

La valeur des examens

**Etude
docimologique
réalisée au Liban
par E.Valin**

unesco

PREFACE

La docimologie ou science des examens est une discipline récente. Jusqu'au XIX^e siècle, les études post-primaires étaient pratiquement réservées aux enfants des classes aisées et ne représentaient pas un sacrifice considérable pour les familles. La situation s'est transformée à mesure que se démocratisait l'enseignement : d'une part, le volume réduit des établissements obligeait les pouvoirs publics à sélectionner les candidats à l'admission ; d'autre part, les élèves issus de classes sociales à revenus modestes étaient poussés à conquérir titres et diplômes dans les plus brefs délais.

Dès le début de ce siècle, l'attention avait déjà été attirée sur le caractère incertain des examens classiques. Et, entre les deux guerres mondiales, des travaux américains, anglais et français, utilisant les ressources de la statistique, confirmèrent la faiblesse des épreuves dans lesquelles on avait eu jusqu'alors pleinement confiance.

Le retentissement de ces recherches fut tel que la Fondation Carnegie patronna en 1930 la réunion de conférences internationales sur les examens. Des opinions très opposées furent émises au cours de ces rencontres, les délégués de langue anglaise manifestant leurs préférences pour les épreuves du type « test », la plupart des délégués de formation française restant plus attachés aux « compositions » classiques.

Les nécessités de l'après-guerre ont accru l'intérêt porté aux méthodes docimologiques depuis 1945. En dehors des Etats-Unis d'Amérique, depuis longtemps outillés à cet égard, le mouvement s'est largement répandu au Royaume-Uni où la sélection en vue des études secondaires et des études supérieures a imposé le recours à des épreuves de plus en plus conformes aux principes statistiques ; des progrès remarquables ont été accomplis, particulièrement sur l'initiative du Scottish Council for Research in Education et de la National Foundation for Educational Research in England and Wales. En France, l'INOP (Institut national d'orientation professionnelle) a accompli une œuvre de pionnier en la matière et les résultats de ses travaux commencent à pénétrer dans l'enseignement public ; l'organisation des épreuves d'admission à l'ensemble des centres d'apprentissage de l'Académie de Paris est un témoignage significatif de cette évolution.

L'Unesco n'a pas manqué de porter un intérêt actif au mouvement pour l'amélioration des épreuves scolaires. En 1957 et en 1958, l'Institut de l'Unesco pour l'éducation (Hambourg) a organisé des rencontres sur les examens et l'évaluation dans l'enseignement ; la synthèse de ces travaux a fait l'objet d'une publication.¹ L'Unesco ne se borne d'ailleurs pas à considérer avec sympathie les aspects théoriques de ce problème. On pourra constater, à la lecture de l'ouvrage que nous présentons, que les experts envoyés par elle dans le cadre de l'assistance technique, ne se contentent pas de transmettre les pratiques traditionnelles de l'Occident dans les régions où ils s'efforcent de promouvoir l'éducation ; à l'exemple de M. Vallin, ils apportent à l'enseignement de ces pays les acquisitions récentes de la docimologie. Ce souci est particulièrement opportun dans les régions en cours de développement, où les ressources limitées rendent impérieuse la recherche d'un maximum de rendement et où chaque intelligence judicieusement éveillée constitue une précieuse valeur d'avenir.

1. Institut de l'Unesco pour l'éducation (Hambourg), *L'évaluation en éducation*, rapport rédigé par F. Hotyat, Hambourg, 1958, 72 p. miméo.

TABLE DES MATIERES

INTRODUCTION	5	Caractéristiques métrologiques	20
CHAPITRE PREMIER. IMPORTANCE DES EXAMENS		<i>Fréquence des notes attribuées</i>	20
Qu'est-ce qu'un examen?	7	<i>Sélectivité de la notation</i>	22
L'examen-révélateur	7	<i>Pondération théorique et pondération réelle</i>	22
L'examen-stimulant	8	Jugement et mesure	24
Caractères propres aux pays en voie de développement accéléré	8	La double correction	24
Limites de caractère général	8	<i>L'équation personnelle</i>	24
CHAPITRE 2. ORGANISATION DE L'ENSEI- GNEMENT ET DES EXAMENS AU LIBAN		<i>L'accord sur le classement</i>	24
L'enseignement	10	<i>L'accord sur la note</i>	25
Le Centre de recherches pédagogiques et la mission de l'Unesco	10	Valeur de la décision	25
Examens ayant fait l'objet de notre étude	11	CHAPITRE 6. LES FACTEURS DE LA NOTATION	
CHAPITRE 3. VALEUR DIAGNOSTIQUE DE L'EXAMEN		Le correcteur	26
Homogénéité de l'examen	13	<i>Fréquence et dispersion des notes</i>	26
<i>Liaisons écrit-oral et surestimations à l'oral</i>	13	<i>Fluctuations dans le temps</i>	26
<i>Pour ou contre l'oral?</i>	13	<i>Style de notation</i>	26
Fidélité de l'examen	15	La copie	26
<i>Liaisons entre les écrits des deux sessions</i>	15	<i>Influence de la qualité</i>	26
<i>Surestimation à la deuxième session</i>	15	<i>Critères de la qualité</i>	29
<i>Comparaison des seuils d'admission (1re et 2e sessions)</i>	15	<i>Influence du sujet traité</i>	29
<i>Pour ou contre la deuxième session?</i>	15	Conclusions	30
CHAPITRE 4. VALEUR PRONOSTIQUE DE L'EXAMEN		CHAPITRE 7. REFORME FONDAMENTALE ET REMEDES PARTIELS	
Du primaire au secondaire	18	Pour une réforme fondamentale	31
Choix des critères de réussite scolaire	18	<i>Orientation plutôt que sélection</i>	31
Résultats et conclusions	19	<i>Preuve plutôt qu'épreuve</i>	31
CHAPITRE 5. VALEUR METROLOGIQUE DE LA NOTATION		Suggestions pour un examen de fin d'études	32
Le problème central: la notation	20	<i>Pour un diagnostic valable</i>	32
		<i>Pour une notation plus fidèle</i>	34
		Suggestions pour un concours d'entrée	35
		<i>Une batterie de tests appropriés</i>	35
		<i>L'apport des tests</i>	35
		<i>Tests d'intelligence et développement économique et social</i>	36
		CHAPITRE 8. RESUME ET CONCLUSIONS	38
		BIBLIOGRAPHIE	39

LISTE DES TABLEAUX ET FIGURES

Tableau 1. Valeur pronostique du C.E.P.	19	Figure 1. Comparaison des notes écrites et orales par matière (sur 20)	14
Tableau 2. Ecart-types des notes sur 20	22	Figure 2. Comparaison des notes de l'écrit de la 1re et de la 2e session	16
Tableau 3. Pondération effective des épreuves	22	Figure 3. Fréquences des notes sur 20 des trois épreuves écrites	21
Tableau 4. Corrélations entre correcteurs	24	Figure 4. Discrimination entre admis et refusés	23
Tableau 5. Pourcentages d'accords et de désaccords	25	Figure 5. Exemples de mauvaise et de bonne sélectivité («Admis-Refusés»)	23
Tableau 6. Indulgence ou sévérité élective	29	Figure 6. Caractéristiques personnelles de la notation	27
Tableau 7. Corrélations entre tests de connais- sances et épreuves du C.E.P. correspondantes	36	Figure 7. Variation des moyennes intercorrecteurs et intracorrecteurs	28
Tableau 8. Corrélations entre tests d'intelligence et tests ou épreuves scolaires	36	Figure 8. Tripartition des résultats de l'écrit	33
Tableau 9. Validité comparée des tests et des épreuves de connaissances scolaires	36		

Depuis qu'ils existent, les examens scolaires ont toujours préoccupé les spécialistes. Jusqu'à la première guerre mondiale, ils n'avaient fait l'objet que de délibérations. C'est en 1922 que Piéron et Laugier leur consacrèrent, en France, les premières études, réalisées selon les méthodes objectives de la science expérimentale. La docimologie¹ était née. Quelques années plus tard, elle prenait un essor international avec l'enquête de la Fondation Carnegie qui réunissait des participants américains, allemands, anglais, écossais, français et suisses.

Les résultats de cette enquête sur l'examen du baccalauréat furent publiés en 1936. D'une part, les opinions exprimées par les représentants français reflétaient généralement une grande confiance dans les examens, et un certain conservatisme opposé à l'introduction de méthodes objectives de contrôle. D'autre part, les études expérimentales de correction multiple poursuivies notamment par Hartog en Angleterre, Starch et Elliott aux Etats-Unis d'Amérique, Laugier et Weinberg en France aboutirent à des résultats dont la similitude ne manquait pas d'être troublante, puisqu'ils montraient tous que les divers correcteurs d'une épreuve écrite différaient notablement dans leurs marges de notation et dans les notes attribuées à la même copie.

Depuis, de nombreuses études sont venues étayer ces conclusions, qui ne sont plus aujourd'hui contestées. Par contre, les solutions proposées - comme l'emploi de méthodes objectives - l'ont été davantage et le sont encore².

Cependant, sous la pression d'une opinion publique de plus en plus sensibilisée au problème, la double correction s'imposa, et prit même, en quelques pays, un caractère obligatoire, au moins pour les candidats arrivant immédiatement au-dessous du seuil d'admission.

Après la dernière guerre, d'autres problèmes surgissaient, qui mettaient les précédents en veilleuse. En effet, l'accroissement considérable de la population scolaire, en France plus particulièrement, donnait alors une actualité aiguë à la question de l'organisation même des examens, trop longs, trop

onéreux et qui amputaient de plus en plus l'année scolaire.

La recherche d'une solution adaptée fit couler beaucoup d'encre ces dernières années. Fallait-il supprimer l'examen ou se contenter de l'amender? Aujourd'hui encore, les propositions ne manquent pas. Rares sont celles qui réussissent à franchir les obstacles opposés par la tradition. L'examen cherche ses buts et ses moyens dans le cadre d'une éducation nouvelle et sous la pression de l'accroissement de la natalité dans tel pays évolué, de l'extension de la scolarisation dans tel autre pays en voie de développement.

L'étude que nous présentons n'a d'autre ambition que d'apporter la confirmation que la plupart des conclusions des études docimologiques réalisées en Amérique et en Europe sont également valables dans des pays où s'est implantée une culture différente de la culture originelle et de mettre en lumière certains problèmes spécifiques à ces pays. Elle vise aussi et surtout à tirer d'une analyse objective des données, des recommandations pratiques pour une réforme des examens.

Ses limites temporelles et géographiques doivent évidemment nous inciter à la prudence quant à la généralisation de nos conclusions et recommandations. Il nous semble cependant que la similitude des examens libanais que nous avons étudiés avec ceux d'autres pays, ainsi que le parallélisme de la plupart des problèmes qu'ils soulèvent, permettent dans une assez large mesure, l'extension de nos conclusions et la généralisation de certaines des réformes proposées.

De toute façon, elle prouve une fois de plus qu'une étude objective des examens est possible et que des méthodes telles que celles que nous avons utilisées sont de nature à aiguiller les tentatives de réforme aussi bien que d'en contrôler la valeur.

1. Le terme « docimologie » (du grec « dokime » = épreuve) a été proposé par M. Piéron pour désigner l'étude systématique des examens.
2. (P.E.) Vernon, *Secondary school selection. A British Psychological Society inquiry*, London, Methuen, 1957, 216 p.

IMPORTANCE DES EXAMENS

QU'EST-CE QU'UN EXAMEN ?

Plutôt que de tenter une définition précise de l'examen, nous allons indiquer comment on le conçoit selon qu'on le considère « d'en haut » ou « d'en bas », selon qu'on l'administre, qu'on y prépare ou qu'on le subit.

Voici d'abord l'opinion d'un haut fonctionnaire de l'enseignement : « Institués pour sanctionner ce qui a été fait pendant la durée des études, pour peser la somme de connaissances et déceler le degré d'intelligence de chaque candidat, les examens apparaissent comme des instruments de contrôle et de mesure auxquels se trouvent soumis méthodes d'éducation, procédés d'enseignement et techniques de préparation ».¹

Ainsi, de ce point de vue, l'examen est un instrument de mesure destiné à contrôler à la fois la valeur du candidat et la valeur de la pédagogie à laquelle ce candidat a été soumis. L'examen est donc envisagé comme un test, un *révélateur* de la formation donnée comme de la formation reçue.

Par ailleurs, il est de constatation courante que, pour la plupart des maîtres, les examens constituent le but même de leur enseignement. Comment en serait-il autrement puisque le pourcentage de succès à l'examen est généralement considéré comme le critère public de la réussite professionnelle de l'éducateur ?

Les élèves, de leur côté, voient plutôt l'examen sous l'angle sportif. Ils l'envisagent généralement comme une compétition, un obstacle à franchir pour « décrocher » un diplôme considéré par certains comme un certificat de fin d'études, et par d'autres comme une autorisation de poursuivre des études à un niveau supérieur.

Selon ces deux optiques, l'examen, vu d'une salle de classe, apparaît donc comme un but à atteindre, une motivation, un *stimulant*.

Révélateur ? Stimulant ? Nous allons voir que ces propriétés dépassent de beaucoup le cadre restreint de l'individu.

doit-il témoigner ? D'une somme de connaissances, de l'aptitude à les utiliser intelligemment, du degré d'intelligence du candidat, ou de tout cela à la fois ? On ne sait trop si la question préoccupe toujours les responsables du choix des épreuves, et les responsables de l'estimation des performances.

D'un autre point de vue, plus pragmatique, l'examen doit-il contrôler la formation acquise ou la capacité d'en acquérir une autre ? En fait, il ne semble pas y avoir de différences radicales entre l'examen terminal et le concours d'entrée, entre l'examen-diagnostic et l'examen-pronostic.

Toutes ces questions, ainsi que le problème de la valeur du contrôle, feront l'objet des chapitres suivants, et nous essaierons d'y répondre.

Ce qui nous préoccupe pour l'instant ce sont des aspects de l'examen moins connus du grand public mais qui constituent pour les spécialistes d'importantes sources de renseignements. Je veux parler non seulement de l'intérêt prévisionnel, pronostique des statistiques des examens, aussi bien dans la planification et la productivité de l'enseignement que dans celles du développement économique, mais aussi de la « réactivité » de l'examen à certaines influences que l'exploitation statistique des effectifs et des résultats permet de mettre en évidence. En voici un exemple :

L'analyse des résultats dans un certain nombre de tests de connaissances scolaires appliquée au concours d'entrée dans l'enseignement secondaire officiel libanais en 1957 a fait apparaître, chez les garçons, un bagage de connaissances légèrement supérieur à celui des filles, sauf en langue française. Pour expliquer cette relative supériorité des filles en français nous n'avons pu invoquer aucun des facteurs habituels. Nous avons finalement trouvé une hypothèse explicative lorsqu'un dépouillement approximatif du milieu confessionnel eut révélé que 71% des filles et seulement 25% des garçons de notre échantillon étaient issus d'un milieu confessionnel dans lequel la langue française est plus répandue.

L'EXAMEN - REVELATEUR

L'examen est d'abord et surtout un instrument de mesure, un moyen de contrôle, un témoignage, un réactif. Mais que doit-il mesurer, contrôler ? De quoi

1. Jean-Baptiste Piobetta, *Examens et concours*. Paris, Presses Universitaires de France, 1943, 116 p.

Sans nous étendre longuement sur ce point, nous devons montrer de quelle façon la sphère d'influence de l'examen atteint, au-delà du domaine psychologique et éducatif, le domaine social et économique.

Source de motivations pour le maître, aiguillon pour l'élève, l'examen, devenu le but final de l'enseignement, préconditionne et oriente cet enseignement. A tel point qu'on ne peut concevoir qu'une réforme de l'enseignement ou une réforme pédagogique prenne racine sans réforme correspondante de l'examen terminal. Réciproquement, toute réforme d'un examen entraîne aussitôt, qu'on le veuille ou non, un rajustement de l'enseignement préparatoire à cet examen.

Si l'on se plaint parfois qu'il y ait trop de littéraires et trop peu de scientifiques, n'est-ce pas parce que l'examen, dans son esprit, dans sa forme et dans son contenu, favorise davantage les littéraires? Si l'on déplore quelquefois que l'enseignement forme plus de gens instruits que de gens cultivés, n'est-ce pas parce que l'examen fait davantage appel aux connaissances empilées dans la mémoire qu'à des formes de pensée et d'expression plus personnelles?

A travers l'institutionnel, l'examen façonne aussi les mœurs, allant, dans certains cas, jusqu'à les corrompre! On sait fort bien que, parfois, l'importante valeur sociale du diplôme, son caractère, très fréquent de « permis d'accès » à un certain niveau de la hiérarchie sociale, de certificat « d'ayant droit » pour certains postes ou certains grades administratifs, mettent en œuvre annuellement, à la période des examens, quelques influences, interventions ou pressions d'origine sociale, raciale, politique ou religieuse. Sans prendre parti sur le fond du problème, il faut bien constater que ces influences faussent, parfois considérablement, la sélection des élites.

Signalons enfin que, sur le plan économique, par la sélection qu'il opère, par le contingent annuel de potentialités humaines qu'il libère sur le marché du travail, l'examen, sanction d'une formation, label de garantie, a une influence qualitative et quantitative déterminante sur le développement économique du pays.

CARACTERES PROPRES AUX PAYS EN VOIE DE DEVELOPPEMENT ACCELERE

Il y a cependant quelques caractéristiques de l'enseignement - communes, croyons-nous, à la majorité des pays en voie de développement - qui retentissent sur l'examen pour en restreindre l'efficacité.

L'école, au lieu d'être une préparation à la vie et au travail, devient trop souvent une fin en soi. Une fois résolu le problème crucial de l'entrée dans un établissement scolaire en général (privé, si la

famille est aisée, public, si elle ne l'est pas) l'enfant y prend racine pour de longues années. Qu'importe s'il redouble une ou plusieurs fois, s'il stagne plusieurs années dans les classes préparatoires aux examens : l'essentiel est qu'il n'en sorte pas avant d'avoir obtenu, quel que soit son âge, le sacro-saint diplôme.¹ Il en résulte que les incapables et les bien doués voisinent, les premiers freinant le développement intellectuel des seconds, encombrant les classes d'une gangue de retardés chroniques et de candidats multi-récidivistes,² et que les candidats que rassemble l'examen sont de valeur moyenne médiocre.

Schématiquement, nous pourrions résumer la situation dans cette formule : l'élève est un apprenti candidat, le candidat un futur diplômé, le diplômé un futur candidat au diplôme immédiatement supérieur.

De plus, le caractère officiel du diplôme confère curieusement, aux yeux de l'opinion, à son titulaire un droit d'accès, quasi obligatoire, à un poste d'Etat, plus administratif que technique. Ainsi s'explique la « fonctionnarité » des pays sous développés, cet engouement pour les emplois administratifs et le désintérêt pour les professions agricoles et industrielles, que la situation mineure de l'enseignement technique et professionnel, d'une part, la concurrence confessionnelle, ethnique et politique sur le plan des pouvoirs publics d'autre part, aggravent encore localement.

Si l'on n'y prend garde, la valorisation excessive du diplôme et surtout le fait qu'il ouvre un accès direct à la fonction publique quelle qu'elle soit, aboutissent à une hypertrophie du secteur tertiaire, à une pléthore de « cols blancs », à un simulacre de développement qui est la négation même du véritable développement.

LIMITES DE CARACTERE GENERAL

Bien qu'une certaine valeur y soit attachée, on n'achète plus de nos jours un diplôme. Mais n'a-t-on pas dénoncé autrefois - et ne dénonce-t-on pas encore parfois aujourd'hui - les fraudes et tricheries en cours d'épreuve, les actes de favoritisme et tentatives de corruption du jury. On a justement critiqué ces manœuvres sans, cependant, prendre suffisamment conscience qu'elles résultent de la subjectivité du contrôle, de l'imprécision du jugement et de la mesure, de l'aléa de la décision. Car les véritables limites de l'examen ne sont pas ses vices parasitaires : elles sont essentiellement d'ordre technique.

1. Comme les épreuves d'examen portent sur des connaissances acquises, emmagasinées, devrions-nous dire, les chances de réussite croissent paradoxalement en fonction du retard scolaire.
2. Au certificat d'études primaires libanais, nous avons dénombré 82% de candidats ayant 1 an et plus de retard, et 97% au Baccalauréat (première partie).

Si l'on n'achète pas un diplôme, on ne le « gagne » pas non plus, bien qu'on entende souvent dire d'un candidat qu'il va tout de même « tenter sa chance », « qu'il a eu de la chance » ou « qu'il n'a pas eu de chance ». Y aurait-il une part de hasard, une marge d'erreur dans le contrôle, dans la mesure ? N'est-ce pas pour tenter de la réduire qu'on a offert successivement aux candidats ces chances supplémentaires que sont (ou que devraient être) la session de rachat, le choix du sujet à traiter et la garantie d'une double *correction* ?

Nous touchons là le vrai problème, mais ces remèdes sont-ils efficaces ? En d'autres termes, l'examen, dans ses formes et son contenu actuels, satisfait-il aux critères qu'on est en droit d'exiger d'un instrument de mesure ?

Demandons-nous plus simplement d'abord - sans poser d'emblée le problème sur le plan métrologique - si l'instrument répond à ce qu'on lui demande, c'est-à-dire s'il permet d'effectuer un inventaire satisfaisant (dans le cas d'un examen de fin d'études) ou une prévision valable (dans le cas d'un examen ou concours d'entrée).

ORGANISATION DE L'ENSEIGNEMENT ET DES EXAMENS AU LIBAN

L'ENSEIGNEMENT :

On sait que le Liban, carrefour culturel entre l'Orient et l'Occident, est une mosaïque ethnique, politique et confessionnelle surprenante. L'organisation de l'enseignement est à l'image de cette diversité. On peut distinguer l'enseignement *officiel*, à la charge et sous la responsabilité de l'Etat, et l'enseignement *privé*, aux effectifs plus importants, qui se subdivise en *Privé national*, groupant des établissements qui appartiennent à des communautés religieuses chrétiennes ou musulmanes, et en *Privé étranger*, fonctionnant avec l'aide et sous le contrôle des missions culturelles étrangères.

Il n'est pas aisé de déterminer la proportion exacte des effectifs scolaires de ces trois types d'établissements. Toutefois, d'après les statistiques qui nous ont été communiquées ou que nous avons établies personnellement, nous pouvons estimer grossièrement qu'en 1958, sur 8 élèves qui fréquentaient les écoles primaires et secondaires, 2 étaient dans une école officielle, 3 dans un établissement privé national et, probablement 3 dans une école privée étrangère.

En fait, ces trois types d'établissements, qui jouissent d'une autonomie pédagogique quasi totale, correspondent statistiquement soit à des populations de niveau socio-économique différent (privé et officiel), soit à des populations de confession différente (privé national et étranger).

L'enseignement officiel, auquel nous nous sommes plus particulièrement intéressé comprend un cycle primaire de 5 ou 6 années qui répartit ensuite ses élèves, soit vers l'enseignement technique (5 à 10%), soit vers l'enseignement complémentaire (40 à 50%), soit vers l'enseignement secondaire (4 ou 5%), dont la création date seulement de quelques années, soit enfin vers des établissements privés.

L'enseignement complémentaire (3 ans) permet le passage dans le technique supérieur (3 ans) ou dans le secondaire terminal (3 ans).

Trois universités et un établissement d'enseignement supérieur accueillent ensuite les bacheliers. Ce sont l'American University of Beirut, l'Université St Joseph, l'Ecole des lettres (françaises) et la jeune université libanaise. Des bourses nationales, étrangères ou internationales permettent à un certain nombre d'étudiants de se spécialiser en Amérique ou en Europe.

LE CENTRE DE RECHERCHES PEDAGOGIQUES ET LA MISSION DE L'UNESCO

Le Centre de recherches pédagogiques de Beyrouth est un organisme libanais dépendant du Ministère de l'éducation nationale. Créé, équipé et animé par la mission d'Assistance technique de l'Unesco au Liban, il comprend un service de documentation et un service de recherches. Ce dernier, dans lequel nous fûmes appelé à travailler en 1957, a pour rôle essentiel de déterminer et de résoudre les problèmes pédagogiques et psychopédagogiques qui se posent au Ministère de l'éducation nationale.

Au cours de notre mission de deux années, le Centre de recherches a eu l'occasion d'aborder, parmi d'autres, les problèmes d'orientation et de sélection scolaires et professionnelles. Et c'est dans le cadre d'une étude des principaux paliers d'orientation de l'enseignement officiel que nous avons été naturellement conduit à effectuer un examen des examens, afin d'éprouver la valeur de ces compétitions scolaires qui, à la fin des cycles primaire et secondaire notamment, déterminent brutalement, dans un sens ou dans un autre, la vie scolaire et professionnelle de toute une population de jeunes.

Est-il besoin de préciser qu'il ne s'agit pas ici de faire le procès de tel ou tel examen libanais, d'ailleurs en tout point semblable aux examens européens si critiquables et tant critiqués. Nous voudrions seulement, à propos de recherches plus larges entreprises à Beyrouth¹ essayer de définir l'importance des examens ainsi que leurs limites et, par une approche méthodologique appropriée, d'en tester l'efficacité afin de suggérer quelques directions de réformes.

Nos méthodes ont été celles de la psychologie appliquée : la méthode des tests et la méthode statistique, doublées et complétées par des enquêtes, des visites aux centres d'examens, des entretiens avec des organisateurs, des correcteurs, des examinateurs et des candidats.

Si révolutionnaires qu'elles puissent paraître, nos conclusions et recommandations ne sont donc pas des constructions de notre imagination : elles sont

1. Ces recherches, et en particulier les études docimologiques dont nous rendons compte ici ont bénéficié de la collaboration technique de M. J. Maroun, du Centre de recherches pédagogiques de Beyrouth. Nous l'en remercions vivement.

objectivement fondées sur les résultats rigoureux d'une étude objective.

EXAMENS AYANT FAIT L'OBJET DE NOTRE ETUDE

Outre les diplômes américains, anglais et français auxquels ils peuvent éventuellement prétendre, les jeunes Libanais voient leurs études générales sanctionnées officiellement par le certificat d'études primaires (C.E.P.) à la fin du cycle primaire, par le brevet à la fin du cycle complémentaire, par le baccalauréat à la fin du cycle secondaire.

Le certificat d'études primaires est nécessaire pour entrer dans un établissement complémentaire ou secondaire. Le brevet est exigé pour être admis dans une classe de seconde d'un établissement secondaire ; il ne l'est cependant pas pour être inscrit sur la liste des candidats au baccalauréat.

Les deux principaux paliers d'orientation se situent au niveau du certificat (numériquement important) et au niveau du baccalauréat (socialement prestigieux).

Le *Certificat d'études primaires* réunit chaque année près de 20.000 candidatures provenant, pour les 2/3, d'écoles privées. La proportion de garçons est double de celle des filles.

Les épreuves toutes écrites ou pratiques se déroulent au mois de mai, dans cinq districts géographiques.

Elles portent sur les matières suivantes :

Matière	Durée	Coefficient
Style arabe	1 heure	3
Calcul	1 heure	2
Histoire et géographie	45 minutes	2
Leçons de choses	30 minutes	1
Dessin (pour les garçons) ou Couture (pour les filles)	1 heure	1
Français ou anglais : dictée avec questions	1 heure	3
		12

Toutes les copies sont centralisées à Beyrouth où elles sont notées par un seul jury, sous la responsabilité du Service des examens. La liste des « certifiés » est arrêtée après totalisation des notes, compte tenu des coefficients. Sont déclarés reçus les candidats ayant obtenu au moins 60 points sur 120. En fait, cette note limite inférieure est abaissée généralement à 56.

Dans l'ensemble du pays, le pourcentage de réussite est en moyenne de 50% environ. Ce pourcentage

est sensiblement le même, qu'il s'agisse de garçons ou de filles, d'élèves d'écoles officielles ou d'écoles privées.¹

Toutefois, nos analyses ont fait apparaître des différences significatives dans les pourcentages de réussite des différentes régions du pays.

Le *baccalauréat* libanais centralisé à Beyrouth groupe en deuxième partie 700 à 800 candidats, masculins dans une forte proportion (90%). Du point de vue de leur provenance scolaire, nous avons relevé, depuis 1956, environ 10% d'élèves d'établissements officiels, 30% d'élèves du « privé national », 30% du « privé étranger » et 30% de candidats libres, plus âgés, qui se préparent seuls ou à l'aide de cours du soir.²

Le baccalauréat se compose de deux parties : la première partie comprend les séries « littéraire », « scientifique », « langues anciennes » ; la seconde partie, accessible après obtention de la première, propose aux candidats les séries « philosophie » et « mathématiques ».

C'est cette dernière partie que nous présenterons plus en détail puisqu'elle a fait l'objet de l'analyse dont nous rendons compte aux chapitres suivants.

Chacune des deux séries comporte des épreuves écrites et des épreuves orales. Les épreuves écrites sont les suivantes :

	Durée	Coefficient
<i>Série « Philosophie »</i>		
Dissertation(en arabe) sur un sujet de philosophie	4h.	3
<i>Physique-chimie-sciences naturelles-hygiène</i>	3h.	2
Dissertation (en français ou en anglais) sur un sujet de philosophie	4h.	3
<i>Série « Mathématiques »</i>		
Dissertation (en arabe) sur un sujet de philosophie	3h.	2
Mathématiques	4h.	3
Physique-chimie	4h.	3

1. De cette égalité de pourcentages de réussite on ne doit pas conclure à une égalité de niveau de ces deux catégories de candidats car, si l'échantillon des candidats des écoles officielles est représentatif, celui des écoles privées ne l'est pas : certains élèves des écoles privées s'abstiennent en effet de poser leur candidature au C.E.P., le diplôme n'étant pas exigé, dans leur établissement, pour effectuer des études secondaires.
2. L'article 20 du chapitre II du Décret 7004 (1er octobre 1946), qui énonce les pièces à produire au moment de l'inscription, exempte le candidat de plus de 21 ans de la présentation du livret scolaire.

Les épreuves orales, parmi lesquelles on retrouve les matières de l'écrit, sont au nombre de 7 pour la série « Philosophie » et de 6 pour la série « Mathématiques », mais le total des coefficients est identique dans les deux séries.

Le candidat doit obtenir au moins 80 points ¹ à l'écrit pour être admissible à l'oral, et au moins

190 au total des points de l'écrit et de l'oral pour être définitivement reçu. Une deuxième session est prévue chaque année en octobre pour les ajournés et empêchés de la session de juillet.

1. En fait, la limite de « repêchage » se situe, suivant les années, à 78, 76, 74 ou même 72 en 1958.

VALEUR DIAGNOSTIQUE DE L'EXAMEN

L'un des buts de l'examen scolaire, le principal toujours, l'unique souvent, est de permettre un bilan des connaissances acquises. Ce but est-il atteint? Comment est-il atteint? Les données que nous avons pu recueillir et traiter, à partir des notes obtenues au baccalauréat libanais (deuxième partie) vont nous fournir des éléments de réponses. Nous envisagerons successivement l'homogénéité de l'examen, c'est-à-dire l'accord entre les bilans établis pour l'écrit et pour l'oral, et la fidélité de l'examen, c'est-à-dire la concordance entre les résultats dans la partie écrite des deux sessions.

HOMOGENEITE DE L'EXAMEN

Liaisons écrit-oral et surestimations à l'oral

Si l'ensemble de l'examen est homogène, c'est-à-dire si l'écrit et l'oral classent de la même façon les candidats, on pourra conclure que ces deux parties de l'examen ont une valeur diagnostique relative comparable. En est-il ainsi? Les coefficients de corrélation calculés entre les notes globales, à l'écrit et à l'oral, de chaque candidat ayant subi ces deux parties de l'examen sont significatifs mais faibles (0,40).¹ Cela indique une parenté, certes, mais une parenté assez lâche entre le diagnostic par les épreuves écrites et le diagnostic par les épreuves orales.

L'homogénéité de l'examen est donc très relative. Il est vrai que le contenu des parties écrite et orale diffère sensiblement puisque trois matières seulement leur sont communes. Toutefois, on constate que, dans une même matière, les corrélations entre l'écrit et l'oral sont encore plus faibles (0,14 en philosophie arabe, 0,20 en philosophie française, et 0,27 en mathématiques et en sciences où elles sont tout de même statistiquement significatives).

Le diagnostic par l'écrit diffère donc sensiblement du diagnostic par l'oral, mais il diffère moins pour les matières scientifiques que pour les épreuves de philosophie.

Si nous considérons maintenant la répartition des points obtenus à l'écrit et à l'oral dans les mêmes matières, nous nous apercevons que les épreuves de mathématiques et de sciences situent de pareille façon la valeur de l'ensemble des candidats, alors qu'en philosophie arabe ou étrangère, les étudiants sont nettement surestimés par l'oral. (figure 1).

La stabilité du diagnostic en mathématiques et en sciences ne peut être due - cette hypothèse sera confirmée par la suite - qu'à la plus grande objectivité de la notation qui s'effectue dans ces deux épreuves, à l'aide d'un barème, ce qui diminue notablement la marge d'erreur sur la note. Peut-être aussi - mais ce serait à vérifier - les examinateurs «scientifiques» ont-ils un sens métrologique plus aigu, une meilleure aptitude à noter dans les épreuves orales de leur spécialité.

Pour ou contre l'oral

Nous avons parlé de surestimation à l'oral plutôt que de sous-estimation à l'écrit, parce que nous accordons plus de crédit à la notation de l'épreuve écrite (à l'aide d'un barème ou par double correction) qu'au jugement de l'examineur à l'oral. D'ailleurs, des études sur la notation d'épreuves orales, rares il est vrai, semblent indiquer des désaccords statistiquement plus importants pour l'oral que pour l'écrit. Trimble² trouve par exemple une corrélation moyenne de 0,32 entre les notes attribuées par trois examinateurs à chacun des vingt-cinq étudiants interrogés oralement, alors que les corrélations trouvées dans des analyses de notation multiple d'épreuves écrites se situent autour de 0,60, valeur que confirme la présente étude.

Nos recommandations iront donc dans le sens de la réduction ou même de la suppression de l'oral. Nos arguments sont d'ordre technique mais il en est d'un ordre tout différent que nous citerons simplement pour mémoire : l'oral à la chaîne est une formalité, un luxe inefficace puisqu'il n'élimine finalement que 2 à 8% des candidats. Il est injuste puisque tous les candidats ne traitent pas les mêmes questions. Enfin sa notation, souvent subjective, peut faire intervenir, consciemment ou non, des sympathies ou antipathies d'ordre confessionnel, racial voire sexuel.

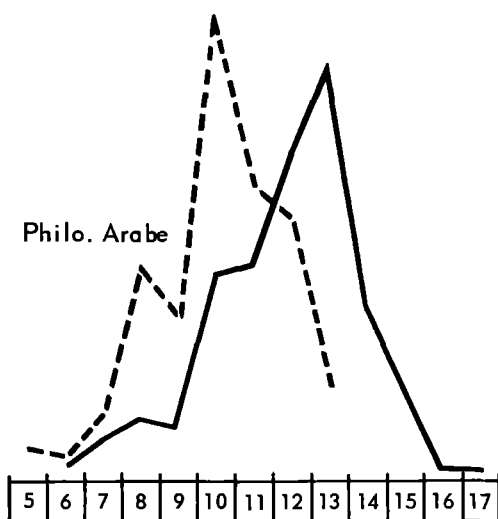
Les défenseurs de l'oral souligneront cependant son double intérêt dans le contrôle complémentaire de la valeur des candidats, et la rectification des erreurs de l'écrit par élimination de ceux qui ont

1. Le coefficient de corrélation r serait égal à 1,00 dans le cas où la réussite à l'oral serait exactement identique à la réussite à l'écrit.
2. (O.C.) Trimble « The oral examination : its validity and reliability », In: *School and Society* (New York), vol. 39, n° 1009, 1934, pp. 550-552.

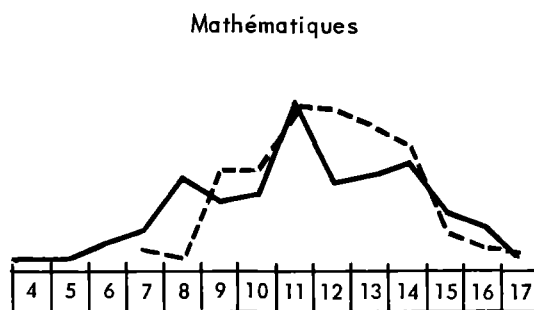
Figure 1

COMPARAISON DES NOTES ECRITES ET ORALES
PAR MATIERE (sur 20)

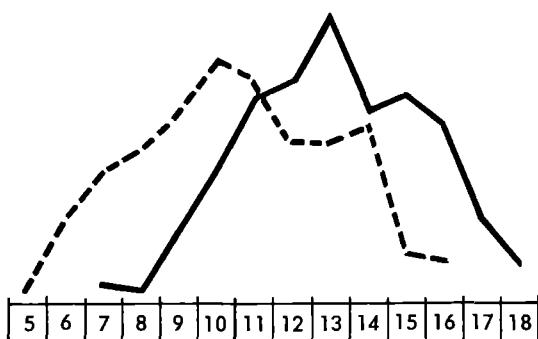
Série Philosophie



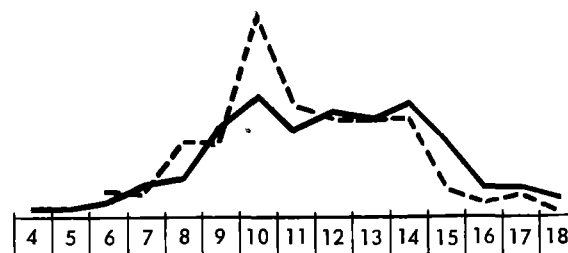
Série Mathématiques



Philo. Française



Sciences



----- Ecrit
———— Oral

bénéficié d'une indulgence excessive. Nous avons pensé qu'une session de contrôle par écrit, remplirait le même office, mais seulement sur certaine catégorie limitée de candidats dont le diagnostic des capacités demande à être précisé en vue de l'admission ou de l'ajournement.

FIDELITE DE L'EXAMEN

Liaisons entre les écrits des deux sessions

Puisque nos préférences techniques sont orientées vers l'écrit, assurons-nous qu'il est fidèle. Demandons-nous si, tout comme la balance indique le même poids au cours de deux pesées successives d'un même objet, l'examen écrit donne la même estimation de la valeur du candidat à quelque temps d'intervalle.

Nous avons pu effectuer ce contrôle de la manière suivante :

On sait que la quasi totalité des refusés de la première session subissent un deuxième écrit, celui de la deuxième session. Il nous a donc été possible de calculer la liaison statistique existant, pour ces mêmes candidats ajournés en juillet, entre les performances de leur premier et de leur deuxième écrit.

Le fait que cette année-là (1958), par suite d'événements insurrectionnels, les deux sessions avaient été rapprochées,¹ supprimait totalement un facteur qui risquait de perturber notre étude de fidélité, à savoir la somme de connaissances nouvelles acquises pendant les vacances séparant habituellement les deux sessions.

Les résultats obtenus peuvent être interprétés de la façon suivante :

1. La liaison entre les diagnostics concernant l'écrit de la première session et celui de la deuxième est d'autant plus étroite que la notation des épreuves est plus objective (La corrélation est de 0,12 pour l'épreuve de philosophie arabe, et de 0,52 pour celle de mathématiques).

2. Même lorsque la notation est satisfaisante, la liaison n'est pas suffisamment élevée pour qu'on puisse considérer que la deuxième session fasse double emploi dans l'appréciation de la valeur relative des candidats. (La corrélation pour l'ensemble écrit de la série « mathématiques »² est de 0,56, elle n'est que de 0,29 pour la série « philosophie »).

Surestimation à la deuxième session

Il s'agit bien de « valeur relative », car nous sommes en droit de penser qu'en valeur « absolue » (si l'on nous permet cette expression) les candidats de la deuxième session sont les moins bons de la population des candidats de la première session, puisqu'ils constituent le groupe des refusés de cette première session. Or, ce fait n'apparaît pas quand on compare la distribution des notes de l'écrit de la première session et des notes de l'écrit de la

deuxième session : leur superposition ne révèle guère de différence qualitative entre les deux échantillons, surtout dans la série « philosophie » (voir figure 2).

Les épreuves de la deuxième session étaient-elles plus faciles ? Nous ne le croyons pas, car la fréquence des notes éliminatoires n'est pas moins élevée à cette session qu'à la première.

Vaut-il mieux penser que la surestimation des notes de la deuxième session amène le niveau moyen de ces « ajournés » au niveau de l'ensemble des candidats de la première ?

C'est effectivement ce que révèle la comparaison des moyennes des *mêmes candidats* (les ajournés de la première session) dans leurs épreuves écrites des deux sessions.

La plus forte différence de moyennes enregistrée est celle de l'épreuve de philosophie française (6,7 points en faveur de la deuxième session). En mathématiques et en sciences cette différence est moins forte (4,07 et 4,53).

Comparaison des seuils d'admission (première et deuxième sessions) :

Nous allons aborder cette même question de façon différente en considérant, non plus la moyenne des notes de ces candidats mais la note limite inférieure qu'il aurait fallu atteindre à la première session (la notation et la difficulté des épreuves des deux sessions étant supposées semblables) pour recevoir un lot de diplômés égal à celui de deux sessions réunies. Dans ces conditions hypothétiques, nos calculs montrent qu'il aurait fallu descendre jusqu'à 59,5 en « philosophie », jusqu'à 64,5 en « mathématiques ». Ces totaux, correspondant à des moyennes de 7,5 et 8 sur 20, soulignent assez la médiocrité des diplômés d'octobre. Ils indiquent aussi que les correcteurs ne jugent que par rapport à la qualité moyenne générale de l'échantillon de copies qui leur est fourni, et que lorsque cet échantillon est de médiocre qualité les notes attribuées se dispersent tout de même autour de 10 sur 20.

Pour ou contre la deuxième session ?

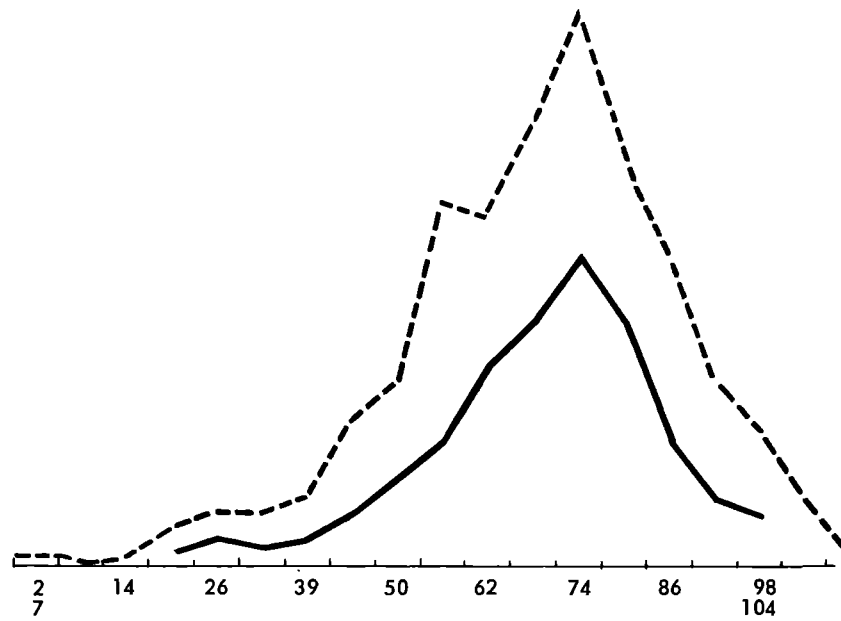
La deuxième session a été instituée principalement pour limiter les aléas d'une session unique ; on offre ainsi au candidat malchanceux ou trop sévèrement jugé en juillet une nouvelle occasion de « tenter sa chance », un peu comme, en athlétisme, on permet plusieurs essais au spécialiste du saut en hauteur. Cependant l'examen n'est pas, il ne devrait pas être, une compétition, mais un instrument de contrôle et c'est à notre avis, une erreur de considérer la seconde session comme un nouvel essai.

1. Quinze jours seulement ont séparé, à la fin du mois d'octobre, la deuxième session de la première.
2. Rappelons qu'en série « mathématiques », les épreuves de mathématiques et de sciences qui représentent 120 points sur les 160 de l'écrit, sont notées suivant un barème.

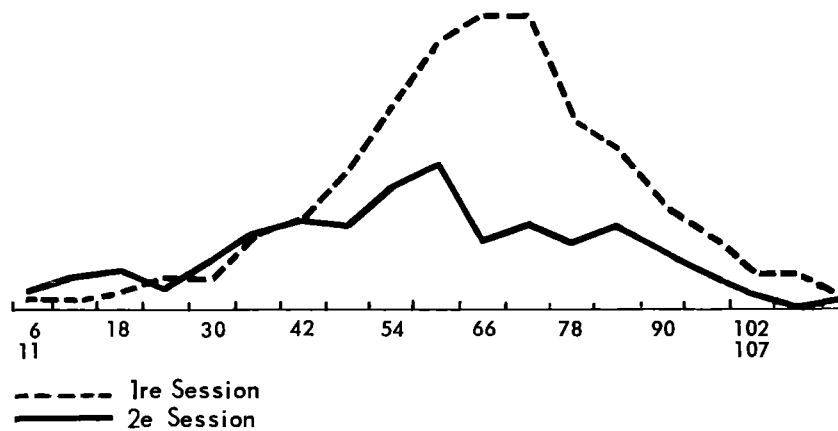
Figure 2

COMPARAISON DES NOTES DE L'ECRIT
DE LA 1^{re} ET DE LA 2^e SESSION

Série Philosophie



Série Mathématiques



C'est cette même conception de « l'examen-tir au nougat », et de sa sanction par « tout ou rien » qui fait envisager indépendamment les résultats de chaque essai. Ne devrait-on pas plutôt, de la même façon que le chimiste tient compte de la moyenne des résultats de ses dosages, « asseoir » le diagnostic sur la somme ou la moyenne des deux sessions et tirer ainsi parti du bénéfice de la double mesure ?

Selon les règles actuelles de la deuxième session, on admet un candidat qui obtient un total de points correspondant à une moyenne de 10,5 sur 20 et cela même si sa moyenne de juillet était de 6 sur 20. Selon la formule que nous envisageons, ce candidat ne serait admis que s'il avait obtenu au moins 9,5 sur 20 de moyenne à la première session.

La session de contrôle dont nous avons déjà parlé à propos de l'oral et qui jouerait également le rôle de la deuxième session traditionnelle mais sur les candidats moyens et non inférieurs, serait donc complémentaire en ce sens que les notes qu'on y obtiendrait s'ajouteraient à celles de la précédente.

Notons que, ce lot de candidats étant de qualité moyenne par rapport à l'ensemble des candidats de juillet, les notes qui seraient attribuées lors de cette session complémentaire ne seraient plus entachées

de l'erreur de surestimation qui caractérise, comme nous l'avons souligné, la notation des copies médiocres de l'actuelle deuxième session.

Ainsi se trouveraient réduites, dans des proportions non négligeables, à la fois les sources d'erreur d'échantillonnage sur la valeur réelle du candidat, et les erreurs d'estimation de cette valeur.

Enfin, l'un des principaux avantages de notre analyse est d'avoir révélé la surestimation des candidats de seconde session, *même lorsque cette session vient immédiatement après la première*, comme c'est le cas dans notre étude. Cet argument renforce notre opinion suivant laquelle, si les notes d'octobre sont supérieures à celles de juillet, cette augmentation est due beaucoup plus à une survalorisation subjective des performances d'octobre qu'à une improbable augmentation des capacités des candidats pendant les vacances de l'intersession.

Pour cette raison, si nous retenons la nécessité d'une session complémentaire destinée aussi bien à récupérer les candidats passables malchanceux qu'à éliminer les médiocres favorisés par la chance, il ne nous paraît pas obligatoire de la placer en octobre.

VALEUR PRONOSTIQUE DE L'EXAMEN

DU PRIMAIRE AU SECONDAIRE :

En octobre 1957, le problème qui nous a été posé au Liban était celui de l'orientation vers l'enseignement secondaire officiel, des élèves titulaires du certificat d'études primaires.

Faute d'un effectif de spécialistes suffisant, l'orientation ne peut malheureusement pas être pratiquée sur les 18.000 candidats au C.E.P. ni même sur les 9.000 titulaires annuels de ce diplôme. La solution de dépannage adoptée jusqu'à présent était la sélection des candidats par concours d'entrée de type traditionnel.

On peut tout de suite se demander si un second examen est bien nécessaire, deux mois après celui du certificat, et si le C.E.P. ne répond pas lui-même aux deux principaux objectifs de l'examen-concours d'entrée à savoir :

Procéder à un inventaire des connaissances de base indispensables à la poursuite des études secondaires.

Fournir des indications sur la capacité de poursuivre ces études secondaires.

En ce qui concerne le premier de ces objectifs, on peut penser que, conçu, réalisé et corrigé de la même façon que la partie écrite du baccalauréat que nous venons d'analyser, le C.E.P. a des qualités et des défauts diagnostiques sensiblement analogues.

C'est plus particulièrement le second objectif qui nous intéresse maintenant : Quel est le pouvoir prédictif du C.E.P. ? ¹

En termes statistiques, la question peut s'exprimer de cette façon : existe-t-il une liaison entre les performances au C.E.P. et les performances de ces enfants au cours de leur première année dans l'enseignement secondaire ?

Les événements qui ont troublé le pays à cette époque, nous ont empêché de vérifier cette hypothèse sur une aussi grande échelle que nous l'aurions voulu. Nous avons pu néanmoins travailler sur deux sections ² d'une école de filles de Beyrouth, dont nous avons d'abord relevé les notes dans les principales épreuves du certificat.

CHOIX DES CRITERES DE REUSSITE SCOLAIRE :

A ces notes du certificat, il fallait ensuite faire correspondre, pour chaque élève, les notes obtenues

au cours de la première année dans l'enseignement secondaire.

Mais quelles notes ? Les moyennes ou les compositions ? Celles des quatre premiers mois (période improprement appelée premier trimestre) ou celles des quatre derniers mois (deuxième trimestre) ?

Nous avons décidé, de propos délibéré, de ne tenir compte que des notes du second « trimestre », le premier correspondant à la période instable d'installation, de mise en route, d'adaptation.

Nous avons tenu cependant à nous assurer auparavant de la parenté et de la stabilité des notes de l'année en calculant les corrélations existant, d'une part, entre les moyennes et compositions de chaque trimestre et, d'autre part, entre les moyennes (ou les compositions) des deux trimestres.

En fait, il y a bien homogénéité des notations (moyennes et compositions) dans chaque matière : la corrélation moyenne est de 0,88 en arabe, 0,93 en français et 0,53 pour les mathématiques.

Il est vraisemblable que l'objectivité relative des notations en mathématiques annihile l'effet de halo qui doit influencer probablement la notation en arabe et en français. Ce halo, ³ par lequel le professeur projette son jugement global sur l'élève dans la notation, a certainement pour effet d'homogénéiser les notes de chaque élève et, dans une certaine mesure, de les stabiliser dans le temps.

Car, excepté en mathématiques, les moyennes et compositions de l'année sont effectivement stables (0,94 en arabe, 0,78 en français, 0,56 en mathématiques en moyenne).

Ces résultats nous ont conduit à retenir plusieurs critères différents de validation :

Pour l'arabe, la somme « moyennes et compositions » du deuxième trimestre.

Pour le français, même critère.

Pour les mathématiques, les moyennes du deuxième trimestre.

Le total annuel des notes a été pris en considération

1. On trouvera plus de détails techniques sur cette question dans : E. Valin. « Orientation scolaire au Liban », in B.I.N.O.P. (mai-juin) n° 3, 1959, Paris, pp. 175-181.
2. Soit au total 53 élèves. Ces deux sections étant sous la responsabilité des mêmes professeurs, nous les avons fusionnées dans l'étude.
3. R. Bonnardel, « Application de la méthode d'analyse factorielle de Thurstone à l'étude de la notation des copies d'examen », in : *Travail humain* (Paris), 1946, pp. 150-167.

pour la seule validation de l'ensemble des épreuves du certificat d'études primaires.

RESULTATS ET CONCLUSIONS :

Voici les corrélations obtenues dans chaque matière entre les critères indiquées ci-dessus et les notes correspondantes du C.E.P.

Tableau 1 : Valeur pronostique du C.E.P.

Arabe	0,63
Français	0,40
Calcul	0,03
Ensemble (A + F + C)	0,35
Total C.E.P.	0,35

Sans attacher d'importance excessive à la valeur absolue de ces corrélations, en raison de l'erreur d'échantillonnage, il est permis de dégager de leur valeur relative les interprétations suivantes :

1. On ne peut prédire avec assurance la réussite en première année du secondaire à partir des résultats au certificat d'études.

2. Les trois matières principales (arabe, calcul, français) ont le même pouvoir prédictif que la note générale. Cela revient à dire que les matières secondaires de l'examen du C.E.P. (histoire et géographie, leçons de choses, dessin ou couture) n'apportent aucune contribution au pronostic.

3. Le meilleur pronostic est celui de la réussite dans l'apprentissage de la connaissance et le maniement de la langue maternelle. Par contre, aucune prédiction n'est permise en ce qui concerne le calcul. Il est vrai que, dans ce domaine, des différences importantes existent entre les programmes du primaire et les programmes du secondaire. De plus, et cela mérite d'être signalé, l'enseignement des mathématiques (et des sciences) est donné en français et non plus en arabe comme dans la classe de préparation au certificat.

En résumé, le certificat d'études primaires, tel qu'il est pratiqué, ne paraît pas posséder la valeur diagnostique et pronostique suffisante pour être utilisé dans l'orientation ou la sélection au seuil des études secondaires.

VALEUR METROLOGIQUE DE LA NOTATION

LE PROBLEME CENTRAL : LA NOTATION

Nous avons pu nous rendre compte, au cours des chapitres précédents, que l'obstacle essentiel à notre analyse tenait à l'évaluation quantitative de la valeur des candidats, c'est-à-dire la notation.

Quel est donc le contenu de la note attribuée ?

Que chiffre cette note et comment le chiffre-t-elle ?

1. Les notes sont des estimations de la performance dans une série d'épreuves, mais cette série d'épreuves est-elle vraiment représentative de la valeur réelle du candidat, est-elle vraiment adaptée au niveau des candidats ? C'est là un problème d'*échantillonnage*, de choix et de nombre d'épreuves, de choix et de nombre des questions de l'épreuve.

Certes la question est plus importante pour les candidats moyens que pour les sujets nettement supérieurs ou nettement inférieurs qui ont de fortes probabilités de manifester leur supériorité ou leur infériorité sur l'échantillon limité d'épreuves qui leur est proposé. Mais les candidats moyens étant précisément les plus nombreux, il importe de chercher à limiter l'erreur d'échantillonnage des épreuves, soit par l'examen du passé scolaire du candidat, soit par l'augmentation du nombre et du contenu des épreuves, soit encore par la session complémentaire que nous avons envisagée.

2. La notation, c'est-à-dire la *quantification* des performances est-elle le produit du jugement humain, syncrétique et subjectif, ou le résultat d'une mesure analytique objective, ou encore les deux à la fois ?

Ce sont ces facteurs que nous allons essayer de circonscrire par la méthode statistique, d'une part, en examinant sur les notes finales et sur la double notation de certaines épreuves écrites du Baccalauréat, leurs caractéristiques métrologiques et, d'autre part, en analysant qualitativement les différents facteurs importants du complexe correctionnotation.

CARACTERISTIQUES METROLOGIQUES DE LA NOTATION

Ces caractéristiques seront d'abord étudiées sur les notes finales des épreuves écrites, c'est-à-dire, sur les notes ayant effectivement contribué, par sommation, à la réussite ou à l'échec du candidat. Ces

notes résultent, soit de l'application d'un barème (en mathématiques et en sciences), soit d'un compromis décidé par un jury restreint sur les résultats d'une double correction indépendante (en philosophie arabe et en philosophie étrangère)¹.

Fréquence des notes attribuées

Lorsque, les notes des trois épreuves écrites étant ramenées au coefficient 1, nous procédons à leur pointage sans nous occuper ni de leur origine ni de leur destination, nous obtenons les courbes de la figure 3.

Ces courbes de fréquence sont dentelées pour la série « Philosophie », plus curvilignes pour la série « Mathématiques ». Pour cette dernière, on remarquera seulement que les notes les plus fréquemment attribuées sont le 10 et le 11 chez les admis, et le 7 et le 8 chez les refusés. Ces deux sommets ou modes se retrouvent dans la courbe de l'ensemble.

Les courbes de la « Philosophie » sont plus intéressantes : elles présentent notamment des pointes à 0, 5, 8, 10 et des dépressions à 1, 4, 6, 9 et 15. Mais nous allons voir que, selon la qualité de la copie, il s'agit soit d'une attraction pour certaines notes-jalons, soit d'une hésitation à attribuer certaines autres notes.

En effet, les courbes de fréquences établies séparément pour les admis et pour les refusés indiquent que :

Sur une copie moyenne, on note plus volontiers 10 que 9, la note 10 étant considérée comme la note-repère de la copie moyenne.

Sur une bonne copie, au contraire, on note moins volontiers 15 que 14 : c'est que la note 15 marque sans doute, dans l'esprit du correcteur, le passage dans la catégorie « Très Bien ».

Sur une copie médiocre, on met plus souvent 7 ou 8.

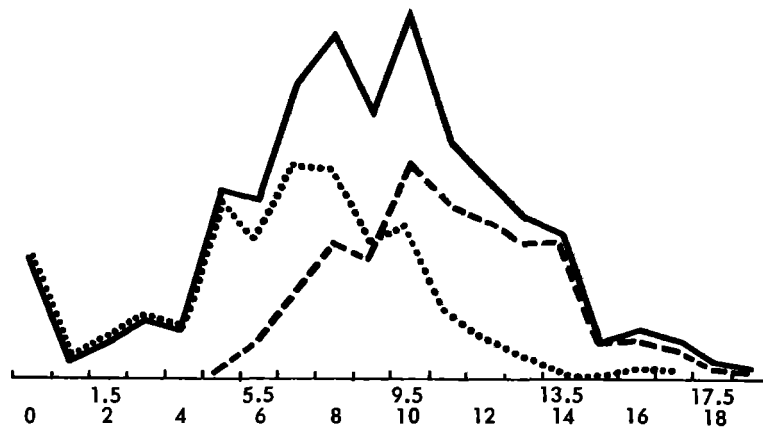
Tout se passe comme si les correcteurs effectuaient une cascade d'estimations à trois échelons :

1. inférieur à 10, égal à 10, supérieur à 10.
2. puis au dessous de 10 : inférieur à 5, égal à 5, supérieur à 5 (dans ce cas, ils notent 7 ou 8).
3. au dessus de 10 : inférieur à 15, égal à 15 (dans ce cas ils notent souvent 14) ou supérieur à 15.

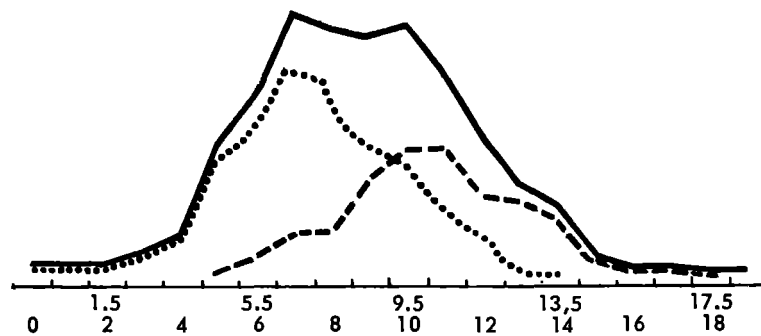
1. Pour l'année sur laquelle a porté notre enquête ce compromis n'a jamais abouti à la moyenne arithmétique des deux notes mais généralement à l'adoption de la plus favorable des deux.

FREQUENCES DES NOTES SUR 20 DES TROIS EPREUVES ECRITES

Série Philosophie



Série Mathématiques



----- Admis
 Refusés
 ——— Ensemble

La grande fréquence des 5 ne s'explique que par l'élimination inconditionnelle qu'entraînent les notes inférieures à cette note.

Signalons enfin qu'en raison du fait qu'elle s'effectue à l'aide d'un barème, la notation des épreuves de mathématiques et de sciences est nettement plus fine (on y relève environ 40 à 50% de notes non entières).

Elle différencie donc mieux les performances. Au contraire, l'estimation globale des épreuves de philosophie (3 ou 4% de notes non entières) ne permet pas une aussi bonne différenciation.

Sélectivité de la notation

Quel est le pouvoir de discrimination entre les élèves des notes de chaque épreuve ? L'écart-type ou « standard deviation » est un indice de dispersion des notes intéressant à considérer. Il est d'autant plus petit que la dispersion des notes autour de la moyenne est faible.

Tableau 2. Ecart-types des notes sur 20

	Philo arabe	Philo française	Sciences
Série « Philo-sophie » (N=446)	1,91	2,71	4,18
Série « Mathématiques » (N=280)	2,08	3,50	3,32

Des valeurs d'écart-type obtenues par calcul et consignées dans le tableau 2 il ressort, entre autres, que, dans la série « Philosophie », il faut deux fois moins de points en philosophie arabe qu'en sciences pour gagner une place au classement. En série « Mathématiques », il faut 1,5 fois moins de points en philosophie qu'en sciences pour gagner une place. Ainsi s'explique, au moins partiellement, la faveur dont jouit l'épreuve de philosophie parmi les candidats.

Quel est maintenant le pouvoir de discrimination des notes, non plus d'un candidat à l'autre, mais entre le groupe des admis et le groupe des refusés ?

Nous avons posé le problème théorique de la façon suivante :

Dans chaque épreuve la courbe de distribution des notes des admis est en décalage plus ou moins accusé par rapport à celle des refusés (voir figure 4). L'intersection des deux courbes correspond en abscisse à la note X à partir de laquelle on a plus de chances d'être admis que refusé à l'écrit. Nous avons recherché ces notes X et calculé les probabilités de réussite pour les candidats qui obtiennent, dans l'épreuve considérée, une note inférieure à X.

Ce sont encore les épreuves de mathématiques et de sciences qui départagent le mieux les admis et les refusés. Par exemple, le candidat qui obtient moins de 29 sur 60 dans l'épreuve de mathématiques n'a que 14% de chances d'être admis alors que celui qui obtient moins de 16,5 sur 40 en philosophie arabe en a 29%. (Voir figure 5).

L'analyse de l'ensemble de nos résultats nous permet de dire en définitive que c'est principalement à cause d'une mauvaise note en sciences qu'on échoue dans la série « Philosophie », et que c'est surtout à cause de mauvaises notes en mathématiques et en sciences qu'on échoue dans la série « Mathématiques ».

Pondération théorique et pondération réelle

On sait que des coefficients prévus dans les textes législatifs qui réglementent le baccalauréat ont pour but de répartir des poids différents entre les trois épreuves écrites. Est-ce que ces coefficients jouent effectivement le rôle auquel ils sont destinés ?

Nous venons de voir (tableau 2) qu'il était plus difficile de gagner une place en sciences qu'en philosophie arabe, c'est donc qu'en fait les notes de sciences ont plus d'importance, plus de poids, que les notes de philosophie arabe.

Par une méthode appropriée¹ nous avons pu trouver le poids réel de chaque épreuve dans l'ensemble des épreuves écrites.

Tableau 3. Pondération effective des épreuves

	Coef. théorique	Coef. réel
Série « Philosophie »		
Philo arabe	3	2,06
Philo française	3	2,93
Sciences	2	3,01
	8	8,00
Série « Mathématiques »		
Philo arabe	2	1,35
Mathématiques	3	3,41
Sciences	3	3,24
	8	8,00

Ainsi, dans la série « Philosophie », on arrive à ce paradoxe que l'épreuve de sciences, dotée officiellement du plus faible coefficient est celle qui, en fait, pèse le plus dans le total de l'écrit et détermine le plus souvent la réussite ou l'échec.

1. Le problème a été traité de façon mathématique par DELAPORTE « Remarques sur la pondération des notes scolaires », dans *Biotypologie*, Paris, vol. 6, 1938, pp. 271-273.

Figure 4

DISCRIMINATION ENTRE ADMIS ET REFUSES

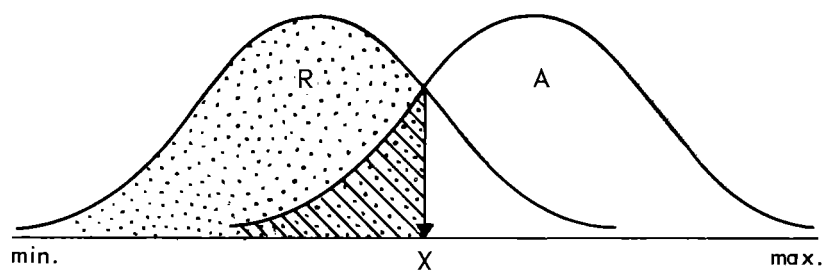
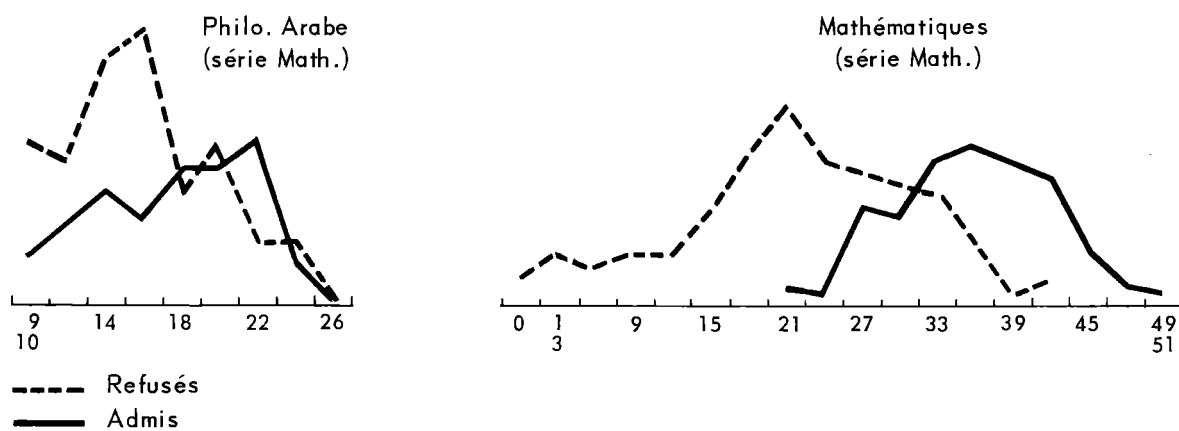


Figure 5

EXEMPLES DE MAUVAISE ET DE BONNE SELECTIVITE (ADMIS-REFUSES)



Quittons un moment l'exploration statistique pour nous livrer à une interprétation des résultats auxquels elle nous a conduit.

Tous ces résultats sans exception établissent un clivage entre, d'une part, les notes de mathématiques et de sciences, d'une valeur diagnostique et pronostique supérieure parce que leur notation est plus fine, plus sensible, plus sélective, plus fidèle, et, d'autre part, les notes de philosophie dont les propriétés métrologiques sont incontestablement moins bonnes.

Il est clair que, dans le domaine mathématique et scientifique où les critères de correction et de notation peuvent être facilement établis, l'accord se réalise sans trop de difficulté sur un barème commun.

De plus, ce domaine se prêtant plus que d'autres à une notation analytique, le fractionnement de la note globale en un certain nombre de « postes de notation » réduit d'autant la place laissée au jugement, responsable de l'erreur de mesure.

Ces différents facteurs de précision dans la mesure n'existent pas dans les matières littéraires et philosophiques, où les critères sont souvent discutés et toujours discutables, où le jugement ne semble pouvoir être que global, syncrétique, et où l'appréciation personnelle de la performance sur l'échelle de 0 à 20 laisse place à des erreurs d'estimation plus larges.

C'est ainsi qu'on peut décrire deux pôles du complexe correction-notation, se repoussant l'un l'autre comme des pôles magnétiques de même signe :

1. la correction-notation analytique et objective dans laquelle on mesure plus qu'on juge ;

2. la correction-notation globale et subjective qui implique plus de jugement que de mesure.

La première donne, nous l'avons vu à propos des mathématiques et des sciences, d'excellents résultats. C'est vers elle que toute notation doit tendre.

La seconde est intéressante à considérer pour ses défauts que nous allons maintenant examiner en détail.

LA DOUBLE CORRECTION

C'est précisément parce qu'on ne s'accorde pas toujours très bien sur ce qui est « bon » et ce qui l'est moins, que les épreuves de philosophie arabe et française sont soumises à une double correction.

Les notes attribuées par les deux correcteurs vont être analysées de trois points de vue différents : Etude des variations entre les moyennes des différents correcteurs¹.

Examen de la plus ou moins grande similitude de classement des mêmes copies par deux correcteurs différents.

Analyse des variations intercorrecteurs des notes accordées aux mêmes copies.

L'équation personnelle

La technique de l'analyse de la variance² a été utilisée pour tester séparément, en ce qui concerne la philosophie arabe et la philosophie française, l'hypothèse selon laquelle les moyennes des notes attribuées par les correcteurs ne diffèrent pas significativement les unes des autres. Si l'on préfère, l'analyse de la variance permet de répondre à cette question : est-ce que les 9 correcteurs notent statistiquement autour de moyennes voisines ?

La réponse est négative aussi bien pour l'épreuve de philosophie arabe que pour celle de philosophie française et il n'est pas possible de considérer que les 9 correcteurs ont statistiquement noté autour de la même moyenne³.

L'accord sur le classement

Est-ce que les deux correcteurs à qui l'on confie séparément le même paquet de copies, rangent ces copies de la même façon selon leur mérite ? Le calcul des coefficients de corrélation nous permet de nous en rendre compte.

Tableau 4. Corrélations entre correcteurs

Philosophie arabe (72 copies par paquet)		Philosophie étrangère (50 copies par paquet)	
entre correcteurs ⁴ <i>r</i>		entre correcteurs <i>r</i>	
A et B	0,31	A' et B'	0,83
B et C	0,75	A' et B'	0,61
C et D	0,39	A' et B'	0,78
D et B	0,82	D' et E'	0,69
E et F	0,51	E' et F'	0,72
F et A	0,50	F' et G'	0,88
G et D	0,57	G' et H'	0,76
E et H	0,71	H' et I' (N=47)	0,77
H et I	0,47	I' et A' (N=22)	0,82
I et G	0,69		

Les corrélations moyennes sont plus fortes sur les copies de philosophie étrangère ($r=0,76$ en moyenne) que sur celles de philosophie arabe ($r=0,57$ en moyenne) où l'on observe des cas de divergences

1. On peut considérer que la répartition des copies entre les correcteurs est effectuée au hasard.

2. Voir notamment Q. Mc NEMAR - *Psychological Statistics*, 2nd ed., New York, Wiley, 1955, pp. 235-266.

3. La fraction de variance totale explicable par les variations intercorrecteurs est de 4% en philosophie arabe et de 5% en philosophie française.

4. On remarquera que les correcteurs B et D ont corrigé 3 paquets de 72 copies alors que les autres n'en ont corrigé que 2.

troublantes, entre A et B, d'une part, C et D, d'autre part. Par contre, d'après la moyenne de leurs corrélations, on peut dire que A' et G' sont les meilleurs juges.

Cependant on peut être bon juge de l'appréciation relative de la valeur des copies et être trop sévère ou trop indulgent lorsqu'il s'agit de quantifier cette appréciation par une note sur 20.

Nous allons nous en apercevoir dans ce qui suit.

L'accord sur la note

Demandons-nous en premier lieu si, sur les mêmes copies, la différence entre la note de l'un des correcteurs et la note de l'autre est nulle. Si elle est nulle pour chaque copie, elle doit être également nulle pour toutes les copies. C'est ce que nous avons testé, par une technique appropriée¹.

D'après les résultats, on peut considérer qu'en philosophie arabe, 4 correcteurs sur 9 sont en désaccord avec leurs autres collègues sur la façon de noter, ce sont :

Les correcteurs C et E dont les notes sont généralement supérieures à celles des deux autres cocorrecteurs.

Le correcteur F, par contre, note au-dessous de ses collègues A et E.

Enfin H donne généralement des notes plus faibles que E, mais plus fortes que I.

En philosophie étrangère, deux correcteurs se distinguent : B' semble relativement plus indulgent et H' plus sévère.

Si l'on analyse de plus près les accords et les désaccords sur la note les résultats ne sont guère encourageants.

Tableau 5. Pourcentages d'accords et de désaccords

Importance du désaccord	Philosophie arabe	Philosophie étrangère
0 pt.	20%	15%
1 pt.	30%	35%
2	21,5%	19%
3	14%	17%
4	9%	8,5%
5	3,5%	3%
6	1%	1%
7	1%	1%
8 points	-	0,5%

L'accord se réalise donc une fois sur 5 ou 6 copies, et l'on trouve jusqu'à 7 et même 8 points d'écart sur quelques copies, heureusement rares ! On frémit quand on pense que la réussite ou l'échec est décidé à partir de telles notes, triplées chacune, et qui représentent, pour les deux épreuves de philosophie réunies, les 5/8 de l'examen écrit².

VALEUR DE LA DECISION

Nous allons clore ce chapitre de métrologie par l'étude de la limite entre le groupe de refusés et le groupe d'admis à l'oral, limite que nous appelons « seuil de décision ».

Lorsqu'il s'agit d'un concours, d'un examen de sélection à la suite duquel on décide d'accepter (ou d'éliminer) environ 27% des candidats, la délimitation des réussites supérieures (ou inférieures) est relativement aisée et n'est pas théoriquement entachée d'une erreur considérable.

Mais généralement dans les examens, la coupure se situe à la moyenne générale théorique de 10 sur 20 qui, d'ailleurs, peut se trouver plus ou moins en-deçà ou au-delà de la moyenne arithmétique générale de l'ensemble des résultats. Or, cette zone de césure groupe une masse de candidats dont les totaux (ou les moyennes générales) diffèrent peu, et sont entachés d'une erreur qui introduit une part de hasard dans la décision.

Cette erreur, tout comme l'erreur sur une seule épreuve, mais dans de moindres proportions, comprend :

L'erreur d'échantillonnage sur la valeur réelle du candidat, erreur due au fait que cette valeur est testée sur telle série d'épreuves plutôt que sur telle autre série.

L'erreur de « mesure » sur sa moyenne générale. Nous ne reviendrons pas sur la première de ces erreurs dont nous avons parlé au début du présent chapitre. Nous avons vu par quels moyens elle pouvait être notablement réduite.

Quant à la seconde, nous pouvons estimer de façon approchée sa valeur maximale. En effet, l'erreur sur la moyenne générale étant plus faible, par suite des compensations statistiques que l'erreur spécifique à chaque épreuve, nous avons recherché sur l'épreuve la plus entachée d'erreur, c'est-à-dire sur l'épreuve de philosophie arabe, quelle était la marge d'écart la plus probable des copies moyennes (notées 9, 9,5, 10, 10,5 ou 11 par double correction).

Sur ces copies les désaccords de plus de 3 points sont rares (14,5%). Dans la grande majorité des cas, l'erreur ne dépasse pas la marge de $\pm 1,5$ dans l'épreuve la plus infidèle. Il semble donc raisonnable d'admettre que tout candidat ayant obtenu à l'écrit une moyenne générale comprise entre 8,5 et 11,5 sur 20 devrait être soumis à une session de contrôle.

Cette zone peut évidemment être réduite lorsque l'erreur sur la décision tend vers zéro, c'est-à-dire lorsqu'on maximise à la fois l'ampleur de l'investigation et l'objectivité de la notation.

1. Pour le t de student, voir Q. Mc NEMAR - *op. cit.*, pp. 216-234.

2. Voir tableau 3 : Pondération effective des épreuves.

LES FACTEURS DE LA NOTATION

Nous avons effectué cette analyse de facteurs de notation, en nous fondant d'une part sur les facteurs dépendant du *correcteur* lui-même : équation personnelle, attraction pour certaines notes, étalement des notes, fluctuation dans le temps, style de notation et, d'autre part, sur les facteurs dépendant de la *copie* : nous avons essayé de voir sur quoi et pourquoi les correcteurs sont en désaccord et si ce désaccord est aussi important sur les bonnes, les moyennes ou les mauvaises copies.

LE CORRECTEUR

Fréquence et dispersion des notes

Ce phénomène d'attraction pour certaines notes nous était apparu statistiquement dans le chapitre précédent. Nous allons voir qu'il possède des caractéristiques personnelles à chaque correcteur, et d'autres communes à certains correcteurs.

Ainsi, à l'examen des histogrammes des notes de chaque correcteur, il apparaît notamment : Une bimodalité, avec sommets à 8 et à 10 (figure 6 A) ou à 5 et 10 (figure 6 C') très fréquente en philosophie arabe, également présente en philosophie étrangère.

Chez certains correcteurs, une grande fréquence de 5, seuil d'élimination (voir 6 E' et 6 C'), alors que d'autres correcteurs ne semblent tenir aucun compte de ce seuil.

Chez les plus indulgents, la rareté du 15.

Du point de vue de la dispersion de ses notes, chaque correcteur varie dans l'étendue de la gamme des notes qu'il utilise et cette gamme diffère d'un correcteur à l'autre. On constate d'ailleurs que la moyenne des différences interindividuelles d'écart-type (0,33) est la même que la moyenne des différences intraindividuelles d'écart-type (0,33).

Fluctuations dans le temps (1er et 2e paquets)

Le jour et l'heure de la correction n'étant pas indiqués sur les copies, il nous a été impossible d'analyser les variations de corrections dans la même journée, de 8 à 14 heures. Mais nous connaissions l'ordre de correction des paquets et nous savions que ceux-ci avaient été corrigés à deux jours d'intervalle. Nous nous sommes assurés qu'il n'y avait pas

trop de fluctuations importantes entre les deux paquets d'un même correcteur.

La plupart des différences de moyennes entre le 1er et le 2e paquet n'étant pas significatives, on peut considérer – et cela est rassurant – que les correcteurs notent sensiblement autour de la même moyenne à deux jours d'intervalle. Les différences sont en moyenne d'un demi ou trois quarts de point. Mais des différences supérieures à 1 point peuvent être observées (figure 7) pour les correcteurs C, H, B', C', D', E'.

On constate cependant pour ces correcteurs, en philosophie étrangère en particulier, une sorte d'auto-régulation consciente ou non, les correcteurs les plus originaux tendant généralement à revenir à la moyenne de leurs confrères, au cours de la correction de leur second paquet de copies.

Ce fait nous paraît devoir attirer l'attention sur la nécessité, pour les correcteurs sans barème, d'une période d'adaptation au cours de laquelle, corrigeant indépendamment « à blanc » quelques copies, ils ajusteraient ensuite en commun leurs points de vue sur les notes à attribuer.

Style de notation

D'autres critères paraissent également intervenir dans la notation. Tout se passe en effet comme si certains se proposaient de situer la valeur de la copie sur l'échelle de 0 à 20.

D'autres, au contraire, auraient tendance à par-tager l'ensemble des copies en « bonnes » et en « médiocres » (voir la figure 6 C').

D'autres enfin seraient surtout préoccupés de séparer du lot les plus mauvaises (voir 6 C) ou les meilleures (6 C').

En conclusion, il semblerait possible, à partir des données que nous venons d'analyser, d'esquisser une typologie des correcteurs.

LA COPIE

Influence de la qualité

Avant de fixer sa note, le correcteur juge globalement la valeur de la copie. Son impression générale le conduit à qualifier cette copie de « bonne », de « moyenne » ou encore de « mauvaise ». Dans ces conditions, on

Figure 6

CARACTERISTIQUES PERSONNELLES
DE LA NOTATION

Distribution des notes de quelques correcteurs

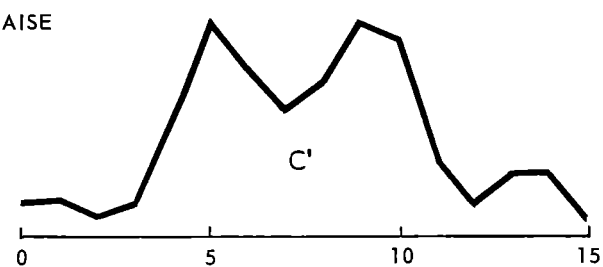
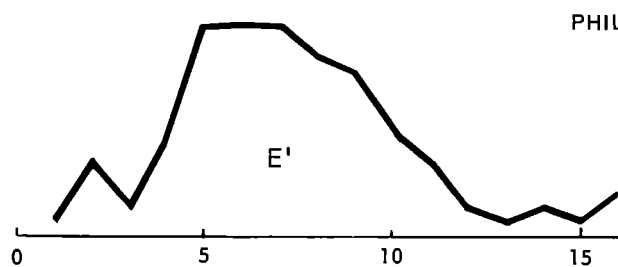
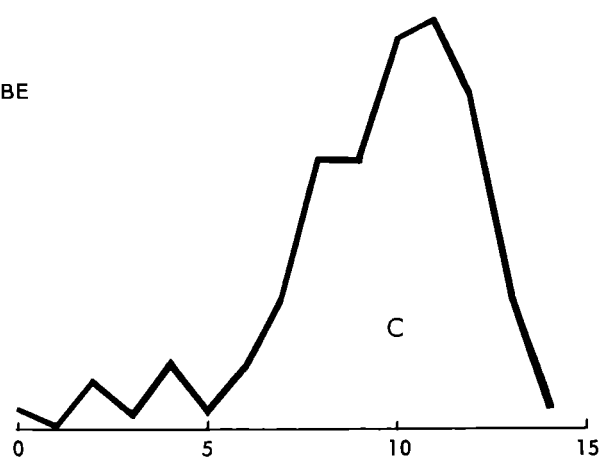
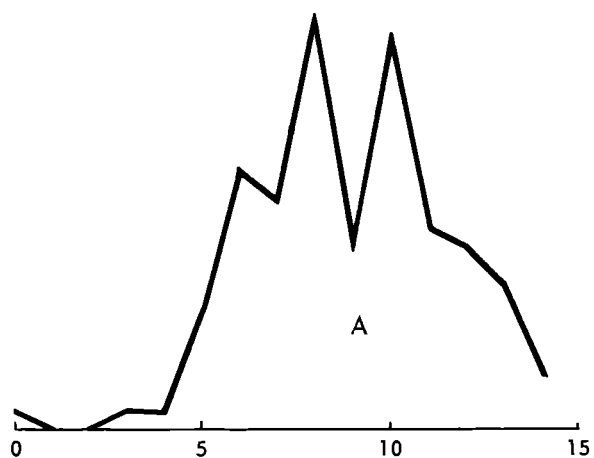
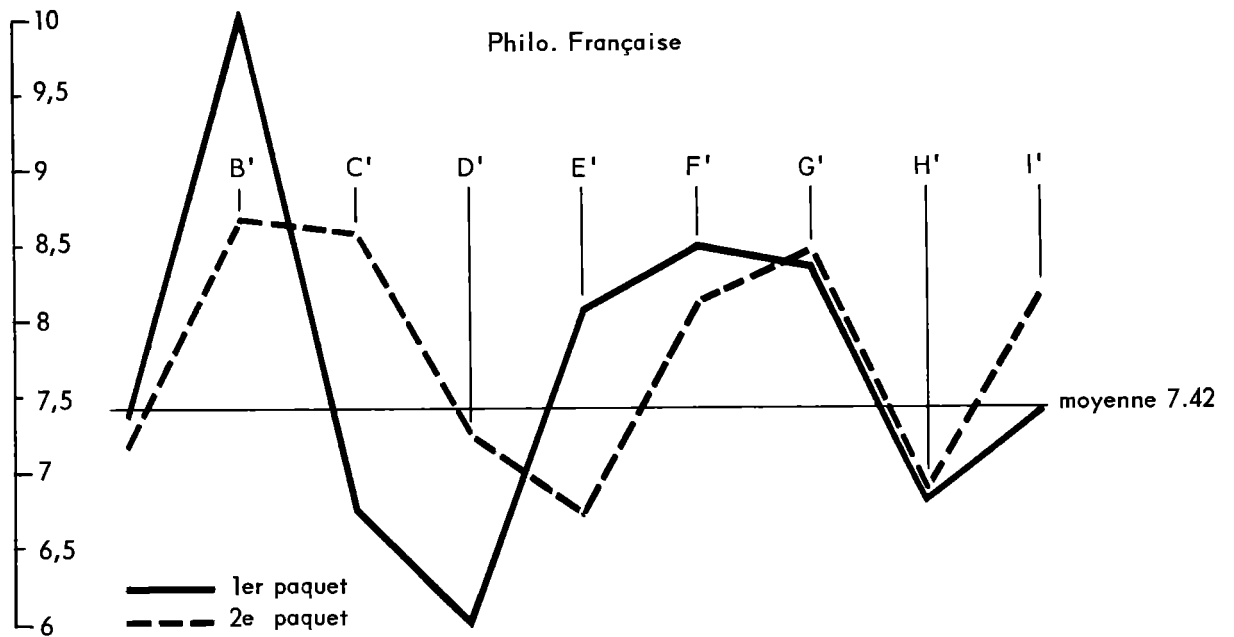
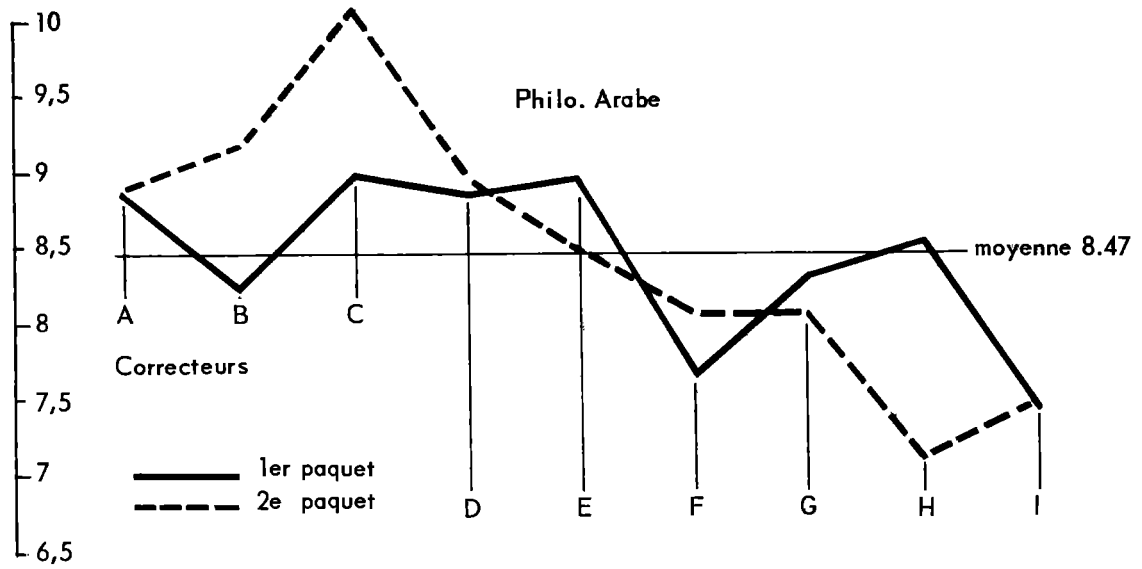


Figure 7

VARIATION DES MOYENNES INTERCORRECTEURS
ET INTRACORRECTEURS



peut se demander si l'accord entre les deux juges-correcteurs n'est pas meilleur sur les bonnes ou les mauvaises copies que sur les moyennes. C'est ce que nous avons voulu mettre en évidence en répartissant les copies, sur la base de la note moyenne de la double correction, en trois groupes comprenant environ 20% de meilleures, 60% de moyennes et 20% de moins bonnes, et en effectuant le décompte des accords et des désaccords dans chacun de ces groupes.

Notre hypothèse se vérifie pour la philosophie arabe, les accords à un point près sont plus fréquents sur les bonnes (62%) et les mauvaises copies (62%). Mais la philosophie étrangère infirme l'hypothèse : les divergences les plus fréquentes sont de 0 à 1 point sur les mauvaises copies (58%), 1 et 2 points sur les moyennes (48%) et 2 et 3 points sur les meilleures (50%).

Pourquoi cela ? Il y a certainement désaccord sur les critères mais il y a aussi une raison technique car il est évident que l'amplitude du désaccord possible est beaucoup plus limitée sur une copie de valeur nulle que sur une copie qui vaut 15, par exemple.

Ici intervient un autre facteur qui est l'attitude du correcteur, dont l'indulgence ou la sévérité relative peut s'exercer électivement sur certaine qualité de copies. Pour ne citer qu'un exemple, nous avons constaté que dans vingt-deux cas, le correcteur D a attribué des notes supérieures à celles qui avaient été données aux mêmes copies par le correcteur C. D'autre part, dans vingt et un cas, le contraire s'est produit : C a attribué des notes supérieures à celles de D.

Tableau 6. Indulgence ou sévérité élective
(les fréquences portent sur la note supérieure
attribuée soit par C, soit par D)

Note supérieure attribuée	... par D	... par C
(points)	(fréquence)	(fréquence)
16	2	
15	1	
14	2	
13	2	
12	3	
11	0	
10	3	3
9	1	3
8	2	4
7	3	1
6	1	3
5		2
4	1	3
3		0
2		1
1	1	1
0		
Total		21

De l'examen de ce tableau, il ressort que D tend à attribuer des notes supérieures - à celles qu'attribue C - aux copies au-dessus de la moyenne tandis que C tend à attribuer des notes supérieures - à celles qu'attribue D - aux copies au-dessous de la moyenne.

Critères de la qualité

Nous avons fait allusion, il y a un instant, au désaccord sur les critères ; il est certain que les correcteurs n'accordent pas, consciemment ou à leur insu le même poids relatif, par exemple, à la forme et au fond de la dissertation.

Pour en avoir une idée plus précise, nous avons dépouillé une vingtaine de copies choisies par nous de telle façon que 10 de ces copies aient réalisé l'accord complet des deux correcteurs sur des notes basses, moyennes et assez bonnes, et qu'une dizaine d'autres reflètent les cas de désaccords extrêmes entre les correcteurs les plus divergents¹.

Autant que nous avons pu en juger, il apparaît que les divergences sont dues à l'importance relative accordée à la forme et au fond, d'une part, à l'originalité et au classicisme du fond, d'autre part.

Par exemple si le contenu de la copie est classique, sans originalité, les correcteurs s'accordent volontiers lorsque la dissertation est bien structurée et que les idées s'enchaînent logiquement. Dans le cas contraire, les divergences apparaissent ou n'apparaissent pas selon que l'un des correcteurs, ou les deux, attribuent plus d'importance à la recherche des éléments du cours professé qu'à l'originalité des idées.

On pourrait schématiquement résumer ces facteurs dans le tableau à double entrée suivant :

FOND		
	-	+
Forme +	Accord (ou faible désaccord)	Accord ou désaccord selon critères : originalité, classicisme
Forme -	Accord	Désaccord ou accord selon critères : Fond-Forme

L'analyse des différents sujets de dissertation proposés aux candidats va d'ailleurs confirmer notre opinion.

Influence du sujet traité

En philosophie étrangère, les trois sujets qui étaient proposés aux candidats et parmi lesquels ils avaient le choix, étaient les suivants :

1. Fixer le sens du mot « Psychologie » ;

1. Ces choix ont été opérés à la lumière des résultats de nos précédentes analyses statistiques.

2. Commenter cette pensée d'Alain : « Imaginer, c'est percevoir, mais mal ».

3. Quel rôle l'économie peut-elle jouer dans la réforme sociale ?

Les deux premiers sujets, plus classiques, ont été choisis chacun par 36% des candidats, mais il nous intéresse davantage de savoir si l'un de ces sujets offre un terrain moins favorable que les autres à l'accord des correcteurs.

On constate effectivement, pour le troisième sujet, le moins classique, les plus forts pourcentages (39%) de gros désaccords (3, 4, 5 et 6 points) et aussi les plus faibles pourcentages d'accords complets (11%).

Le désaccord moyen est de 1,7 points sur les deux premiers sujets, et de 2,2 points sur le troisième.

Il est donc permis de penser que, plus le sujet s'éloigne de la question de cours qui lui correspond et plus il fait place aux idées personnelles, moins il favorise l'accord entre les juges.

CONCLUSIONS

Nous pouvons maintenant résumer nos principales conclusions.

1. La double correction réalise seulement 1 fois sur 5 ou 6, l'accord complet des juges.

2. Les variations intraindividuelles des correcteurs sont sensiblement de même amplitude que leurs variations interindividuelles.

3. Ces dernières proviennent, dans une certaine mesure, d'un manque d'accord préalable sur les objectifs mêmes de la notation.

4. Elles trouvent également leur source dans les divergences sur les critères de jugement, parfois considérables, surtout lorsque la copie est empreinte d'une certaine originalité.

5. Il semble que, d'une façon générale, la quantification du jugement ne s'effectue pas directement sur l'échelle de 0 à 20 mais dérive d'une cascade d'estimations aboutissant, sinon inconsciemment du moins, en fait, à une fixation de la note par rapport à une échelle en 5 points, correspondant aux notes 5, 7-8, 12-13 et 15.

Il est vraisemblable également qu'autour de chacune de ces 5 notes-jalons, les variations sont ensuite plus ou moins aléatoires. Elles correspondent précisément aux plus forts pourcentages de désaccord (1 ou 2 points).

Ces différentes constatations expliquent pourquoi, en philosophie, nous avons orienté nos propositions de réforme de la notation vers l'établissement préalable d'un barème portant sur chacun des deux critères, « Fond » et « Forme », estimé sur une échelle à 5 échelons déterminés qualitativement par un jury de correcteurs.

REFORME FONDAMENTALE ET REMEDES PARTIELS

POUR UNE REFORME FONDAMENTALE

Orientation plutôt que sélection

Tout examen opère une sélection : soit qu'on retienne seulement les élèves les mieux doués (c'est le cas du concours d'entrée), soit qu'on élimine les sujets insuffisants (c'est le cas de l'examen de fin d'études). Or, toute sélection implique un rejet, un déchet. Tout se passe en fait, comme si, par l'examen, la société faisait choix de ses protégés et se désintéressait des autres.

Peut-on ainsi admettre, que, surtout dans les pays en voie de développement, toutes les potentialités humaines ne soient pas utilisées, de la plus haute à la plus faible, de la plus générale à la plus spécifique ?

La solution, nous la percevons dans une réforme fondamentale de l'examen qui, au lieu de crible, devrait faire office de plaque tournante, de table d'orientation.

Le diagnostic large et sûr qu'on pourrait tirer de l'examen devrait essentiellement aboutir à une orientation de l'élève vers la voie scolaire et professionnelle qui lui convient le mieux.

« Diagnostic large », cela signifie que d'autres investigations que celles des seules acquisitions scolaires sont nécessaires. Certes l'examen, dans ses formes traditionnelles, fait parfois appel à la compréhension des lois et problèmes, à des facteurs intellectuels. Jamais cependant il ne fait intervenir certaines aptitudes et certains intérêts spécifiques qui conditionnent l'efficacité d'un individu.

À supposer même qu'il le fasse, on n'en tire actuellement aucun profit, puisque le succès est décidé sur le total des points, sur la valeur moyenne.

C'est là une absurdité que dénonce notamment M. LAUGIER¹ : « Additionner la note de mathématiques avec celle de version latine est presque aussi irrationnel qu'additionner la taille de l'homme et sa pression artérielle ».

D'autre part, en ce qui concerne l'orientation de l'enfant « la note 18 en mathématiques comporte des indications positives précises, et 2 en français comporte des contre-indications catégoriques ; mais l'on ne peut tirer, semble-t-il, aucune conclusion utile du fait qu'il a une moyenne de 10 ».

Cependant, aussi différencié qu'il puisse être, le diagnostic du niveau et des caractéristiques des connaissances scolaires n'est jamais qu'un des

éléments déterminants de l'orientation. Celle-ci doit également tenir compte du niveau et des caractéristiques de l'intelligence générale, des aptitudes et des intérêts du sujet. C'est dire l'importance que nous accordons aux tests autant pour leur champ d'investigation dans ces domaines, que pour leurs qualités métrologiques.

Preuve plutôt qu'épreuve

« Diagnostic large », avons-nous dit, mais aussi « diagnostic sûr ». «

« On ne saurait continuer à utiliser pour mesurer la valeur d'un candidat, un mètre en caoutchouc dont la longueur change avec la main qui l'emploie » écrivait CLAPAREDE.

D'abord est-ce bien la valeur du candidat que l'on mesure avec ce mètre en caoutchouc ? Nous avons déjà posé cette question sur le plan de l'échantillonnage. Ajoutons ici qu'à l'épreuve-obstacle, qui éprouve plus qu'elle ne prouve, il faut préférer l'épreuve bien adaptée, de contenu varié et de niveau gradué, c'est-à-dire l'épreuve-test.

Mais ce premier problème résolu, comment arriver à une mesure plus précise ? En utilisant des instruments de mesure fidèles d'une part, et d'autre part en multipliant les mesures.

L'instrument est d'autant plus fidèle qu'il varie moins avec celui qui l'emploie. Cela revient à dire que la fidélité croît d'autant plus que le correcteur-juge, notant selon son impression globale, fait place au correcteur-pointeur, notant par référence à un barème objectif.

Il est certain que la possibilité d'établir des barèmes analytiques et précis – nulle dans les épreuves orales, et c'est la raison pour laquelle nous les rejetons – est plus réduite dans les épreuves écrites de lettres et de philosophie que dans les épreuves scientifiques et mathématiques ; mais cette possibilité existe néanmoins² et nous en donnerons un exemple par la suite.

1. H. LAUGIER, Une puissante recherche scientifique (conférence prononcée à la Sorbonne le 9 juin 1955), dans *Les cahiers rationalistes*, n° 148, 1955.

2. A l'aide d'un tel barème et avec un traitement statistique simple des notes, VERNER, MARTIN, SIMS ont obtenu en moyenne des écarts inférieurs à 4,5 points sur 100 pour 8 correcteurs différents, sur 33 copies (cf. *Journal of Educational Research*, September 1933).

En second lieu, la précision de la mesure est fonction du nombre de mesures. C'est pourquoi la double correction nous semble nécessaire dans les épreuves de littérature ou de philosophie, même lorsqu'un barème de notation plus ou moins analytique et précis aura été appliqué au moins pour les copies des candidats voisins de la zone de décision.

SUGGESTIONS POUR UN EXAMEN DE FIN D'ETUDES

Le véritable examen probatoire d'orientation dont nous venons d'esquisser les grandes lignes représente à nos yeux l'idéal vers lequel on doit tendre.

Mais en attendant, les examens et les concours existent et il est prudent, étant donné leur grande résonance dans le public, de ne pas trop bousculer un certain traditionalisme. Aussi vaut-il peut-être mieux préparer graduellement l'évolution favorable par un léger traitement curatif et roboratif.

Pour un diagnostic valable

Voici par exemple quelques suggestions que nous avons faites en vue d'une réforme du Baccalauréat libanais, compte tenu de l'organisation actuelle de l'enseignement et de l'examen dans ce pays.

1. Création d'une troisième option « Philosophie-Sciences », qui élargirait et assouplirait le choix des candidats et répondrait à un besoin réel : les futurs étudiants en médecine, presque tous candidats en « Philosophie » sont mal préparés à recevoir l'enseignement scientifique de leur faculté.

2. Valorisation de l'écrit : l'écrit comporterait 5 épreuves au lieu de 3, et l'oral 2 au lieu de 6 ou 7. Parmi les 5 épreuves écrites de chaque série, les deux principales seraient affectées théoriquement du coefficient 3 et les trois autres du coefficient 2, soit un total de coefficients égal à 12, ce qui représente un maximum théorique de 240 points à l'écrit.

3. Tripartition des candidats selon les résultats de l'écrit :

A l'issue des cinq épreuves écrites, les candidats seraient classés en trois catégories (voir figure 8), sur la base du total des points obtenus, compte tenu des coefficients :

Catégorie A : Les candidats ayant obtenu un total égal ou supérieur à 138¹ seraient admis à l'oral allégé.

Catégorie B : Les candidats ayant obtenu un total compris entre 102² et 138 seraient admis à se présenter à la session complémentaire.

Catégorie C : Les candidats ayant obtenu un total égal ou inférieur à 102 seraient ajournés. Ils ne pourraient se présenter à nouveau que l'année suivante

Nous avons pu estimer respectivement à 35% - 40% l'importance numérique de la catégorie A, à 20% -

30% celle de la catégorie B et à 35% - 40% celle de la catégorie C.

4. Oral allégé pour la catégorie A :

Cet oral ne comporterait qu'une épreuve, à savoir : une explication de texte en langue étrangère (coefficient 2).

L'admission définitive serait prononcée exclusivement pour ceux qui auraient obtenu un total général égal ou supérieur à 158 points.

On remarquera que, dans la situation la plus critique de cette catégorie, celle du candidat ayant obtenu 138 à l'écrit, il faut au moins obtenir 20 points (c'est-à-dire la moyenne) à l'oral pour être définitivement reçu. Seuls les candidats ayant plus de 13 de moyenne à l'écrit pourraient se permettre de ne pas prendre l'oral au sérieux, mais ces cas seraient rares.

5. Session complémentaire pour la catégorie B :

Il s'agit de candidats moyens, qui nécessitent un deuxième sondage.

Deux semaines environ après la première session, ces candidats subiraient un écrit complémentaire allégé portant sur les trois matières principales (deux épreuves affectées du coefficient 3 et la troisième du coefficient 2, soit un total de coefficients égal à 8).

Sur la base du *total des écrits des deux sessions*, ces candidats seraient divisés en deux catégories.

Catégorie B₁ : Les candidats ayant obtenu un total général (8 épreuves) égal ou supérieur à 200 points (total correspondant à une moyenne de 10 sur 20) seraient admis à l'oral allégé, le même que celui de la catégorie A. Seraient déclarés définitivement reçus les candidats dont le total général des deux écrits et de l'oral allégé est égal ou supérieur à 220.

Catégorie B₂ : Les candidats ayant obtenu un total général (8 épreuves) inférieur à 200 seraient ajournés sans exception.

6. Des épreuves plus représentatives :

Considérant, pour des raisons techniques indiquées au chapitre précédent, qu'il est préférable de juger un candidat sur plusieurs questions que sur une seule, et que la somme des trois notes sur dix est moins entachée d'erreur que la seule note sur 30 d'une épreuve unique,

Considérant d'autre part que le choix de la question par le candidat est injuste parce qu'il ne place pas tous les candidats devant les mêmes obstacles, nous proposons de restreindre le champ des sujets et d'en augmenter le nombre. Par exemple, chaque épreuve pourrait comporter trois questions obligatoires sur des sujets limités et un problème (ou une dissertation) également obligatoire.

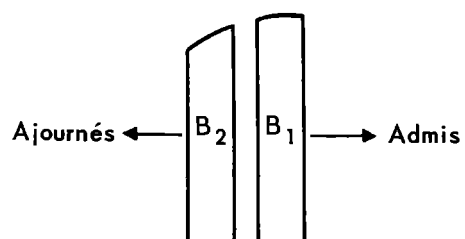
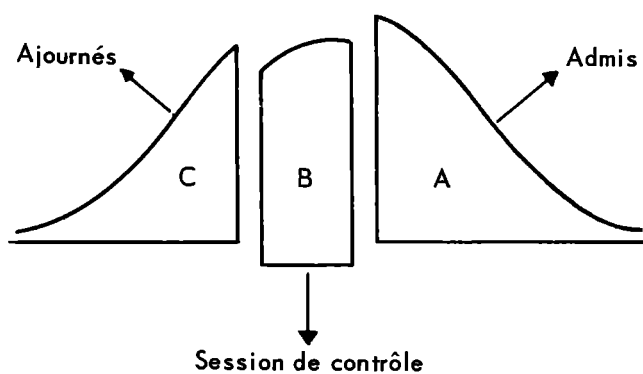
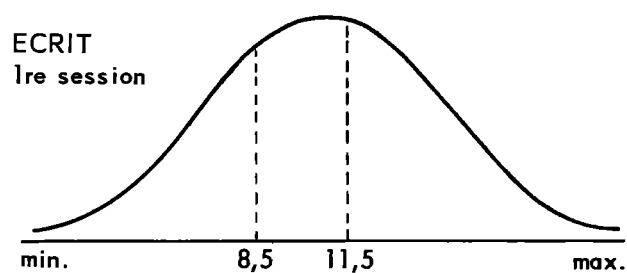
1. Ce total correspond à une moyenne de 11,5 sur 20.

2. Ce total correspond à une moyenne de 8,5 sur 20.

Figure 8

TRIPARTITION DES RESULTATS DE L'ECRIT :

SUGGESTION POUR
L'EXAMEN TERMINAL



Enfin, plutôt que de faire appel au seul savoir mémorisé, aux connaissances "de conserves", il serait préférable d'introduire des sujets portant sur le maniement des mécanismes de pensée et d'expression et sur l'application des connaissances acquises et des instruments de connaissance.

Pour une notation plus fidèle

En accord avec les conclusions de nos analyses, nous envisagerons la notation dans ses phases successives :

Etablissement et polissage du barème. La solution générale au problème de la notation consiste, rappelons-le, à substituer à la quantification d'un jugement global sur une échelle à 20 échelons, une série de notations analytiques relativement indépendantes à échelons peu nombreux.

La première étape aurait pour objet, d'une part, d'établir ces différents secteurs soumis séparément à l'appréciation du correcteur, et d'autre part, de préciser leurs échelons de quantification. Elle serait préparée par les commissions spécialisées après lecture individuelle de quelques copies.

La commission procéderait ensuite à la notation de quelques copies en commun et à l'aide du barème établi, jusqu'à complète disparition des ambiguïtés sur les différents points de ce barème.

En mathématiques et en sciences, l'établissement du barème n'offre que quelques difficultés mineures. Il n'en est pas de même en lettres.

Voici cependant, à titre simplement indicatif, un exemple de barème de dissertation, pouvant servir d'illustration du procédé de notation analytique unifiée¹.

On noterait séparément le fond et la forme de la façon suivante :

Le fond serait jugé, selon son contenu, et l'une des 5 notes suivantes lui serait attribuée : 1, 4, 7, 10 et 13, d'après les indications du barème :

- Note 1. Nul, hors sujet, paraphrase du sujet
4. Quelques idées appropriées, sans originalité
7. Dissertation classique (cours bien assimilé)
10. Dissertation classique avec quelques idées originales
13. Dissertation originale.

A ces notes on retrancherait ou l'on ajouterait selon la présentation des idées :

- 2 ou -1. Le plan n'apparaît pas. Les idées s'enchaînent mal
0. Présentation assez cohérente des idées
- +1 ou +2. Plan net, devoir bien construit, idées bien enchaînées

L'échelle correspondant à la Forme comprendrait 5 échelons d'appréciation du style et de l'orthographe²:

1. Charabia, cousu de fautes d'orthographe
2. Nombreuses incorrections, impropriétés, fautes d'orthographe

3. Style à peu près correct, sans trop de fautes d'orthographe

4. Style correct, pas de faute d'orthographe

5. Style personnel, pas de faute d'orthographe

La présentation matérielle et l'écriture interviendraient dans ces notes par -1, 0 ou +1, selon la qualité du graphisme, la présence ou l'absence de taches et de ratures.

L'ensemble du schéma de notation peut être figuré comme suit :

Exemple de notation analytique pour une épreuve de lettres						
FOND	Contenu des idées	1	4	7	10	13
	Présentation des idées	-2	-1	0	+1	+2
	TOTAL « FOND »	= . . . Max. 15				
FORME	Style et orthographe	1	2	3	4	5
	Présentation matérielle	=1	0	+1		
	TOTAL « FORME »	= . . Max. 6				

Exceptionnellement, un total général «Fond + Forme» supérieur à 20 ou inférieur à 0 serait transformé en 20 ou en 0³.

Notation proprement dite. Pour introduire plus d'homogénéité dans la correction, il est vivement recommandé de répartir les copies au hasard entre les correcteurs.

On peut par exemple procéder, comme au jeu de cartes, à une distribution circulaire.

Ajustement des distributions. Les copies ayant été notées, chaque correcteur établira l'histogramme de ses notes, par pointage de chaque note attribuée après sommation des notes partielles sur l'échelle de 0 à 20. On comparera ensuite ces distributions afin de vérifier que toutes sont axées sur la même note centrale. Si quelques-uns des histogrammes présentaient un décalage trop important par rapport aux autres, il conviendrait de corriger par + ou -2, ou + ou -3 chaque note attribuée par ces correcteurs de façon que leurs histogrammes reviennent dans la norme.

Pondération des notes. Si les distributions des notes sur 20 des épreuves écrites ont sensiblement le même étalement, on pourra, sans risquer trop d'erreurs, appliquer les coefficients prévus.

1. Nous précisons bien qu'il ne s'agit que d'une illustration du principe que nous préconisons. Le barème présenté ici n'est évidemment ni unique ni omnivalent.
2. On pourrait noter séparément le style et l'orthographe, mais l'expérience prouve que l'excès d'analyse rend la correction laborieuse sans apporter un gain appréciable de fidélité.
3. Quel que soit le barème, il importe de prévoir, pour une copie moyenne, une note correspondant à 10 sur 20.

(S'il n'en était pas ainsi, on devrait faire appel aux techniciens de la statistique qui procéderaient à un étalonnage de chaque épreuve, transformeraient les notes brutes en notes standardisées sur lesquelles les coefficients prévus pourraient être appliqués avec une parfaite efficacité).

SUGGESTIONS POUR UN CONCOURS D'ENTREE

En octobre 1957, nous avons réalisé au Liban, à titre expérimental et avec l'approbation du Ministère de l'Education Nationale, un concours national d'entrée dans les établissements officiels d'enseignement secondaire.

Ce concours, qui s'adressait aux nouveaux titulaires du certificat d'études primaires était d'une conception très différente de celle du concours classique à la fois par sa forme, son contenu et sa notation.

Partant de l'hypothèse suivant laquelle on pouvait valablement prédire la réussite dans les études secondaires¹ à partir de deux séries d'informations sur les connaissances scolaires primaires d'une part, et sur les potentialités intellectuelles d'autre part, nous avons utilisé une batterie composée d'épreuves scolaires et d'épreuves psychologiques.

Une batterie de tests appropriés

Ces épreuves, nous les avons voulu communes aux candidats et candidates à l'admission dans tous les établissements officiels secondaires, et uniformes dans leurs conditions d'application de correction et de notation. C'est pourquoi nous avons utilisé des tests (test de connaissances scolaires et tests d'intelligence) que nous avons spécialement préparés, étudiés et mis au point pour garantir ces conditions, et satisfaire aux critères d'un concours d'entrée.

1. Les tests de connaissances scolaires : nous avons, avec la collaboration de professeurs et d'instituteurs spécialisés, littéraires et scientifiques, construit trois épreuves (arabe, calcul, français) destinées à tester, non seulement la somme des connaissances primaires acquises, mais aussi leur application intelligente dans des exercices ou problèmes de difficulté graduée. Pour la notation de chaque test, un barème précis, objectif, pondéré, a été établi.

2. Tests d'intelligence : deux épreuves originales, correspondant chacune aux deux grands aspects de la structure de l'intelligence mis en évidence par les factorialistes anglo-saxons, ont été préparées sous notre direction technique par les psychologues techniciens du Centre de recherches pédagogiques de Beyrouth.

L'une, dite « d'intelligence verbale », comporte six types différents d'exercices non scolaires impliquant

la recherche de relations plus ou moins complexes entre des phrases, des mots ou des nombres.

L'autre, d'un poids plus faible, ne présente au contraire que des problèmes non verbaux c'est-à-dire portant sur des relations entre figures ou symboles géométriques.

3. Trois étalonnages² des notes obtenues par l'ensemble des candidats ont été effectués :

Un étalonnage de chacune des épreuves

Un étalonnage pour l'ensemble des tests de connaissances scolaires

Un étalonnage pour l'ensemble des tests d'intelligence.

Par référence à ces étalonnages, il est aisé de transformer les notes brutes en notes comparables échelonnées de 0 à 10, sur lesquelles on peut ensuite appliquer les coefficients de pondération prévus.

4. Afin de contrôler notre expérimentation, nous avons établi une fiche psycho-pédagogique pour chaque candidat admis. Elle est composée de façon que puissent y être consignés :

Des renseignements d'identité et des renseignements sur la famille,

Des renseignements médicaux et sociaux,

Les résultats obtenus au certificat d'études primaires,

Les résultats obtenus dans notre examen d'entrée par tests,

Les résultats scolaires trimestriels et annuels ainsi que les observations des professeurs de l'enseignement secondaire.

L'apport des tests

Nous n'entrerons ici ni dans les détails de l'organisation et des résultats de cette expérience ni dans les études comparatives, pédagogiques et psychologiques qu'elle a permises.

Du seul point de vue docimologique elle a confirmé, s'il en était encore besoin, la valeur diagnostique³ et pronostique des tests. En effet :

Constatons d'abord que le diagnostic différentiel des connaissances n'est pas le même selon que l'on considère le classement dans le test ou le classement dans l'épreuve du C.E.P. Voici les corrélations obtenues sur l'échantillon de population dont il a déjà été question (4.1).

1. P.E. VERNON, *Secondary school selection ; A British Psychological Society inquiry*, London, Methuen, 1957 216 pages.

2. Précisons qu'il s'agit d'étalonnages en 11 catégories normalisées comprenant respectivement 4%, 4%, 8%, 11%, 15%, 16%, 15%, 11%, 8%, 4%, 4%, des candidats.

3. H. LAUGIER, H. PIERON, Mme H. PIERON, E. TOULOUSE, D. WIENBERG, *Etudes docimologiques sur le perfectionnement des examens et concours*, Etude critique de la valeur sélective du Certificat d'études primaires et comparaison de cet examen avec une épreuve par tests. Paris, Le travail humain, 1935, 90 pages.

Tableau 7. Corrélations entre tests de connaissances et épreuves du C.E.P. correspondantes

Corrélations entre ...	Garçons	Filles
Test et épreuve d'arabe	0,48	0,45
Test et épreuve de calcul	0,39	0,57
Test et épreuve de français	0,48	0,36
Les trois tests et les trois épreuves (A + C + F)	0,45	0,38

Toutes ces liaisons sont significatives. Il existe donc bien, pour chaque matière, un facteur commun couvrant - mais partiellement - la réussite dans l'épreuve du certificat et dans le test correspondant. Il reste cependant que, les corrélations n'étant pas très fortes, les deux types d'épreuves diffèrent sensiblement dans leur façon d'ordonner les niveaux de connaissances. Cela ne saurait nous surprendre étant donné leur contenu différent, quantitativement et qualitativement.

Les différences qualitatives sont précisément mises en relief par les liaisons constatées entre les tests d'intelligence et l'ensemble de trois épreuves principales du C.E.P. d'une part et, d'autre part, entre les tests d'intelligence et l'ensemble des tests scolaires.

Tableau 8. Corrélations entre tests d'intelligence et tests ou épreuves scolaires

Corrélations entre ...	Garçons	Filles
Les tests d'intelligence et les trois épreuves (A + C + F) du C.E.P.	0,11	0,00
Les tests d'intelligence et les trois tests (A + C + F)	0,42	0,31
Tests d'intelligence et note globale au C.E.P.	0,21	0,12

Seules sont significatives les corrélations entre tests d'intelligence et tests de connaissances, ce qui indique la part des processus intellectuels impliqués dans nos tests de connaissances. Par contre l'absence de corrélation entre les tests d'intelligence et les trois épreuves principales du certificat révèle que celles-ci ne font pas suffisamment appel aux capacités intellectuelles des candidats.

D'ailleurs si l'on considère la parenté existant entre la totalité des épreuves du C.E.P. et l'ensemble de nos tests, on s'aperçoit qu'elle est faible ; les corrélations n'atteignent que 0,29 chez les garçons et 0,30 chez les filles.

Les deux types d'examen divergent donc, sinon

totalement, du moins assez profondément, dans leur diagnostic. Il reste à savoir lequel des deux a la meilleure validité, lequel permet le mieux d'estimer les possibilités des élèves dans l'enseignement secondaire qu'ils sont appelés à recevoir.

Voici en fonction des critères de réussite scolaire définis au chapitre 4.2, les corrélations obtenues avec les tests et avec les épreuves classiques.

Tableau 9. Validité comparée des tests et des épreuves de connaissances scolaires

	Tests scolaires	Epreuves du C.E.P.
Arabe (réussite après un an d'enseignement secondaire)	0,74	0,63
Français " "	0,74	0,40
Calcul " "	0,31 ¹	0,03 ¹
Ensemble « scolaire » (A + C + F)	0,67	0,35

1. L'explication de la faiblesse de ces validités est donnée au chapitre 4.3.

La comparaison de ces chiffres fait suffisamment ressortir la supériorité de nos tests scolaires sur les épreuves classiques, du point de vue de la valeur prédictive.

Mais il n'est pas moins intéressant de constater que nos tests d'intelligence ont, eux aussi, leur validité propre, égale ou supérieure même à la validité de certaines épreuves du Certificat (0,40 pour le test-verbal, 0,46 pour le test non verbal, 0,56 pour l'ensemble de ces deux tests).

En définitive, l'ensemble de notre batterie de tests est bien plus valide¹ que l'examen du certificat :

Ensemble de la batterie de tests 0,70

Total général des épreuves du certificat . 0,35

De cette série de résultats nous dégagerons deux conclusions :

1. Du point de vue du diagnostic, les tests de connaissances scolaires sont supérieurs aux épreuves scolaires traditionnelles.

2. Des tests psychologiques appropriés apportent à la prédiction de la réussite ultérieure une contribution dont il serait regrettable de ne pas tirer profit dans un concours d'entrée.

Tests d'intelligence et développement économique et social

Nous avons vu que la validité de nos tests d'intelligence, pour être moins forte que celle des tests de

1. On trouvera des résultats de validation à plus longue échéance dans les revues spécialisées et notamment dans : M. REUCHLIN, « L'orientation dans l'enseignement du second degré » in : B.I.N.O. P. (Paris), numéro spécial, 1959, pp. 150-158.

connaissances scolaires, est cependant loin d'être négligeable.

Or, de semblables recherches, entreprises en Afrique, ont abouti à des résultats décevants. Par exemple, sur des populations de jeunes scolarisés africains, comparables à notre échantillon libanais du point de vue de l'âge, DORMEAU¹ constate qu'on n'a pu observer aucune corrélation significative entre les résultats aux divers types de tests expérimentés et les réussites scolaires. Il est donc permis de penser que le seuil d'acculturation à partir duquel les tests d'intelligence peuvent apporter une contribution au pronostic de la réussite scolaire, se situe avant le certificat d'études

primaires à Beyrouth, et après cet examen dans certaines régions d'Afrique.

Cette observation serait peut-être de nature à conduire à des hypothèses plus générales.

Il semble en tout cas que les tests d'intelligence d'orientation et sélection scolaires ne puissent pas être systématiquement appliqués, dans des pays différemment évolués, sur des populations scolaires comparables à d'autres points de vue.

1. G. DORMEAU, «La recherche scientifique en psychologie et les problèmes de main-d'œuvre», dans *Colloque d'Abidjan Dakar (décembre 1959) sur la recherche scientifique et technique et le développement économique et social des pays africains*.

RESUME ET CONCLUSIONS

Les examens qui jalonnent la vie scolaire des individus revêtent partout, et particulièrement dans les pays en voie de développement, une importance considérable, individuelle et nationale.

Depuis quelques dizaines d'années cependant, des spécialistes ont montré que la valeur réelle des examens ne correspondait ni à leur importance effective, ni surtout à celle qu'on leur accorde généralement.

La présente étude est un témoignage de la très relative efficacité diagnostique et pronostique de nos examens traditionnels adoptés par des pays de culture originellement différente.

L'analyse statistique des résultats des examens de l'enseignement officiel primaire et secondaire libanais a permis de mettre en évidence deux groupes de facteurs perturbateurs généralement observés en d'autres lieux, à savoir :

1. Les erreurs d'échantillonnage sur la valeur réelle du candidat ;
2. Les erreurs d'estimation et de quantification des performances.

D'autres facteurs, plus spécifiques aux pays en voie de développement accéléré faussent considérablement la sélection et la promotion des mieux doués. Ils tiennent principalement à la conception du rôle de l'école et à la valeur de placement dans la fonction publique du diplôme.

D'un point de vue général, nos suggestions concourent à faire de l'examen une série d'épreuves d'orientation, révélatrices du niveau et des caractéristiques des connaissances et des potentialités intellectuelles, et fidèles, c'est-à-dire notées selon des critères métrologiques définis.

A défaut de cette réforme radicale, nous avons

indiqué comment, dans l'immédiat, l'examen terminal traditionnel pouvait être modifié dans sa forme et son contenu. L'examen pourrait, sans graves inconvénients, être allégé de sa partie orale et de sa seconde session. La tripartition rationnelle des performances à l'écrit et la création d'une session de contrôle pour les candidats du groupe médian assureraient à la fois la rectification des indulgences de l'écrit (rôle de l'oral actuel) et le repêchage des malchanceux ou des candidats sévèrement notés (rôle de la seconde session traditionnelle). Cette façon de procéder aurait l'avantage d'être rapide et économique puisqu'elle porterait sur un effectif 5 fois moindre que celui de l'oral et de la 2e session réunis.

Il nous semble aussi que la sommation des notes des deux sessions (la première et la complémentaire) requirait d'autant l'erreur d'échantillonnage sur la valeur réelle de ces candidats moyens.

Nous avons également tenté d'apporter une contribution à l'amélioration des procédés de notation, dans le sens d'une plus grande objectivité.

Enfin, nous avons montré comment le concours d'entrée pouvait accroître sa validité par la méthode des tests.

Ces réformes ne sont que quelques moyens parmi d'autres de rationaliser et de rendre plus équitable l'attribution d'un diplôme ou l'accession à certains enseignements, mais elles vont plus loin : appliquées dans un véritable esprit compréhensif elles doivent avoir une heureuse influence aussi bien sur la pédagogie appliquée dans les classes préparatoires aux examens, que sur la promotion des élites.

BIBLIOGRAPHIE

- American Council on Education. *Educational measurement*, edited by E.F. Lindquits and others, Washington, D.C., 1951. 819 p.
- American Council on Education. Cooperative Study of Evaluation in General Education. *General education: explorations in evaluation, the final report*. Washington, D.C., 1954. 302 p.
- Cambon, J. «L'examen d'entrée dans les centres d'apprentissage de l'Académie de Paris». *Bulletin de l'Institut national d'orientation professionnelle* (Paris), numéro spécial 1958, pp. 185-192.
- Clark, H. «The effect of a candidate's age upon teachers' estimates and upon his chances of gaining a grammar school place». *British Journal of Educational Psychology* (London), vol. XXVI, part III, November 1956, pp. 207-217.
- Conference on Examinations, 1st., Eastbourne, England, 1931. *Proceedings*, edited by Paul Monroe. New York, Bureau of Publications, Teachers College, Columbia University, 1931. 316 p.
- Conference on Examinations, 3rd., Dinard, France, 1938. *Proceedings*, edited by Paul Monroe. New York, Bureau of Publications, Teachers College, Columbia University, 1939. 330 p.
- Dale, R.R. *From school to university*. London, Routledge and Kegan Paul, 1954. 258 p.
- Emmet, W.G. *An inquiry into the prediction of secondary school success*. London, University of London Press, 1942. 58 p.
- Guerbet-Seux, A.M.; Reuchlin, M. «Etude sur l'examen d'entrée en sixième dans cinq établissements scolaires parisiens». *Bulletin de l'Institut national d'orientation professionnelle* (Paris), numéro spécial 1958, pp. 9-19.
- Hartog, P.J.; Rhodes, E.C. *An examination of examinations*. 2nd ed. London, Macmillan, 1936. 81 p.
- International Examinations Inquiry. English Committee. *The marking of English essays*. London, Macmillan, 1941. 165 p.
- International Examinations Inquiry. French Committee. *La correction des épreuves écrites dans les examens, enquête expérimentale sur le baccalauréat*. Paris, La maison du livre, 1936. 387 p.
- Jeffery, G.B. *External examinations in secondary schools, their place and function*. London, Harrap, 1958. 128 p.
- Laugier, H. et al. *Etudes docimologiques sur le perfectionnement des examens et des concours*. Paris, Le travail humain, 1935. 90 p.
- Laugier, H.; Weinberg, D. «Le facteur subjectif dans les notes d'examen». *L'année psychologique* (Paris), vol. XXVIII, Ire partie, 1927. pp. 236-244.
- ; ——— «Le facteur subjectif dans les notes d'examen, deuxième contribution». *L'année psychologique* (Paris), vol. XXXI, Ire partie, 1930, pp. 229-241.
- McClelland, W.W. *Selection for secondary education*. 3rd ed. London, [Published for the Scottish Council for Research in Education by the] University of London Press, 1949. 264 p.
- McIntosh, D.M. *Promotion from primary to secondary education*. London, University of London Press, 1949. 151 p.
- Mialaret, G. «Les applications de la psychologie au domaine scolaire». *Traité de psychologie appliquée*, livre VII, pp. 1469-1531. Paris, Presses Universitaires de France, 1959.
- Michigan State College. Board of Examiners. *Comprehensive examinations in a program of general education*. East Lansing, Michigan State College Press, 1949. 165 p.
- National Foundation for Educational Research in England and Wales. *The allocation of primary school leavers to courses of secondary education. First interim report*, by A.F. Watts and Patrick Slater, London, Newnes, 1951. 68 p.
- . *The allocation of primary school leavers to courses of secondary education. Second interim report: «Secondary school entrance examinations»*, by A.F. Watts, D.A. Pigeon and A. Yates. London, Newnes, 1953. 77 p.
- . *The allocation of primary school leavers to courses of secondary education. Third interim report: «Admission to grammar schools»*, by A. Yates and D.A. Pigeon. London, Newnes, 1957. 260 p.
- National Union of Teachers. *Transfer from primary to secondary schools*. London, Evans Bros., 1949. 190 p.
- New York (State). University. *A design for improving early secondary education in New York State*. Albany, 1954. 118 p.
- Nisbet, J.; Buchan, J. «The long-term follow-up of assessments at age eleven». *British Journal of Educational Psychology* (London), vol. XXIX, part I, February 1959, pp. 1-8.

- Nuttin, J. ; Swinnen, K. *Overgang naar het middelbaar onderwijs*. Leuven (Louvain), Universitaire Uitgaven, [1956]. 180 p.
- Oliver, R.A.C. *An experimental examination in general studies*. Manchester, England, Joint Matriculation Board, 1955. 36 p.
- Parkyn, G.W. *Success and failure at the university*. Wellington, New Zealand Council for Educational Research, 1959. 2 v.
- Peel, E.A. «Assessment of interest in practical topics». *British Journal of Educational Psychology* (London), vol. XVIII, part I, February 1948, pp. 41-47.
- Petch, J.A. *Fifty years of examining: The Joint Matriculation Board, 1903-1953*. London, Harrap, 1953. 226 p.
- Pilliner, A.E.G. «The position and size of the border-line group in an examination». *British Journal of Educational Psychology* (London), vol. XX, part II, June 1950, pp. 133-136.
- Piobetta, J. *Le baccalauréat*. Paris, Baillière, 1937. 1.060 p.
- Reuchlin, M. «Le problème pratique des pronostics». *Traité de psychologie appliquée*, livre III, pp. 395-421. Paris, Presses universitaires de France, 1954.
- Ross, C.C. *Measurement in today's schools*. 3rd ed. revised by J.C. Stanley. New York, Prentice Hall, 1954. 485 p.
- Ruarus, D.G. ; Van Raalte, C.R. *De kunst van examineren*. Leiden, Nederlandsche Uitgever-smaatschappij, 1953. 79 p.
- Ruarus, D.G. ; Van Raalte, C.R. *De kunst van het examen doen*. Leiden, Nederlandsche Uitgever-smaatschappij, 1953. 79 p.
- Sanders, C. *Student selection and academic success in Australian universities*. Sydney, Commonwealth Office of Education, 1948. 158 p.
- Scottish Council for Research in Education. *The prognostic value of university entrance examinations in Scotland*. London, University of London Press, 1936. 197 p.
- «Symposium on the effects of coaching and practice in intelligence tests». *British Journal of Educational Psychology* (London), vol. XXIII, part III, November 1953. vol. XXIV, part II, June 1954. 5 articles.
- «Symposium on the selection of pupils for different types of secondary education». *British Journal of Educational Psychology* (London), vol. XVII, part II, June 1947. vol. XX, part I, February 1950. 9 articles.
- Valentine, C.W. ; Emmet, W.G. *The reliability of examinations*. London, University of London Press, 1932. 196 p.
- Vernon, P.E. *Secondary school selection: a British Psychological Society inquiry*. London, Methuen, 1957. 216 p.
- Walker, A.S. *Pupils' school records*. London, Newnes, 1955. 199 p.

Nous signalons l'existence du système des
BONS DE L'UNESCO
Afin de remédier aux difficultés d'ordre monétaire
que soulèvent les achats à l'étranger
de livres, films, équipements de laboratoire, etc.,
l'Unesco a conçu une sorte de monnaie internationale,
le BON UNESCO
Les BONS DE L'UNESCO fournissent
aux écoles, universités, professeurs et étudiants
d'un grand nombre de pays
la possibilité de se procurer aisément le matériel
dont ils ont besoin pour leurs études
ou leurs recherches

Les BONS DE L'UNESCO sont en vente
dans la plupart des Etats membres
où il existe un contrôle des changes
Pour de plus amples renseignements,
s'adresser, dans chaque pays,
à la commission nationale pour l'Unesco
ou, directement,
au siège de l'Organisation.

Le BON DE VOYAGE UNESCO,
nouvelle application
du système des BONS DE L'UNESCO,
vise à écarter les difficultés de change
qui entravent souvent
les déplacements entrepris à des fins éducatives
ou culturelles :
sortes de chèques de voyage internationaux,
les BONS DE VOYAGE UNESCO,
fournissent aux étudiants,
aux professeurs et aux chercheurs
les devises dont ils peuvent avoir besoin
pour poursuivre leurs études
ou leurs travaux à l'étranger.

Toutes précisions utiles sont données dans le dépliant

LES BONS DE L'UNESCO
ainsi que dans le dépliant

L'UNESCO PRESENTE LE BON DE VOYAGE UNESCO

*où l'on trouvera la liste des organismes nationaux responsables
de la répartition et de l'émission des bons et les banques où
ceux-ci peuvent être échangés contre les devises nécessaires.*

les bons de l'unesco

permettent d'acheter :

livres, périodiques,
photocopies,
microfilms,
reproductions d'œuvres d'art,
diagrammes, globes terrestres,
cartes géographiques ;
partitions musicales,
disques,
films éducatifs sous forme de :
a) copies positives et
contre-types ;
b) négatifs originaux et
contre-types ;
et pellicule vierge de 16 mm.
pour tirage de ces films ;
matériel scientifique pour
l'enseignement et la recherche
notamment :
instruments et matériel
d'optique, balances et poids,
verrerie de laboratoire,
appareils de mesure électrique
et acoustique,
appareils d'analyse et de
contrôle, etc.

*Ces dépliants seront
adressés aux personnes qui
en feront la demande au*

**Service
des bons de l'Unesco
place de Fontenoy
Paris-7^e
France**

PUBLICATIONS DE L'UNESCO: AGENTS GÉNÉRAUX

- Afghanistan:** Panuzai Press Department, Royal Afghan Ministry of Education, KABUL.
- Albanie:** N. Sh. Botimeve, «Naim Frasheri», TIRANA.
- Allemagne (Rép. féd.):** R. Oldenbourg Verlag, Unesco-Vertrieb für Deutschland, Rosenheimerstrasse 145, MÜNCHEN 8.
- Antilles françaises:** Librairie J. Bocage, Rue Lavoisier, B.P. 208, FORT-DE-FRANCE (Martinique).
- Antilles néerlandaises:** G.C.T. Van Dorp & Co., (Ned. Ant.) N.V., WILLEMSTAD (Curaçao, N.A.).
- Argentine:** Editorial Sudamericana, S.A. Alsina 500, BUENOS AIRES.
- Australie:** Melbourne University Press, 369 Lonsdale Street, MELBOURNE C.1 (Victoria).
- Autriche:** Verlag Georg Fromme & Co., Spengergasse 39, WIEN V.
- Belgique:** Office de publicité, S.A., 16, rue Marcq, BRUXELLES; N.V. Standaard Boekhandel, Belgiëlei 151, ANTWERPEN. Pour *Le Courrier*: Louis de Lannoy, 22, place de Brouckère, BRUXELLES.
- Birmanie:** Burma Translation Society, 529-531 Merchant Street, RANGOON.
- Bolivie:** Librería Selecciones, avenida Camacho 369, casilla 972, LA PAZ.
- Brésil:** Fundação Getúlio Vargas, 186 Praia de Botafogo, caixa postal 4081, RIO DE JANEIRO.
- Bulgarie:** Raznoiznos, 1 Tzar Assen, SOFIA.
- Cambodge:** Librairie Albert Portail, 14, avenue Boulloche, PHNOM-PENH.
- Canada:** L'imprimeur de la Reine, OTTAWA (Ont.).
- Ceylan:** Lake House Bookshop, P.O. Box 244, Lady Lochore Building, 100 Parsons Road, COLOMBO 2.
- Chili:** Editorial Universitaria, S.A., avenida B. O'Higgins 1058, casilla 10220, SANTIAGO.
- Chine:** The World Book Co., Ltd., 99 Chungking South Road, Section 1, TAIPEH, Taiwan (Formosa).
- Colombie:** Librería Central, Carrera 6-A n.º 14-32, BOGOTÁ.
- Corée:** Korean National Commission for Unesco, P. O. Box Central 64, SEOUL.
- Costa Rica:** Imprenta y Librería Trejos, S.A., apartado 1313, SAN JOSÉ.
- Cuba:** Librería Económica, Pte. Zayas 505-7, apartado 113, LA HABANA.
- Danemark:** Ejnar Munksgaard, Ltd., 6 Nørregade, KOBENHAVN K.
- République Dominicaine:** Librería Dominicana, Mercedes 49, apartado de correos 656, CIUDAD TRUJILLO.
- Équateur:** Casa de la Cultura Ecuatoriana, Núcleo del Guayas, Pedro Moncayo y 9 de Octubre, casilla de correo n.º 3542, GUAYAQUIL.
- Espagne:** Librería Científica Medinaceli, Duque de Medinaceli 4, MADRID 14. Pour *Le Courrier*: Ediciones Iberoamericanas S.A., Pizarro 19, MADRID.
- États-Unis d'Amérique:** Unesco Publications Center, 801 Third Avenue, NEW YORK 22, N.Y., et sauf pour les périodiques: Columbia University Press, 2960 Broadway, NEW YORK 27, N.Y.
- Ethiopie:** International Press Agency, P.O. Box 120, ADDIS ABABA.
- Finlande:** Akateeminen Kirjakauppa, 2 Keskuskatu, HELSINKI.
- France:** Librairie de l'Unesco, place de Fontenoy, PARIS-7^e; Vente en gros: Unesco, Section des ventes, place de Fontenoy, PARIS-7^e.
- Grèce:** Librairie H. Kauffmann, 28, rue du Stade, ATHENES.
- Haïti:** Librairie «A la Caravelle», 36, rue Roux, B.P. 111, PORT-AU-PRINCE.
- Hong-kong:** Swindon Book Co., 25 Nathan Road, KOWLOON.
- Hongrie:** Kultura, P.O. Box 149, BUDAPEST 62.
- Inde:** Orient Longmans Private Ltd., 17, Chittarajah Avenue, CALCUTTA 13; Indian Mercantile Chamber, Nicol Road, BOMBAY 1; 36a Mount Road, MADRAS 2; Gunfoundry Road, HYDERABAD 1; Kanson House, 24/1 Asaf Ali Road, P.O. Box 386, NEW DELHI 1. Sous-dépôts: Oxford Book & Stationery Co., 17, Park Street, CALCUTTA 16; et Scindia House, NEW DELHI.
- Indonésie:** Bappit Pusat «PERMATA», Djl. Nusantara 22, DJAKARTA.
- Irak:** MacKenzie's Bookshop, BAGHDAD.
- Iran:** Commission nationale iranienne pour l'Unesco, avenue du Musée, TEHERAN.
- Irlande:** The National Press, 2 Wellington Road, Ballsbridge, DUBLIN.
- Israël:** Blumstein's Bookstores Ltd., 35 Allenby Road et 48 Nahlat Benjamin Street, TEL AVIV.
- Italie:** Libreria Commissionaria Sansoni, via Gino Capponi 26, casella postale 552, FIRENZE.
- Jamaïque:** Sangster's Book Room, 91 Harbour Street, KINGSTON; Knox Educational Services, SPALDINGS.
- Japon:** Maruzen Co., Ltd., 6, Tori-Nichome, Nihonbashi, P.O. Box 605, Tokyo Central, TOKYO.
- Jordanie:** Joseph I. Bahous & Co., Dar-ul-Kutub, Salt Road, P.O. Box 66, AMMAN.
- Liban:** Librairie Antoine, A. Naufal et Frères, B.P. 656, BEYROUTH.
- Libéria:** J. Momolu Kamara, 69 Front & Gurley Streets, MONROVIA.
- Luxembourg:** Librairie Paul Bruck, 33, Grand-Rue, LUXEMBOURG.
- Malaisie (Fédération de) et Singapour:** Federal Publications Ltd., Times House, River Valley Road, SINGAPORE.
- Malte:** Sapienza's Library, 26 Kingsway, VALLETTA.
- Maurice:** Nalanda Co. Ltd., 30, Bourbon Street, PORT-LOUIS.
- Maroc:** Centre de diffusion documentaire du B.E.P.I., B.P. 211, RABAT.
- Mexique:** Editorial Hermes, Ignacio Mariscal, 41, México D.F. MÉXICO.
- Monaco:** British Library, 30, boulevard des Moulins, MONTE-CARLO.
- Nicaragua:** Librería Cultural Nicaragüense, calle 15 de Septiembre n.º 115, MANAGUA.
- Nigeria:** C.M.S. (Nigeria) Bookshops, P.P. Box 174, LAGOS.
- Norvège:** A. S. Bokhjórnet, Lille Grensen 7, OSLO.
- Nouvelle-Zélande:** Unesco Publications Centre, 100 Hackthorne Road, CHRISTCHURCH.
- Pakistan:** The West-Pak Publishing Co. Ltd., Unesco Publications House, P.O. Box 374, 56-N Gulberg Industrial Colony, LAHORE.
- Panama:** Cultural Panameña, Avenida 7.ª n.º TI-49 apartado de correos 2018, PANAMÁ.
- Paraguay:** Agencia de Librerías de Salvador Nizza Yegros, entre 25 de Mayo y Mcal. Estigarribia, ASUNCIÓN.
- Pays-Bas:** N. V. Martinus Nijhoff, Lange Voorhout 9, 's-GRAVENHAGE.
- Pérou:** «ESEDAL - Oficina de Servicios» Dpto. de Venta de Publicaciones, Av. Tacna 359, Ofc. 51, Casilla 577, LIMA.
- Philippines:** Philippine Education Co. Inc., 1104 Castillejos Quiao, P.O. Box 620, MANILA.
- Pologne:** Ośrodek Rozwozu Technicznej Wydawnictw Naukowych PAN, Pałac Kultury i Nauki, WARSZAWA.
- Portugal:** Dias & Andrade Lda., Livraria Portugal, rua do Carmo 70, LISBOA.
- République arabe unie:** La Renaissance d'Egypte, 9 Sh. Adly-Pasha, CAIRO (Egypt).
- Fédération de Rhodésie et du Nyassaland:** The Book Centre, First Street, SALISBURY (Southern Rhodesia).
- Roumanie:** Cartimex, Str. Aristide Briand 14-18, P.O. Box 134-135, BUCURESTI.
- Royaume-Uni:** H. M. Stationery Office, P.O. Box 569, LONDON S.E. 1.
- Salvador:** Manuel Navas & Cia., 1.ª avenida Sur 37, SAN SALVADOR.
- Singapour:** Voir: Malaisie (Fédération de).
- Suède:** A/B C.E. Fritzes Kungl. Hovbokhandel, Fredsgatan 2, STOCKHOLM 16. Pour *Le Courrier*: Svenska Unescorådet, Vasagatan 15-17, STOCKHOLM C.
- Suisse:** Europa Verlag, Ramistrasse 5, ZÜRICH; Payot, 49, rue du Marché, GENÈVE.
- Tchécoslovaquie:** Artia Ltd., 30 Ve Smečkáč, PRAHA 2.
- Thaïlande:** Suksapan Panit, Mansion 9, Rajdamnern Avenue, BANGKOK.
- Turquie:** Librairie Hachette, 469, Istiklal Caddesi, Beyoglu, ISTANBUL.
- Union Sud-Africaine:** Van Schaik's Bookstore (Pty) Ltd., Libri Building, Church Street, P.O. Box 724, PRETORIA.
- URRS:** Mezhdunarodnaja Kniga, MOSKVA G-200.
- Uruguay:** Unesco, Centro de Cooperación Científica para América Latina, bulevar Artigas 1320-24, casilla de correo, 859, MONTEVIDEO. Oficina de Representación de Editoriales, plaza Cagancha 1342, 1^{er} piso, MONTEVIDEO.
- Viet-nam:** Librairie papeterie Xuân-Thu, 185-193, rue Tu-Do, B.P. 283, SAIGON.
- Yougoslavie:** Jugoslovenska Knjiga, Terazije 27, BEOGRAD.