113.774	1.136	115.400	148	115.548
		5.375	+ 116.699	+ 1.268	123.342	293	123.635
Trois profils de dotation en UNION					Total en FicU	Lignes volantes	Total 1970

En pratique, et heureusement, parmi les 116 699 élèves avec double bagage de base (F1+F3), presque tous (113 774, soit 97%) sont dotés en complément par le bagage F4.

Il y a aussi l'absence du bagage F5 sur certaines lignes. En considérant les quatre bagages importés (F1 F3 F4 F5), une ligne peut être garnie, ou non, de chacun des quatre bagages. Quinze profils de dotation sont possibles, et sont retrouvés dans le fichier fossile 1970.