





RAPPORT MÉTHODOLOGIQUE
PROGRAMME NATIONAL D'ÉVALUATION DES
ACQUIS DES ÉLÈVES DU TRONC COMMUN
PNEA 2016



TABLE DES MATIÈRES



I. PROGRAMME NATIONAL D'ÉVALUATION DES ACQUIS (PNEA)	5
1. Objectifs des évaluations PNEA	5
2. Niveaux scolaires et domaines cibles	6
3. Programmation et périodicité	6
4. Construction des tests	6
4.1. <i>Élaboration des cadres de référence</i>	7
4.2. <i>Conception des tests</i>	8
4.3. <i>Analyse psychométrique</i>	8
4.3.1. <i>Indice de difficulté d'un item</i>	9
4.3.2. <i>Pouvoir de discrimination d'un item</i>	9
4.3.3. <i>Fidélité du test</i>	9
5. Développement des questionnaires	9
5.1. <i>Déterminants de l'environnement socio-éducatif</i>	10
5.2. <i>Composantes de l'environnement socio-éducatif</i>	10
5.2.1. <i>Le climat scolaire</i>	11
5.2.1.1. <i>Le climat relationnel</i>	11
5.2.1.2. <i>Le climat de sécurité</i>	11
5.2.1.3. <i>Le climat de justice</i>	11
5.2.1.4. <i>Le climat d'appartenance</i>	11
5.2.2. <i>Les pratiques éducatives et de gestion</i>	11
5.2.2.1. <i>La qualité de l'enseignement</i>	11
5.2.2.2. <i>Le système de reconnaissance</i>	11
5.2.2.3. <i>Le système d'encadrement</i>	11
5.2.2.4. <i>L'emphase sur la réussite</i>	12
5.2.2.5. <i>Les opportunités d'investissement</i>	12
5.2.2.6. <i>Le temps consacré à l'enseignement</i>	12
5.2.2.7. <i>Le leadership éducatif</i>	12
5.2.3. <i>Problèmes sociaux et scolaires</i>	12
5.2.3.1. <i>Problèmes sociaux</i>	12
5.2.3.2. <i>Problèmes scolaires</i>	12
5.2.4. <i>Échelle de mesure de l'environnement socio-éducatif</i>	13
5.2.5. <i>Validité des questionnaires</i>	14
5.2.5.1. <i>Validité de contenu</i>	14
5.2.5.2. <i>Validité du construit</i>	14
6. Plan d'échantillonnage	14
7. Exploitation et analyse des résultats	15
7.1. <i>Analyse bi-variée</i>	16
7.2. <i>Analyse multi-niveaux</i>	17
II. PROGRAMME NATIONAL D'ÉVALUATION DES ACQUIS DES ÉLÈVES DES TRONCS COMMUNS DU SECONDAIRE QUALIFIANT (PNEA2016)	19
1. Objectifs de l'évaluation PNEA 2016	19
2. Niveaux scolaires et domaines cibles	19
3. Construction des tests	19
3.1. <i>Test de mathématiques</i>	19
3.2. <i>Test des Sciences de la Vie et la Terre</i>	21
3.3. <i>Test de Physique-Chimie</i>	23
3.4. <i>Test de la langue Arabe</i>	23
3.5. <i>Test de la langue Française</i>	26
3.6. <i>Test de d'Histoire-Géographie</i>	27

4. Développement des questionnaires	28
4.1. Validité de contenu	28
4.2. Validité du construit	28
4.2.1. Analyse factorielle exploratoire	29
4.2.2. Analyse de fidélité	29
4.2.3. Résultats.....	30
4.3. Climat scolaire.....	30
4.4. Pratiques éducatives et de gestion	32
4.5. Problèmes sociaux et scolaires	33
5. Plan d'échantillonnage	35
6. Collecte et exploitation des données	36
6.1. Collecte des données	36
6.2. Exploitation des données	37
6.2.1. La pondération au niveau de l'établissement	37
6.2.2. La pondération au niveau de la classe	37
6.2.3. La pondération au niveau de l'élève	38
7. Passation des tests et administration des questionnaires	38
 ANNEXES	42
 RÉFÉRENCES BIBLIOGRAPHIQUES	45

I. PROGRAMME NATIONAL D'ÉVALUATION DES ACQUIS (PNEA)

Les évaluations-bilans standardisées (autres que les examens certificatifs) sont des outils pour le pilotage du système éducatif. Leur méthodologie de construction s'appuie sur les méthodes de mesure en éducation et les modèles psychométriques. Elles concernent des échantillons représentatifs d'établissements, de classes et d'élèves et sont organisées en fin d'année scolaire. Ces évaluations permettent, entre autres, d'agir sur les curricula notamment les programmes scolaires, l'organisation pédagogique, la formation des enseignants et les pratiques enseignantes.

Parmi les objectifs assignés aux évaluations bilan, on relève l'évaluation des performances de l'ensemble du système éducatif, la comparaison des performances entre établissements, la responsabilisation des acteurs éducatifs notamment les enseignants.

D'ailleurs, les résultats de ces tests sont utilisés, avec d'autres paramètres, comme indicateurs de la qualité de l'enseignement, de l'efficacité globale des politiques et des pratiques éducatives, des performances des enseignants et en tant qu'indicateur de niveau de qualité d'un système éducatif ou d'un établissement scolaire donné.

Ainsi, l'article 72 du levier 12 de la Vision stratégique 2015-2030, recommande :

« La mise en œuvre des curricula, des programmes et des formations doit être soumise systématiquement à une évaluation régulière. Celle-ci devra s'étendre aux plans de réforme, aux acquis scolaires et de formation, aux manuels scolaires et outils didactiques, et aux pratiques pédagogiques et de formation. Elle s'effectue en référence aux normes nationales et internationales ».

Ainsi, et au vu de ses attributions et de ses prérogatives en matière d'évaluation du système d'éducation et de formation, l'Instance Nationale d'Évaluation auprès du Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique (CSEFRS), a mis en place un Programme National d'Évaluation des Acquis (PNEA). Ce programme vise l'instauration d'un référentiel national en évaluation des acquis scolaires, à même d'aider à piloter les réformes pédagogiques et, d'une façon générale, éclairer les prises de décisions relatives aux politiques éducatives. Une première

édition du PNEA a eu lieu en 2008 et a ciblé les élèves des cycles primaire et secondaire collégial.

Le PNEA est un dispositif d'évaluation qui, en plus des instruments de mesure développés (cadres de références, tests, questionnaires, psychométrie, méthodes d'échantillonnage, guides), vise l'évaluation des performances des élèves marocains en termes de connaissances, d'aptitudes et de compétences en vue d'améliorer la qualité des apprentissages et des formations.

Les finalités du PNEA se résument ainsi :

- Le monitoring et l'aide au pilotage du système éducatif dans son ensemble ;
- L'instauration d'une gestion du système éducatif axée sur les résultats ;
- La promotion de la culture de l'évaluation et la reddition des comptes à tous les niveaux ;
- Le développement de la redevabilité des établissements scolaires ;

La responsabilisation des enseignants afin d'améliorer leurs pratiques d'enseignement.

1. Objectifs des évaluations PNEA

Le programme PNEA vise à :

- Évaluer les niveaux linguistique, scientifique et culturel des élèves ;
- Évaluer les compétences des élèves en fin de la scolarité primaire et obligatoire ;
- Cerner les effets «Établissement», «Enseignant » et «Directeur » sur les acquis ;
- Apprécier la qualité du climat scolaire et son impact sur la réussite scolaire ;
- Préciser les liens entre les variables de contexte et le rendement scolaire ;
- Évaluer l'impact des réformes pédagogiques sur la qualité des apprentissages ;
- Mettre à disposition des responsables des données fiables aidant à la prise de décision ;
- Promouvoir la recherche scientifique et pédagogique.

2. Niveaux scolaires et domaines cibles

Le Programme National d'Évaluation des Acquis se focalise sur les niveaux de 6^{ème} année primaire, 3^{ème} année du secondaire collégial ainsi que les tronc communs du secondaire qualifiant et porte sur l'évaluation des acquis en mathématiques, sciences et en langues.

Le choix de ces niveaux scolaires tient compte des considérations suivantes :

- L'évaluation des acquis en fin du primaire est d'un intérêt crucial pour le système éducatif. En effet, au terme de la scolarité primaire, les élèves sont censés maîtriser les compétences de base à savoir l'écriture, la lecture et le calcul et ce, pour pouvoir suivre leur scolarité ultérieure. Il serait donc judicieux de procéder, dans le cadre du PNEA, à des évaluations de ces compétences en fin de l'école primaire.
- L'évaluation des compétences des élèves en fin de scolarité obligatoire s'avère capitale aussi bien pour la continuité des études (ou la formation professionnelle) que pour une intégration réussie dans la vie active et sociale. Mais, il faut, au préalable, construire un socle de compétences que les élèves devraient maîtriser au terme de la scolarité obligatoire. Ce socle doit porter sur les compétences relatives à la maîtrise des langues, l'appropriation des éléments de la culture scientifique et technologique, la maîtrise des TIC, les compétences sociales et civiques ainsi que l'acquisition de l'esprit d'autonomie et d'initiative.

L'acquisition de ce socle de compétences doit faire l'objet d'une évaluation par le PNEA ;

- L'évaluation des acquis des élèves des tronc communs du secondaire qualifiant est intégrée au PNEA car ce niveau scolaire - année d'orientation en amont et en aval – est décisif pour le projet éducatif de chaque lycéen.
- En plus de cette ossature permanente et périodique, d'autres niveaux scolaires ou disciplines peuvent être intégrés au programme sous requêtes des instances décisionnelles du Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique.

3. Programmation et périodicité

Les opérations de chaque évaluation PNEA, de la préparation jusqu'à la publication des résultats, s'étalent sur trois ans.

Première année : Élaboration du dispositif d'évaluation et passation du pré-test ;

Deuxième année : Finalisation du plan d'échantillonnage et passation de l'étude principale ;

Troisième année : Analyse des résultats, rédaction et publication du rapport d'évaluation.

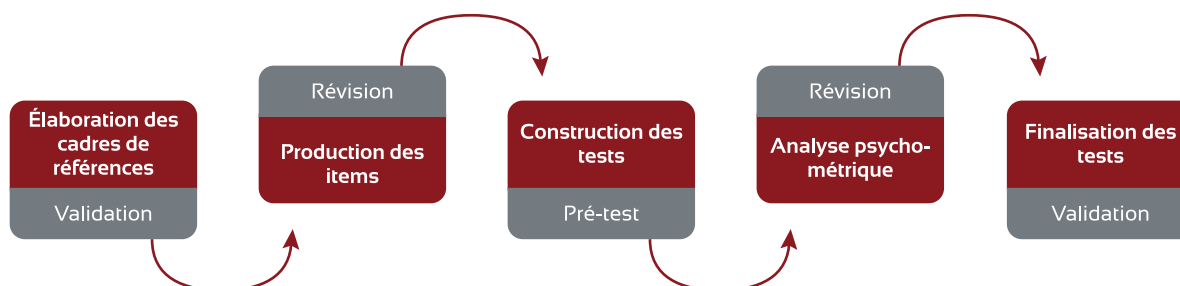
La même étude d'évaluation sera reconduite d'une façon périodique tous les 4 ans et ce à l'instar des études internationales similaires, notamment TIMSS.

Schéma de la périodicité

	2015	2016	2017	2018
PNEA Tronc commun	1 ^{ère} année : - Dispositif méthodologique - Passation des tests	2 ^{ème} année : - Echantillonnage - Passation de l'étude principale	3 ^{ème} année : - Analyse - Publications - Résultats	
PNEA Primaire et Collège		1 ^{ère} année : - Conception des tests	2 ^{ème} année : - Echantillonnage - Passation de l'étude principale	3 ^{ème} année : - Analyse - Résultats

4. Construction des tests

Le processus de construction des tests PNEA se déroule ainsi :



4.1. Élaboration des cadres de référence

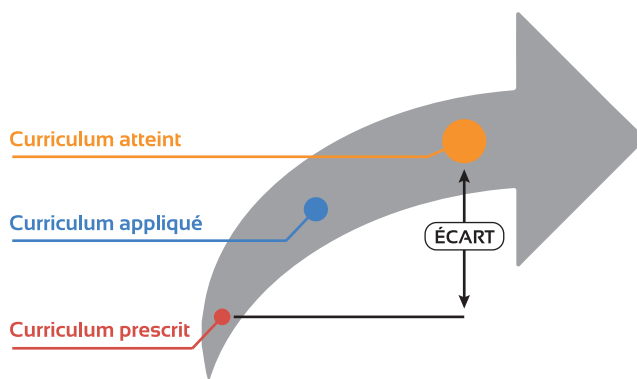
Préalablement à toute étude PNEA, un cadre d'évaluation est défini. L'élaboration de ce document fondateur revient à un groupe d'experts pédagogiques du domaine évalué sous supervision de l'INE.

Les experts pédagogiques sont sélectionnés en fonction de leur expérience dans l'enseignement, leur maîtrise du curriculum national et leur participation à l'élaboration des cadres de référence des examens certificatifs.

Le référent des évaluations PNEA est le curriculum national prescrit : il s'agit d'évaluer l'écart entre le curriculum atteint (acquis des élèves) et le curriculum officiel.

Il est pratiquement impossible de formuler toutes les questions/problèmes susceptibles de mesurer les objectifs/compétences d'un programme scolaire annuel et, par conséquent, il s'avère nécessaire de sélectionner un échantillon représentatif de l'univers objet de l'évaluation.

On cherche donc à établir la structure la plus proche du curriculum en recourant à un modèle conceptuel à deux dimensions : la première est relative aux aires de contenus que le test comportera, la seconde se

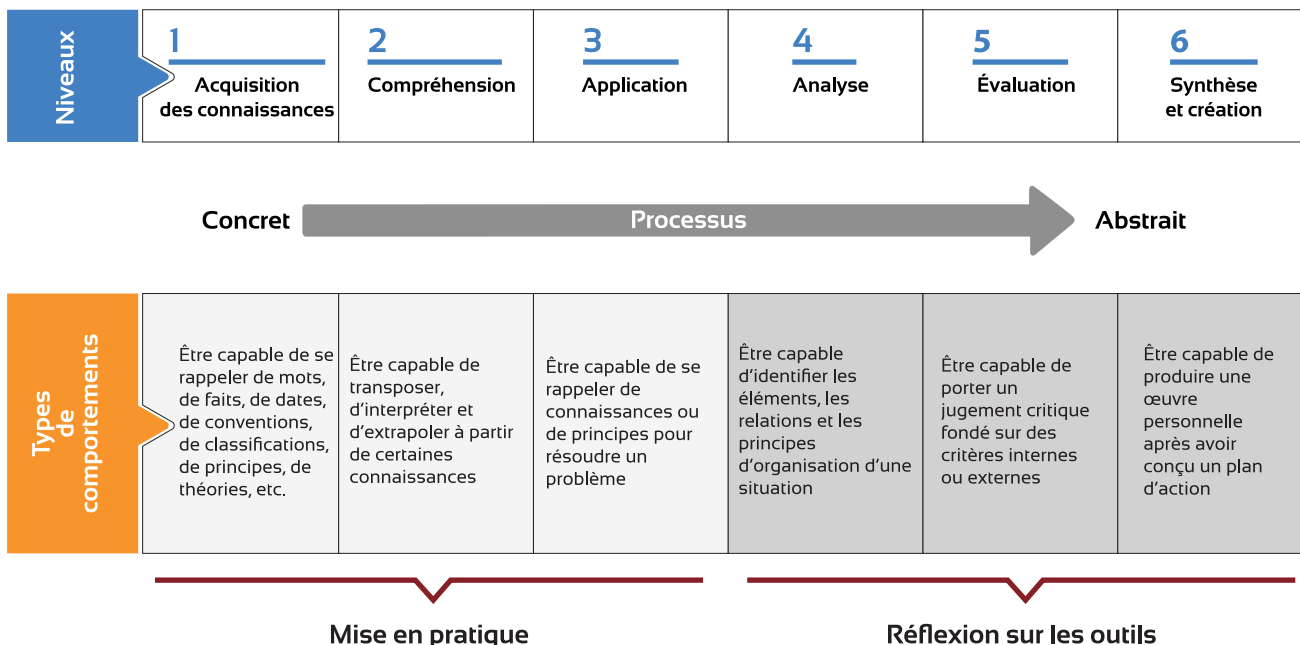


rapporte aux manifestations des comportements.

Dans le domaine pédagogique, les manifestations sont définies par des modèles typologiques (taxonomie cognitive) alors que les aires de contenus sont des composants homogènes des disciplines : chaque item ou famille d'items se situe par rapport à une aire de contenus (domaine de contenu) et une manifestation spécifique (niveau cognitif).

Pour déterminer les niveaux cognitifs, les évaluations PNEA recourent à une adaptation de la taxonomie de Bloom¹ révisée selon le domaine d'évaluation cible.

Taxonomie de Bloom révisée



Concrètement, la construction des cadres de référence passe par la réalisation successive des

opérations cliniques du curriculum ci-dessus :

1- Benjamin S. Bloom. Taxonomy of Educational Objectives, 1956, trad. : Taxonomie des objectifs pédagogiques, vol.1 : Domaine cognitif, Presses de l'Université du Québec, 1975. Benjamin S. Bloom, 1956. Taxonomy of Educational Objectives: The Classification of Educational Goals. Longman. Voir aussi Robert J. Marzano, Designing a new taxonomy of educational objectives, Corwin Press, 2001,

Analyse de contenu du curriculum prescrit

Répartition des programmes scolaires en domaines de contenu

Enumération des objectifs/compétences assignés à chaque domaine de contenu

Détermination de l'importance relative de chaque domaine de contenu

Définition des niveaux cognitifs/taxonomiques de chaque domaine de contenu

Enumération des objectifs/compétences assignés à chaque niveau cognitif

Détermination de l'importance relative de chaque niveau cognitif

Elaboration des tableaux de spécification

L'importance relative est mesurée par le pourcentage de l'enveloppe horaire annuelle consacrée par le curriculum à chaque domaine de contenu et à chaque niveau cognitif.

Une fois une première version des cadres de références élaborée, elle est présentée par l'Instance Nationale d'Évaluation pour examen et enrichissement à un groupe élargi d'inspecteurs et d'enseignants choisis sur la base de leur expérience et leur maîtrise des programmes scolaires.

Enfin, les cadres de référence révisés sont soumis à une validation de contenu avec le Département de l'Éducation Nationale.

4.2. Conception des tests

En plus de la conformité aux dispositions des cadres de références, les concepteurs des items sont tenus de respecter les principes suivants :

- La congruence entre les items et les objectifs d'évaluation ;
- L'indépendance des items du test ;
- La diversification des réponses (QCM, Appariement, Réponse ouverte, ... ;
- La plausibilité des choix des réponses QCM ;
- La conception des items nouveaux, variés, réalistes et signifiants ;
- Un nombre d'items assurant la représentativité (longueur du test).

Ainsi, les concepteurs des items développent un ensemble de questions/problèmes portant sur les objectifs assignés à chaque croisement du tableau de spécification (domaines de contenu/niveaux cognitifs) accompagnés de leur guide de correction.

Ces items font l'objet de discussions avec des enseignants compétents et expérimentés qui enseignent le domaine et le niveau scolaires concernés.

C'est dans cette banque d'items que sont puisés les questions/problèmes formant deux tests équivalents.

Avant la construction proprement dite des tests, on doit préalablement répondre aux questions suivantes :

- Combien de temps va durer le test ?
- Dans quel format les items seront présentés ?
- Combien y aura-t-il d'items ?
- De quelle difficulté seront-ils ?
- A quel type de réponse s'attendre ?

Les deux formes équivalentes de tests ainsi élaborées, elles sont ensuite expérimentées auprès d'un échantillon d'élèves judicieusement choisis.

4.3. Analyse psychométrique

Les résultats de l'expérimentation font l'objet d'une analyse psychométrique pour pouvoir :

- Obtenir des indicateurs sur la précision du test ;
- Améliorer la formulation des items ;
- Remplacer les items trop faciles ou trop difficiles et les items non-discriminants ;
- Reformuler les leurs non attractifs et les items ambigus ;
- Valider le test après modifications et calibrer les tests futurs.

Les théories des tests offrent un cadre conceptuel pour apprécier la valeur des résultats obtenus au

moyen d'un instrument de mesure. Chaque théorie s'appuie sur une conception particulière de la mesure et sur une série de postulats pour la nature des données.

Quels que soient les postulats, il est important de réaliser que chaque théorie n'est qu'un modèle simplifié de la réalité. Chaque modèle s'ajuste plus ou moins bien à la réalité qu'il cherche à décrire. Toutefois, deux modèles restent les plus utilisés dans ce type d'évaluation à savoir la théorie classique des tests (ou du score vrai) et la Théorie de Réponse à l'Item (TRI).

Jusqu'à présent, les évaluations PNEA se contentent de la théorie classique des tests.

4.3.1. Indice de difficulté d'un item

La difficulté des items a un grand impact sur le score au test. C'est pourquoi le choix des items doit tenir compte de la proportion des répondants susceptibles de les réussir ou de les échouer : un item présentant un taux de réussite trop bas ou trop élevé est peu informatif.

De plus, connaître l'indice de difficulté (ou p-indice) des items d'un test peut également permettre de les hiérarchiser dans l'ordre de leur passation. On mettra alors les items les plus faciles en début de test pour une progression harmonieuse dans la passation du test.

Dans la pratique, on choisit les p-indices de façon à obtenir une distribution normale des scores au test.

Pour les évaluations PNEA, on élimine tous les items dont le taux de réussite est supérieur à 85% ou inférieur à 10%.

4.3.2. Pouvoir de discrimination d'un item

Lorsqu'une démarche d'évaluation veut distinguer des individus en fonction d'un critère donné (leurs performances scolaires par exemple), on opte pour des items qui possèdent un pouvoir de discrimination élevé.

Pour calculer l'indice de discrimination d'un item, on répartit l'ensemble des élèves en trois groupes selon leur score au test : les 27% ayant les scores les plus élevés (E), les 27% ayant les scores les plus faibles (F) et les 46% ayant les scores intermédiaires.

On calcule pour les groupes E et F respectivement les proportions de réussite à l'item R_E et R_F .

L'indice de discrimination est défini ainsi : $D = R_E - R_F$.

L'indice de discrimination peut aussi être mesuré par les corrélations inter-items ou item-test.

Dans les évaluations PNEA, on décide ainsi :

- Tous les items dont $D \geq 0,20$ sont retenus ;
- Les items dont $0,10 \leq D < 0,20$ sont reformulés ;
- Les items dont $D < 0,10$ sont remplacés.

4.3.3. Fidélité du test

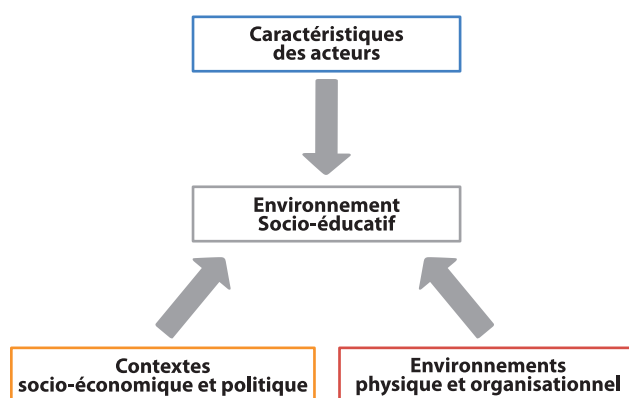
La fidélité renvoie à la précision avec laquelle un instrument mesure l'aptitude ou l'habileté, qu'on cherche à estimer. En théorie classique, elle est définie comme étant la corrélation entre le score observé et le score vrai. Cette définition, traduit une relation étroite entre l'erreur de mesure et la notion de fidélité. En effet, plus forte est cette corrélation, plus réduites sont les erreurs de mesure et meilleure sera la fidélité.

Dans les études PNEA, la fidélité est approchée par la consistance interne qui permet d'évaluer jusqu'à quel point chacun des items d'un test mesure la même dimension. En ce sens, elle reflète le degré d'homogénéité des items constituant le test.

Plusieurs indicateurs permettent de mesurer la fidélité d'un test en termes de consistance interne, dont le plus utilisé est l'Alpha de CRONBACH. Compris entre 0 et 1, plus cet indice est élevé, plus l'instrument de mesure est fidèle. En général, l'homogénéité du test est jugée satisfaisante lorsque l'Alpha est supérieur ou égal à 0,70.

5. Développement des questionnaires

L'environnement socio-éducatif est soumis à l'influence de plusieurs facteurs, en particulier les caractéristiques des acteurs éducatifs (directeurs, enseignants et élèves), les conditions socio-économiques et politiques ainsi que les environnements physique et organisationnel des établissements scolaires.



Afin d'identifier l'impact de ces facteurs sur le rendement scolaire des élèves, et partant, l'évaluer, trois questionnaires sont développés à savoir :

- Un questionnaire destiné aux élèves ;
- Un questionnaire destiné aux enseignants ;
- Un questionnaire destiné aux directeurs (ou au personnel administratif).

Ces questionnaires visent à obtenir le maximum d'informations permettant de donner plus d'éclairage sur une gamme variée de facteurs susceptibles d'influencer d'une façon directe ou indirecte les acquis scolaires des élèves.

5.1. Déterminants de l'environnement socio-éducatif

La qualité d'un environnement socio-éducatif est tributaire de trois grandes sources d'influence. Celles-ci, sont, comme l'a démontré Michel Janosz, les caractéristiques des acteurs, les aspects physique et organisationnel de l'environnement de l'établissement ainsi que les conditions socioéconomiques et sociopolitiques². Dans le cadre des évaluations PNEA, le choix des variables de ces déterminants est basé sur un ensemble de supports fiables et valides notamment les questionnaires PIRLS, TIMSS et TALIS avec adaptation au contexte marocain et en introduisant des questions en rapport avec ce contexte.

Ainsi, le questionnaire «Élève», contient, entre autres, des questions relatives aux caractéristiques sociodémographiques (Genre, Age, Milieu de résidence,...) , au vécu familial (Age des parents, Niveau d'instruction des parents, Profession des parents, Taille du ménage, Type de logement, Langues parlées à la maison, Disponibilité des supports éducatifs,...) ainsi qu'aux antécédents scolaires (Pré-scolarité, Redoublement, Orientation scolaire ...).

Parmi les questions figurant dans les questionnaires «Enseignant» et «Directeur», on peut mentionner celles portant sur les caractéristiques sociodémographiques (Genre, Age, Milieu de résidence, État matrimonial,...), le niveau de formation (Diplômes, Formation initiale et continue,...), l'expérience professionnelle (Cadre, Grade, Ancienneté, Responsabilités,...) et les conditions d'exercice (Distance à l'école, Logement de fonction,...).

En plus des caractéristiques des acteurs éducatifs, l'environnement physique (Lieu, Architecture,

Latrines, Laboratoire, Terrain de sport...) et organisationnel (Structure pédagogique, Ratio Élèves/Enseignant, Ressources humaines, Supports didactiques, Conseil de gestion ...) de l'école influe certainement sur la qualité de l'environnement socio-éducatif.

Notons enfin que le contexte socioéconomique (Ressources allouées à l'éducation, Motivation des ménages, Infrastructure de base, ...) et les politiques éducatives (Modèles pédagogiques, Appui social aux élèves, Formation des enseignants, Développement des TICE, ...) influent largement sur la qualité de l'environnement socio-éducatif.

5.2. Composantes de l'environnement socio-éducatif

Certes, il existe plusieurs modèles pour évaluer l'environnement socioéducatif. L'élaboration des questionnaires PNEA se base sur un modèle théorique canadien de l'environnement socioéducatif qu'on a adapté au contexte éducatif marocain³.

Ce modèle considère qu'on ne peut juger de la qualité de l'adaptation psychosociale d'une personne qu'en tenant compte de manière simultanée, d'une part du potentiel d'adaptation de l'individu (ses habiletés biologiques, cognitives, socio-affectives, comportementales, etc.) et, d'autre part du potentiel éducatif de son environnement familial, scolaire et social.

L'école est donc censée offrir aux acteurs éducatifs un milieu sécurisant, favorable à la pratique de leur métier d'apprenant ou d'enseignant ou de gestionnaire.

Par ailleurs, l'école doit être un lieu qui favorise l'intégration sociale des individus en satisfaisant leurs besoins d'attachement, de relations interpersonnelles, de développement des habiletés sociales et d'appartenance à une communauté.

L'environnement socio-éducatif de l'école, au sens de Janosz, est constitué de trois composantes⁴ :

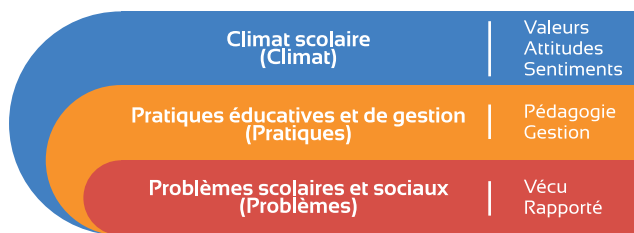
- Le climat scolaire ;
- Les pratiques éducatives et de gestion ;
- Les problèmes scolaires et sociaux.

2- Michel Janosz, Patricia Georges et Sophie Parent « L'environnement socioéducatif à l'école secondaire : Un modèle théorique pour guider l'évaluation du milieu », Revue canadienne de psychoéducation, Volume 27, Numéro 2, 1998. Les auteurs ont adapté la batterie d'indicateurs de Gary D. Gottfredson. « Effective school battery. User's Manual, 1984, University of Maryland.

3- Michel Janosz et Patricia Georges. Op cit.

4- Michel Janosz op cit. Voir aussi Michel Janosz et Christiane Bouthilier « Rapport de validation du questionnaire sur l'environnement socioéducatif des écoles secondaires (QES – Secondaire) », Montréal 2007.

Composantes de l'environnement socioéducatif



5.2.1. Le climat scolaire

Le climat scolaire traduit essentiellement les valeurs, les attitudes et les sentiments régnant dans le milieu scolaire. En fait, un bon climat scolaire crée une disposition favorable aux apprentissages.

Le climat scolaire est appréhendé à travers cinq dimensions⁵ :

5.2.1.1. Le climat relationnel

Ce climat est indicateur de l'atmosphère qui règne entre les élèves, entre les élèves et les enseignants, entre les enseignants et l'administration et entre les élèves et l'administration.

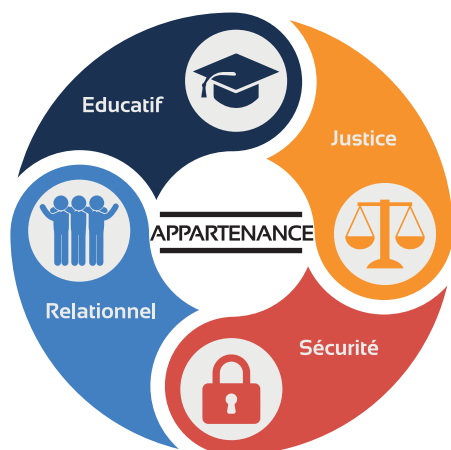
La qualité du climat relationnel est tributaire de la qualité des contacts interpersonnels, du respect mutuel et de l'assurance du soutien d'autrui.

5.2.1.2. Le climat de sécurité

Un climat prévisible et calme, où règnent la sécurité et la tranquillité, est nécessaire à la concentration qu'exigent les activités d'apprentissage.

La qualité du climat de sécurité se manifeste à travers les sentiments de confiance des acteurs éducatifs et leur perception des risques et de victimisation.

Le climat scolaire et ses différentes facettes



5.2.1.3. Le climat de justice

Le sentiment de justice à l'école passe par i) la reconnaissance de la légitimité des règles et leur

application judicieuse, ii) l'existence d'un système d'évaluation équitable et, iii) le mérite ou la sanction dépendent des comportements et non des personnes.

5.2.1.4. Le climat d'appartenance

Un climat d'appartenance se manifeste par i) un sentiment de fierté à fréquenter l'école, ii) l'importance accordée à l'école et, iii) l'adhésion aux valeurs qu'elle véhicule.

Le lien très fort entre le « climat scolaire » et la réussite scolaire est largement établi par la recherche. Travailler sur la notion de « climat scolaire » revient donc à travailler sur des questions touchant à la construction et à la transmission des connaissances.

Notons, toutefois, la difficulté à changer les attitudes, les perceptions et les sentiments des gens. Par contre, les pratiques éducatives et de gestion, elles aussi associées au rendement scolaire, se prêtent davantage à l'intervention.

5.2.2. Les pratiques éducatives et de gestion

Les recherches empiriques reconnaissent à certaines pratiques éducatives et de gestion une grande influence sur la qualité des apprentissages notamment les dimensions suivantes :

5.2.2.1. La qualité de l'enseignement

Dans les écoles les plus performantes, les enseignants tiennent compte des stratégies d'apprentissage utilisées par leurs élèves, adaptent leurs méthodes pédagogiques aux besoins des apprenants et cherchent à développer des habiletés métacognitives.

5.2.2.2. Le système de reconnaissance

Les écoles qui recourent à des renforçateurs positifs plutôt qu'aux sanctions enregistrent moins de problèmes et de turbulence.

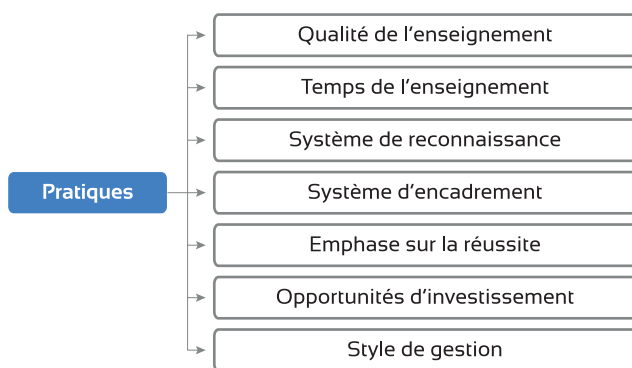
Plus les élèves sont récompensés pour leurs performances, plus ils perçoivent la valorisation de l'effort fourni dans le milieu scolaire.

5.2.2.3. Le système d'encadrement

Il s'agit des règles et procédures régissant l'ordre et la discipline à l'école et qui doivent être, formulées par écrit, cohérentes, accessibles et appliquées avec rigueur et transparence.

Un système d'encadrement de qualité est une condition nécessaire à l'instauration d'un climat d'ordre et de sécurité, indispensables aux apprentissages.

5- Janosz et Patricia Georges, op.cit, p. 294



5.2.2.4. L'emphase sur la réussite

Les recherches ont démontré, qu'à potentiel équivalent, les élèves des enseignants persuadés de leur douance relative obtiennent de meilleurs résultats que ceux dont les enseignants sont persuadés de leur moindre capacité (Effet Pygmalion).

5.2.2.5. Les opportunités d'investissement

L'école devrait offrir des occasions d'investissement scolaires en dehors de la classe notamment des activités parascolaires intéressantes, valorisées et accessibles et ce, afin de développer les compétences sportives, artistiques et sociales des élèves.

5.2.2.6. Le temps consacré à l'enseignement

Dans les écoles les moins efficaces, il y a beaucoup de perte de temps avant et pendant les cours. Une bonne planification des leçons est de nature à limiter le gaspillage du temps d'enseignement : plus les élèves sont concentrés sur les activités d'apprentissage, moins ils risquent de se disperser.

5.2.2.7. Le leadership éducatif

Une bonne direction d'école présente clairement ses attentes, distribue judicieusement les tâches et valorise ses enseignants et ses élèves en les impliquant à la prise de décisions.

Les bonnes pratiques de gestion impliquent une bonne communication entre l'école et les parents d'élèves : plus les parents s'impliquent dans l'école, plus les élèves sentent l'importance sociale accordée à l'éducation.

Il s'agit de règles et de procédures régissant l'ordre et la discipline au sein de l'école et qui doivent être cohérentes, accessibles, formulées par écrit et appliquées avec rigueur et transparence.

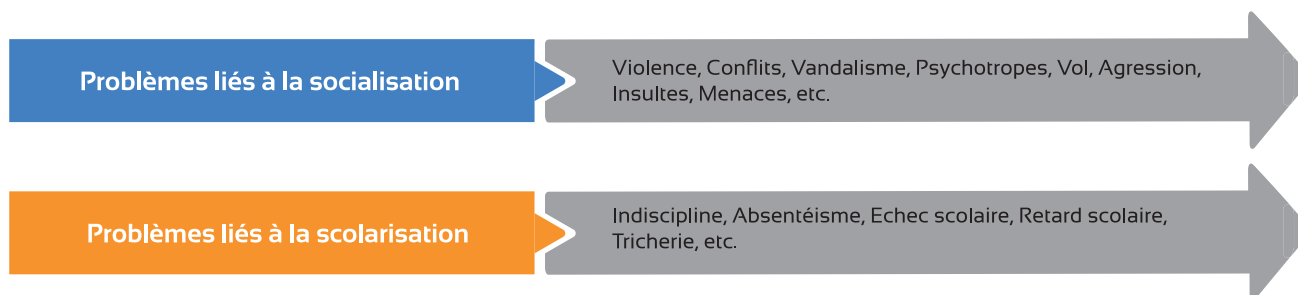
5.2.3. Problèmes sociaux et scolaires

Par ailleurs, la plupart des établissements scolaires connaissent des problèmes d'incivilités. Ainsi, se posent les questions suivantes :

- Les incivilités sont-elles exogènes à l'école ?
- Sont-elles dues au seul contexte économique et social des élèves ?
- Le contexte scolaire lui-même influe-t-il l'exposition des élèves aux incivilités ?

5.2.3.1. Problèmes sociaux

Si les facteurs socio-économiques expliquaient les incivilités à l'école, alors tous les établissements recrutant dans le même type de milieu socio-économique devraient connaître un niveau d'incivilités semblable. Or, ce n'est pas le cas. C'est dire la nécessité de distinguer les problèmes liés à la socialisation et les problèmes endogènes à l'école.



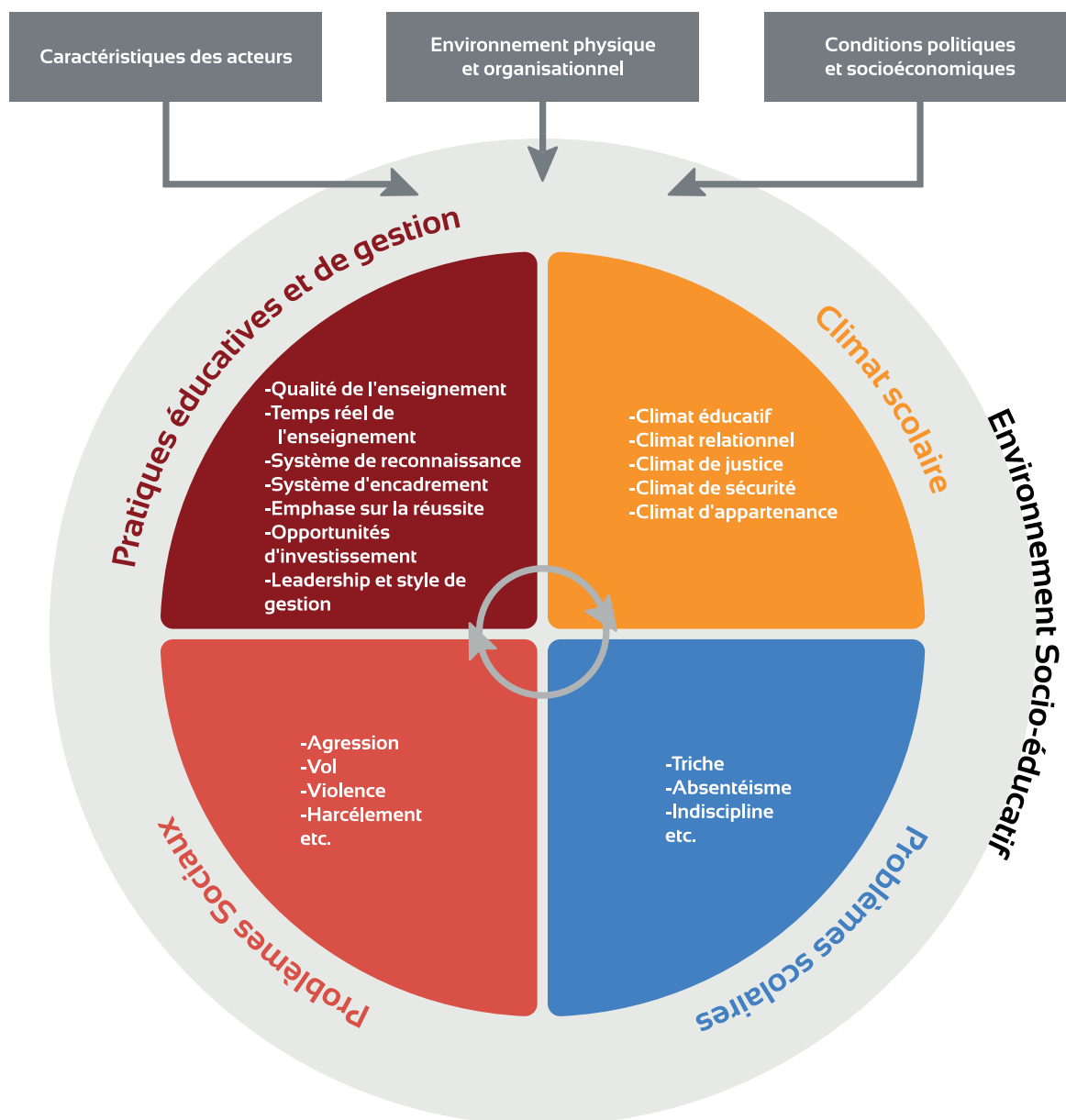
Le fait d'impliquer tous les membres de la communauté scolaire pour passer d'une culture de la passivité à une culture de la réactivité face aux problèmes d'incivilités favorise l'acquisition des compétences civiques chez les élèves.

Parmi les problèmes sociaux, on note la violence, les conflits, le vandalisme, les psychotropes, le vol, les agressions, les insultes, les menaces, etc.

5.2.3.2. Problèmes scolaires

Les problèmes inhérents à l'école comprennent la tricherie, l'indiscipline, l'échec scolaire, etc.

Force est de noter que les composantes de l'environnement socioéducatif sont, d'une part inter-reliées et interagissent entre elles et, d'autres part influencées par les caractéristiques des acteurs éducatifs, l'environnement physique et organisationnel ainsi que les conditions politiques et socioéconomiques.



5.2.4. Échelle de mesure de l'environnement socio-éducatif

Une échelle de mesure représente un phénomène latent (un construit) supposé mesurable directement par plusieurs questions regroupées ensemble. On doit donc s'assurer que les énoncés du questionnaire que l'on associe à un construit le sont vraiment et que les échelles de mesure correspondantes présentent la validité et la fidélité requises.

Pour le modèle théorique adopté pour le développement des questionnaires PNEA, les échelles de mesure sont associées respectivement au climat scolaire et aux pratiques éducatives et de gestion. Ainsi, chaque dimension du climat scolaire ou des pratiques éducatives et de gestion est appréhendée par 4 à 8 questions dont les réponses sont : Totallement d'accord - Peu d'accord - Peu

désaccord - Totallement désaccord

Pour les questions qui vont dans le sens positif de ce qui est souhaitable, les réponses sont codifiées ainsi :

Totallement d'accord = 4, Peu d'accord = 3, Peu désaccord = 2, Totallement désaccord = 1

Dans le cas contraire, la codification des réponses est inversée de manière à ce qu'un score élevé représente toujours une perception positive. Dans ce cas, on aura :

Totallement d'accord = 1, Peu d'accord = 2, Peu désaccord = 3, Totallement désaccord = 4

Le score de chaque dimension est la moyenne des scores aux énoncés le mesurant.

Contrairement au climat scolaire, les questions

correspondantes aux problèmes sociaux et scolaires ne sont pas regroupées en échelles de mesure. Ainsi, chacune de ces questions est considérée comme un indicateur renseignant un aspect des comportements d'indiscipline perçus et rapportés, ou des problèmes de violence perçus et subis ou de l'accessibilité aux drogues ainsi que le sentiment de sécurité.

Pour les comportements d'indiscipline scolaire perçus et agis ou des problèmes de violence perçus et subis, le choix des réponses est :

Jamais (1), Rarement (2), Quelquefois (3), Souvent (4), Toujours (5).

Pour le rapport des problèmes subis, le choix des réponses se présente ainsi :

Oui (1), Non (2), Ce n'est jamais arrivé (3).

5.2.5. Validité des questionnaires

5.2.5.1. Validité de contenu

La validité du contenu des questionnaires est assurée lors des ateliers de travail avec des experts pédagogiques (Enseignants, Inspecteurs et Directeurs d'établissements).

Ensuite, les questionnaires sont administrés à un échantillon d'élèves, à leurs enseignants et aux directeurs de leurs établissements scolaires (Expérimentation).

5.2.5.2. Validité du construit

La validité du construit porte sur deux composantes de l'environnement socio-éducatif à savoir le climat scolaire et les pratiques éducatives et de gestion. Les résultats de l'étude expérimentale des échelles de mesure correspondantes aux dimensions de ces deux composantes font partie d'une analyse factorielle pour s'assurer de la validité des échelles de mesure choisies. Cette analyse se fait séparément pour les deux composantes dans chacun des trois questionnaires.

Etant donné que les questionnaires sont conçus de telle sorte à tenir compte des perceptions de tous les acteurs (Élèves, Enseignants et Directeurs), on retient donc les énoncés qui contribuent à une équivalence entre les échelles de mesure entre les trois questionnaires et on retire ceux qui brouillent le sens des facteurs latents.

Pour se donner une idée de la fidélité des échelles de mesure, on examine leur consistance interne en calculant Alpha de CRONBACH. Ainsi, une échelle de mesure est jugée acceptable dès que sa consistance interne dépasse 0,70.

6. Plan d'échantillonnage

La population de référence de chaque évaluation PNEA est définie par l'effectif des élèves du niveau scolaire concerné.

Les petites unités scolaires comptant moins de huit élèves dans le niveau scolaire ciblé sont exclues, sans toutefois que le taux d'exclusion global dépasse 5% de la population cible.

La sélection des échantillons doit satisfaire certaines exigences en termes de taille, de stratification et de méthodes de sélection et, ce, afin d'en assurer la représentativité.

Le choix des méthodes de sélection dépend d'un ensemble de contraintes liées aux objectifs de l'évaluation, à la précision recherchée et aux coûts induits.

Le plan d'échantillonnage adopté dans le cadre des études PNEA est un plan d'échantillonnage à plusieurs degrés (Établissements, Classes, Élèves), stratifié explicitement par région et par type d'établissement (Public/Privé) et implicitement par milieu (Urbain/Rural).

En tenant compte des objectifs fixés, ce plan d'échantillonnage probabiliste à plusieurs degrés est l'approche la plus efficace pour obtenir des données précises, généralisables et à coût raisonnable.

Ce plan d'échantillonnage permet d'avoir une précision des estimateurs au moins égale à celle d'un échantillon aléatoire simple de taille 400 élèves.

Moyenne : $m \pm 0.1*s$ avec une confiance de 95%

Proportion : $p \pm 5%$ avec une confiance de 95%

Corrélation : $r \pm 0.1$ avec une confiance de 95%.

Certes, à taille égale, le plan d'échantillonnage PNEA est moins précis qu'un plan d'échantillonnage aléatoire simple et par conséquent, la taille des échantillons PNEA sera plus importante pour avoir la même précision.

A l'instar des études TIMSS et PIRLS, la taille de l'échantillon PNEA est déterminée ainsi :

• Taille de l'échantillon établissements = $a_0 * [1 + IC(MCS-1)] / MCS + a_0 * 0,5 / MCS$

• La taille n de l'échantillon des élèves = $n = a * MCS$

Avec :

a_0 : Taille fixée à 400 élèves dans la conception du plan d'échantillonnage

IC : Coefficient de corrélation intra-classe

MCS : Taille moyenne des classes.

Le tirage se fait concrètement selon les étapes ci-dessous :

- Les établissements de chaque strate explicite sont triés selon les variables de la stratification implicite;
- Au sein de chaque strate explicite triée, l'échantillon des établissements est sélectionné par un tirage systématique avec probabilités proportionnelles à la taille (PPS) ;
- Dans chaque établissement retenu, on élabore une liste exhaustive de ses classes et on sélectionne une classe par un tirage aléatoire simple ;
- Tous les élèves d'une classe échantillonnée font partie de l'échantillon « élèves » soumis à l'évaluation.

Les résultats des évaluations sont soumis à une variabilité qui dépend notamment des erreurs d'échantillonnage. Il est possible d'estimer statistiquement ces erreurs et de produire des intervalles de confiance sur les différents estimateurs calculés.

Collecte et nettoyage des données

En plus des tests et des questionnaires, des guides de passation, de correction et de contrôle de qualité sont élaborés en vue d'uniformiser et de standardiser les opérations d'évaluation.

Par ailleurs, une formation en cascade est assurée à tous les intervenants : après être formés, les formateurs régionaux assurent la formation des coordonnateurs provinciaux qui à leur tour forment les coordonnateurs locaux (directeurs d'établissements concernés).

Les préparatifs logistiques de la passation des tests et l'administration des questionnaires s'opèrent en étroite concertation avec le Département de l'Éducation Nationale qui doit, en principe, apporter les réponses adéquates aux questions suivantes :

- Quelles sont les ressources humaines disponibles pour la passation du dispositif ?
- Quelle période en fin d'année scolaire est la plus convenable pour l'évaluation ?
- Comment s'assurer de la conformité des salles et du matériel au dispositif d'évaluation ?
- Comment va se dérouler l'inscription des candidats et l'enregistrement de leur présence ?
- Comment installer le dispositif dans les salles et le récupérer ?
- Quels sont les endroits sécurisés où le matériel peut être stocké ?
- Quels dysfonctionnements peuvent se produire et quelles sont les procédures à appliquer pour y faire face ?

Concernant la correction, les correcteurs sont amenés à corriger les questions ouvertes uniquement. Une correction double d'un échantillon de ces questions permet de s'assurer de la qualité de la correction réalisée.

Quant au nettoyage des données collectées, on commence par vérifier si l'identifiant dans chaque base de données est unique.

Ensuite, on vérifie pour chaque variable la cohérence des valeurs prises avec le code-book pour détecter les aberrations éventuelles. Ces dernières doivent être corrigées en consultant le support original.

Il est également nécessaire de vérifier les incohérences entre les variables qui donnent lieu à des combinaisons peu plausibles.

Le remplacement des valeurs manquantes se fait essentiellement par la valeur moyenne de l'ensemble des réponses à cette variable (imputation par la moyenne).

7. Exploitation et analyse des résultats

Les variables initiales ou brutes ne permettent pas toujours de répondre aux besoins d'analyse et aux questions de recherche posées. Leur recodage ou la création de nouvelles variables sont deux opérations qui sont préalables à toute exploitation des résultats.

Le recodage consiste à créer une variable qualitative catégorielle en procédant à un découpage en classes d'une variable quantitative. On peut également recoder une variable, déjà catégorielle, en regroupant les classes initiales. Aussi, on peut avoir besoin de faire l'inverse, c'est à dire transformer une variable ordinale avec différentes modalités en une variable numérique en recodant ces dernières en valeurs numériques discrètes.

Le recodage par transformation est également utilisé. Il s'agit par exemple de la transformation centrée-réduite qui consiste à centrer une variable par rapport à sa moyenne et de la réduire par rapport à son écart type. Ceci permet d'avoir une variable normalisée ou standardisée (score standardisé par exemple).

La création de nouvelles variables permet de dériver une nouvelle variable à partir d'une autre, en utilisant des conditions ou des opérations arithmétiques (soustraction, division, calcul de moyenne, etc.).

La fusion des bases de données relatives aux scores aux tests et aux questionnaires de contexte est rendue possible grâce à l'unicité de l'identifiant de chaque base : l'identifiant élève permet d'apparier la base des scores avec les variables du questionnaire « Élève » et la base, ainsi obtenue, est fusionnée avec celle du questionnaire « Enseignant » à l'aide de l'identifiant classe. Enfin, on fusionne cette dernière

base avec celle du questionnaire « Directeur » et ce, en faisant appel à l'identifiant établissement. Ces opérations sont effectuées pour chaque niveau scolaire et pour chaque matière objet de l'évaluation.

Après avoir effectué une analyse descriptive qui donne une vision sur le profil des participants aux évaluations PNEA en l'occurrence les élèves, les enseignants et les directeurs d'établissements scolaires, on recourt à l'analyse bi-variée et à l'analyse multi-niveaux pour déterminer l'effet des variables de contexte sur le rendement scolaire des élèves.

7.1. Analyse bi-variée

L'objectif de l'analyse bi-variée est d'examiner l'existence d'une corrélation entre la variable expliquée (le rendement scolaire) et les variables potentiellement explicatives prises séparément (variables de contexte). L'effet absolu mesure l'apport indépendant de la variable au rendement des élèves.

Les tests de Khi-2 permettent de savoir si la répartition des effectifs dans un tableau de contingence est significativement différente de celle de la table calculée sous l'hypothèse d'indépendance des deux variables croisées.

Le degré de signification de chacune des relations fonctionnelles entre le rendement scolaire (variable expliquée) et les variables de contexte (variables explicatives) est évalué en ayant recours au test statistique de Khi-2 : on considère que toute relation dont la probabilité (p-value) inférieure à un seuil de 5%, est significative.

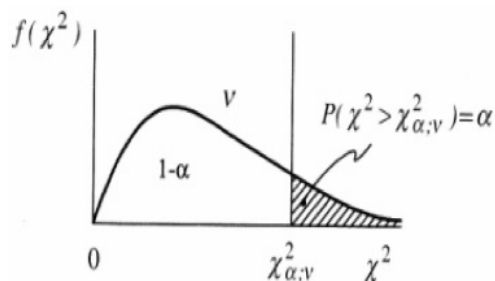
Le test de Khi-2 permet de mesurer la significativité de la relation entre deux caractères. L'idée générale de ce test est de calculer la statistique de Khi-2 (observée) qui représente la somme des écarts entre effectifs observés et effectifs théoriques, d'un tableau de contingence et de la comparer à la valeur théorique d'une loi de probabilité Khi-2 de degré de liberté u et un risque (le risque de rejeter l'hypothèse nulle alors qu'elle est vraie).

Ainsi, pour tester s'il y a une liaison significative entre deux caractères (par exemple le score et le genre), on pose l'hypothèse H_0 : « Il n'y a pas de relation entre les deux caractères » et on détermine la valeur Khi-2 observée du tableau étudié.

La statistique χ^2 donne une mesure de l'écart existant entre les fréquences théoriques attendues F_e et celles observées dans l'échantillon F_o : plus χ^2 sera grand, plus le désaccord sera important et la coïncidence parfaite si $\chi^2=0$ où $\chi^2 = \sum \frac{(F_o - F_e)^2}{F_e}$

Afin d'être en mesure d'interpréter au mieux le résultat du Khi-2, il faut également choisir le seuil de signification α . Le seuil de signification nous indique

le niveau avec lequel nous pouvons affirmer qu'il existe effectivement une différence statistiquement significative. Lorsque le Khi-2 franchit le seuil, la différence devient alors significative. Le seuil nous indique la probabilité de commettre une erreur en prenant la décision.



On peut résumer le processus de décision par le biais du tableau suivant :

	Ho est vrai	Ho est Faux
On accepte Ho	Bonne décision avec une probabilité de 95%	Erreur de deuxième espèce
On rejette Ho	Erreur de première espèce avec une probabilité de 5%	Bonne décision

D'une manière concrète, on procède ainsi :

» Formulation des hypothèses

H_0 = {les deux variables qualitatives sont indépendantes}

contre H_1 = {les deux variables qualitatives sont dépendantes}

• Calcul de la statistique

$$D = \sum_{i,j} \frac{(N_{ij} - np_{ij})^2}{np_{ij}} \quad np_{ij} = \frac{\left(\sum_j n_{ij}\right)\left(\sum_i n_{ij}\right)}{n} = \frac{n_i n_j}{n}$$

Où n_{ij} est l'effectif des individus possédant la modalité i de la 1^o variable et la modalité j de la 2^o variable.

n_i est l'effectif des individus possédant la modalité i de la 1^o variable et n_j l'effectif des individus possédant la modalité j de la 2^o variable.

• Ho vraie

$$D \longrightarrow \chi^2_{(1-\alpha)(c-1)}$$

Où l est le nombre de modalités de la 1^o variable (nombre de lignes du tableau de contingence) et c le nombre de modalités de la 2^o variable (nombre de colonnes du tableau de contingence)

» **Détermination du seuil critique**

$$\alpha = P [\text{rejeter } H_0 \text{ quand } H_0 \text{ vraie}]$$

$$= P [D > h \text{ quand } H_0 \text{ vraie}]$$

$$= P \left[\chi_{(1-\alpha)(e-1)}^2 > h \right]$$

α étant donné, on en déduit la valeur de h

» **Règle de décision**

$d > h$ on rejette H_0

$d < h$ on «accepte» H_0

Dans les évaluations PNEA, on opte pour un seuil de significativité de 5%.

7.2. Analyse multi-niveaux

Cerner la contribution directe et indirecte de l'école ou de la classe sur le rendement scolaire des élèves, en tenant compte de leurs caractéristiques individuelles, familiales, sociodémographiques et économiques, est l'un des objectifs des évaluations PNEA.

Il s'agit de déterminer à quel point la fréquentation d'un établissement ou d'une classe donné(e) fait une différence dans la réussite scolaire des élèves, au-delà de leurs caractéristiques personnelles et sociales.

Les méthodes multi-niveaux, initialement développées dans les sciences de l'Éducation (Goldstein, 1995), où le rendement scolaire d'un élève dépend, non seulement de ses caractéristiques, mais aussi des paramètres propres à son environnement scolaire (école, professeur, classe), permettent de s'affranchir, d'une part de l'erreur écologique, qui consiste à interpréter, au niveau individuel, les résultats d'une modélisation effectuée à un niveau agrégé et, d'autre part, de l'erreur atomiste qui ignore le contexte dans lequel évolue l'individu et étend à la dimension du contexte un ensemble d'effets individuels.

Ainsi, les modèles multi-niveaux permettent d'isoler l'effet spécifiquement attribuable à l'école ou à la classe en séparant les différents niveaux de l'analyse (l'élève, la classe, l'école) et permettent d'estimer ce qui relève de chacun dans l'explication des variations observées dans les mesures.

Le principe consiste à décomposer la variance observée dans la mesure d'une variable dépendante en autant de parties distinctes qu'il y a de niveaux d'analyse : on distingue la proportion de la variance observée attribuable à la variabilité des scores entre les élèves, celle attribuable à la variabilité des scores entre les classes et, finalement, celle attribuable à la variabilité des scores entre chacune des écoles.

Les effets ainsi obtenus s'énoncent souvent en part de variance expliquée : on estime le pourcentage

de la variance expliquée par le fait que les élèves appartiennent à un établissement ou à une classe donné(e). Cette détermination de la proportion de la variance totale qui s'explique par les variations entre les écoles ou entre les classes (appelée corrélation intra-classe) constitue la première étape d'une analyse multi-niveaux.

On distinguera les modèles à constante aléatoire (l'effet des variables explicatives est le même dans chacun des contextes, seul change l'effet moyen de chaque contexte par rapport à l'ensemble), et les modèles à pente aléatoire ou l'effet d'une variable est considéré différent selon l'unité contextuelle analysée.

Ne pas prendre en compte un effet multi-niveaux revient notamment à sous-estimer la variance des effets contextuels, et donc surestimer leur significativité.

Le modèle multiniveaux à constantes aléatoires

Niveau 1 $Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij}$

Niveau 2 $\beta_{0j} = \gamma_{00} + u_{0j}$

Équation complète $Y_{ij} = \gamma_{00} + \beta_1 X_{ij} + u_{0j} + e_{ij}$

Les composants de la variance :

$$Var(e_{ij}) = \sigma_e^2$$

$$Var(u_{0j}) = \sigma_{u0}^2$$

Variance totale : $Var(y_{ij}) = \sigma_e^2 + \sigma_{u0}^2$

Le modèle multiniveaux complet : constantes et pentes aléatoires

Niveau 1 $Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij}$

Niveau 2 $\beta_{0j} = \gamma_{00} + u_{0j}$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Équation complète

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + u_{0j} + u_{1j} X_{ij} + e_{ij}$$

Les composants de La variance :

$$Var(e_{ij}) = \sigma_e^2$$

$$Var(u_{0j}) = \sigma_{u0}^2$$

$$Var(u_{1j}) = \sigma_{u1}^2$$

$$Cov(u_{0j}, u_{1j}) = \sigma_{u10}$$

$$Var(Y_{ij} | X_{ij}) = \sigma_{e0}^2 + (\sigma_{u0}^2 + \sigma_{u1}^2 X_{ij}^2 + 2\sigma_{u10} x_{ij})$$

Les variables explicatives utilisées dans l'investigation des déterminants des acquis scolaires portent sur les caractéristiques extrascolaires, le climat scolaire, les pratiques éducatives et de gestion et les problèmes sociaux et scolaires.

Comme tout autre modèle statistique, les paramètres estimés doivent être sujets à des tests

de significativité. A cette fin, les statisticiens ont développé une mesure similaire à la statistique de Fisher utilisée dans les régressions par les Moindres Carrés Ordinaires (MCO). Cette mesure appelée « la déviance » permet d'assurer la pertinence de l'introduction des variables supplémentaires, ce qui est à même de comparer les estimations entre les différents modèles.

Analytiquement, la déviance se définit comme étant la différence entre la vraisemblance de deux modèles. Si L1 et L2 sont respectivement la vraisemblance des modèles 1 et 2, avec L2, le modèle contenant des variables supplémentaires, la déviance D est définie comme suit :

$$D = -2 \cdot \log(L1) - 2 \cdot \log(L2)$$

D suit une loi de Khi-2 avec m degré de liberté égal au nombre de variables supplémentaires. En comparant la valeur D avec le fractile de la loi de Khi-2(m), on rejette ou non l'hypothèse de la pertinence de rajouter des variables supplémentaires par rapport au modèle initial.

Les logiciels statistiques utilisés au sein de l'Instance Nationale d'Évaluation, notamment SPSS et STATA incorporant un module d'estimation des données hiérarchisées, permettent de calculer, entre autres, la déviance D.

II. PROGRAMME NATIONAL D'ÉVALUATION DES ACQUIS DES ÉLÈVES DES TRONCS COMMUNS DU SECONDAIRE QUALIFIANT (PNEA 2016)

L'étude PNEA 2016 cible les élèves des tronc communs de l'enseignement secondaire qualifiant. Le choix de ce niveau scolaire se justifie, d'une part, par le fait qu'il représente une année d'orientation en amont et en aval et, d'autre part les élèves de ce niveau ne sont sujets à aucune évaluation certificative.

Les domaines d'évaluation concernés sont la langue arabe, le français, l'histoire-géographie, les mathématiques, la physique-chimie et les sciences de la vie et la terre.

Cette étude d'évaluation intervient dans un contexte caractérisé par :

- L'adoption de la Vision stratégique de la réforme 2015-2030 ;
- Le recentrage du nouveau curriculum sur les apprentissages fondamentaux ;
- Le processus de l'autonomisation des établissements scolaires en cours et son corollaire l'évaluation axée sur les résultats.

1. Objectifs de l'évaluation PNEA 2016

L'étude PNEA 2016 vise à :

- Évaluer les niveaux linguistique, scientifique et culturel des élèves ;
- Cerner l'«Effet élève » et l'«Effet établissement » sur les acquis ;
- Apprécier la qualité du climat scolaire et son impact sur la réussite scolaire ;
- Préciser les liens entre les variables de contexte et le rendement scolaire ;
- Évaluer l'impact des réformes pédagogiques sur la qualité des apprentissages ;
- Mettre à disposition des responsables et acteurs des données de résultats fiables aidant à la prise de décision ;
- Promouvoir la recherche scientifique et pédagogique.

2. Niveaux scolaires et domaines cibles

Le Programme National d'Évaluation des Acquis (PNEA 2016) se focalise sur l'évaluation des

acquis des élèves des tronc communs « Lettres et Sciences Humaines », « Sciences », « Technique » et « Originel » et ce, en Arabe, en Français, en Histoire-Géographie, en Physique-Chimie et en Sciences de la Vie et la Terre.

3. Construction des tests

Un dispositif d'évaluation a été spécialement conçu pour les besoins de l'étude PNEA 2016. Il s'agit de dix cadres de référence établis sur la base des curricula nationaux des six champs disciplinaires et ce, pour quatre tronc communs (Lettres et Sciences Humaines, Sciences, Technique et Originel). Ces cadres de référence ont été développés et validés, sous la supervision et l'encadrement de l'INE, par un groupe d'experts, spécialistes dans les six domaines disciplinaires susmentionnés.

Par la suite, vingt tests ont été construits à raison de deux tests par discipline (forme A et forme B) et ont été expérimentés auprès d'un échantillon aléatoire de 3.924 élèves répartis sur le territoire national.

Après l'analyse psychométrique des résultats de l'expérimentation, 10 tests répondant aux critères de précision requis ont été finalisés et validés.

Pour des raisons de coût, de précision et d'objectivité ainsi que pour éviter la correction et la saisie, les tests sont composés de questions fermées (questions à choix multiples, questions de liaison ou questions à réponses courtes) et des productions écrites en Arabe et en Français.

3.1. Test de mathématiques

L'évaluation des acquis des élèves du tronc commun en mathématiques repose sur deux dimensions : les domaines de contenus et les niveaux cognitifs. Les contenus évalués se rapportent au calcul numérique et fonctions, la géométrie et la statistique.

L'apprentissage des mathématiques passe nécessairement par l'acquisition des connaissances, leur application directe ainsi que le développement de la capacité de raisonnement.

Afin d'évaluer les compétences chez les élèves du tronc commun en mathématiques, trois niveaux cognitifs sont considérés, à savoir : la connaissance, l'application et le raisonnement.

Tableau 1. Tableau de spécification des mathématiques par domaine de contenus et niveau cognitif

Sciences/Technique	Calcul numérique	Géométrie	Statistique	Total
Connaissance	21,00%	11,55%	2,45%	35%
Application	27,00%	14,85%	3,15%	45%
Raisonnement	12,00%	6,60%	1,40%	20%
Total	60%	33%	7%	100%

Source : Cadres de références des mathématiques, INE-2010.

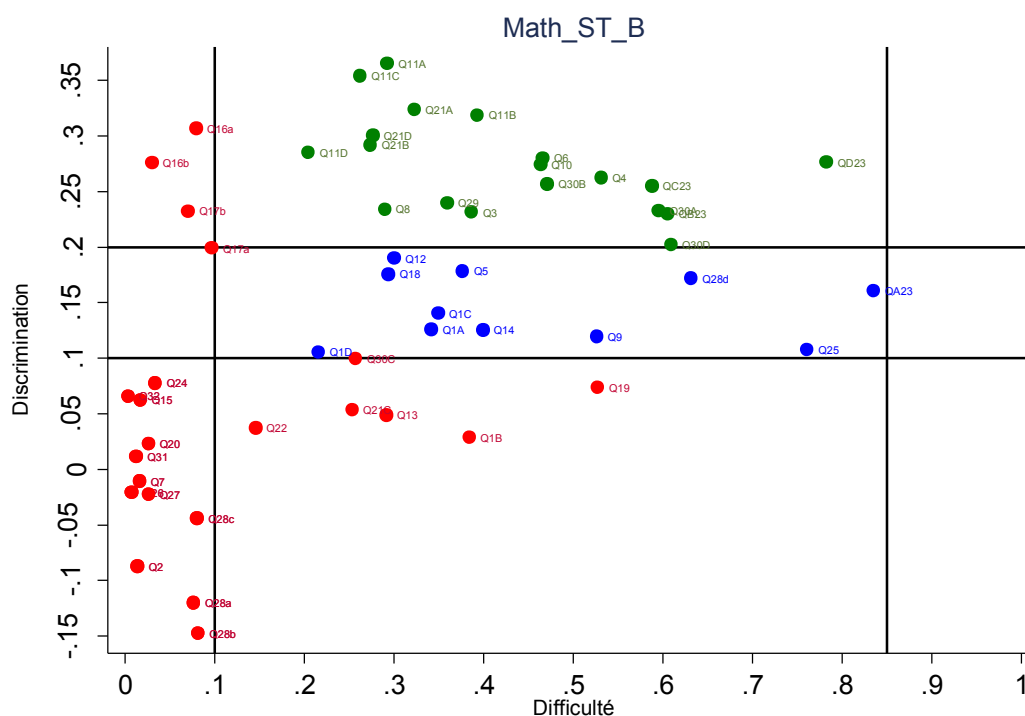
Tableau 2. Tableau de spécification des mathématiques par domaine de contenus et niveau cognitif

Lettres/Original	Calcul numérique	Géométrie	Statistique	Total
Connaissance	24%	8%	8%	40%
Application	30%	10%	10%	50%
Raisonnement	6%	2%	2%	10%
Total	60%	20%	20%	100%

Source : Cadres de références des mathématiques, INE-2010.

Pour les tronc communs «Sciences» et «Technique», la consistance interne du test A des mathématiques ($\alpha=0.53$) est inférieure à celle du test B ($\alpha=0.61$), cette dernière forme du test a été retenue pour être améliorée.

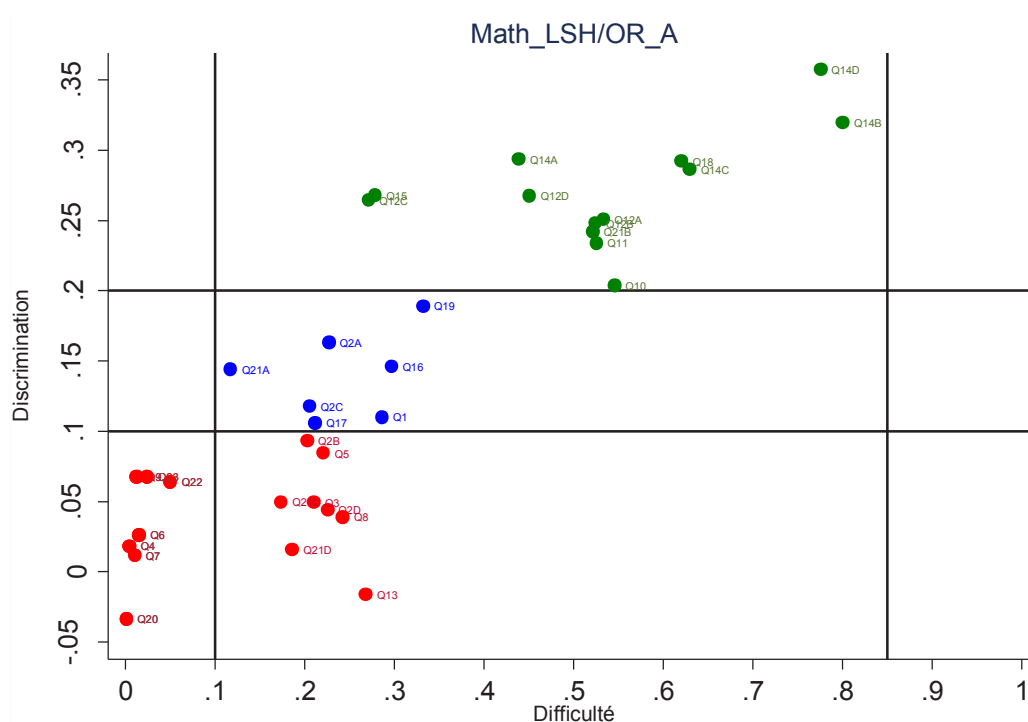
Graphique 1. Analyse des items du test B des mathématiques (tronc communs Sciences et Technique)



Source : Résultats des Pré-tests, INE, 2011

Quant aux tronc communs « Lettres et Sciences Humaines », la consistance interne des deux tests des mathématiques est assez faible ($\alpha=0.50$ pour la forme A et $\alpha=0.49$ pour la forme B). C'est la forme A qui a fait l'objet de redressement.

Graphique 2. Analyse des items du test B des mathématiques (trons communs Lettres et Originel)



Source : Resultats des Pré-tests, INE, 2011

Les items qui ne satisfont pas aux critères minimaux de difficulté, de discrimination et de fidélité ont été soit reformulés (items en bleu) ou totalement remplacés (items en rouge). Les items conformes (en vert) ont été maintenus.

«Environnement» et «Reproduction chez les plantes» pour le tronc commun «Sciences» et «Eau source de la vie» et «Homme et Environnement» pour les trons communs «Lettres et Sciences Humaines» et «Originel».

3.2. Test des Sciences de la Vie et la Terre

Les domaines de contenus que couvre le programme des Sciences de la Vie et la Terre sont

Les niveaux cognitifs visés sont la connaissance, l'application et le raisonnement.

Tableau 3. Tableau de spécification des SVT par domaine de contenus et par niveau cognitif

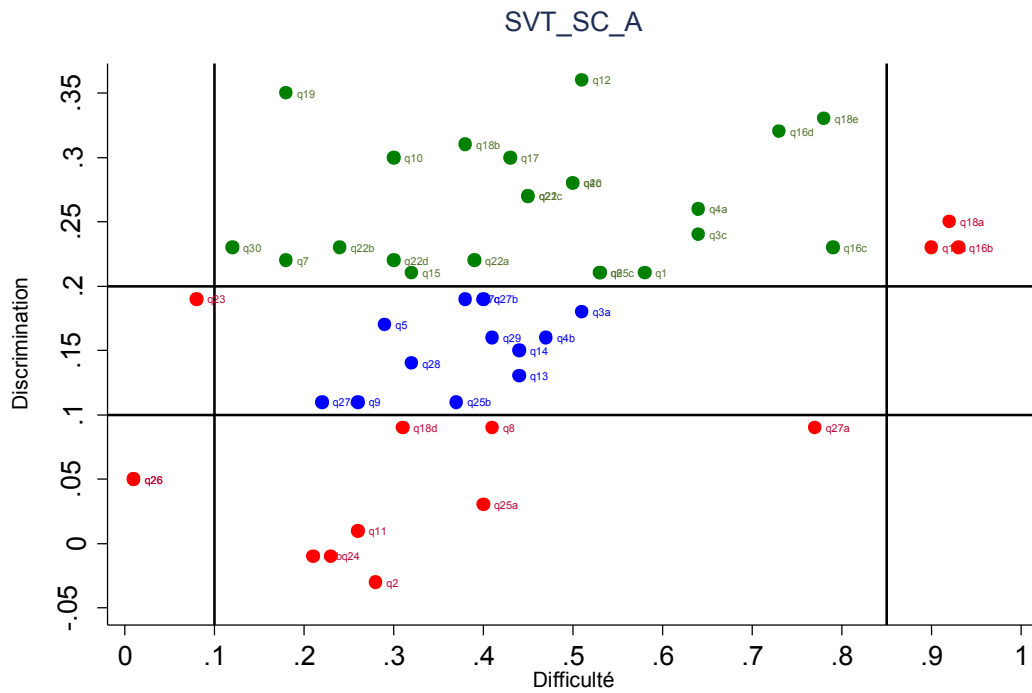
	Sciences			Lettres / originel		
	Environnement	Reproduction	Total	Eau source de la vie	Homme et Environnement	Total
Connaissance	15%	15%	30%	20%	20%	40%
Application	20%	20%	40%	20%	20%	40%
Raisonnement	15%	15%	30%	10%	10%	20%
Total	50%	50%	100%	50%	50%	100%

Source : Cadres de référence des Sciences de la Vie et de la Terre, INE-2010.

La consistance interne du test A est $\alpha=0.68$ pour le tronc commun « Sciences ». Quatre items font

baisser l'Alpha de CRONBACH, mais légèrement. C'est ce test qui a été sélectionné pour être adapté.

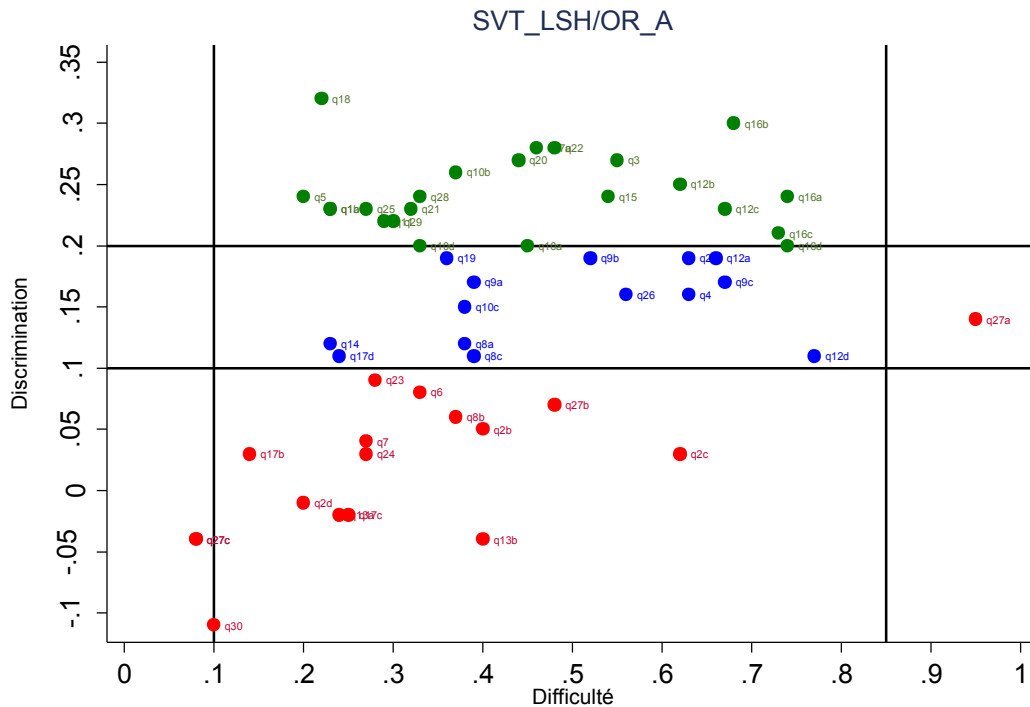
Graphique 3. Analyse des items du test A des Sciences de la Vie et la Terre (TC. Sciences)



Source : Resultats des Pré-tests, INE, 2011

Quant à la consistance interne du test A des troncs communs «Lettres et Sciences Humaines» et «Original», Alpha de CRONBACH est de 0.63, celui du test B est de 0.59. Le test A est retenu.

Graphique 4. Analyse des items du test A des Sciences de la Vie et la Terre (Lettres/Original)



Source : Resultats des Pré-tests, INE, 2011

Les items de la bande inférieure ont été écartés tandis que ceux du milieu ont été reformulés.

3.3. Test de Physique-Chimie

Le programme de physique-chimie (PC) des tronc communs scientifique et technique vise à renforcer chez les apprenants trois compétences essentielles:

la connaissance, l'application et le raisonnement, et ce dans trois domaines de contenu: la mécanique, l'électricité et la chimie. Le degré d'importance de chaque domaine de contenu d'une part, et de chaque niveau cognitif de l'autre, est illustré dans le tableau ci-après.

Tableau 4. Tableau de spécification de PC par domaine de contenu et par niveau cognitif

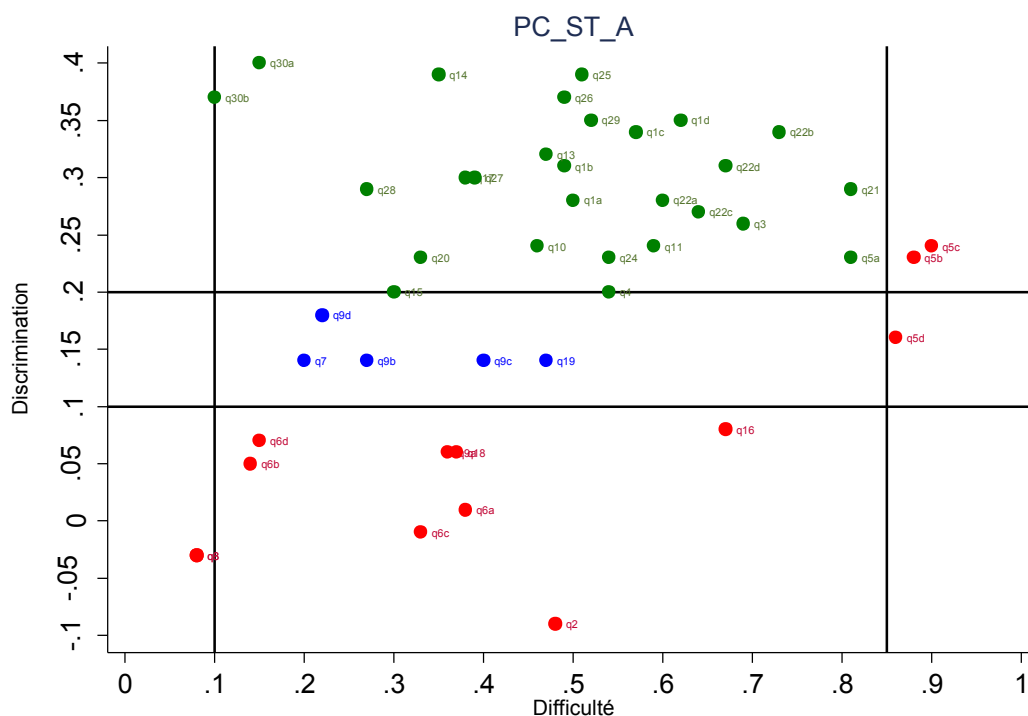
	Mécanique	Electricité	Chimie	Total
Connaissance	12,0%	12,0%	16,0%	40%
Application	10,5%	10,5%	14,0%	35%
Raisonnement	7,5%	7,5%	10,0%	25%
Total	30%	30%	40%	100%

Source : Cadres de référence de Physique-Chimie, INE-2010.

Le degré d'homogénéité du pré-test A est jugé très suffisant car sa consistance interne mesurée par Alpha de CRONBACH ($\alpha = 0.74$), est supérieure au seuil minimal requis. L'examen de la fidélité des

items montre que la consistance interne du test n'est influencée que par une seule question qui impacte négativement l'Alpha de CRONBACH.

Graphique 5. Analyse des items du test A de Physique-Chimie (Sciences/Technique)



Source : Résultats des Pré-tests, INE, 2011

L'analyse des items révèle que le pré-test de physique-chimie dans sa version A, contient trois questions très faciles et une seule très difficile. Pour ce qui est de la discrimination, le test comprend six questions non discriminantes avec des indices inférieurs à 0.10 et trois autres items incohérents (indices de discrimination négatifs). L'examen de la fidélité des items, montre qu'une seule question

impacte négativement la consistance interne du test.

3.4. Test de la langue Arabe

Le programme de la langue arabe des tronc communs couvre trois domaines de contenu : la lecture, les activités réflexives de la langue et

l'«expression et rédaction». S'ajoute à cela l'étude des œuvres pour les tronc communs « Lettres et sciences humaines » et « Originel ».

Ce programme vise également à renforcer, chez les apprenants, des compétences relatives à la

connaissance, à la compréhension, à l'application, à l'analyse, à la synthèse et à l'évaluation. Les tableaux suivants affichent les degrés d'importance de chaque domaine de contenu et de chaque niveau cognitif selon le volume horaire qui lui est consacré.

Tableau 5. Tableau de spécification de l'arabe par domaine de contenu et niveau cognitif

	Lecture	Activités réflexives sur la langue	Etude des Œuvres	Expression et Rédaction
Lettres et Sciences Humaines				
Connaissance	8,00%	6,25%	4,44%	**
Compréhension	12,00%	7,50%	4,44%	**
Analyse	12,00%	**	6,67%	**
Synthèse	4,00%	**	2,22%	**
Évaluation	4,00%	**	2,22%	**
Application	**	6,25%	**	**
Total	40%	20%	20%	20%
Originel				
Connaissance	8,00%	12,50%	2,22%	**
Compréhension	12,00%	15,00%	2,22%	**
Analyse	12,00%	**	3,33%	**
Synthèse	4,00%	**	1,11%	**
Évaluation	4,00%	**	1,11%	**
Application	**	12,50%	**	**
Total	40%	40%	10%	10%
Sciences et Technique				
Connaissance	6,25%	7,81%	-	-
Compréhension	12,50%	9,38%	-	-
Application	-	7,81%	-	-
Analyse	18,75%	-	-	-
Synthèse	6,25%	-	-	-
Évaluation	6,25%	-	-	-
Total	50%	25%		25%

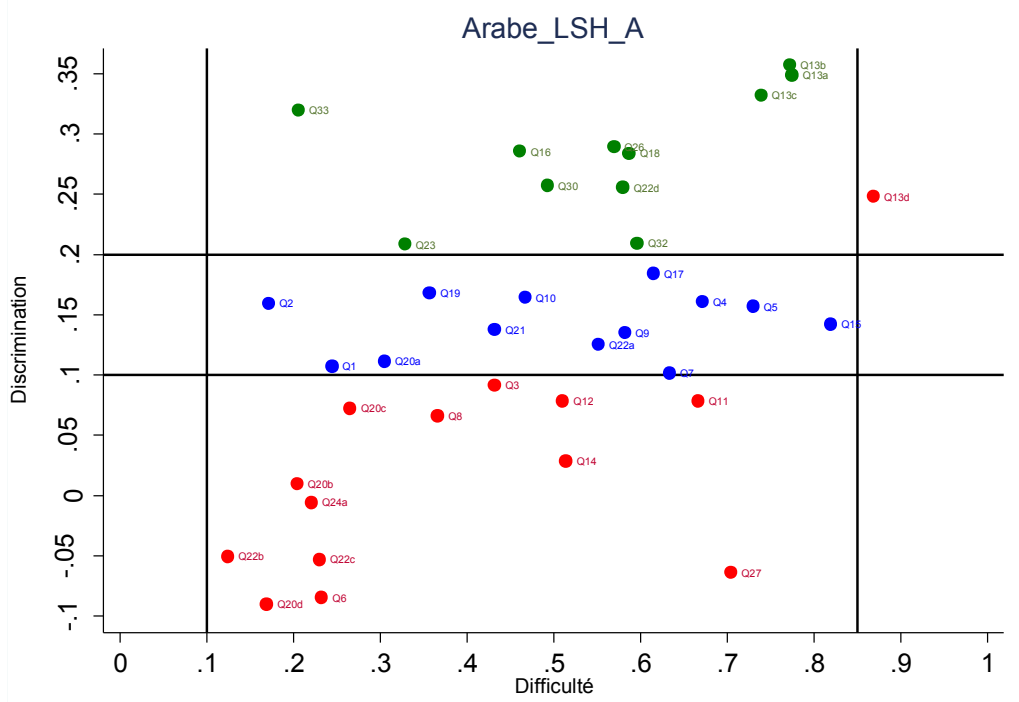
Source : Cadres de référence de la langue arabe, INE-2010.

L'homogénéité du test A des tronc communs «Lettres et Sciences Humaines» est considérée comme insatisfaisante car la valeur du coefficient Alpha de CRONBACH n'est que 0.53.

L'analyse de l'indice de fidélité montre qu'onze items réduisent légèrement la consistance interne de ce test car leurs indices α -i sont supérieurs au coefficient α du test.

L'analyse du graphique 6 montre que les indices de difficulté respectent les normes requises à l'exception d'un item qui est très facile. Les items peu discriminants ont été soit reformulés ou carrément remplacés.

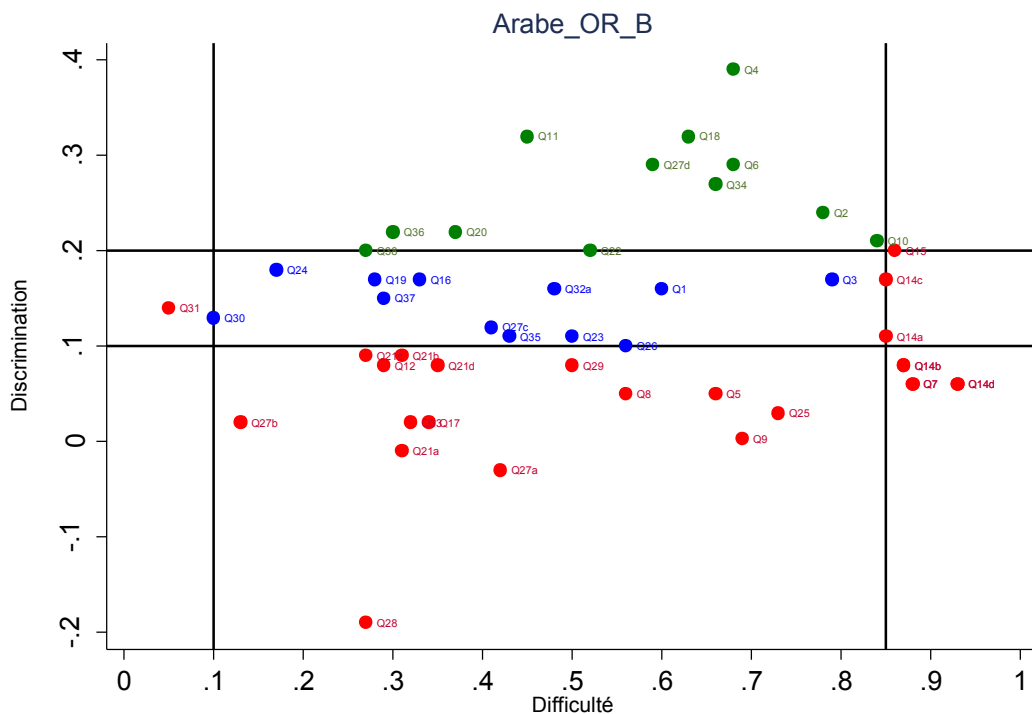
Graphique 6. Analyse des items du test A de la langue Arabe (Lettres et Sciences Humaines)



Source : Resultats des Pré-tests, INE, 2011

Quant au tronc commun « Original », c'est le consistence interne est relativement plus élevée test B qui était retenu pour amélioration car sa ($\alpha_B=0,57>\alpha_A=0,51$).

Graphique 7. Analyse des items du test A de la langue Arabe (Original)



Source : Resultats des Pré-tests, INE, 2011

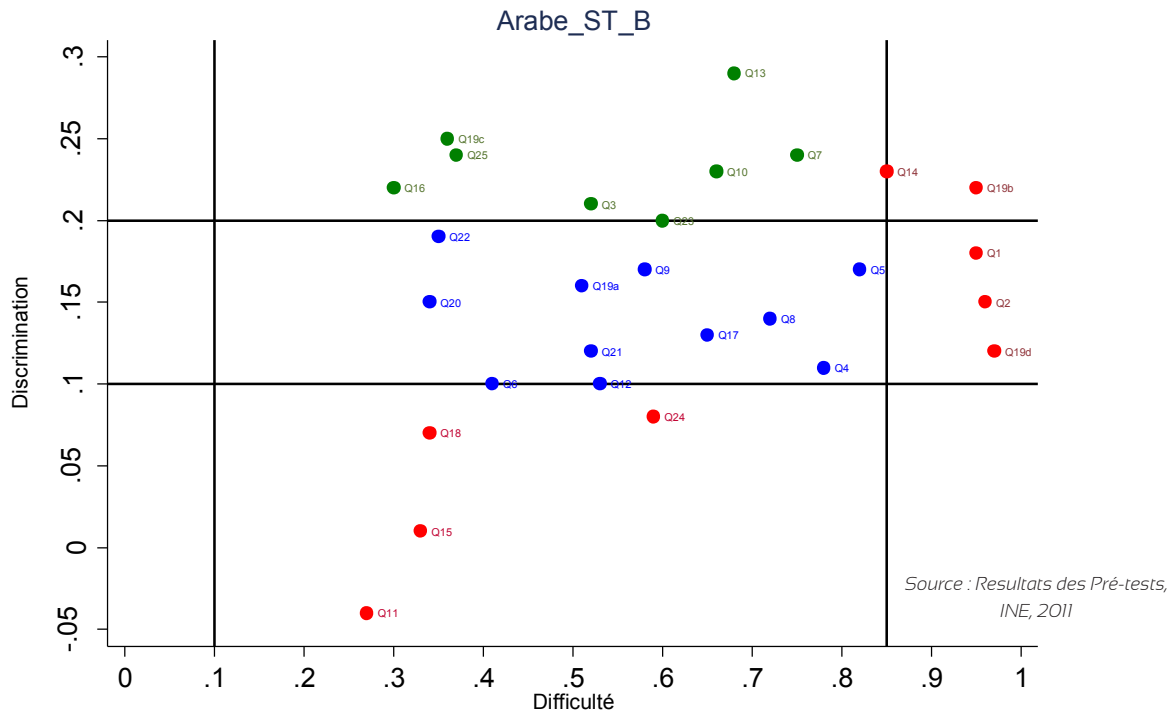
L'analyse du graphique 7 représentant les indices de difficulté et de discrimination fait ressortir qu'un seul item est jugé très difficile, et quatre items trop faciles.

En ce qui concerne le pouvoir discriminant, trois items ont des indices de discrimination négatifs et partant, nuisent à l'homogénéité du test : ils ont été éliminés.

En ce qui concerne les indices de fidélité, on constate qu'aucun item ne réduit la consistance interne de ce test puisqu'ils affichent tous un α -i inférieurs à l'Alpha de CRONBACH.

Le test B du pré-test de la langue arabe des tronc communs « Scientifique » et « Technique » affiche une homogénéité interne insuffisante avec un Alpha de CRONBACH de 0,47.

Graphique 8. Analyse des items du test B de la langue Arabe (Sciences/Technique)



L'analyse des items révèle que quatre questions sont très faciles, vu que leurs taux de réussite dépassent 85%. De plus, un seul item influe négativement sur la consistance interne car son α -i dépasse Alpha de CRONBACH du test.

3.5. Test de la langue Française

Les programmes de français dans l'enseignement secondaire qualifiant ciblent le développement la lecture et de la compétence rédactionnelle chez les lycées.

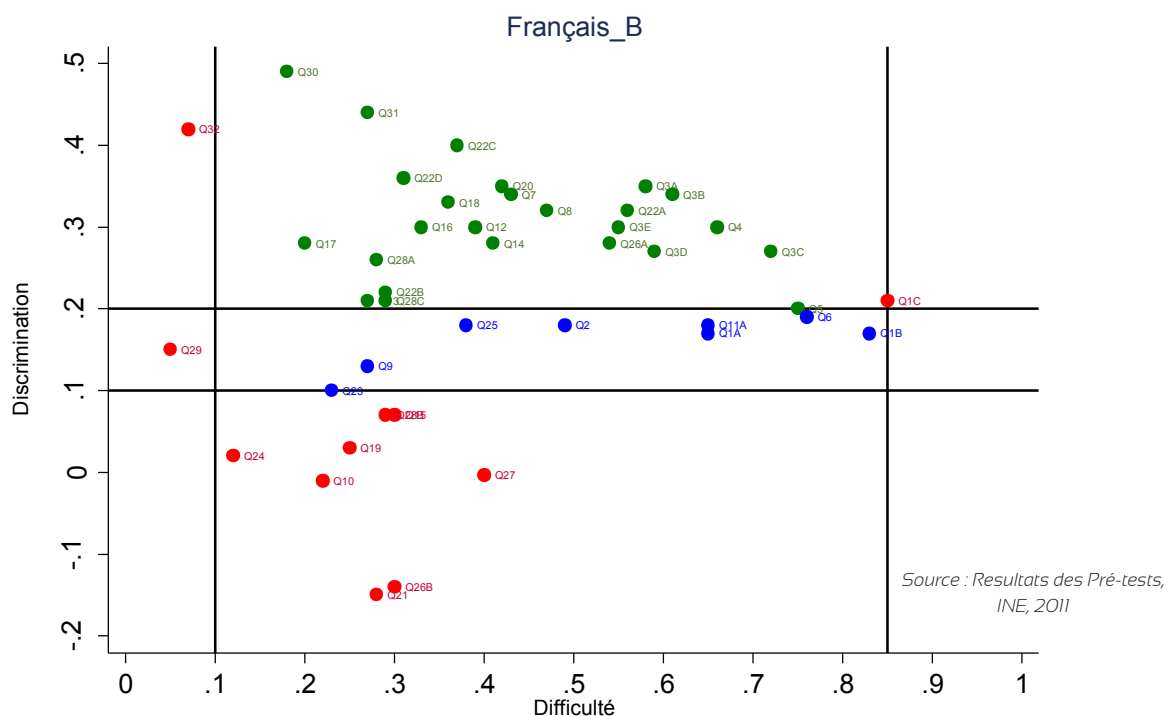
L'homogénéité du test B est insatisfaisante car la valeur du coefficient alpha de CRONBACH n'est que 0,44.

Tableau 6. Tableau de spécification du Français par domaine de contenus et niveau cognitif

	Compréhension de l'écrit	Production écrite
Connaissance	5%	-
Compréhension	10%	-
Analyse	25%	-
Synthèse	5%	-
Évaluation	5%	-
Réécriture	-	20%
Texte narratif	-	30%
Total	50%	50%

Source : Cadres de référence de la langue française, INE-2010.

Graphique 9. Analyse des items du test B de la langue Française



Le graphique 9 montre que deux items sont jugés trop difficiles et un seul trop facile.

En ce qui concerne le pouvoir discriminant, huit items ont été écartés car leurs indices de discrimination sont inférieurs à 0.10, tandis que sept autres ont été reformulés.

D'ailleurs, quatre items ont un coefficient α -i supérieur au coefficient Alpha de CRONBACH du test et, par conséquent, réduisent la fidélité du test : ces items ont été écartés.

3.6. Test de d'Histoire-Géographie

L'enseignement de l'Histoire-Géographie au secondaire qualifiant vise à développer, chez l'élève,

l'intelligence sociale et le sens critique, mettre à sa disposition les outils cognitifs et méthodologiques, le préparer à résoudre les problèmes et le rendre capable de comprendre les mécanismes de l'espace géographique.

Ainsi, ces compétences visées par l'enseignement de l'Histoire-Géographie portent sur l'utilisation des notions, la lecture et l'analyse des documents et la rédaction d'un article.

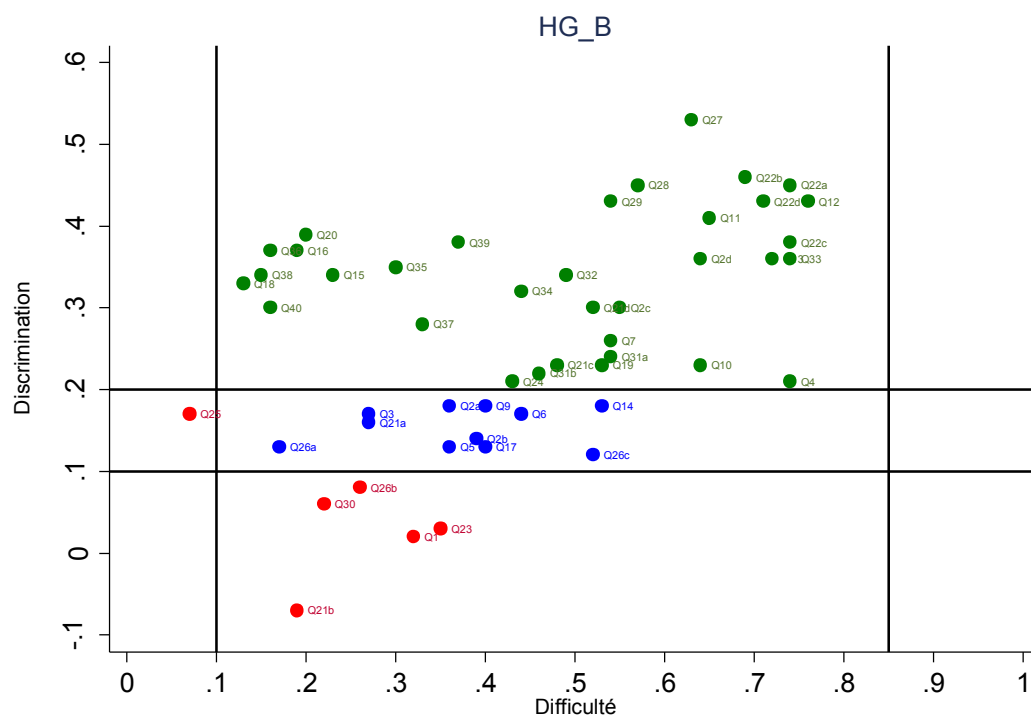
La valeur de l'Alpha de CRONBACH, aussi bien du test A que du test B de l'Histoire-Géographie est suffisante car elle dépasse 0.80, valeur largement suffisante pour avoir une bonne consistance interne.

Tableau 7. Tableau de spécification d'Histoire-Géographie par domaine et par niveau cognitif

	Utilisation des notions	Lecture et analyse des documents	Rédaction d'un article	Total
Histoire				
L'Europe de l'Ouest et évolution vers la modernité	7,5%	10%	7,5%	25%
Le monde islamique entre la transformation et la tradition	7,5%	10%	7,5%	25%
L'Europe occidentale et mise en place de la modernité durant le 17 ^{ème} et le 18 ^{ème} siècle.	7,5%	10%	7,5%	25%
La situation du monde islamique : tentatives de réforme et début de la domination européenne	7,5%	10%	7,5%	25%
Total	30%	40%	30%	100%
Géographie				
La terre : composants de l'environnement naturel	7,5%	10%	7,5%	25%
L'Homme et l'exploitation du domaine	7,5%	10%	7,5%	25%
L'écosystème et sa perturbation	7,5%	10%	7,5%	25%
Les efforts humains pour rétablir l'équilibre écologique et le développement durable	7,5%	10%	7,5%	25%
Total	30%	40%	30%	100%

Source : Cadres de référence de l'histoire-géographie, INE-2010.

Graphique 10. Analyse des items du test B de la langue Française



Les résultats de l'analyse psychométrique montrent que seule une question est très difficile, cinq items ne sont pas discriminants et aucune question ne réduit la consistance interne du test B.

4. Développement des questionnaires

Le niveau des acquis des élèves est tributaire de plusieurs facteurs. Qu'ils soient personnels, familiaux ou scolaires, ces facteurs influent, d'une manière ou d'une autre, les apprentissages. Afin de cerner ceux qui ont un impact significatif sur les performances scolaires des élèves du tronc commun du secondaire qualifiant, trois questionnaires ont été élaborés. Le premier est destiné aux élèves, le deuxième aux enseignants et le troisième aux directeurs des établissements. Chacun de ces questionnaires est divisé en trois sections.

Dans le questionnaire «Élève», la première section renseigne sur les caractéristiques personnelles, familiales et scolaires de celui-ci.

Les questionnaires «Enseignant» et «Directeur» portent sur les caractéristiques personnelles et professionnelles de ces derniers.

4.1. Validité de contenu

Afin de valider le contenu des trois questionnaires, l'INE a organisé un atelier, les 7 et 8 décembre 2015, réunissant des experts et des acteurs du système d'éducation et de formation, dont notamment des enseignants, des directeurs d'établissements, des inspecteurs, des conseillers d'orientation ainsi que

des formateurs.

Les participants se sont organisés en six groupes examinant chacun, et de façon détaillée, une thématique précise des questionnaires. Ainsi, le premier groupe a abordé la partie consacrée aux caractéristiques personnelles, familiales et scolaires des élèves. Les deuxième et troisième groupes se sont penchés sur l'étude des questions concernant les aspects personnels, professionnels et éducatifs des enseignants et directeurs, respectivement. Le quatrième groupe a examiné le volet portant sur les pratiques éducatives et organisationnelles, le cinquième a traité celui du climat scolaire et le sixième celui des problèmes sociaux et scolaires.

La méthode de travail a consisté à ce que chaque groupe fasse une lecture attentive de la partie du questionnaire dont il a la charge et note, pour chaque question, les remarques soulevées et les améliorations proposées. Cette étape s'est déroulée la première journée. Lors de la seconde, les rapporteurs des six groupes ont présenté les résultats de leurs travaux qui ont fait l'objet d'une discussion collective.

Toutes les propositions pertinentes, n'affectant pas les fondements théoriques et qui ont fait l'objet d'un consensus partagé par l'ensemble des participants, ont été prises en considération.

4.2. Validité du construit

Cette phase consiste à évaluer la validité des questionnaires via l'analyse factorielle exploratoire et l'analyse de fidélité.

Suivant le cadre théorique définissant l'environnement éducatif, les items portant sur les pratiques et sur le climat scolaire ont été regroupés en des échelles mesurant chacune d'elles une dimension bien précise de ces deux composantes.

Pour vérifier si les échelles retenues sont valides et s'appliquent au contexte marocain, une analyse factorielle a été menée afin de vérifier l'adéquation du modèle théorique avec les données empiriques issues de l'expérimentation.

Aussi, faut-il que chaque énoncé composant une échelle concoure effectivement à mesurer l'aspect visé par cette échelle, d'où le recours à l'analyse de fidélité qui a servi de moyen pour tester la fidélité du construit.

Pour mener ces deux analyses, une expérimentation de la version des questionnaires validés durant l'atelier susmentionné a été conduite dans quatre provinces de la région Rabat-Salé-Kénitra que sont : Rabat, Salé, Khémisset et Témara. Elle a concerné un

échantillon de 1167 élèves du tronc commun, ainsi que 233 enseignants et 42 directeurs d'établissements.

L'expérimentation des questionnaires s'est effectuée en ligne via le logiciel libre Lime Survey. Le choix de ce mode de passation est motivé par ses avantages multiples. En effet, il permet de recevoir les données en temps réel et partant, minimiser le coût et le temps engendrés par la collecte et la saisie traditionnelles des données. Le suivi de l'opération d'expérimentation des questionnaires a été assuré par des superviseurs formés à cette fin.

Avant leur administration auprès des établissements sélectionnés pour l'expérimentation, les questionnaires ont été testés dans un premier temps dans le lycée Abdellah Guennoun à Rabat. L'objectif était de s'assurer de la faisabilité de l'opération en termes d'adéquation du matériel et de connexion internet. Ainsi, les problèmes soulevés à ce stade ont permis d'améliorer la plateforme des questionnaires et de mieux cerner les conditions de passation.

Tableau 8. Taux de remplissage des questionnaires (Étude expérimentale)

	Taille échantillon	Participants	Taux
Élèves	1277	1167	91%
Enseignants	236	233	99%
Directeurs	42	40	95%

Source : étude expérimentale, PNEA 2016.

L'opération de collecte a duré une semaine et les données ainsi obtenues ont fait l'objet d'une analyse factorielle exploratoire et une analyse de fidélité.

4.2.1. Analyse factorielle exploratoire

Avant d'entamer l'extraction des facteurs, il faut s'assurer que les données se prêtent à l'analyse factorielle. En effet, les variables doivent satisfaire les conditions suivantes :

- Aucune variable ne doit être une combinaison linéaire parfaite d'une ou de plusieurs autres variables. Pour tester cela, on examine le déterminant de la matrice des items qui doit être très petit, sans pour autant être égal à 0, indiquant ainsi que la matrice est inversible;
- Les corrélations entre les items ne doivent pas être nulles. Pour tester cette hypothèse, on fait appel au test de sphéricité de Bartlett qui est un test de χ^2 et dont la probabilité doit être inférieure à 0.05 ;
- Les variables retenues doivent constituer un ensemble cohérent et une mesure adéquate du concept visé. Le test de Kaiser-Meyer-Olkin (KMO) permet de vérifier cette hypothèse. Dans ce cadre, un KMO supérieur ou égal à 0.90 indique un excellent ajustement des items aux facteurs latents, une valeur entre 0.80 à 0.90 indique un

bon ajustement, de 0.70 à 0.80, un ajustement moyen, de 0.60 à 0.70, un ajustement faible et pour les valeurs inférieures à 0.60, l'ajustement est insuffisant.

Le nombre de facteurs est déterminé par les échelles adoptées en suivant le modèle théorique développé par Janosz (1998). Néanmoins, ces facteurs doivent afficher une valeur propre supérieure ou égale à 1.

Pour décider du maintien d'une variable dans la solution factorielle, deux critères sont généralement examinés : la qualité de représentation et la saturation.

- Qualité de représentation : appelée également communauté, elle représente la partie de la variance partagée avec les autres variables. Pour qu'une variable soit retenue, elle doit avoir une communauté supérieure à 0.2 ;
- Saturation : mesure le degré d'association d'une variable à un facteur. Ainsi, pour interpréter un facteur, on recherche les variables qui ont les saturations les plus élevées en valeur absolue (>0.30).

4.2.2. Analyse de fidélité

La fidélité renvoie à la précision avec laquelle un instrument mesure une aptitude, une habileté ou

une échelle. Plusieurs méthodes permettent de mesurer cet aspect :

- Fidélité test-retest : permet d'évaluer la stabilité du construit dans le temps à travers la corrélation entre deux mesures prises à deux temps ;
- Fidélité entre formes parallèles ou équivalentes : indique à quel point deux ensembles différents d'items peuvent mesurer le même concept ;
- Consistance interne : permet d'évaluer jusqu'à quel point chacun des items mesure la même dimension. En ce sens, elle reflète le degré d'homogénéité des items constituant un instrument de mesure. Etant moins onéreuse et plus simple à mettre en œuvre, car elle ne nécessite qu'un seul test qui est administré une seule fois, cette méthode est la plus utilisée pour évaluer la fidélité d'un instrument de mesure.

L'Alpha de Cronbach est l'indice que l'on utilise le plus souvent pour évaluer la consistance interne d'un instrument. Compris entre 0 et 1, plus il est élevé, plus l'instrument de mesure est fidèle au sens de la consistance interne. Pour être consistant, l'Alpha d'une échelle de mesure doit au moins être égale à 0,70.

L'analyse factorielle porte sur les questionnaires destinés aux élèves et enseignants uniquement. Celui destiné aux directeurs d'établissements ne peut se prêter à de telles analyses en raison du nombre insuffisant d'observations (40 directeurs d'établissements).

4.2.3. Résultats

L'examen des résultats de l'analyse factorielle témoigne de l'adéquation de la solution globale, et ce pour les deux questionnaires. En effet, le déterminant de la matrice de corrélation est très petit indiquant qu'elle est inversible. Aussi, les corrélations entre les items ne sont pas nulles puisque le test de Bartlett est significatif avec une probabilité inférieure à 0.05 (p value= 0.000). De même, le test de Kaiser-Meyer-Olkin, qui avoisine la valeur de 0.90 dans le questionnaire enseignant et la dépasse dans celui de l'élève, indique un excellent ajustement des items aux facteurs latents.

D'autre part, toutes les variables affichent une qualité de représentation dépassant le seuil minimal requis (0.2).

4.3. Climat scolaire

Selon le cadre théorique adopté, le climat scolaire est constitué de cinq échelles : (i) le climat éducatif, (ii) le climat relationnel, (iii) le climat de sécurité, (iv) le climat de justice et (v) le climat d'appartenance.

Composée de 11 items, l'échelle du climat éducatif du questionnaire élève a un Alpha de Cronbach égal à 0.72. Cinq items, qui ne lui sont pas associés, ont été retirés, ce qui a impacté ce coefficient à la hausse avec une valeur de 0.76. N'étant pas corrélés au facteur reflétant le climat éducatif dans le questionnaire enseignant, les mêmes items ont également été retirés de ce questionnaire, donnant ainsi lieu à un Alpha de 0.84 et permettant la comparabilité de l'échelle entre les deux questionnaires.

Composé de 8 items dans le questionnaire élève et 9 items dans celui destinés aux enseignants, le climat relationnel affiche un alpha relativement faible pour le premier (0.69) et acceptable pour le second (0.81). Ainsi, en se basant sur le questionnaire enseignant pour lequel l'échelle du climat relationnel est plus fidèle, les trois variables reflétant les relations positives entre les élèves d'une part et entre ceux-ci et les enseignants de l'autre, sont attribuées au même facteur pour lequel ces variables affichent des corrélations statistiquement significatives (>0.30). Deux variables indiquant des relations négatives entre les élèves ou entre ces derniers et les enseignants ont été supprimées. Celle renseignant la relation entre les élèves et la direction de l'établissement a été, quant à elle, reformulée.

L'échelle mesurant le climat de sécurité contenait au début cinq variables dans les deux questionnaires, élève et enseignant. Sa fidélité en termes de consistance interne est de l'ordre de 0.65 pour le premier et 0.78 pour le second. Deux questions ne lui sont pas associées car elles ont une saturation inférieure à 0.30. L'une d'elle a été reformulée, l'autre retirée car les autres variables retenues mesurent l'information qu'elle est sensée capter.

Tableau 9. Résultats de l'analyse factorielle exploratoire sur le climat scolaire

Sous dimensions	Questions (enseignant)	Communauté	Solution factorielle					Alpha de Cronbach	
			1	2	3	4	5		
Climat éducatif	- يبذل تلامذتي جهدا كبيرا لتحسين مستواهم الدراسي	,691	,193	,790	,098	,000	,140	,794	
	- يطمح تلامذتي إلى التفوق في دراستهم	,631	,095	,711	,091	,004	,330	,792	
	- ينجز تلامذتي واجباتهم الدراسية	,452	,228	,577	,183	,172	,063	,822	
	- يفكر بعض تلامذتي في متابعة الدراسات العليا	,209	,038	,409	,046	,107	,163	,846	
	- يشجع الآباء أبناءهم على بذل الجهد للنجاح	,721	,155	,501	,082	,300	,591	,797	
	- تساعد الأسر أبناءها لمتابعة الدراسة	,612	,075	,460	,116	,316	,531	,815	
Climat relationnel	- يتعاون تلامذة هذه المؤسسة فيما بينهم بشكل إيجابي	,487	,255	,590	,153	,218	-,045	,664	
	- يحترم تلامذة هذه المؤسسة بعضهم البعض	,629	,337	,657	,145	,219	-,118	,608	
	- يتواصل تلامذة هذه المؤسسة مع أساتذتهم بتلقائية	,409	,317	,500	,189	,143	-,046	,749	
	- يوجد انسجام وتفاهم بين أساتذة هذه المؤسسة	,365	,327	,334	,367	,083	,063	,734	
	- يتعامل معي أساتذة هذه المؤسسة باحترام	,351	,022	,180	,521	,120	-,178	,782	
Climat scolaire	Climat de sécurité	- أشعر بالأمن داخل المؤسسة	,717	,346	,235	,155	,711	,108	,675
		- أشعر بالأمن بجوار المؤسسة	,536	,393	,216	,083	,503	,272	,813
		- يُحسّ الأساتذة و الإداريون بالأمن داخل المؤسسة	,667	,298	,217	,215	,690	,090	,714
Climat de justice	- يتّسم تعامل الإدارة معي بالعدل والمساواة	,547	,128	,145	,689	,169	-,079	,863	
	- يتعامل الأساتذة مع التلامذة بالعدل والمساواة	,550	,296	,165	,638	-,026	,166	,845	
	- تتعامل الإدارة مع التلامذة بالعدل والمساواة	,742	,220	,059	,779	-,012	,289	,787	
	- تتعامل الإدارة مع الأساتذة بالعدل والمساواة	,804	,278	,046	,816	,215	,115	,775	
Climat d'appartenance	- أحب متابعة التدريس بهذه المؤسسة	,502	,566	,289	,189	,227	-,103	,911	
	- معظم التلامذة يحبون الدراسة في هذه المؤسسة	,496	,670	,147	,083	,130	,036	,912	
	- معظم الأساتذة يحبون التدريس في هذه المؤسسة	,765	,829	,155	,107	,205	,026	,899	
	- أعضاء الإدارة يحبون العمل في هذه المؤسسة	,796	,848	,134	,198	,122	,060	,898	
	- ظروف التدريس في المؤسسة ملائمة للتصنيف	,682	,668	,229	,281	,316	,058	,901	
	- مستوى التحصيل الدراسي في هذه المؤسسة جيد	,621	,652	,333	,195	,206	,062	,904	
	- أغلب أساتذة هذه المؤسسة متحمسون للعمل	,551	,538	,232	,337	,079	,297	,912	
	- أغلب إداريي هذه المؤسسة يحبون عملهم	,592	,599	,172	,370	,065	,248	,908	

Source : étude expérimentale, PNEA 2016.

Les items composant le climat de justice sont tous corrélés significativement à un même facteur dont la consistance interne est satisfaisante, et ce, aussi bien pour le questionnaire élèves que pour celui concernant les enseignants. Leurs Alpha de Cronbach sont de l'ordre de 0.87 et 0.86 respectivement.

Onze items sont supposés former l'échelle du climat d'appartenance qui affiche un niveau de fidélité

excellent, atteignant 0.85 dans le questionnaire élève et dépassant 0.90 dans celui des enseignants. Un item, étant corrélé à un autre facteur et constituant une redondance avec un autre item, a été retiré. Dans le questionnaire élève, deux items influent négativement sur l'Alpha de Cronbach et possède une structure complexe reflétée par leur corrélés à plus d'un facteur. Ces deux items ont été supprimés des deux questionnaires afin d'obtenir une échelle comparable entre les élèves et les enseignants.

4.4. Pratiques éducatives et de gestion

Suivant le modèle théorique de Janosz (1998), cette dimension est représentée par six échelles reflétant les pratiques éducatives et une renseignant les pratiques organisationnelles. Il s'agit de la qualité de

l'enseignement, le temps consacré à l'enseignement, les opportunités d'investissement, le leadership et le style de gestion, le système de reconnaissance, l'emphase sur la réussite ainsi que le système d'encadrement.

Tableau 10. Résultats de l'analyse factorielle exploratoire sur les pratiques éducatives

Sous dimensions	Questions (élève)	Communaute	Solution factorielle							Alpha de Cronbach
			1	2	3	4	5	6	7	
Qualité de l'enseignement	- يستخدم/تستخدم أستاذ(ة) هذه المادة طرقا في التدريس تجعل المادة سهلة الفهم	,614	,747	,123	,078	,021	,135	-,023	-,025	,876
	- يشرح/تشرح أستاذ(ة) هذه المادة في بداية كل درس أهمية المضامين التي سيتم التطرق إليها	,613	,761	,145	,079	,016	,063	,004	-,014	,876
	- يشركونا/ تشركونا أستاذ(ة) هذه المادة أثناء شرح الدرس	,513	,698	,126	,070	,058	-,024	,049	,052	,880
	- يربط أستاذ(ة) هذه المادة المعارف القبلية بالدرس	,556	,722	,115	,103	,049	-,069	,021	,096	,879
	- يكيف أستاذ(ة) هذه المادة شرح الدرس مع مستوى التلامذة	,596	,740	,144	,075	-,004	,104	,001	,048	,877
	- يحرص أستاذ(ة) هذه المادة على أن يفهم التلامذة دروسهم	,748	,832	,146	,058	,038	,133	,014	,013	,872
	- يتأكد/ تتأكد أستاذ(ة) مادة اللغة الفرنسية من استيعابنا للدروس	,702	,802	,174	,053	,040	,086	,048	,019	,873
	- يقوم/ تقوم أستاذ(ة) هذه المادة بالتصحيح الجماعي للفروض	,282	,500	,073	,073	,072	-,104	,082	-,006	,894
	- يكلفنا/ تكلفنا أستاذ(ة) هذه المادة إنجاز بحوث في المنزل	,444	,615	,082	,112	,001	-,043	,075	-,069	,888
	- لا أتمكن من فهم هذه المادة لكون الأستاذ(ة) لا يبذل/تبذل جهدا كافيا في الشرح	,412	,266	-,082	,019	,519	,271	,054	,132	,910
Temps consacré à l'enseignement	- لا أتمكن من فهم هذه المادة لعدم انضباط التلامذة في القسم	,459	-,050	-,039	,027	,717	-,024	,023	,053	,752
	- لا أتمكن من فهم هذه المادة لعدم انضباطي في القسم	,364	-,030	-,059	-,029	,521	,160	,025	,271	,807
	- يقوم التلامذة بالتشويش خلال شرح أستاذ(ة) هذه المادة للدرس	,449	-,012	,013	-,012	,713	-,040	,030	-,086	,804
	- تتم إضاعة الوقت في بداية الدرس وبين الحصص	,452	,131	-,020	-,002	,670	,135	,009	-,027	,764
	- الوقت الذي يمضيه أستاذ(ة) هذه المادة في ضبط سلوك التلامذة يؤثر سلبا على الوقت المخصص للدروس	,469	,043	-,018	-,010	,739	,000	,026	-,102	,750
Opportunités d'investissement	- تهتم المؤسسة بالأنشطة الرياضية	,343	,174	,124	,475	-,063	,042	,088	,085	,798
	- تهتم المؤسسة بالأنشطة الثقافية	,472	,120	,168	,675	-,002	-,020	,095	-,022	,774
	- تنظم المؤسسة خرجات علمية	,288	,060	,173	,492	,022	-,040	,021	-,173	,796
	- ينخرط التلامذة بحماس في الأنشطة الرياضية	,365	,114	,121	,572	,013	,030	,035	,079	,811
	- يساهم التلامذة في الأنشطة الثقافية	,495	,082	,134	,719	,003	-,052	,043	-,054	,776
	- الأنشطة الموازية التي تنظمها المؤسسة مفيدة	,534	,098	,202	,724	-,024	,054	,046	,129	,771

	Sous dimensions	Questions (élève)	Communauté	Solution factorielle							Alpha de Cronbach	
				1	2	3	4	5	6	7		
Pratiques éducatives	Opportunités d'investissement	- معظم الأساتذة يساهمون في الأنشطة الموازية	,491	,114	,160	,704	-,002	,006	,033	,023	,811	,773
		- تنظم الإدارة حصص الدعم التربوي لفائدة التلامذة	,470	,175	,644	,112	-,007	,038	-,003	-,168		,838
	Leadership éducatif Style de gestion	- تخبر الإدارة ولي أمرك بسيرك الدراسي	,401	,133	,418	,159	,004	-,003	,593	-,075	,852	,844
		- تستدعي الإدارة ولي أمرك للاستفسار عن غياباتك	,284	,090	,241	,130	,017	,059	,496	,071		,859
		- تساعد الإدارة التلامذة في حل مشاكلهم الدراسية	,605	,194	,752	,133	-,032	,040	,111	-,061		,817
		- تنصت الإدارة إلى التلامذة قبل اتخاذ القرارات التي تعنيهم	,628	,131	,741	,143	-,073	,050	,066	-,127		,818
		- توفر الإدارة فرصا للتلامذة ليعبروا عن آرائهم	,668	,141	,795	,167	-,075	,067	-,011	-,088		,816
		- ينتخب التلامذة من يتحدث باسمهم مع الإدارة	,334	,159	,552	,104	-,036	-,037	-,019	-,007		,845
		- تعمل الإدارة جهدها لتوفير مستلزمات التعلم	,550	,174	,677	,210	,011	,006	,083	,183		,829
		- يعوض الأساتذة الحصص التي يتغيبون فيها	,409	,219	,558	,129	-,062	-,007	,056	,049		,840

Source : étude expérimentale, PNEA 2016.

Pour le questionnaire élève, tous les items ont une qualité de représentation acceptable dépassant 0.2. Les quatre premières échelles parmi les sept susmentionnées sont confirmées par l'analyse factorielle avec des niveaux de fidélité importants, allant de 0.80 à 0.89. Cependant, un item parmi ceux composant l'échelle «qualité d'enseignement» affecte négativement l'Alpha de Cronbach et est plus corrélé avec l'échelle «temps consacré à l'enseignement». De même, un item supposé appartenir à l'échelle «opportunité d'investissement» a rejoint celle concernant le «leadership et style de gestion».

Les trois échelles restantes ont, quant à elles, des niveaux de fidélité faibles, les items les composant sont corrélés à d'autres échelles ou ont une qualité de représentation faible (<0.2). Dans le premier cas, ces items sont gardés pour faire partie des échelles auxquelles ils sont associés, dans le second, ces items sont retirés.

Concernant le questionnaire enseignant, trois échelles parmi les sept précitées sont validées par l'analyse factorielle avec une fidélité dépassant 0.78. Il s'agit du temps consacré à l'enseignement, aux opportunités d'investissement, au leadership et au style de gestion. Pour ce qui est de la qualité

de l'enseignement, cette échelle a une fidélité faible du point de vue de la consistance interne (Alpha de Cronbach < 0.5). Elle est composée de 9 items dont cinq sont corrélés à d'autres facteurs.

Tout comme le questionnaire élève, les échelles concernant le système de reconnaissance, l'emphase sur la réussite et le système d'encadrement ont une fidélité faible et les items les composant sont associés à d'autres facteurs. Ainsi, de nouveaux items ont été développés pour faire partie des dites échelles.

4.5. Problèmes sociaux et scolaires

Pour le questionnaire élève, toutes les questions affichent une très bonne qualité de représentation puisque l'indice de communauté est supérieur à 0.2. La consistance interne répond, elle aussi, aux exigences requises dans la mesure où chaque Alpha de Cronbach est supérieur à 0.80, et ce pour les deux dimensions sociale et scolaire.

La solution factorielle est composée de trois facteurs. Le premier regroupe les questions rattachées aux problèmes vécus, le deuxième rassemble les questions sur les problèmes agis et le troisième concerne les problèmes subis.

Tableau II. Résultats de l'analyse factorielle exploratoire sur les problèmes sociaux et scolaires

Sous dimensions	Questions (élève)	Communauté	Solution factorielle			Alpha de Cronbach	
			1	2	3		
Problèmes sociaux	Vécus	ما وتيرة حدوث ما يأتي : تلاميذ يدخنون السجائر داخل المؤسسة	,710	,684	,190	,193	,955
		ما وتيرة حدوث ما يأتي : تلامذة يتعاطون المخدرات داخل المؤسسة	,750	,710	,207	,179	,955
		ما وتيرة حدوث ما يأتي : تلامذة يستهلكون المشروبات الكحولية داخل المؤسسة	,689	,679	,318	,111	,955
		ما وتيرة حدوث ما يأتي : تباع المخدرات داخل المؤسسة	,679	,673	,301	,144	,955
		ما وتيرة حدوث ما يأتي : تباع المخدرات بالقرب من المؤسسة	,547	,619	,157	,264	,955
		ما وتيرة حدوث ما يأتي : تلامذة يتعرضون للشتيم أو التهديد داخل المؤسسة	,707	,709	,085	,348	,954
		ما وتيرة حدوث ما يأتي : أساتذة يتعرضون للشتيم أو التهديد داخل المؤسسة	,690	,734	,176	,233	,955
		ما وتيرة حدوث ما يأتي : تلامذة يتعرضون للاعتداء داخل المؤسسة .	,732	,735	,108	,352	,954
		ما وتيرة حدوث ما يأتي : أساتذة يتعرضون للاعتداء داخل المؤسسة	,690	,687	,234	,205	,955
		ما وتيرة حدوث ما يأتي : تلامذة يتعرضون للسرقة داخل المؤسسة	,625	,610	,033	,406	,955
	ما وتيرة حدوث ما يأتي : أساتذة يتعرضون للسرقة داخل المؤسسة	,550	,619	,204	,182	,955	
	ما وتيرة حدوث ما يأتي : ينعدم الأمن بجوار المؤسسة	,406	,512	,064	,297	,956	
	ما وتيرة حدوث ما يأتي : تلامذة لا يحافظون على سلامة و نظافة مرافق المؤسسة	,376	,500	,072	,183	,957	
	Agis	ما وتيرة قيامك بما يأتي : عدم المحافظة على سلامة ونظافة مرافق المؤسسة	,599	,190	,691	,208	,956
		ما وتيرة قيامك بما يأتي : التهديد أو الشجار داخل المؤسسة	,737	,227	,763	,311	,955
		ما وتيرة قيامك بما يأتي : تدخين السجائر داخل المؤسسة	,813	,221	,783	,284	,955
		ما وتيرة قيامك بما يأتي : تعاطي المخدرات داخل المؤسسة	,828	,236	,774	,267	,955
		ما وتيرة قيامك بما يأتي : استهلاك المشروبات الكحولية داخل المؤسسة	,744	,299	,738	,167	,955
		ما وتيرة قيامك بما يأتي : الشجار بجوار المؤسسة	,718	,150	,710	,388	,955
		ما وتيرة قيامك بما يأتي : استعمال وسائل التواصل الاجتماعي للإساءة للغير	,725	,204	,735	,372	,955
Subis	ما وتيرة تعرضك لما يأتي : الشتم داخل المؤسسة	,654	,290	,312	,639	,955	
	ما وتيرة تعرضك لما يأتي : التهديد أو الاعتداء داخل المؤسسة	,741	,322	,353	,661	,954	
	ما وتيرة تعرضك لما يأتي : السرقة داخل المؤسسة	,697	,245	,256	,733	,955	
	ما وتيرة تعرضك لما يأتي : التهديد أو الاعتداء بجوار المؤسسة	,802	,255	,269	,798	,954	
	ما وتيرة تعرضك لما يأتي : السرقة بالقرب من المؤسسة	,770	,203	,248	,786	,955	
	ما وتيرة تعرضك لما يأتي : التحرش داخل المؤسسة	,715	,254	,272	,660	,955	
	ما وتيرة تعرضك لما يأتي : التحرش بجوار المؤسسة	,735	,171	,264	,697	,955	
	ما وتيرة تعرضك لما يأتي : الإساءة عبر وسائل التواصل الاجتماعي	,654	,229	,380	,662	,954	
Problèmes scolaires	Vécus	ما وتيرة حدوث ما يأتي : تلامذة يلجؤون إلى الغش في الامتحانات	,435	,449	,156	,075	,845
		ما وتيرة حدوث ما يأتي : يتساهل الأساتذة مع التلامذة الذين يغشون	,379	,478	,264	,042	,855
		ما وتيرة حدوث ما يأتي : تلامذة يتغيبون عمدا عن حصص الدرس	,551	,565	,151	,066	,838
		ما وتيرة حدوث ما يأتي : أساتذة يتغيبون عمدا عن حصص الدرس	,579	,575	,273	,123	,838

Sous dimensions	Questions (élève)	Communauté	Solution factorielle			Alpha de Cronbach	
			1	2	3		
Problèmes scolaires	Vécus	ما وتيرة حدوث ما يأتي : تلامذة يتأخرون عمدا عن حصص الدرس	,554	,564	,198	,047	,835
		ما وتيرة حدوث ما يأتي : أساتذة يتأخرون عمدا عن حصص الدرس	,602	,602	,238	,135	,839
	Agis	ما وتيرة قيامك بما يأتي : الغش في فروض المراقبة المستمرة والامتحانات	,523	,256	,584	,166	,841
		ما وتيرة قيامك بما يأتي : إزعاج زملائك في القسم	,584	,228	,666	,189	,842
		ما وتيرة قيامك بما يأتي : التغيب عن حصص الدرس بدون عذر	,638	,264	,726	,149	,839

Source : étude expérimentale, PNEA 2016.

En ce qui concerne le questionnaire enseignant, les items renseignant les « problèmes agis » affichent des communautés relativement faibles. L'analyse descriptive de ces items a montré qu'ils ne sont pas discriminants. De plus, ils ne sont pas cohérents avec les autres items composant les problèmes sociaux et scolaires puisqu'ils affectent l'Alpha de Cronbach à la baisse. Les problèmes vécus et subis sont, quant à eux, confirmés dans la solution factorielle et forment deux facteurs distincts.

5. Plan d'échantillonnage

Vu l'importance que revêt l'étude PNEA 2016 en tant qu'outil d'évaluation de la qualité des apprentissages des lycéens marocains, le choix de l'approche méthodologique du plan d'échantillonnage constitue une étape importante dans la mise en place de ce dispositif.

L'Instance Nationale d'Évaluation a préparé le plan d'échantillonnage en tenant compte de différents critères méthodologiques notamment le niveau

d'inférence, le niveau de comparabilité et le degré de précision des indicateurs.

Le choix retenu est un plan d'échantillonnage stratifié à deux degrés. La stratification explicite concerne le niveau régional et le type d'établissement (privé ou public). Au sein de chaque strate explicite (sauf pour la strate « privé »), une stratification implicite par milieu (Rural/Urban) a été effectuée.

De plus, chaque tronc commun est considéré comme un environnement statistique indépendant. Ainsi, un plan d'échantillonnage est établi pour chaque tronc commun séparément. Une estimation optimisant le nombre d'établissements retenus est menée pour tenir compte du nombre de troncs communs présents dans chaque établissement.

Cet échantillonnage présente un scénario de tirage d'une classe unique par tronc commun et par établissement échantillonné.

Tableau 12. Répartition des établissements par région, milieu et type d'enseignement

Région	Public		Privé	Total
	Rural	Urbain		
Tanger-Tetouan-Al Hoceima	8	44	7	59
Oriental	4	44	1	49
Fès-Meknès	7	44	6	57
Rabat-Salé-Kénitra	6	44	13	63
Béni Mellal-Khénifra	14	31	1	46
Grand Casablanca-Settat	4	55	19	78
Marrakech-Safi	10	45	7	62
Drâa-Tafilalet	18	27	1	46
Souss-Massa	11	39	16	66
Provinces du sud	5	42	8	55
Maroc	87	415	79	581

Tableau 13. Répartition des classes par région, milieu et type d'enseignement

Région	Public		Privé	Total
	Rural	Urbain		
Tanger-Tetouan-Al Hoceima	16	81	8	105
Oriental	8	87	1	96
Fès-Meknès	14	80	6	100
Rabat-Salé-Kénitra	12	85	13	110
Béni Mellal-Khénifra	27	65	1	93
Grand Casablanca-Settat	8	96	19	123
Marrakech-Safi	20	78	7	105
Drâa-Taflalet	35	56	1	92
Souss-Massa	22	73	17	112
Provinces du sud	10	80	8	98
Maroc	172	781	81	1034

Tableau 14. Estimation des effectifs par région, milieu et type d'enseignement

Région	Public		Privé	Total
	Rural	Urbain		
Tanger-Tetouan-Al Hoceima	490	2800	145	3435
Oriental	281	3129	22	3432
Fès-Meknès	499	2907	147	3553
Rabat-Salé-Kénitra	425	3029	287	3741
Béni Mellal-Khénifra	937	2450	31	3417
Grand Casablanca-Settat	291	3519	446	4257
Marrakech-Safi	738	3036	176	3949
Drâa-Taflalet	1181	2083	23	3287
Souss-Massa	794	2769	336	3898
Provinces du sud	304	2574	144	3021
Maroc	5938	28294	1757	35989

Les effectifs de l'échantillon se répartissent ainsi :

- 16.952 pour le tronc commun « Lettres & Sciences Humaines », soit 47% ;
- 16.021 pour le tronc commun « Sciences », soit 45% ;
- 1.886 pour le tronc commun « Technique », soit 5% ;
- 1.130 pour le tronc commun « Originel », soit 3%.

Les résultats présentés tiennent compte du nombre minimal d'élèves à retenir pour chaque approche. Celui-ci est déterminé en fonction d'une erreur d'échantillonnage préalablement établie : $\pm 0,2s$ pour les moyennes et $\pm 0,10\%$ pour les pourcentages avec un intervalle de confiance de l'ordre de 95%.

6. Collecte et exploitation des données

En plus des tests et des questionnaires, des guides de passation, de correction et de contrôle de qualité sont élaborés en vue d'uniformiser et de standardiser les opérations d'évaluation.

Par ailleurs, une formation en cascade est assurée à tous les intervenants : après avoir été formés,

les formateurs régionaux assurent la formation des coordonnateurs provinciaux, qui à leur tour, forment les coordonnateurs locaux (directeurs d'établissements concernés).

Les préparatifs logistiques de la passation des tests et l'administration des questionnaires s'est opérée en étroite concertation avec le Département de l'Éducation Nationale.

6.1. Collecte des données

Cette opération consiste à effectuer des contrôles sur les données brutes et à en corriger les anomalies. Le choix de la solution informatique pour la passation des questionnaires et des tests a permis de réduire significativement les erreurs liées à la saisie manuelle des données, dans la mesure où on a enregistré une quasi-absence de valeurs aberrantes.

En cela, le nettoyage des bases de données s'est principalement porté sur le traitement des doublons. Ces derniers, se présentent lorsque l'identifiant n'est pas unique. Dans ce cas, si toute la ligne est répétée

on garde un seul enregistrement et on supprime les autres. Par contre, si seul l'identifiant est dupliqué et les autres items/variables sont différents alors on retient l'enregistrement pour lequel l'élève a obtenu le score le plus élevé, dans le cas des tests, et celui dont le nombre de variables renseignées est le plus grand, dans le cas des questionnaires.

Cette étape a été précédée par la suppression des observations qui sont vides, la correction des identifiants erronés ainsi que la vérification de la participation des établissements de l'échantillon, et ce à travers l'appariement des listes des échantillons avec les bases de données. Lorsqu'un code établissement est inexistant, une recherche dans les clés USB, qui ont servi à la passation, est menée, et ce en se référant à la région à laquelle appartient cet établissement. Si cette recherche révèle que ledit établissement existe et que les données y afférentes sont enregistrées dans les clés qui ont été fournies, alors il s'agit d'une omission lors de la consolidation des données. Les établissements ainsi omis sont, dans ce cas, ajoutés aux bases de données. Dans l'autre cas de figure, les établissements sont considérés comme étant non participants. La même démarche a été adoptée pour les enseignants. Les élèves n'ayant pas passé le test d'une matière donnée entrent, quant à eux, dans la catégorie des non participants pour cette matière. Leur nombre ainsi que celui des établissements non participants sont pris en considération lors de l'ajustement des poids d'échantillonnage.

6.2. Exploitation des données

Après le nettoyage des bases de données, on procède au recodage des réponses aux items en attribuant 1 à la réponse correcte et 0 aux autres choix. Cela permet de calculer le taux de réussite de chaque item. Pour calculer les scores, le calcul préalable des poids s'impose.

L'échantillonnage se fait selon trois niveaux : établissement, classe et élève. A chaque niveau correspond un poids basique d'échantillonnage qui est l'inverse de la probabilité de sélection à ce niveau, ajusté pour la non-participation. Le poids total d'échantillonnage de chaque élève est la résultante du produit des poids des trois niveaux.

Les quatre troncs communs de l'étude constituent quatre univers indépendants. De ce fait, l'échantillonnage est effectué pour chacun d'eux indépendamment des trois autres. Il s'en suit que le calcul des poids est fait par tronc commun.

6.2.1. La pondération au niveau de l'établissement

Etant donné que pour chaque strate explicite, les établissements ont été sélectionnés proportionnellement à leur taille, le poids basique de

l'établissement k est donné par :

$$B_{et}^k = \frac{T}{n * t_k}$$

Avec n le nombre d'établissements de l'échantillon de la strate, t_k est la taille de l'établissement k , correspondant au nombre d'élèves de cet établissement :

$$T = \sum_{k=1}^N t_k$$

Avec N est le nombre total des établissements de la strate.

Pour les troncs communs technique et originel, le poids basique de chaque établissement vaut 1 puisque tous les établissements produisant ces types d'enseignement ont été échantillonnés.

Lorsqu'un établissement de l'échantillon ne participe pas, son poids basique doit être ajusté afin de compenser la réduction de la taille de l'échantillon due à la non-participation. Cet ajustement est exprimé par la formule suivante :

$$A_{et} = \frac{n_p + n_{np}}{n_p}$$

Avec n_p est le nombre d'établissements échantillonnés et participant à l'étude, n_{np} est le nombre d'établissements qui appartiennent à l'échantillon mais n'ayant pas participé.

Le poids de l'établissement est le produit de son poids basique et le coefficient d'ajustement :

$$Pa_{et}^k = B_{et}^k * A_{et}$$

6.2.2. La pondération au niveau de la classe

Si un établissement k est sélectionné, une classe est choisie, avec une probabilité égale, parmi toutes les classes de cet établissement. Son poids basique correspond à l'inverse de sa probabilité de sélection et est donc égale au nombre total de ses classes, noté C^k :

$$B_{cl}^k = C^k$$

Comme une seule classe est tirée parmi toutes les classes d'un établissement échantillonné, la non-participation de celui-ci implique celle de la classe lui appartenant. Ainsi, l'ajustement à ce niveau est le même effectué au premier niveau. Partant, toutes les classes de l'établissement k de l'échantillon ont pour poids:

$$Pa_{cl}^k = B_{cl}^k = C^k$$

6.2.3. La pondération au niveau de l'élève

Le troisième niveau d'échantillonnage consiste à choisir les élèves qui vont participer à l'étude. Puisque l'échantillonnage à l'intérieur d'une classe est exhaustif, c'est-à-dire tous les élèves de cette dernière sont choisis, le poids basique de chacun d'entre eux est égal à 1. Ainsi pour les élèves de la classe j de l'établissement k on note :

$$B_{el}^{jk} = 1$$

Ce poids est ajusté pour la non-participation par la formule ci-après :

$$A_{el}^{jk} = \frac{n_p^{jk} + n_{np}^{jk}}{n_p^{jk}}$$

Avec n_p^{jk} est le nombre d'élèves participants dans la classe j de l'établissement k , c'est-à-dire les élèves qui ont, pour une matière donnée, un score et n_{np}^{jk} le nombre d'élèves de cette classe qui sont censés avoir un score mais qui n'ont pas pris part à l'évaluation. Partant, le poids ajusté des élèves est donné par :

$$Pa_{el}^{jk} = B_{el}^{jk} * A_{el}^{jk} = A_{el}^{jk}$$

Le poids total d'échantillonnage des élèves de la classe j appartenant à l'établissement k est le produit des trois poids ajustés:

$$P^{jk} = Pa_{et}^k * Pa_{cl}^k * Pa_{el}^{jk}$$

7. Passation des tests et administration des questionnaires

Une solution informatique a été conçue sur une plateforme enregistrée sur une clé USB intégrant l'ensemble des outils nécessaires à la passation des tests et l'administration des questionnaires et ce, dans le but de :

- Permettre aux élèves la passation Offline des tests et des questionnaires en minimisant les risques et les problèmes liées au matériel informatique disponible dans les salles multimédia ;
- Fonctionner avec un minimum d'infrastructure et de pré-requis informatiques ;
- Assurer la sécurité et la confidentialité des tests ainsi que les réponses et les données personnelles des acteurs ;
- Avoir des données consolidées dans les meilleurs délais.

Avant d'être adoptée, cette solution a été testée dans plusieurs établissements de la région Rabat-Salé-Kénitra.

Suite à cette expérience, il s'est avéré que la mise en place de cette solution au niveau de chaque

établissement nécessite :

- Une salle multimédia équipée d'au moins 15 ordinateurs opérationnels ;
- Une instauration d'un clavier en lettres arabe et latin ;
- La disponibilité d'une souris fonctionnelle pour chaque ordinateur ;
- L'existence d'un port USB fonctionnel dans chaque ordinateur ;
- L'installation d'un mode d'exploitation XP et plus.

En outre, une journée de présentation de la solution et de formation a été organisée le 30 mars 2016, en faveur des coordonnateurs régionaux du Programme GÉNIE.

Au cours du mois avril 2016, les équipes du Programme GÉNIE ont veillé à la préparation des salles multimédia et leur dotation en matériels et logiciels nécessaires à la passation. Une clé USB test a été transmise à chaque coordonnateur provincial pour tester la solution au niveau des établissements de la direction provinciale dont il relève.

Un guide des opérations de l'étude principale, spécifiant les missions de chaque intervenant, a été élaboré et adressé aux coordonnateurs régionaux. Ces derniers ont transmis aux coordonnateurs provinciaux et aux directeurs d'établissements concernés le guide des opérations lors d'une journée de formation organisée dans chaque académie.

Pour la passation Offline de l'étude principale de l'évaluation PNEA2016, un lot de 12.000 clés USB a été acquis par le Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique. La solution contenant le code MASSAR de chaque élève ainsi que les instruments de mesure (tests et questionnaires) pour chaque région ont été intégrés dans une clé USB. Ces clés ont été dupliquées par région à raison de 20 clés USB par établissement concerné.

Ainsi, les clés par région portant des étiquettes d'identification ont été mises en lots : le lot de chaque région contient autant d'enveloppes que de directions provinciales. Ces enveloppes contiennent des sous-enveloppes de 20 USB par établissement.

Ces lots ont été remis aux directeurs d'académie le 25 mai 2016 lors d'une journée d'information et de sensibilisation organisée par l'Instance Nationale d'Évaluation.

Pour faciliter la passation des tests et l'administration des questionnaires Offline, les directeurs d'établissements concernés ont été invités à nommer des passateurs parmi les enseignants d'informatique, ou à défaut, des enseignants qui maîtrisent l'outil informatique.

Pour s'assurer du bon déroulement de l'étude, un guide de qualité a été élaboré et partant, les contrôleurs de qualité doivent veiller à la conformité des opérations sur le terrain aux dispositions prescrites dans le guide des opérations. En fait, les conditions de passation doivent être uniformes

et standardisées pour éviter tout biais lié aux différences des circonstances de passation.

L'étude principale PNEA 2016 s'est déroulée du 2 au 20 mai 2016.

La participation à l'étude se présente ainsi :

Tableau 15. Effectifs des élèves participants à l'étude PNEA-2016 (Public & Privé)

Régions	Tronc Commun											
	Lettres			Sciences			Technique			Originel		
	E	P	T	E	P	T	E	P	T	E	P	T
1	1738	1621	93%	1490	1409	95%	143	136	95%	221	181	82%
2	1664	1555	93%	1487	1425	96%	156	150	96%	118	112	95%
3	1720	1516	88%	1633	1508	92%	163	147	90%	107	98	92%
4	1701	1583	93%	1679	1633	97%	235	229	97%	115	110	96%
5	1658	1563	94%	1581	1533	97%	132	127	96%	41	40	98%
6	1606	1474	92%	2067	1985	96%	460	434	94%	128	109	85%
7	1790	1712	96%	1718	1639	95%	245	240	98%	201	174	87%
8	1554	1491	96%	1519	1467	97%	88	86	98%	129	91	71%
9	1905	1748	92%	1637	1579	96%	189	189	100%	145	144	99%
10	661	631	95%	585	565	97%	40	39	98%	49	45	92%
11	775	736	95%	564	544	96%	20	19	95%	-	-	-
12	152	124	82%	132	127	96%	36	34	94%	11	7	64%
National	16924	15754	93%	16092	15414	96%	1907	1830	96%	1265	1111	88%

Tous tronc communs	Effectifs cibles	Participants	Taux
National	36.188	34.109	94%

1. TANGER-TETOUAN-AL HOUCEIMA
2. L'ORIENTAL
3. FES-MEKNES
4. RABAT-SALE-KENITRA
5. BENI MELLAL-KHENIFRA
6. CASABLANCA-SETTA
7. MARRAKECH-SAFI
8. DRAA-TAFILALT
9. SOUSS-MASSA
10. GUELMIM-OUED NOUN
11. LAAYOUNE-SAKIA AL HAMRA
12. DAKHLA-OUED EDDAHAB

Par ailleurs, 4606 enseignants ont rempli le questionnaire «Enseignant» répartis ainsi :

- 852 enseignants d'arabe ;
- 854 enseignants de français ;
- 822 enseignants d'histoire-géographie ;
- 832 enseignants des mathématiques ;
- 479 enseignants des sciences de la vie et la terre ;
- 767 enseignants de la physique-chimie.

Quant aux directeurs d'établissement, 543 parmi 581 ont répondu au questionnaire « Directeur », soit un taux de participation de 93%.

E : Effectifs cibles
P : Effectifs participants
T : Taux de participation

ANNEXES





Variables	Tronc commun des lettres et sciences humaines			Tronc commun scientifique		
	Mathématiques	Arabe	Français	Mathématiques	Arabe	Français
	Coefficient (Ecart-type)	Coefficient (Ecart-type)	Coefficient (Ecart-type)	Coefficient (Ecart-type)	Coefficient (Ecart-type)	Coefficient (Ecart-type)
Genre						
Garçon	Référence	Référence	Référence	Référence	Référence	Référence
Fille	NS	3.80*** (0.28)	3.05*** (0.28)	-0.97*** (0.29)	4.14*** (0.31)	5.58*** (0.41)
Nombre de frères et sœurs	NS	NS	-0.35*** (0.09)	-0.24** (0.10)	NS	-0.44*** (0.13)
Raison d'orientation (les matières principales du tronc m'intéressent)	NS	1.74*** (0.25)	0.88*** (0.25)	3.15*** (0.25)	1.73*** (0.28)	2.06*** (0.37)
milieu de résidence						
Rural	Référence	Référence	Référence	Référence	Référence	Référence
Urbain	NS	1.14*** (0.32)	NS	NS	1.00** (0.45)	1.27*** (0.47)
Plus haut niveau d'éducation de l'un des parents						
Sans niveau	Référence	Référence	Référence	Référence	Référence	Référence
Primaire	NS	-0.65** (0.30)	NS	NS	NS	NS
Collège	-0.89** (0.45)	-1.14*** (0.35)	NS	NS	NS	NS
Qualifiant	NS	-1.11*** (0.37)	0.71* (0.42)	NS	NS	NS
Supérieur	-1.48** (0.67)	NS	2.11*** (0.60)	NS	NS	2.10*** (0.69)
L'élève parle français à la maison (unique ou avec autre langue)	-	-	-	-	-	3.53*** (0.69)
Ressources pédagogiques à la maison	0.77*** (0.18)	0.78*** (0.14)	1.18*** (0.17)	0.84*** (0.12)	1.05*** (0.15)	1.68*** (0.21)
Ressources TIC à la maison	NS	NS	0.75*** (0.15)	0.41*** (0.14)	0.66*** (0.16)	0.81*** (0.20)
Préscolaire	1.13*** (0.32)	0.74*** (0.26)	0.67*** (0.25)	NS	1.98** (0.38)	1.71*** (0.35)
Type d'établissement primaire						
Privé	Référence	Référence	Référence	Référence	Référence	Référence
Public	NS	NS	-5.36*** (0.58)	NS	NS	-5.40*** (0.64)

	Tronc commun des lettres et sciences humaines			Tronc commun scientifique		
	Mathématiques	Arabe	Français	Mathématiques	Arabe	Français
Milieu de l'établissement primaire						
Rural	Référence	Référence	Référence	Référence	Référence	Référence
Urbain	-0.95** (0.41)	NS	NS	-0.80** (0.38)	NS	NS
Milieu du collège						
Rural	Référence	Référence	Référence	Référence	Référence	Référence
Urbain	NS	NS	NS	1.25*** (0.48)	NS	NS
Redoublement	-1.39*** (0.29)	-3.76*** (0.25)	-3.36*** (0.25)	-3.37*** (0.25)	-5.77*** (0.33)	-5.42*** (0.39)
Triche	-0.73** (0.32)	-1.27*** (0.27)	-1.45*** (0.25)	-1.55*** (0.26)	-1.73*** (0.35)	-2.50*** (0.36)
Absentéisme	-1.25*** (0.38)	-0.970*** (0.28)	-0.69** (0.31)	NS	-0.86** (0.41)	NS
Incivilités subies	-0.50*** (0.16)	-0.96*** (0.12)	NS	-0.79*** (0.14)	-0.96*** (0.17)	-0.70*** (0.18)
Climat relationnel	-0.48** (0.23)	-1.17*** (0.18)	-0.68*** (0.15)	-0.66*** (0.19)	-0.89*** (0.19)	-1.33*** (0.23)
Engagement personnel	0.54** (0.24)	0.61*** (0.14)	0.31** (0.14)	0.49** (0.20)	1.27*** (0.18)	1.15*** (0.22)
Climat d'appartenance		-0.77*** (0.17)		-0.56*** (0.20)	-0.36** (0.18)	
Climat de justice	0.90*** (0.21)			0.43** (0.18)		
Climat de sécurité	-0.59*** (0.20)	NS	NS	NS	NS	NS
Opportunité d'investissement	NS	0.33* (0.17)	NS	-0.51*** (0.16)	NS	NS
Qualité de l'enseignement	0.42** (0.18)	0.86*** (0.18)	1.26*** (0.15)	0.93*** (0.29)	NS	2.23*** (0.18)
Système de reconnaissance	NS	NS	NS	0.85*** (0.25)	NS	NS
Système d'encadrement	NS	NS	NS	-0.59*** (0.21)	NS	NS
Temps consacré à l'enseignement	0.42** (0.17)	0.48*** (0.16)	0.60*** (0.14)	0.72*** (0.14)	0.37** (0.19)	1.14*** (0.22)
Certificat professionnel	NS	2.36*** (0.90)	NS	NS	NS	NS
L'ancienneté totale de l'enseignant						
Moins de 20 ans	Référence	Référence	Référence	Référence	Référence	Référence
20 ans et plus	2.54** (1.13)	-2.02*** (0.67)	NS	NS	-1.62* (0.86)	NS
Manque de ressources matérielles	NS	-0.82*** (0.30)	-0.73* (0.38)	NS	NS	NS

Manque du personnel administratif à l'établissement	NS	NS	-1.82** (0.86)	NS	NS	NS
Taille de la classe	NS	NS	NS	-0.10** (0.04)	-0.28*** (0.05)	-0.23*** (0.07)
Constante	42.53*** (0.72)	37.77*** (0.88)	32.61*** (1.11)	38.27*** (1.55)	54.43*** (2.02)	42.32*** (2.61)

*, **, *** indiquent un seuil de significativité respectivement de 10%, 5% et 1 %.

RÉFÉRENCES BIBLIOGRAPHIQUES



1. *Rapport méthodologique – PNEA 2008, CSEFRS, Mai 2009 ;*
2. *Technical Report, TIMSS Advanced 2008;*
3. *Sampling manual, Teaching and Learning International Survey (TALIS), OCDE – 2006;*
4. *School Sampling Manual, TIMSS 2007;*
5. *Michel Janosz et Christiane Bouthilier « Rapport de validation du questionnaire sur l'environnement socioéducatif des écoles secondaires (QES – Secondaire) », Montréal 2007 ;*
6. *Sampling manual, Teaching and Learning International Survey (TALIS), OCDE – 2006;*
7. *Manuel pour la préparation de l'échantillonnage de l'école, Programme International pour le Suivi des Acquis des élèves (PISA2003), OCDE-2002 ;*
8. *Michel Janosz, Patricia Georges et Sophie Parent « L'environnement socioéducatif à l'école secondaire : Un modèle théorique pour guider l'évaluation du milieu », Revue canadienne de psychoéducation, Volume 27, Numéro 2, 1998.*
9. *Benjamin S. Bloom. Taxonomy of Educational Objectives, 1956, trad. : Taxonomie des objectifs pédagogiques, vol.1 : Domaine cognitif, Presses de l'Université du Québec, 1975. Benjamin S. Bloom, 1956. Taxonomy of Educational Objectives: The Classification of Educational Goals. Longman. Voir aussi Robert J. Marzano, Designing a new taxonomy of educational objectives, Corwin Press, 2001,*



CONSEIL SUPÉRIEUR DE L'ÉDUCATION
DE LA FORMATION ET DE LA RECHERCHE SCIENTIFIQUE
Angle Avenues Al Melia et Allal El Fassi, Hay Riad, Rabat - B.P. 6535



www.csefrs.ma