

Etude docimologique de l'examen du baccalauréat en Algérie

Abdelhafid Dahmani & Mohamed Touati-Tliba⁽¹⁾

Abstract:

Cet article présente une étude docimologique de l'examen du baccalauréat. L'Analyse Factorielle Exploratoire (AFE) puis l'Analyse de Corrélation Intra Class (ACIC) sont appliquées pour étudier la validité et la fiabilité des trois principales séries du 2012. L'AFE des notes révèle trois principaux construits assez nets. Ces dimensions, les mêmes pour les trois séries, peuvent être associées aux habilités «*verbale*», «*quantitative*» et «*langues étrangères*». Pour étudier la fiabilité, l'ACIC est appliquée aux trois facteurs révélés (*mesure de cohérence*) puis à la double correction des matières (*mesure d'accord total*). Bien que les mesures de fiabilité semblent varier d'une dimension à l'autre, elles sont jugées, néanmoins, «acceptables» même pour les dimensions «*verbale*» et «*langues étrangères*», où l'alpha de Cronbach est respectivement supérieur à 0.8 et à 0.66 pour les trois séries. De même, l'ACIC de la double correction montre, pour les différentes matières des trois séries, des degrés d'accord entre les correcteurs qui varient de «très satisfaisant» (pour les mathématiques, CCIC > 0.96) à «assez satisfaisant» (pour la philosophie, CCIC > 0.85). Globalement, Il ressort de cette analyse que l'examen du baccalauréat est suffisamment valide et fiable et peut donc être considéré comme mesure crédible de la performance scolaire.⁽²⁾

Mots clés: Docimologie; Validité; Fiabilité; Analyse Factorielle Exploratoire; Coefficient de Corrélation Intra Classe.

(1) Ecole supérieure de commerce-ESC d'Alger. Laboratoire d'études pratiques en sciences commerciales et en sciences de gestion de l'ESC.

(2) Ce travail est effectué dans le cadre du projet PNR N° 08/2011. Les auteurs du présent article expriment leur gratitude à tous les collègues du laboratoire de l'ESC notamment M. Sadeg et H. Mokraoui (chef et membre de l'équipe PNR) et membre de l'équipe PNR) pour les discussions fructueuses durant la réalisation du projet. Les auteurs remercient vivement Mr A. Salhi (Directeur de l'ONEC) pour l'accès aux données nécessaires à cette étude.

Introduction

La notation scolaire est censée mesurer les performances des élèves. Cependant, la fiabilité des examens, notamment à réponses ouvertes et non à choix multiples, a depuis longtemps été mise en cause (Laugier et Weinberg (1936)). Ce problème demeure l'objet de débats et de controverses (Suchaut, 2008). Laugier et Weinberg ont montré que la double correction, supposée augmenter la fiabilité des examens et concours standardisés, est illusoire; *«il faudrait, pour obtenir une note «exacte», 127 correcteurs en philosophie - 78 en composition française - 28 en anglais - 19 en version latine - 16 en physique - 13 en mathématiques»*.⁽³⁾ Pour Merle (2007), *«La philosophie, le français, les sciences économiques et sociales, disciplines évaluées essentiellement par des dissertations ou commentaires, sont particulièrement soumis aux incertitudes de la notation»*. Selon Aymes (1979), cité par Merle, *«l'incertitude de la notation d'une copie a été régulièrement confirmée, notamment en mathématiques, même lorsqu'il existe un barème précis de notation, question par question»*.

Le présent travail est consacré à l'étude docimologique des résultats du baccalauréat (bac) de l'année 2012 pour les séries Sciences expérimentales (S), Lettres et philosophie (L) et Gestion et économie (G). Les résultats empiriques montrent que l'examen du baccalauréat est suffisamment valide et fiable. En effet, l'analyse factorielle exploratoire (AFE) révèle trois principaux construits identiques pour les séries du bac étudiées. Ces dimensions peuvent être associées aux habilités *«verbale»*, *«quantitative»* et *«langues étrangères»*. L'analyse de corrélation intra class (ACIC) appliquée aux facteurs révélés montre une fiabilité acceptable. L'ACIC de la double correction montre notamment un degré d'accord *«assez satisfaisant»* entre les correcteurs des différentes matières y compris pour la philosophie.

Cet article comporte trois sections. La première concerne l'étude descriptive des notes des différentes matières relatives aux trois séries du bac étudiées. La deuxième section présente une analyse factorielle exploratoire. Elle est utilisée pour identifier et analyser les dimensions latentes dans les notes du bac. La troisième section est consacrée à l'analyse de la fiabilité des résultats de l'examen. Enfin, une conclusion résume les résultats essentiels de cette étude.

1. Analyse descriptive des notes des matières du baccalauréat

Cette section est consacrée à l'analyse descriptive des notes des différentes matières relatives aux trois séries du bac S, L et G. Le choix de ces séries est motivé

(3) Selon ces auteurs, *«Une note «exacte» étant une moyenne de notes telle que l'adjonction d'une autre note ne modifie pas sensiblement cette moyenne»*.

par le fait que les trois filières, à elles seules, concernent 82% des 515 907 candidats présents au Bac de l'année 2012. Les effectifs relatifs à l'ensemble des filières du bac sont représentés dans le tableau 1 :

Tableau 1: Répartition des candidats au bac

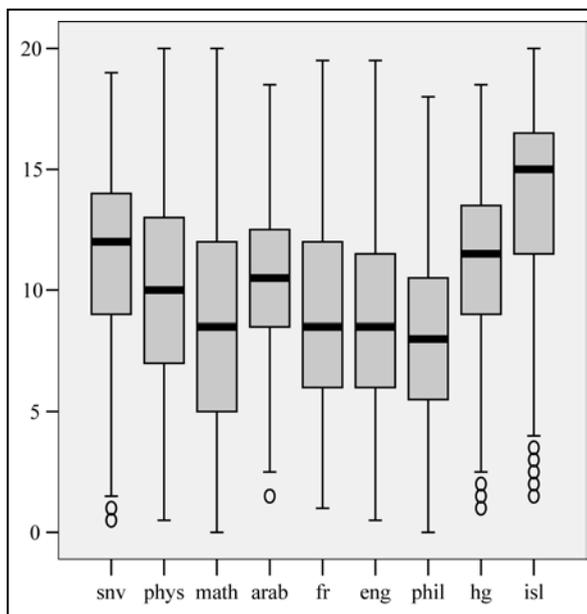
Code série	Série bac	Effectif	Pourcentage
30	L	181 937	35,27%
31		41 335	8,01%
32	G	56 536	10,96%
33	S	186 810	36,21%
34		13 733	2,66%
35		9 383	1,82%
36		11 042	2,14%
37		9 041	1,75%
38		6 090	1,18%
	Total	515 907	100,00%

Baccalauréat Sciences:

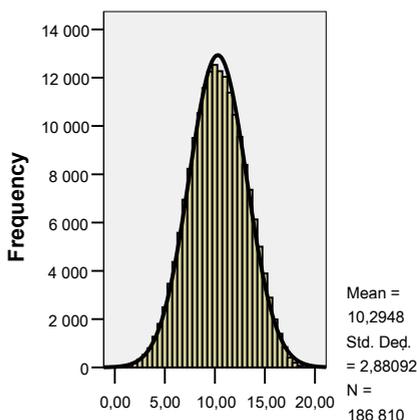
Le tableau 2 montre les moyennes, les écart-types et les coefficients d'asymétrie de l'ensemble des épreuves du bac série S. Nous constatons que les notes moyennes varient d'un minimum de 8.10 pour la philosophie (PHIL) à un maximum de 13.64 pour les sciences islamiques (ISL). En conséquence, les évaluateurs des différentes matières (concepteurs, correcteurs et jurys) de l'examen du bac semblent avoir des appréciations distinctes quant au niveau moyen des candidats. Cette dernière constatation concerne même les matières de spécialité. En effet, la note moyenne de SNV de 11.34 est nettement plus élevée que la moyenne des mathématiques (MATH) qui n'est que de 8.63. De même, pour les écart-types, nous observons que les évaluateurs de la matière MATH utilisent une notations plus discriminatoire (un écart-type de 4.59); par contre pour ARAB l'écart-type n'est que de 2.88. Les *boîtes en moustache*, (voir figure1), basées sur un échantillon aléatoire de 1%, résument graphiquement les caractéristiques essentielles des distributions des notes des différentes matières de la série S.

Tableau 2: Statistiques des notes des matières du bac série S (N=186810)

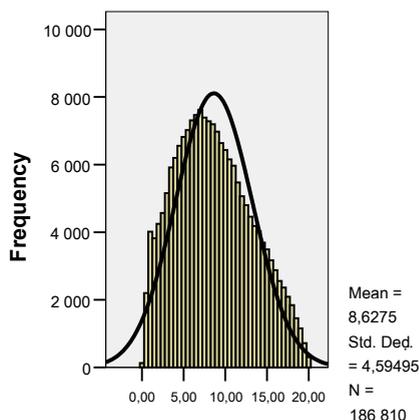
Matière	Abrév	Min	Max	Moyenne	Ecart-type	Coef. Asymétrie
Science	SNV	0	20	11,34	3,67	-0,584
Physique	PHYS	0	20	9,78	4,27	0,013
Mathématiques	MATH	0	20	8,63	4,59	0,3
Langue arabe	ARAB	0	19,5	10,29	2,88	-0,043
Français	FR	0	20	8,99	3,76	0,306
Anglais	ENG	0	20	9,08	3,78	0,354
Philosophie	PHIL	0	18,5	8,1	3,11	0,262
Histoire-Géographie	HG	0	19,5	10,88	3,19	-0,324
Education Islamique	ISL	0	20	13,64	3,93	-0,806

**Figure 1:** Distributions des notes des matières Bac S (Echantillon aléatoire 1%)

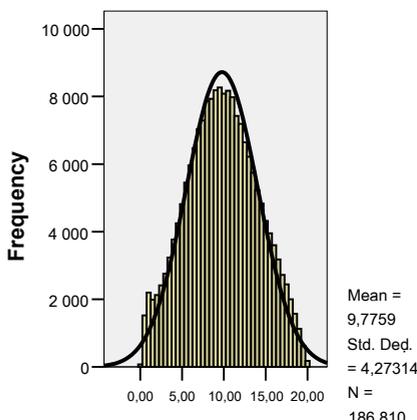
A l'exception de l'arabe et, à degré moindre, la physique (voir les graphes ci-dessous) dont les distributions des notes semblent s'approcher a priori d'une distribution normale, toutes les autres matières montrent des distributions asymétriques (voir le graphe des mathématiques ci-dessous) . Cependant, même pour l'arabe et la physique, les tests de Kolmogorov-Smirnov et de Shapiro-Wilk montrent bien que l'hypothèse de la normalité ne peut guère être retenue.⁽⁴⁾



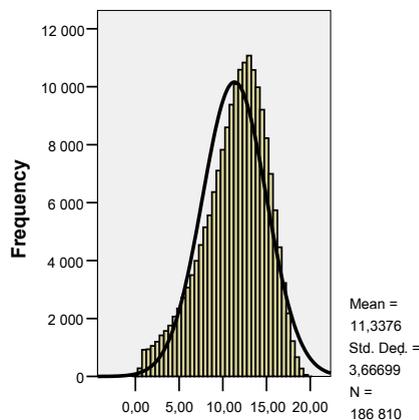
arab Bac S



math Bac S



phys Bac S



snv Bac S

(4) Voir le tableau des tests de Kolmogorov-Smirnov et de Shapiro-Wilk des notes du bac S dans l'annexe n°1. Ces tests sont basés sur un échantillon aléatoire de taille égale à 1% de la taille de la population.

Baccalauréat Lettres:

Les moyennes et les écart-types de l'ensemble des épreuves du bac série L sont résumés dans le tableau 3. Nous remarquons que les notes moyennes varient d'un minimum de 4.27 pour la matière MATH à un maximum de 11.00 pour ISL. En conséquence, les évaluateurs des différentes matières du bac semblent avoir, là aussi, des appréciations différentes quant au niveau moyen des candidats. Cette dernière constatation ne concerne pas les deux matières principales de la spécialité: les notes moyennes des matières ARAB et PHIL, respectivement de 10.11 et de 10.14, sont presque identiques. Pour ce qui est des dispersions, nous observons que, mises à part les matières ISL avec un écart-type élevé (4.47), et ENG, avec un écart-type réduit (2.79), les écart-types des autres matières sont semblables et varient de 3.04 à 3.46. Les *boîtes en moustache* de la figure 2, basées sur un échantillon aléatoire de 1%, résumant graphiquement les caractéristiques essentielles des distributions de notes des différentes matières de la série L.

Tableau 3: Statistiques des notes des matières du bac série L (N=181937)

Matière	Abrév	Min.	Max.	Moyenne	Ecart-type	Coef. Asymétrie
Langue arabe	ARAB	0	19	10,11	3,04	-0,351
Philosophie	PHIL	0	19	10,14	3,46	-0,415
Histoire-Géographie	HG	0	19,5	9,61	3,46	-0,266
Français	FR	0	19,5	7,25	3,10	0,472
Anglais	ENG	0	19,5	6,24	2,79	0,864
Mathématiques	MATH	0	20	4,27	3,13	1,059
Education Islamique	ISL	0	20	11,00	4,47	-0,227

Quelques matières (voir par exemple le graphe de l'arabe ci-dessous) montrent des distributions empiriques de notes qui s'approchent visiblement de la distribution normale. D'autres matières montrent, à l'instar du cas extrême de MATH (voir le graphe des mathématiques ci-dessous) des distributions sévèrement asymétriques. Toutefois, les tests de Kolmogorov-

Smirnov et de Shapiro-Wilk montrent bien que l'hypothèse de la normalité est rejeté pour tous les cas.⁽⁵⁾

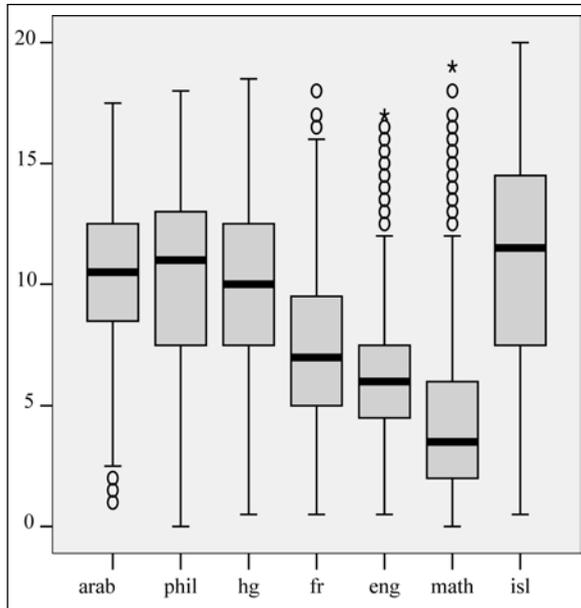
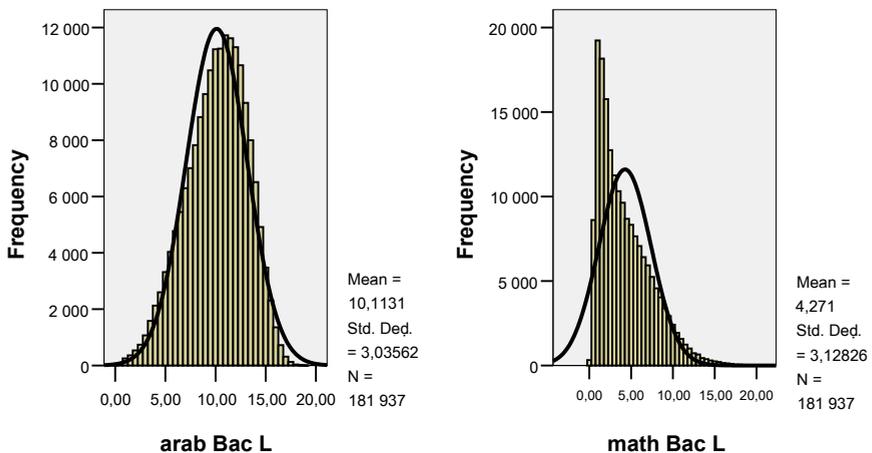
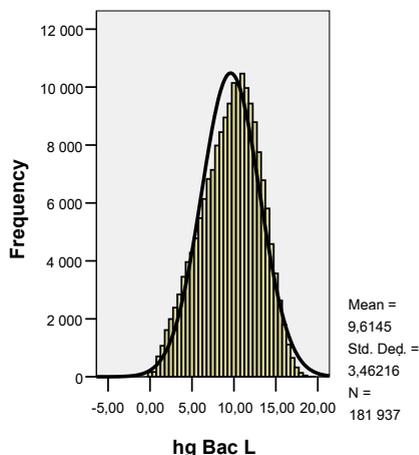
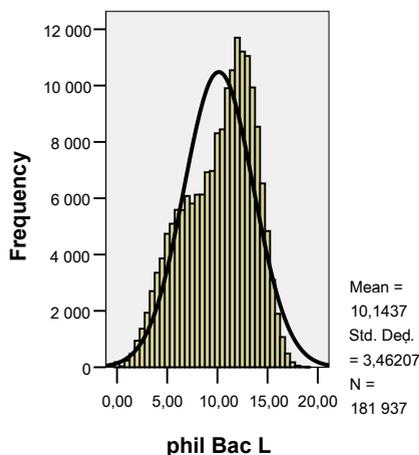


Figure 2: Distributions des notes des matières Bac L (Echantillon aléatoire 1%)



(5) Voir Tableau des tests de Kolmogorov-Smirnov et de Shapiro-Wilk des notes du bac L dans l'annexe n°1.



Baccalauréat Gestion-Economie:

Le tableau 4 montre les principales statistiques de l'ensemble des matières du bac G.

Tableau 4: Statistiques des notes des matières du bac série G (N=56536)

Matière	Abrév	Min	Max	Moyenne	Ecart-type	Coef.Asymétrie
Comptabilité-Finance	COMP FIN	0	20	9,65	5,02	-0,02
Economie-Gestion	ECO GES	0	20	13,77	4,70	-0,83
Mathématiques	MATH	0	19,5	6,78	3,60	0,38
Histoire-Géographie	HG	0	18,5	10,18	2,71	-0,32
Langue arabe	ARAB	0	18	7,98	2,48	0,22
Droit	DROIT	0	20	14,06	4,90	-0,87
Français	FR	0	19	7,34	3,28	0,58
Anglais	ENG	0	19,5	6,60	2,86	0,73
Philosophie	PHIL	0	17,5	5,41	2,78	1,15
Education Islamique	ISL	0	20	11,17	4,16	-0,23

Nous constatons que les notes moyennes varient d'un minimum de 5.41 pour la philosophie à un maximum de 14.03 pour le droit. Bien que pour les matières de spécialité, les notes moyennes sont au-dessus de 9.65, les autres matières, mise à part ISL, montrent des tendances centrales au-dessous de 7.98. Ainsi, les évaluateurs des matières de spécialité semblent être en désaccord avec les autres évaluateurs quant au niveau moyen des candidats. Pour ce qui est des dispersions, nous observons pour les principales matières de spécialité des écarts types au-dessus de 4.7. Cependant, pour les langues les écarts-type sont plus réduits avec un minimum de 2.48 pour l'Arabe. Les *boîtes en moustache*, (voir figure 3), résument graphiquement les caractéristiques essentielles des distributions de notes des matières de la série G.

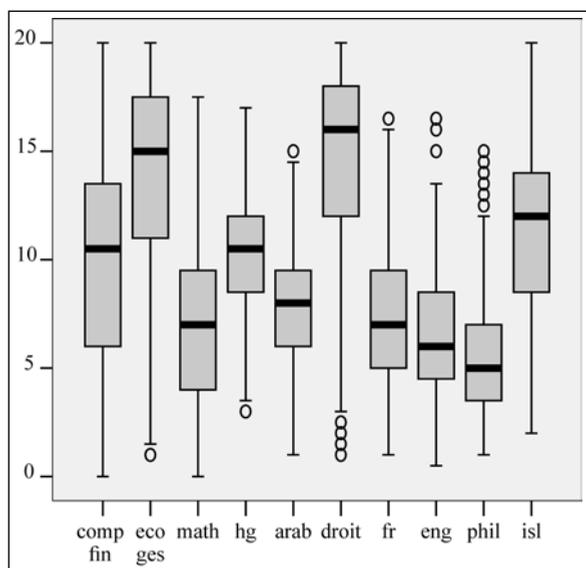
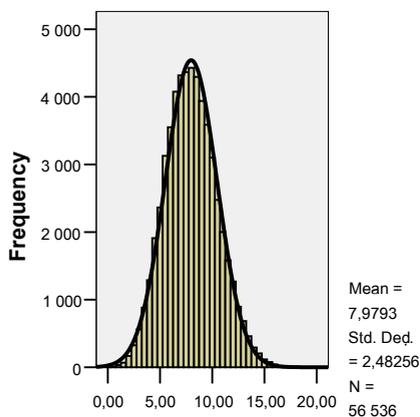


Figure 3: Distributions des notes des matières Bac Lettres (Echantillon aléatoire 1%)

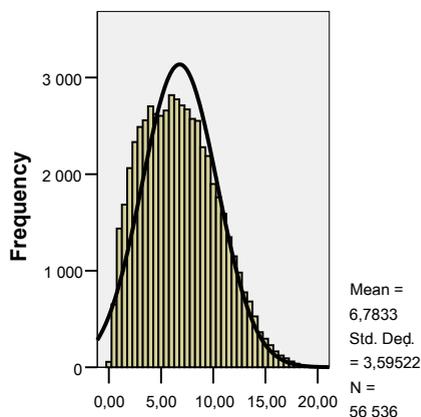
A l'exception de l'arabe, encore une fois, (voir le graphe ci-dessous) dont la distribution des notes semble s'approcher de la distribution normale, les autres matières, surtout de spécialité, montrent des distributions sévèrement asymétriques (voir les graphes de GESECO et de MATH ci-dessous).⁽⁶⁾ Les tests de Kolmogorov-Smirnov et de Shapiro-Wilk montrent ici, comme pour

(6) Pourquoi, quel que soit la série du bac, la distribution des notes de l'arabe, par rapport aux autres matières, est-elle plus proche de la distribution normale? Cette question mérite une investigation.

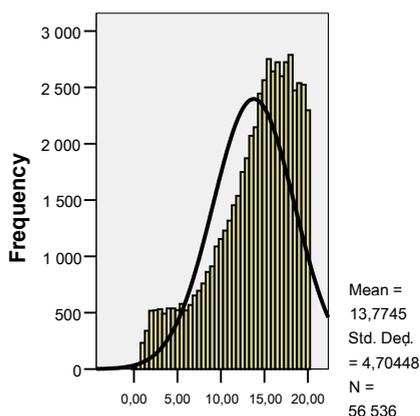
les deux cas précédents, que l'hypothèse de la normalité est rejetée pour l'ensemble des matières.⁽⁷⁾



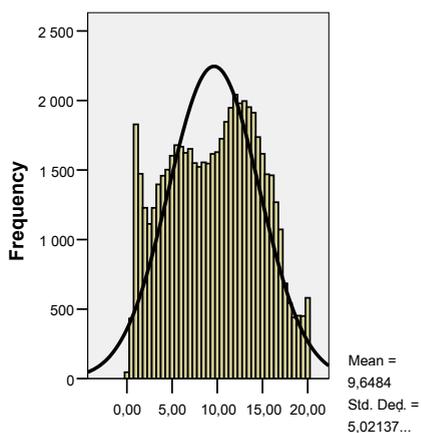
arab Bac G



math Bac G



ecoges Bac G



compfin Bac G

2. Dimensions latentes dans les notes du baccalauréat

Pour identifier les facteurs latents (les dimensions) des notes des différentes matières pour les trois séries du bac (S, L et G), nous utilisons l'AFE. Cette phase d'analyse exploratoire est un préalable à l'étude de la fiabilité

(7) Voir tableau des tests de Kolmogorov-Smirnov et de Shapiro-Wilk des notes du bac G dans l'annexe n°1.

(Cotton et al.,1957). L'analyse factorielle est intimement liée à la notion de validité (Nunnally, (1978), Gorsuch(1983)). Le but principal de l'AFE est l'identification d'un ensemble restreint de construits latents sous-jacents à un ensemble plus important de variables observées. L'objectif recherché, selon Fabrigar et al. (1999), est *«d'arriver à une compréhension conceptuelle plus parcimonieuse d'un ensemble de variables en déterminant le nombre et la nature des facteurs communs nécessaires pour tenir compte de la structure des corrélations entre les variables mesurées»*.

Parmi les questions méthodologiques à considérer, lors de la conduite d'une AFE, les plus citées sont la pertinence des données pour un modèle factorielle, le choix de la méthode d'extraction, le nombre de facteurs à retenir, le procédé de rotation et l'interprétation de la solution factorielle (Ford et al.(1986), Finch & West (1997), Fabrigar et al. (1999), Henson & Roberts (2006)). Les décisions prises concernant ces questions peuvent avoir un impact considérable sur les résultats de l'analyse et sur l'interprétation de ces résultats (Ford et al. (1986), Finch & West (1997)).

Avant l'extraction des facteurs, il est recommandé de vérifier si l'AFE est une approche appropriée.⁽⁸⁾ L'adéquation de la résolution factorielle peut être vérifiée à travers la mesure de Kaiser-Meyer-Olkin (KMO) (Kaiser, 1970) ainsi qu'à travers le test de sphéricité de Bartlett (Bartlett, 1950).⁽⁹⁾ Pour ce qui est de l'extraction, bien que la méthode d'extraction de l'Analyse en Composantes Principales (ACP) soit la plus utilisée (Ford et al. (1986), Pohlmann (2004), Henson & Roberts (2006)), les «puristes» de l'AFE recommandent une méthode d'extraction factorielle plutôt que l'ACP (Ford et al. (1986), Costello & Osborne (2005)). L'ACP n'est pas un vrai modèle factoriel (Finch & West (1997)). Il est important de reconnaître que le but d'identifier des construits latents est différent d'un simple objectif de réduction des dimensions (Fabrigar et al. (1999)). Alors qu'un vrai modèle

(8) Lorsque toutes les corrélations sont inférieures à 0.30, il est alors recommandé de reconsidérer si l'AFE est une approche appropriée pour la question de recherche envisagée (Tabachnick & Fidell (2001)).

(9) La mesure de KMO, représentant le montant de la variance commune entre les variables, est un indice d'adéquation de la solution factorielle. Cette mesure, qui varie entre 0 et 1, indique la pertinence de l'analyse factorielle lorsqu'elle dépasse un certain seuil (souvent la valeur de 0.6 est suggérée). Quant au test de sphéricité de Bartlett, il montre que l'hypothèse d'une matrice de corrélation d'identité est rejetée lorsque le niveau de signification est adéquat (généralement le niveau $\alpha < 5\%$ est exigé).

factoriel distingue explicitement deux sources de variabilité, les facteurs communs qui influencent plus qu'une variable et les facteurs spécifiques qui n'influencent qu'une seule variable, l'ACP ne fait pas cette distinction.⁽¹⁰⁾

Pour obtenir des meilleurs résultats, Fabrigar et al. (1999) et Costello & Osborne (2005) recommandent la méthode d'extraction de maximum de vraisemblance lorsque les données sont issues d'une distribution approximativement normale ou la méthode de factorisation en axe principal lorsque l'hypothèse de normalité n'est pas appropriée. D'autres auteurs (Arrindell & Van der Ende (1985), Velicer & Jackson (1990)) soulignent qu'en pratique, souvent, il n'y a presque pas de différence entre les méthodes d'extractions factorielles et de l'ACP. La similarité des résultats des deux méthodes dépend du nombre et de la fiabilité des variables analysées (Henson & Roberts (2001)).

Parmi les critères de détermination du nombre de facteurs, celui de Kaiser-Guttman est le plus utilisé (Fabrigar et al. (1999), Henson & Roberts (2006), Pohlmann (2004)).⁽¹¹⁾ De même, le critère de Cattell, en dépit de sa nature subjective, car basée sur l'interprétation d'un graphique de valeurs propres, nommé scree plot, demeure fréquemment utilisé (Ford et al. (1986), Fabrigar et al. (1999), Henson & Roberts (2006), Pohlmann (2004)).⁽¹²⁾

Afin de faciliter l'interprétation de la solution factorielle, une rotation des axes est souvent envisagée. Parmi une myriade de procédés de rotation, la rotation Varimax de Kaiser (1958) demeure la plus utilisée (Fabrigar et al. (1999), Pohlmann (2004), Henson & Roberts (2006)). Selon Fabrigar et al. (1999), cette rotation est la plus adéquate parmi les rotations orthogonales.

L'interprétation de la solution factorielle est évidemment importante. Toutefois, elle est assez subjective et dépend des choix méthodologiques précédemment discutés. Pour réduire la subjectivité, une règle couramment utilisée précise que seules les variables avec des chargements supérieurs à

(10) L'ACP analyse la matrice des corrélations tandis que l'AFE analyse une matrice qui remplace chaque unité dans la diagonale principale de la matrice des corrélations par une estimation de la part de la variabilité due aux facteurs communs.

(11) La règle de Kaiser-Guttman consiste à retenir tous les facteurs dont la valeur propre est supérieure à 1,

(12) La règle de Cattell consiste à conserver le nombre de facteurs correspondants aux points situés avant un point d'inflexion du graphique détecté par inspection visuelle.

0,4 sur un facteur devraient être considérées comme importantes et utilisées. Cependant, pour interpréter un facteur, en plus de l'examen des chargements élevés, l'examen des chargements faibles ainsi que les signes de l'ensemble des variables est très informatif (Ford et al. (1986).

Pour les trois séries de bac étudiées, l'adéquation de la résolution factorielle est vérifiée à travers la mesure de KMO qui prend des valeurs très satisfaisantes (0.889 pour le bac S, 0.841 pour le bac L et 0.848 pour le bac G) ainsi qu'à travers le test de sphéricité de Bartlett qui montre un niveau de signification $\alpha < 10^{-3}$ pour les trois cas.

La méthode d'extraction de factorisation en axe principal met en relief, selon la règle de Kaiser-Guttman, deux facteurs principaux pour les trois séries. Sans rotation, ces deux facteurs sont la seule matière «Français», d'un côté, et toutes les autres matières, de l'autre.⁽¹³⁾ Ce dernier facteur, représentant une moyenne, peut être identifié à l'*habilité générale*. De même, avec rotation varimax, l'un des deux facteurs peut être associé à l'*habilité générale* alors que l'autre facteur concerne la dimension *langues étrangères* (FR et ANG).⁽¹⁴⁾ La part de la variabilité expliquée par les deux facteurs est de 57.9% pour le bac S, de 50% pour le bac L et de 43.6% pour le bac G.

Les tableaux 5, 6 et 7 montrent les chargements (*loadings*) des différentes variables sur les deux facteurs identifiés ainsi que les communalités (*communalities*) de ces variables. Dans l'ensemble, les items sont assez bien pris en compte. Cependant pour les cas particuliers des variables à faibles communalités (PHIL bac S, PHIL bac G et MATH bac L), la qualité de la représentation n'est pas suffisante. Ainsi, l'exploration d'un facteur additionnel est suggérée (Costello & Osborne (2005)) d'autant plus que le graphique de Cattell pourrait également indiquer un choix de trois facteurs.⁽¹⁵⁾

(13) Il est à noter que les coefficients de corrélation de la matière FR avec les autres matières, mise à part avec la matière ENG, sont tous relativement faibles. Ils sont inférieurs à 0.38 pour bac S, à 0.3 pour bac L et à 0.2 pour bac G (voie annexe n° 2). A l'école algérienne, le «français» ne semble pas être une matière comme les autres. La matière «FR» semblerait être une mesure du degré d'intégration à une classe sociale plutôt qu'une simple mesure additionnelle d'habilité générale.

(14) Il est intéressant d'expliquer plus ce construit. Pourquoi, pour le bac en Algérie, les langues étrangères forment systématiquement une dimension à part contrairement au cas du bac français? La question mérite une investigation approfondie.

(15) Voir les *scree plots* pour les trois séries en annexe n° 3.

Tableau 5: Rotated Factor Matrix (Varimax)

BAC S	Factor 1	Factor 2	Communalities
SNV	0,80	0,28	0,710
PHYS	0,77	0,32	0,690
ISL	0,71	0,13	0,523
MATH	0,70	0,33	0,607
HG	0,70	0,20	0,528
ARAB	0,64	0,32	0,514
PHIL	0,56	0,14	0,330
ENG	0,33	0,78	0,713
FR	0,17	0,75	0,596

Tableau 6: Rotated Factor Matrix(Varimax)

BAC L	Factor 1	Factor 2	Communalities
ISL	0,773	0,136	0,616
HG	0,731	0,248	0,596
ARAB	0,671	0,227	0,501
PHIL	0,650	0,181	0,456
MATH	0,440	0,203	0,234
FR	0,145	0,727	0,550
ENG	0,296	0,679	0,548

Tableau 7: Rotated Factor Matrix(Varimax)

BAC G	Factor 1	Factor 2	Communalities
DROIT	0,765	0,038	0,586
ECOGES	0,725	0,103	0,536

ISL	0,724	0,049	0,527
COMPFIN	0,643	0,223	0,463
HG	0,601	0,181	0,393
MATH	0,566	0,232	0,374
ARAB	0,535	0,206	0,328
PHIL	0,373	0,070	0,144
ENG	0,207	0,708	0,544
FR	0,069	0,678	0,464

Pour le baccalauréat série S, la méthode d'extraction de factorisation en axe principal met en relief trois facteurs principaux qui expliquent 63.45% de la variabilité totale.⁽¹⁶⁾ La méthode de rotation «varimax» permet d'identifier la nature de ces trois facteurs. Le premier facteur regroupe les matières ARAB, PHIL, HG, ISL et SNV. Il peut être ainsi associé à l'habilité «*verbale*». Le second peut être dénommé facteur d'habilité «*quantitative*»; il englobe les matières MATH et PHYS.⁽¹⁷⁾ Enfin, le troisième facteur concerne essentiellement la dimension «*langues étrangères*» (FR et ANG). Le tableau 8 montre les chargements (loadings) des différentes matières sur les trois facteurs identifiés. Par comparaison aux résultats mis en exergue dans les graphes 6a à 6d des quatre options du bac français série sciences, le graphe 3 ci-après montre une répartition similaire des matières du bac algérien série sciences, à l'exception des deux langues étrangères qui définissent à elles seules ici une troisième dimension.⁽¹⁸⁾ Bien qu'un modèle à trois facteurs semble améliorer la qualité de la représentation par rapport à un modèle à deux facteurs, la variable SNV charge à plus de 0.55 sur deux axes et présente un cas sévère d'item à «chargement croisé» (Costello & Osborne (2005)).

(16) La méthode d'extraction ACP a mis en relief les mêmes trois facteurs principaux et expliquent 75% de la variabilité totale (voir l'annexe n° 3)

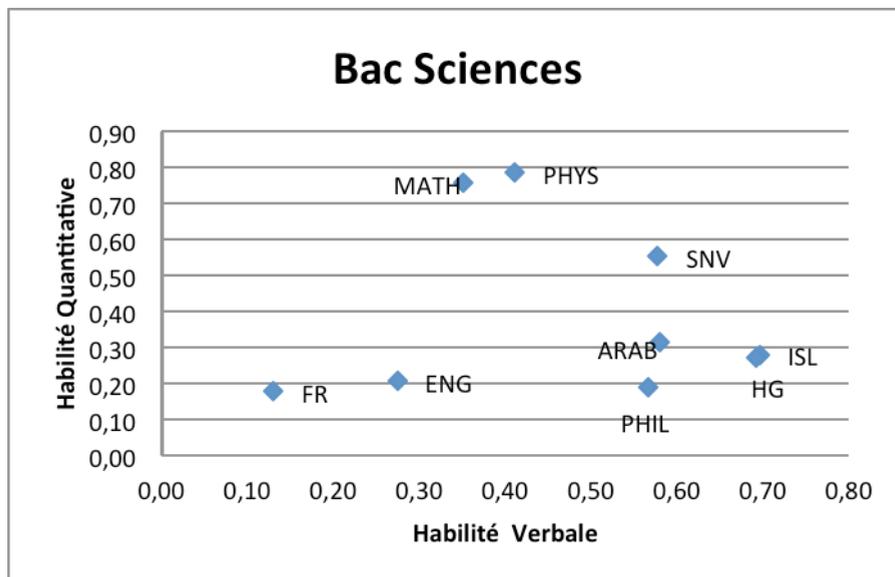
(17) Pour la méthode d'extraction ACP, le deuxième facteur regroupe en plus des matières MATH et PHYS la matière SNV. Ainsi ce facteur pourrait être dénommé facteur d'habilité «scientifique» au lieu de «quantitative» (voir l'annexe n° 3).

(18) Voir Perelmuter et Tomasini,(2005).

Tableau 8: Rotated Factor Matrix(Varimax)

BAC S	Factor1	Factor2	Factor3	Communalities
ISL	0,70	0,28	0,13	0,578
HG	0,69	0,27	0,20	0,589
ARAB	0,58	0,31	0,31	0,529
SNV	0,58	0,55	0,24	0,696
PHIL	0,57	0,19	0,14	0,377
PHYS	0,41	0,79	0,25	0,847
MATH	0,35	0,76	0,27	0,766
ENG	0,27	0,21	0,82	0,789
FR	0,13	0,18	0,70	0,539

Les deux cartes factorielles (figures 4 et 5) sont exposées en deux dimension pour permettre l'appréciation visuelle de la solution factorielle pour la série S.

**Figure 4:** Carte factorielle Bac S

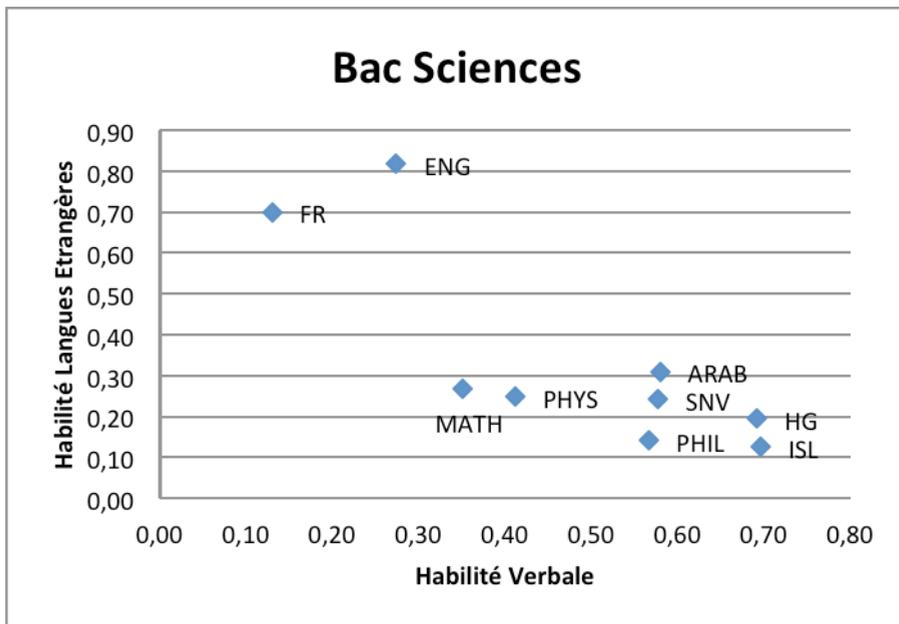


Figure 5: Carte factorielle Bac S

En ce qui concerne le bac L, bien que la méthode d'extraction de l'ACP permet de détecter les mêmes trois facteurs de la série S expliquant 73.8% de la variabilité totale (voir les résultats de l'APC en annexe n° 3), pour la méthode de factorisation en axe principal, il n'est pas adéquat d'extraire trois facteurs. La résolution factorielle précédente à 2 facteurs demeure plus appropriée. La figure 6 montre la carte factorielle correspondante pour la série L.

Comme l'ACP ne distingue pas les sources, communes et spécifiques, de la variabilité d'un item, le construit «habileté quantitative» est représentée par le seul item, MATH, qui forme la dimension. Cependant, dans le modèle factoriel, la variabilité d'un item unique qui n'est pas issue des deux premiers facteurs communs est considérée comme variabilité spécifique. La dimension «quantitative» est ainsi évacuée. Généralement, un facteur à moins de trois items est faible et instable. Toutefois, la solidité d'un facteur peut être assurée même avec un nombre réduit d'items lorsque l'échantillon est très grand (Costello & Osborne (2005)). Bien qu'elle soit très satisfaisante pour le cas du modèle ACP (loadings=0.99), la qualité de la représentation de la variable

MATH n'est pas du tout satisfaisante pour les deux cas des modèles à deux et à trois facteurs (loadings=0.23).

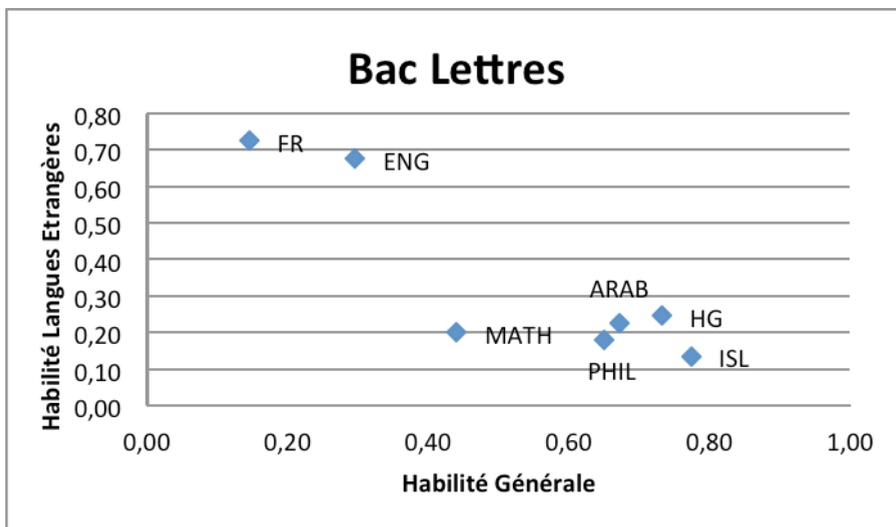


Figure 6: Carte factorielle Bac L

Pour récupérer la dimension quantitative perdue dans les modèles à facteurs, nous remplaçons la variable MATH (moyenne des deux corrections) par les deux corrections de la matière (MATH1 et MATH2). Etant donnée la taille importante de l'échantillon étudié, deux items doivent, en principe, suffire pour rétablir la dimension. En effet, la méthode d'extraction de factorisation en axe principal, permet maintenant d'identifier trois facteurs et la rotation «varimax» permet de révéler leur nature. Le premier regroupe les matières ARAB, PHIL, HG et ISL. Il peut être ainsi associé à l'habilité «verbale». Le second concerne essentiellement la dimension «langues étrangères». Alors que le troisième facteur récupère la dimension perdue de l'habilité «quantitative». Le tableau suivant montre les chargements (loadings) des différentes matières sur les trois facteurs identifiés pour le bac L.

TABLEAU 9: Rotated Factor Matrix(Varimax)

BAC L	Factor1	Factor2	Factor3	Communalities
MOYISL	0,76	0,17	0,12	0,63
MOYHG	0,71	0,20	0,23	0,59
MOYARAB	0,66	0,17	0,21	0,50
MOYPHIL	0,64	0,15	0,17	0,46
MATH2	0,24	0,93	0,13	0,93
MATH1	0,24	0,93	0,13	0,93
MOYFR	0,14	0,09	0,72	0,55
MOYENG	0,29	0,12	0,67	0,55

Comparant le bac lettres en Algérie à l'option (LV1) la plus proche du bac français, série Lettres, le graphe ci-après montre une répartition similaire, à l'exception, encore une fois, des deux langues étrangères qui définissent à elles seules ici aussi une dimension à part (voir le graphe précédent).⁽¹⁹⁾

La figure suivante montre la carte factorielle pour les dimensions *verbale* et de *langues étrangères* du bac L.

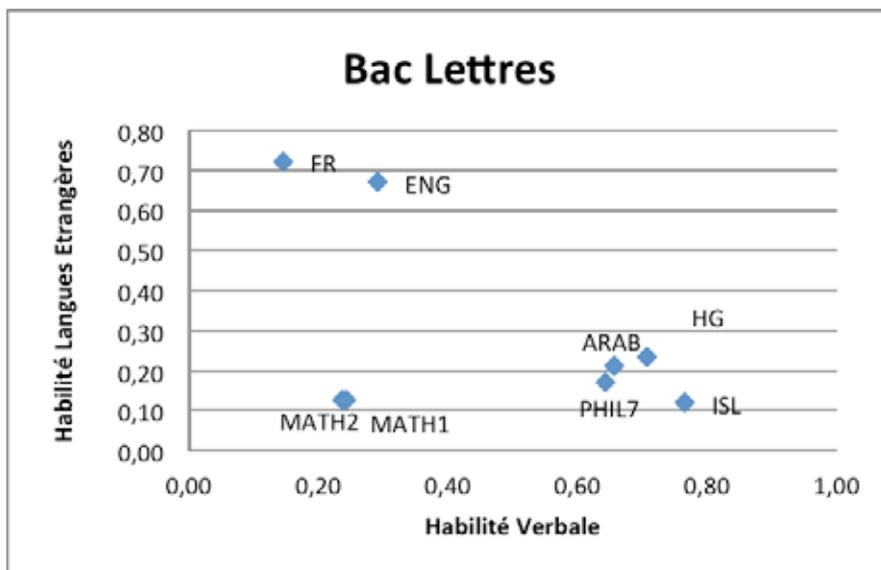


Figure 7: Carte factorielle Bac L

(19) Voir Perelmuter et Tomasini,(2005)

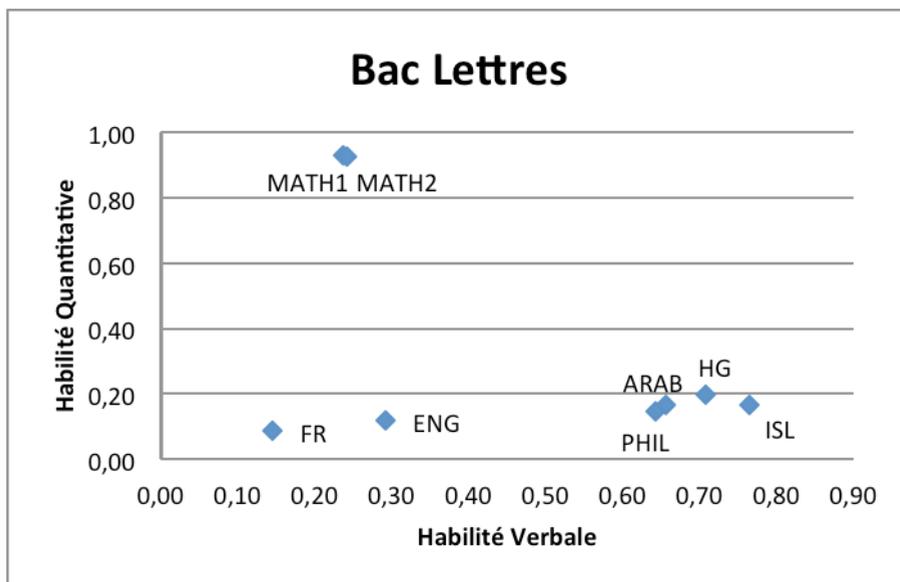


Figure 8: Carte factorielle Bac L

Quant au bac série G, la méthode d'extraction de factorisation en axe principal extrait là aussi trois facteurs qui expliquent 49.7% de la variabilité totale.⁽²⁰⁾ La méthode de rotation «varimax» permet d'observer les mêmes dimensions. Le facteur de l'habilité *verbale* regroupe les mêmes matières que celles communes entre les séries S et L auxquelles il faut ajouter deux matières, DROIT et ECOGES, spécifique au bac G. Le facteur d'habilité *quantitative* comprend en plus des mathématiques (MATH) la matière COMPFIN qui est elle aussi spécifique à ce bac (voir le tableau 10).

(20) La méthode d'extraction ACP a mis en relief les mêmes trois facteurs principaux qui expliquent 63.7% de la variabilité totale (voir l'annexe n° 3)

TABLEAU 10: Rotated Factor Matrix(Varimax)

BAC G	Factor1	Factor2	Factor3	Communalities
ISL	0,71	0,23	0,04	0,57
DROIT	0,70	0,32	0,02	0,59
ECOGES	0,64	0,34	0,08	0,53
HG	0,58	0,22	0,17	0,41
ARAB	0,51	0,20	0,20	0,34
PHIL	0,41	0,05	0,08	0,18
COMPFIN	0,33	0,79	0,14	0,75
MATH	0,29	0,66	0,17	0,55
ENG	0,20	0,09	0,79	0,67
FR	0,05	0,12	0,61	0,39

Il y'a lieu de signaler ici la faible communalité de l'item PHIL (0.18), ceci est due vraisemblablement au fait que cette matière n'est pas suffisamment liée aux autres matières.⁽²¹⁾

En résumé, on note que les dimensions obtenues montrent une certaine stabilité à travers les trois séries de bac (habilité verbale, habilité quantitative et habilité langues étrangères). On note également que les trois facteurs sont en cohérence avec les spécificités des séries du bac.

3. Fiabilité de l'examen du baccalauréat.

Dans ce qui suit l'étude de fiabilité est menée à travers l'analyse de corrélation intra classe (ACIC) d'abord pour chacune des trois construits sous-jacents à l'ensemble des matières de l'examen du bac et ensuite pour la double correction de chaque matière de l'examen. Contrairement au coefficient de corrélation de Pearson qui mesure la relation linéaire entre deux variables qui peuvent ne

(21) Tous les coefficients de corrélation entre PHIL et les autres matières ne dépassent pas 0.3 sauf pour GESECO où le coefficient est de 0.31 (voir annexe n° 2).

partager ni la métrique ni la variance, les coefficients de corrélation intra classe (CCIC) mesurent l'homogénéité de plusieurs variables de même métrique et de même variance.⁽²²⁾ Une typologie des CCIC est présentée dans Shrout & Fleiss (1979) et dans McGraw & Wong (1996) selon le modèle statistique supposé et selon l'utilisation potentielle des résultats.⁽²³⁾ En outre, McGraw & Wong (1996) distinguent clairement les mesures de cohérence (*consistency*) (CCIC(C)) des mesures d'accord total (*total agreement*) (CCIC(A)). Pour les mesures de cohérence, les items peuvent ne pas avoir les mêmes moyennes même pour le cas de cohérence parfaite. Cependant, pour l'accord total, les moyennes des items doivent être les mêmes.

Le coefficient alpha de Cronbach (1951) est le plus important et le plus répandu des CCIC (Cortina, 1993). Il mesure le degré de cohérence de plusieurs variables lorsque la moyenne de ces variables est utilisée.⁽²⁴⁾ Il est d'usage de juger une valeur de alpha supérieure à 0.7 comme adéquate.⁽²⁵⁾ Pourtant, la valeur d'alpha dépend du nombre d'items, du nombre de dimensions et de la corrélation moyenne entre les items (Cortina, 1993). Coton, Campbell & Malone (1957) suggèrent que la statistique d'alpha n'est appropriée que si les items ou variables ne définissent qu'un seul facteur commun. Dans le cas d'une seule dimension, alpha est une mesure de la puissance de cette dernière (Cortina, 1993).

a) Fiabilité à travers l'analyse des principaux construits

L'étude de fiabilité concerne, ici, les trois dimensions dégagées pour chaque série de l'examen du bac où la matière de chaque construit représente un item. Comme les distributions des notes des différentes matières peuvent ne pas avoir les mêmes tendances centrales, les mesures de cohérence sont ici adéquates.

(22) Selon McGraw & Wong (1996), la classe est définie par la métrique et par la variance. Ainsi, le coefficient de corrélation de Pearson est une mesure inter classe.

(23) Les différents modèles statistiques d'ANOVA (modèle à 1 facteur, modèle à 2 facteurs à effets aléatoires et modèle à 2 facteurs à effets mixtes) suggèrent des CCIC différents. De même, lorsque l'unité d'analyse est la notation individuelle une mesure individuelle du CCIC est adéquate. Cependant, lorsque l'unité d'analyse est la moyenne de plusieurs notes une mesure moyenne du CCIC est plus pertinente.

(24) Techniquement, les variables sont supposées être générées par un modèle à 2 facteurs à effets aléatoires (Voir McGraw & Wong, 1996).

(25) Bien qu'un seuil de 0.7 soit acceptable pour le cas de CCIC(C), un seuil plus exigeant est évidemment nécessaire pour le cas de CCIC(A).

Bac Sciences

L'analyse de fiabilité des trois dimensions dégagées précédemment pour le bac série science révèle un niveau assez élevé de fiabilité. Pour la dimension «quantitative» (2 items), le coefficient alpha de Cronbach avoisine 0.89 ce qui est largement satisfaisant. Cependant, ce coefficient est autour de 0.85 pour la dimension verbale (5 items) et de 0.79 pour la dimension langues étrangères, ce qui est acceptable étant donné le type d'examen (réponses ouvertes et non à choix multiples) et la nature des matières (matières non quantitatives). Notons que la dimension verbale est mieux exprimée par la matière SNV suivie de près par HG (voir le tableau d'Alpha de Cronbach si la matière est ignorée ci-dessous).

Tableau 11: CCIC (C) Bac S

Construit	Mesure individuelle	Mesure moyenne
QUANTITATIF (2 items)	0,803	0,891
VERBAL (5 items)	0,520	0,844
LANGUES ETRANGERES (2 items)	0,646	0,785

matière	Alpha de Cronbach (facteur verbal) si la matière est ignorée
SNV	0,798
ARAB	0,815
PHIL	0,837
HG	0,803
ISL	0,807

Bac Lettres

L'analyse de fiabilité des deux dimensions, verbale et de langues étrangères, du bac série lettres montre un niveau raisonnable surtout pour la dimension verbale.⁽²⁶⁾ En effet, le coefficient alpha de Cronbach est de 0.82. Quant à la dimension langues étrangères, ce même coefficient est près de 0.7. Nous

(26) La dimension quantitative est représentée par une seule matière.

signalons que les matières qui expriment le plus la dimension verbale sont HG et ISL et celle qui l'exprime le moins est PHIL. (Voir le tableau d'Alpha de Cronbach si la matière est ignorée ci-dessous).

Tableau 12: CCIC(C) Bac L

Construit	Mesure individuelle	Mesure moyenne
QUANTITATIF (1item)		
VERBAL (4 items)	0,525	0,816
LANGUES ETRANGERES (2 items)	0.534	0.696

matière	Alpha de Cronbach (facteur verbal) si la matière est ignorée
ARAB	0.778
PHIL	0.782
HG	0.752
ISL	0.758

Bac Gestion-Economie

Par rapport aux séries S et L, les matières de spécialité pour la série G sont partagées entre les deux dimensions: verbale et quantitative. En ce qui concerne l'analyse de fiabilité des trois dimensions de cette série, nous constatons que le coefficient alpha de Cronbach demeure là aussi relativement important même s'il est légèrement inférieur aux deux séries précédentes. Le construit verbal, avec un coefficient de Cronbach de 0.8, est mieux résumé par la matière DROIT contrairement à la matière PHIL.

Tableau 13: CCIC(C) Bac G

Construit	Mesure individuelle	Mesure moyenne.
QUANTITATIVE (2item)	0.610	0.758
VERBALE (6 items)	0,399	0,800
LANGUES ETRANGERES (2 items)	0.500	0.667

matière	Alpha de Cronbach (facteur verbal) si la matière est ignorée
ECOGES	0.746
HG	0.774
ARAB	0.786
DROIT	0.738
PHIL	0.804
ISL	0.743

b) Fiabilité à travers l'analyse de la double correction

La double correction est censée augmenter la fiabilité des examens et concours standardisés. Il est d'usage, lors des corrections d'une copie, de recourir à une 3^{ème} correction lorsque la différence entre les deux premières corrections excède un certain seuil. Pour le bac algérien, ce seuil dépend de la matière. Par exemple, pour le bac S (2012), le seuil est fixé à 3.5 points pour les matières de spécialité (SNV, PHYS et MATH) alors qu'il est de 4 points pour les autres matières. Les trois tableaux 16, 17 et 18 présentent la fréquence et la fréquence relative de la troisième correction ainsi que la moyenne des différences absolues (MAD) entre les notes des deux premières corrections pour l'ensemble des matières des séries du bac S, L et G respectivement.⁽²⁷⁾ Il ressort de la lecture de ces trois tableaux que pour certaines matières, à l'instar de la matière ENG, la troisième correction est moins fréquente (<1.1% pour ENG) et le MAD est plus faible (<1.1 pour ENG) alors que pour d'autres matières, à l'instar de PHIL, la troisième correction est plus fréquente (>4.8% pour PHIL) et le MAD est plus important (>1.43 pour PHIL).

(27) La fréquence relative dépend évidemment du seuil mais elle dépend également du coefficient de corrélation des notes des deux correcteurs.

Tableau 14: Fréquences de la troisième correction Bac S

Bac S	Freq. 3 ^{ème} note	N	% 3 ^{ème} Note	MAD
SNV	8583	186810	0,046	1,39
PHYS	6953	186810	0,037	1,28
MATH	5921	186810	0,032	1,18
ARAB	5700	186810	0,031	1,48
FR	2511	186798	0,013	1,14
ENG	2036	186810	0,011	1,09
PHIL	13216	186810	0,071	1,82
HG	8266	186810	0,044	1,62
ISL	5818	186810	0,031	1,45

Tableau 15: Fréquences de la troisième correction Bac L

Bac L	Freq. 3 ^{ème} note	N	%3 ^{ème} Note	MAD
ARAB	4790	181937	0,026	1,43
PHIL	10617	181937	0,058	1,74
HG	10417	181937	0,057	1,68
FR	1159	181934	0,006	0,99
ENG	922	181936	0,005	0,83
MATH	1904	181937	0,010	0,80
ISL	5985	181937	0,033	1,46

Tableau 16: Fréquences de la troisième correction Bac G

Bac G	Freq. 3 ^{ème} note	N	% 3 ^{ème} Note	MAD
COMPFIN	1370	56536	0,024	1,06
ECOGES	2230	56536	0,039	1,19
MATH	947	56536	0,017	1,02
HG	2142	56536	0,038	1,54
ARAB	1452	56536	0,026	1,42
DROIT	1100	56536	0,019	1,15
FR	582	56536	0,010	1,04
ENG	432	56536	0,008	0,93
PHIL	2715	56536	0,048	1,57
ISL	2041	56536	0,036	1,50

Pour étudier la fiabilité de la double correction, les mesures d'accord total sont plus pertinentes. Les tableaux 17,18 et 19 présentent les CCIC(A) de la double correction pour l'ensemble des matières des trois séries et montrent des valeurs qui dépassent le seuil de 0.85. Nous pouvons constater que, pour certaines matières, le CCIC(A) (mesure moyenne) est supérieur à 0.95 (4 matières sur 9 du bac S, 4 matières sur 7 du bac L et 5 matières sur 10 du bac G). Ce qui exprime un degré d'accord assez élevé entre les correcteurs. Bien que le processus de la double correction semble améliorer sensiblement la fiabilité de certaines matières à faible CCIC, à l'instar de PHIL qui passe d'une mesure individuelle de 0.743 à une mesure moyenne de 0.852, les degrés d'accord pour ces matières demeurent modestes.⁽²⁸⁾

(28) Pour 3 matières du bac S, une matière du bac L et 3 matières du bac G, les CCIC(A) sont inférieures à 0.9.

Tableau 17: CCIC(A) de la double correction Bac S

BAC S	Mesure individuelle	Mesure moyenne	Différence
MATH	0,939	0,969	0,03
ENG	0,926	0,962	0,036
PHYS	0,922	0,96	0,038
FR	0,92	0,958	0,038
ISL	0,888	0,941	0,053
SNV	0,882	0,937	0,055
HG	0,806	0,892	0,086
ARAB	0,8	0,889	0,089
PHIL	0,743	0,852	0,109

Tableau 18: CCIC(A) de la double correction BAC L

BAC L	Mesure individuelle	Mesure moyenne	Différence
MATH	0,933	0,965	0,032
ENG	0,914	0,955	0,041
ISL	0,911	0,953	0,042
FR	0,91	0,953	0,043
ARAB	0,83	0,907	0,077
HG	0,818	0,9	0,082
PHIL	0,803	0,891	0,088

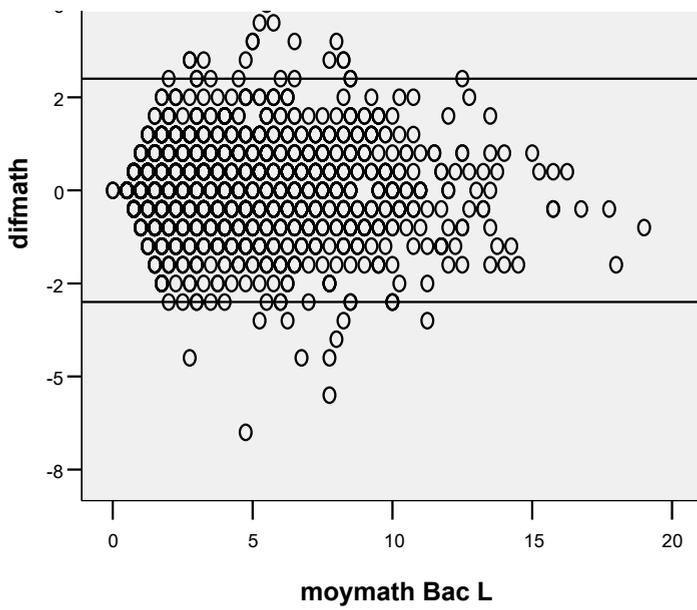
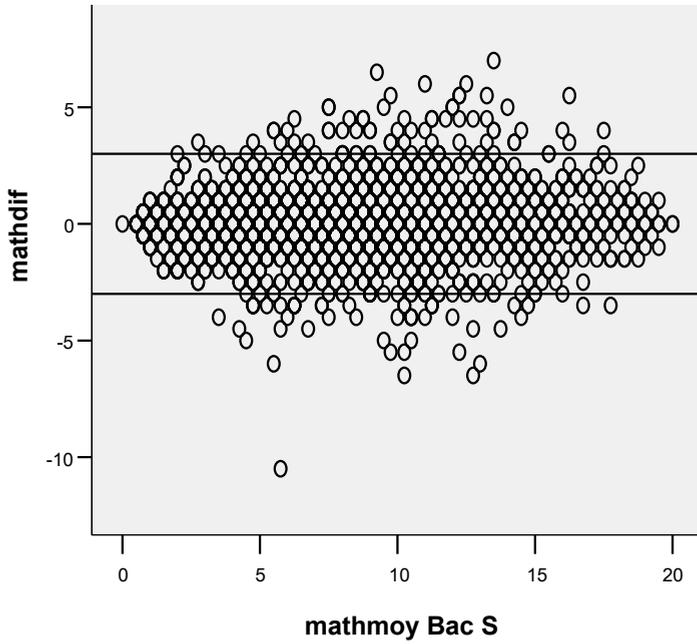
Tableau 19: CCIC(A) de la double correction Bac G

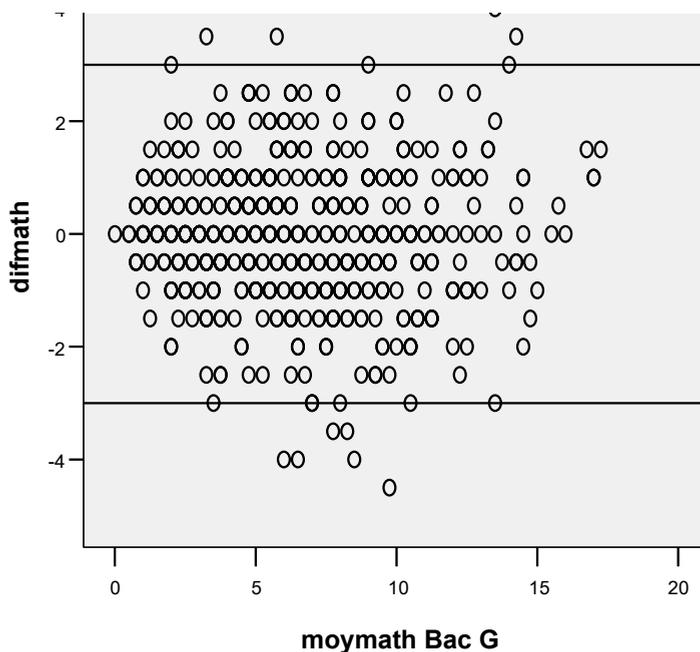
BAC G	Mesure individuelle	Mesure moyenne	Différence
COMPFIN	0,957	0,978	0,021
DROIT	0,948	0,973	0,025
ECOGES	0,938	0,968	0,03
MATH	0,928	0,963	0,035
FR	0,908	0,952	0,044
ENG	0,902	0,948	0,046
ISL	0,893	0,943	0,05
HG	0,76	0,864	0,104
ARAB	0,757	0,862	0,105
PHIL	0,746	0,854	0,108

Présentant des résultats parmi les plus extrêmes, les deux matières MATH et PHIL sont les seules à être discutées et commenter ci-après à travers des graphes appropriés (Cannings et al. (2005)).⁽²⁹⁾ Les mathématiques représentent, pour les trois séries, des CCIC quasi-identiques et très importants (pas moins de 0,96) ce qui exprime un degré d'accord très élevé entre les correcteurs. Même si la double correction améliore légèrement la fiabilité, la correction individuelle offre déjà une fiabilité considérable (Mesure individuelle pas moins de 0,92). Les trois graphes suivants, présentent les différences et les moyennes des notes entre les deux corrections de la matière MATH sur l'axe des ordonnées et l'axe des abscisses respectivement.⁽³⁰⁾ Ces graphes explicitent les conclusions tirées précédemment et mettent en reliefs la fréquence de la troisième correction.

(29) Ces deux matières sont communes entre les trois séries du bac. Les graphes concernant les autres matières figurent en annexe.

(30) Les graphes sont basés sur des échantillons de 1%.



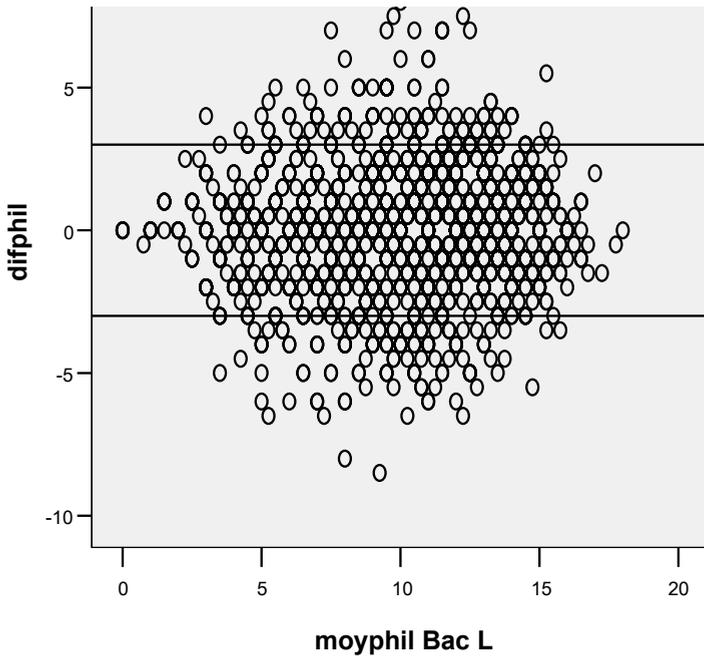
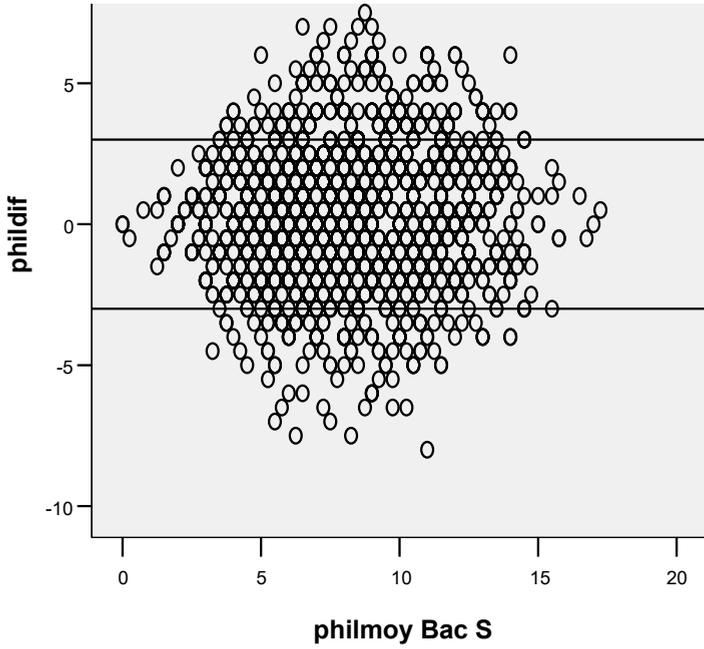


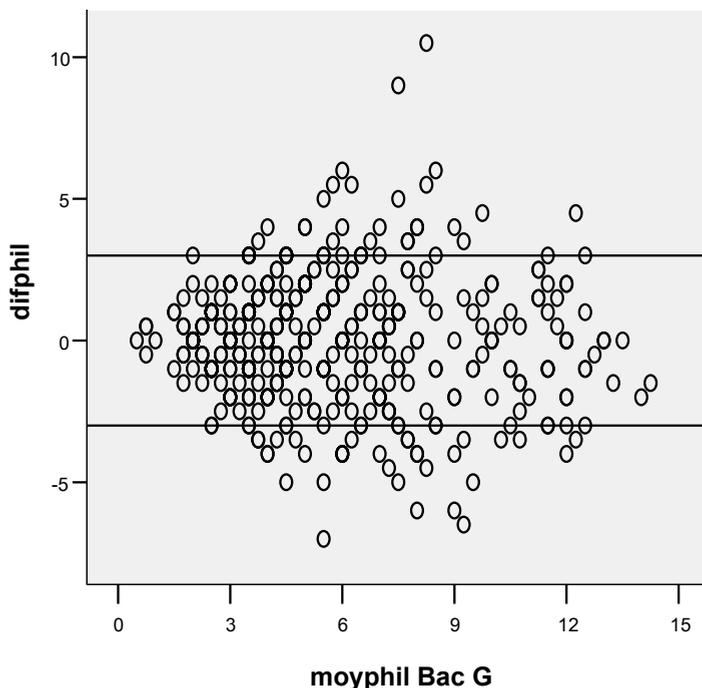
Pour ce qui est de la philosophie, Merle (2007) le dit si bien: “*la compétence philosophique est réputée rebelle à la mesure*». Pour le cas précis des trois séries de bac analysées dans la présente étude, les trois épreuves de PHIL montrent les CCIC les plus faibles par rapport aux autres matières.⁽³¹⁾ Néanmoins, lorsque la moyenne des deux correcteurs est utilisée, au lieu d’une mesure individuelle, les CCIC sont supérieurs à 0.85 pour les trois épreuves et approche 0.9 pour le bac Lettres. Nous pouvons constater que la double correction, pour le cas des épreuves de PHIL, améliore sensiblement la fiabilité.

Les graphes suivants, présentent les différences et les moyennes des notes entre les deux corrections sur l’axe des ordonnées et l’axe des abscisses respectivement.⁽³²⁾ Ces graphes permettent d’apprécier la fréquence de la troisième correction.

(31) Voir l’annexe n° 4 pour comparer avec d’autres matières.

(32) Les graphes sont basés sur des échantillons de 1%.





Conclusion:

L'objectif de cet article est l'étude de la validité et de la fiabilité de l'examen du bac. L'étude est structurée autour de trois axes.

Le premier concerne l'analyse descriptive des notes des trois principales séries du bac. Cette analyse a montré que les évaluateurs des différentes matières (concepteurs, correcteurs et jurys) de l'examen du bac semblent avoir des appréciations distinctes quant au niveau moyen des candidats au baccalauréat. De plus, les notes, à l'exception de l'arabe pour les trois séries et de la physique pour le bac sciences, semblent suivre des distributions différentes et sévèrement asymétriques.

Le second axe concerne l'identification des construits latents (dimensions) des notes des différentes matières. Pour chacune des trois séries étudiées, l'Analyse Factorielle Exploratoire révèle trois construits assez nets. Ces

trois dimensions, les mêmes pour les trois séries, peuvent être associées aux habilités «*verbale*», «*quantitative*» et «*langues étrangères*». La dernière habilité, semble être particulière au bac algérien par rapport au bac français.

Le dernier axe utilise l'Analyse de Corrélation Intra-Classe appliquée aux trois dimensions révélées (*consistency*) et à la double correction (total agreement) pour proposer une mesure de fiabilité. Bien que cette mesure varie d'une dimension à l'autre, elle est dans tous les cas relativement importante même pour les dimension verbale et de langues étrangères où l'alpha de Cronbach est respectivement supérieur à 0.8 et à 0.66 pour les trois séries. L'Analyse de Corrélation Intra-Classe de la double correction montre, pour les différentes matières des trois séries, des degrés d'accord entre les correcteurs qui varient de «très satisfaisant» pour les mathématiques (CCIC > 0.96) à «assez satisfaisant» pour la philosophie (CCIC > 0.85). Ainsi les degrés d'accord les moins importants sont jugés acceptables.

A travers ce qui a été développé dans cette étude, il ressort que l'examen du baccalauréat est suffisamment valide et fiable et peut donc être considéré comme mesure crédible de la performance scolaire. Toutefois, la variété et la nature des distributions des notes ainsi que le processus de la double correction méritent de plus amples investigations.

Bibliographie:

- Arrindell, W. A., & van der Ende, J., (1985), "An Empirical Test of the Utility of the Observations to Variables Ratio in Factor and Components Analysis", *Applied Psychological Measurement*, 9(2), 165-178.
- Bartlett, M. S. (1950), "Tests of significance in factor analysis.", *British Journal of Psychology*, 3(2), 77-85.
- Cattell, R. B. (1966), "The scree test for the number of factors", *Multivariate Behavioral Research*, 1, 245-276.
- Cannings et al, (2005), "Putting double marking to the test: a framework to assess if it is worth the trouble", *Medical Education*; 39, pp 299–308
- Cronbach, L. J., (1951), "Coefficient alpha and the internal structure of tests", *Psychometrika*, Vol.16, 297-334
- Cortina J., (1993), "What Is Coefficient Alpha? An Examination of Theory and Applications", *Journal of Applied Psychology*, Vol. 78, No. 1,98-104.
- Costello & Osborne, (2005), "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis", *Practical Assessment, Research & Evaluation*, Vol. 10, pp. 173-178.
- Cotton, J. W, Campbell, D. T, & Malone, R. D. (1957), «The relationship between factorial composition of test items and measures of test reliability», *Psychometrika*, 22, 347-358.
- Fabrigar, Wegener, MacCallum, & Strahan, (1999), "Evaluating the use of exploratory factor analysis in psychological research", *Psychological Methods*, 4(3), 272-299.
- Finch & West (1997), "The investigation of personality structure: Statistical models", *Journal of Research in Personality*, 31, 439-485.
- Ford, MacCallum & Tait, (1986), "The Application of Exploratory Factor Analysis in Applied Psychology: a Critical Review and Analysis", *Personnel Psychology*, 39(2), 291-314.
- Gorsuch, R., (1983), "Factor Analysis", Hillsdale, NJ, Erlbaum.
- Henson, R. K. & J. K. Roberts (2006), "Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice." *Educational and Psychological Measurement*, 66 (3).

- Kaiser, H. F. (1970), "A Second-Generation Little Jiffy", *Psychometrika*, 35(4), 401-415.
- Kaiser, H. (1958), "The varimax criterion for analytic rotation in factor analysis.", *Psychometrika*, 23, 187-200.
- Laugier & Weinberg, (1936), "Elaboration statistique des données numériques de l'enquête sur la correction des épreuves du baccalauréat", in *La correction des épreuves écrites dans les examens, Paris, Maison du Livre*.
- McGraw, K. & Wong, S. P. (1996), «Forming inferences about some intraclass correlation coefficients», *Psychological Methods*, 1, 30-46
- Merle, (2007), "La notation de la compétence philosophique spécificités de la philosophie et spécificités de l'évaluation", *Côté-Philo* n°14
- Merle, (1998.), « L'évaluation des élèves, enquête sur le jugement professoral», *Revue française de pédagogie, Recherches en psychologie de l'éducation, Vol 122*, pp. 181-183.
- Nunnally, J. C. (1978), «Psychometric theory», New York, McGraw-Hill.
- Perelmuter et Tomasini,(2005), "Étude docimologique du baccalauréat de la session 2003", 172 pp 05-38, MENESR, Direction de l'évaluation et de la prospective.
- Pieron, H. (1963), "Examens et docimologie", Paris: Presses universitaires de France.
- Pohlmann J., (2004), «Use and Interpretation of Factor Analysis in the Journal of Educational Research», *The Journal of Educational Research*, Vol. 98, No. 1, pp. 14-22
- Shrout & Fleiss, (1979), "Intra-class Correlations: Uses in Assessing Rater Reliability", *Psychological Bulletin*, Vol. 86, No. 2, 420-428
- Suchaut, (2008), "La loterie des notes au bac Un réexamen de l'arbitraire de la notation des élèves", *Workingpaper*, IREDU
- Tabachnick & Fidell, (2001), «Using multivariate statistics», Needham Heights, MA, Allyn & Bacon.
- Velicer, W. F. & Jackson, D. N., (1990), "Component Analysis Versus Common Factor-Analysis - Some Further Observations", *Multivariate Behavioral Research*, 25(1), 97-114.
- Williams, B., Brown, T., & Onsmann, A., (2010), "Exploratory factor analysis: A five-step guide for novices.", *Australasian Journal of Paramedicine*, 8(3). _

Annexe 1: Analyse descriptive des notes des matières du baccalauréat

Tests de Kolmogorov-Smirnov et de Shapiro-Wilk des notes du bac Sciences

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
SNV	0,090	1826	0,000	0,970	1826	0,000
PHYS	0,035	1826	0,000	0,991	1826	0,000
MATH	0,064	1826	0,000	0,978	1826	0,000
ARAB	0,043	1826	0,000	0,996	1826	0,000
FR	0,081	1826	0,000	0,979	1826	0,000
ENG	0,070	1826	0,000	0,981	1826	0,000
PHIL	0,074	1826	0,000	0,984	1826	0,000
HG	0,068	1826	0,000	0,989	1826	0,000
ISL	0,123	1826	0,000	0,937	1826	0,000

Tests de Kolmogorov-Smirnov et de Shapiro-Wilk des notes du bac Lettres

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ARAB	0,0635	1786	0,000	0,987	1786	0,000
PHIL	0,1077	1786	0,000	0,967	1786	0,000
HG	0,0588	1786	0,000	0,983	1786	0,000
FR	0,0659	1786	0,000	0,987	1786	0,000
ENG	0,101	1786	0,000	0,957	1786	0,000
MATH	0,1242	1786	0,000	0,908	1786	0,000
ISL	0,0781	1786	0,000	0,972	1786	0,000

Tests de Kolmogorov-Smirnov et de Shapiro-Wilk des notes du bac Gestion et Economie

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
COMPFIN	0,07	529	0,000	0,98	529	0,000
ECOGES	0,12	529	0,000	0,94	529	0,000
MATH	0,07	529	0,000	0,98	529	0,000
HG	0,06	529	0,000	0,99	529	0,003
ARAB	0,06	529	0,000	0,99	529	0,008
DROIT	0,14	529	0,000	0,90	529	0,000
FR	0,09	529	0,000	0,97	529	0,000
ENG	0,09	529	0,000	0,98	529	0,000
PHIL	0,15	529	0,000	0,91	529	0,000
ISL	0,08	529	0,000	0,98	529	0,000

Annexe 2: Coefficients de corrélation entre les matières des séries S, L et G:

BAC S	SNV	PHYS	MATH	ARAB	FR	ENG	PHIL	HG	ISL
SNV	1	0,73	0,68	0,59	0,34	0,47	0,46	0,59	0,60
PHYS	0,73	1	0,81	0,55	0,37	0,48	0,42	0,55	0,54
MATH	0,68	0,81	1	0,53	0,36	0,47	0,39	0,50	0,48
ARAB	0,59	0,55	0,53	1	0,34	0,49	0,42	0,52	0,55
FR	0,34	0,37	0,36	0,34	1	0,65	0,20	0,29	0,22
ENG	0,47	0,48	0,47	0,49	0,65	1	0,31	0,39	0,35
PHIL	0,46	0,42	0,39	0,42	0,20	0,31	1	0,50	0,44
HG	0,59	0,55	0,50	0,52	0,29	0,39	0,50	1	0,58
ISL	0,60	0,54	0,48	0,55	0,22	0,35	0,44	0,58	1

BAC L	ARAB	PHIL	HG	FR	ENG	MATH	ISL			
ARAB	1	0,48	0,53	0,26	0,36	0,34	0,56			
PHIL	0,48	1	0,53	0,23	0,31	0,32	0,52			
HG	0,53	0,53	1	0,29	0,38	0,39	0,60			
FR	0,26	0,23	0,29	1	0,54	0,21	0,20			
ENG	0,36	0,31	0,38	0,54	1	0,27	0,33			
MATH	0,34	0,32	0,39	0,21	0,27	1	0,36			
ISL	0,56	0,52	0,60	0,20	0,33	0,36	1			

BAC G	compfin	ecoges	math	hg	arab	droit	fr	eng	phil	isl
COMPFIN	1	0,50	0,64	0,39	0,35	0,48	0,19	0,25	0,20	0,41
ECOGES	0,50	1	0,40	0,43	0,34	0,64	0,15	0,22	0,31	0,51
MATH	0,64	0,40	1	0,35	0,34	0,40	0,19	0,25	0,15	0,39
HG	0,39	0,43	0,35	1	0,44	0,43	0,16	0,26	0,29	0,48
ARAB	0,35	0,34	0,34	0,44	1	0,38	0,14	0,29	0,23	0,46
DROIT	0,48	0,64	0,40	0,43	0,38	1	0,10	0,19	0,29	0,57
FR	0,19	0,15	0,19	0,16	0,14	0,10	1	0,50	0,07	0,07
ENG	0,25	0,22	0,25	0,26	0,29	0,19	0,50	1	0,15	0,20
PHIL	0,20	0,31	0,15	0,29	0,23	0,29	0,07	0,15	1	0,29
ISL	0,41	0,51	0,39	0,48	0,46	0,57	0,07	0,20	0,29	1

Annexe 3: Analyse factorielle des notes du baccalauréat :
3.1 Méthode d'Extraction: ACP (avec rotation varimax)

Bac sciences

Total Variance Explained

	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,89	54,29	54,29	2,70	30,00	30,00
2	1,12	12,40	66,68	2,27	25,26	55,26

3	0,75	8,34	75,02	1,78	19,76	75,02
4	0,57	6,34	81,36			
5	0,47	5,22	86,58			
6	0,40	4,45	91,03			
7	0,32	3,60	94,63			
8	0,30	3,30	97,92			
9	0,19	2,08	100,00			

Facteurs obtenus avant rotation (Série bac S)

	1	2
SNV	0,84	-0,13
PHYS	0,84	-0,05
MATH	0,81	-0,01
ARAB	0,76	
HG	0,75	-0,23
ISL	0,73	-0,32
ENG	0,68	0,57
PHIL	0,62	-0,29
FR	0,54	0,73

Facteurs obtenus après rotation (Série Bac S)

FACTEURS	1	2	3
MATH	0,86	0,20	0,23
PHYS	0,85	0,27	0,22
SNV	0,74	0,43	0,20
PHIL	0,10	0,84	0,13

HG	0,39	0,70	0,17
ISL	0,46	0,66	0,06
ARAB	0,46	0,54	0,30
FR	0,16	0,08	0,91
ANG	0,28	0,24	0,82

Bac Lettres

	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,33	47,62	47,62	2,57	36,72	36,72
2	1,10	15,70	63,33	1,58	22,54	59,26
3	0,74	10,56	73,88	1,02	14,62	73,88
4	0,53	7,54	81,42			
5	0,46	6,61	88,03			
6	0,45	6,43	94,46			
7	0,39	5,54	100,00			

Extraction Method: Principal Component Analysis.

Facteurs obtenus avant rotation (Série bac lettres)

	1	2	3
HG	0,80	-0,17	-0,08
ISL	0,77	-0,31	-0,14
ARAB	0,76	-0,18	-0,14
PHIL	0,72	-0,24	-0,20
ENG	0,63	0,58	-0,05
FR	0,52	0,73	-0,03
MATH	0,58	-0,10	0,81

Facteurs obtenus après rotation (Série Bac L)

FACTEURS	1	2	3	Communalités
ISL	0,82	0,10	0,15	0,672
PHIL	0,78	0,13	0,07	0,708
HG	0,76	0,22	0,21	0,623
ARAB	0,75	0,20	0,14	0,624
FR	0,10	0,89	0,07	0,742
ANG	0,27	0,81	0,10	0,807
MATH	0,25	0,13	0,96	0,995

Bac Gestion et Economie

	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,05	40,50	40,50	2,45	24,52	24,52
2	1,35	13,46	53,96	2,36	23,59	48,11
3	0,98	9,76	63,72	1,56	15,61	63,72
4	0,78	7,78	71,50			
5	0,69	6,90	78,40			
6	0,55	5,55	83,94			
7	0,47	4,72	88,66			
8	0,45	4,51	93,17			
9	0,35	3,47	96,64			
10	0,34	3,36	100,00			

Extraction Method: Principal Component Analysis.

Facteurs obtenus après rotation (Série Bac G)

DROIT	0,76	-0,26	0,00
ECOGES	0,75	-0,18	0,00
ISL	0,74	-0,25	0,11
COMPFIN	0,73	0,00	-0,45
HG	0,69	-0,06	0,19
MATH	0,67	0,05	-0,52
ARAB	0,64	0,01	0,17
FR	0,32	0,81	0,04
ENG	0,46	0,71	0,16
PHIL	0,44	-0,14	0,64

Facteurs obtenus après rotation (Série Bac G)

FACTEURS	1	2	3
PHIL	0,77	-0,16	0,08
ISL	0,65	0,44	-0,01
HG	0,61	0,33	0,17
DROIT	0,59	0,53	-0,03
ECOGES	0,57	0,53	0,04
ARAB	0,55	0,30	0,21
MATH	0,10	0,83	0,17
COMPFIN	0,20	0,82	0,15
FR	0,00	0,10	0,86
ENG	0,20	0,12	0,83

3.2 Méthode d'Extraction: Principal axis factoring (avec rotation varimax)

Bac sciences

Extraction Sums of Squared Loadings Rotation Sums of Squared Loadings

Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,55	50,52	50,52	2,33	25,93	25,93
2	0,75	8,30	58,83	1,85	20,58	46,51
3	0,42	4,62	63,45	1,52	16,94	63,45
.						
9						

Scree Plot

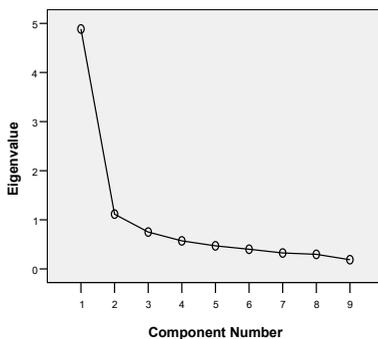


Figure: Bac S

Bac Lettres

Extraction Sums of Squared Loadings Rotation Sums of Squared Loadings

Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,87	40,95	40,95	2,31	33,07	33,07
2	0,65	9,34	50,30	1,21	17,22	50,28
3	0,05	0,75	51,04	0,05	0,76	51,04
.						
7						

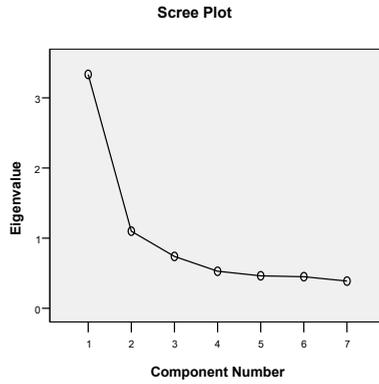


Figure: Bac L

Bac Gestion et Economie

Extraction Sums of Squared Loadings

Rotation Sums of Squared Loadings

Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,58	35,84	35,84	2,40	24,02	24,02
2	0,88	8,79	44,63	1,44	14,42	38,44
3	0,51	5,07	49,70	1,13	11,26	49,70
.						
10						

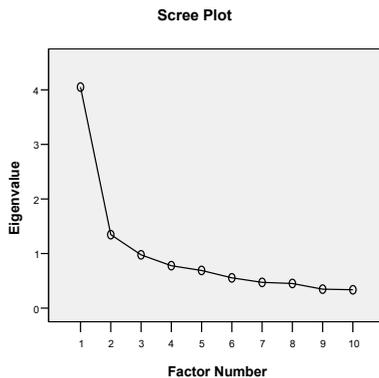
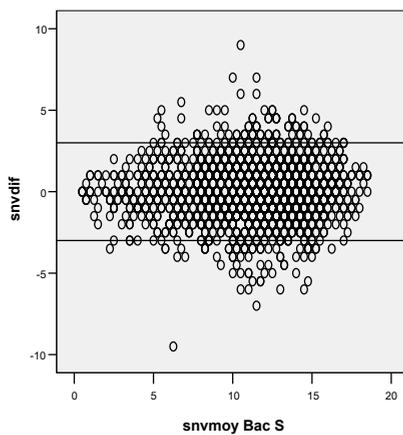
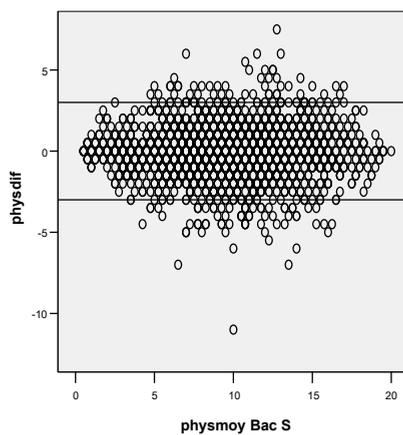
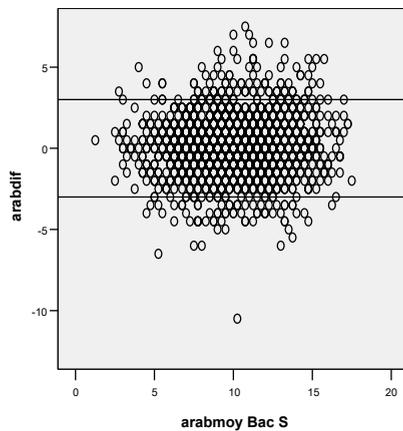
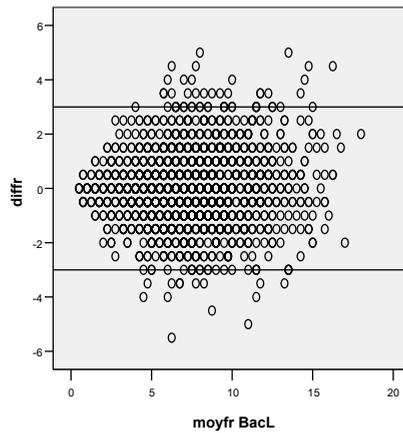
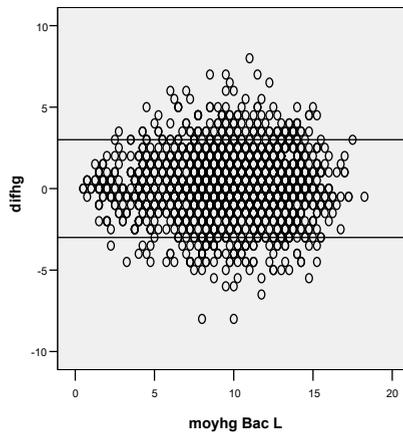
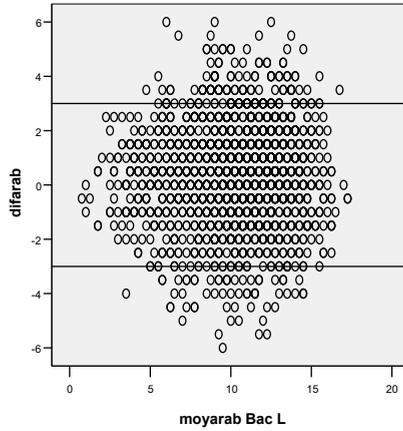


Figure: Bac G

Annexe 4: Analyse de Fiabilité de l'examen du baccalauréat: Double correction**Bac sciences**

Bac Lettres



Bac Gestion et Economie

