

Étude Évaluative D'examen Normalisé De Sciences De La Vie Et De La Terre Au Cycle Secondaire Collegial

Abdellah. El Allaoui

Ministère de l'Éducation et de la Formation Professionnelle,
délégation Meknès, Maroc

Fouzia.Rhazi Filali

Équipe Microbiologie et Santé, Département de Biologie,
Laboratoire de Chimie Biologie Appliquées à l'Environnement,
Université Moulay Ismail Faculté des Sciences, Maroc

El mokhtar. El Hadri

Centre d'Orientation et de Planification de l'Éducation, Rabat

Khalid.Fetteh

Malika. Bouhadi

Ministère de l'Éducation et de la Formation Professionnelle,
délégation Meknès, Maroc

doi: 10.19044/esj.2016.v12n1p283 URL:<http://dx.doi.org/10.19044/esj.2016.v12n1p283>

Abstract

Generally an exam, a performance test or a test to be used for summative or evaluation purposes must be prepared according to a definite plan. The present paper aimed at analyzing and evaluating the certificate exam, the study also revealed the questions representation of Bloom's cognitive levels. This article presents also the statistical and psychometric indices that characterize each of the questions (26 items) that make up a certification exam of the Sciences of Life and Earth in a Moroccan high school in Meknes. Indeed, two hundred test copies have been analyzed using the SPSS and Anitem software. The average of goals achieved in activities of knowledge, analysis, understanding and evaluation is 61%, 52%, 37.5%, 58% respectively, but it is only 19% for "application" activity. The success rate in open questions items is only 38.5% against 61.5% for Multiple Choice Questions (MSQ) and closed questions. The internal homogeneity coefficient is greater than 0.8 ($\alpha = 0.84$), which means that the homogeneity of the instrument has been considered satisfactory. According to the index of difficulty out of the 26 items, two are extremely inferior ($P > 0.8$), while two others are extremely superior ($P < 0.2$). According to the index of discrimination eleven items meet the conditions of validity, three are not discriminating and twelve should be reviewed. The correlation between the

success rate in these items and their difficulty index is high ($R = 0.98$). In light of the findings, we recommended the necessity of having measurement and evaluation experts in writing questions of exams and we highly recommend the necessity of continuous training of teachers in the field of assessment and evaluation.

Keywords: Morocco, Item, Evaluation, Validity, Exam

Résumé

Généralement un examen, une épreuve de rendement ou un test devant servir à des fins d'évaluation certificative ou sommative doivent être préparés selon un plan bien défini. Cet article présente des indices statistiques et psychométriques qui caractérisent chacune des questions (26 items) composant un examen certificatif des Sciences de la Vie et de la Terre dans un collège marocain à Meknès. En effet, deux cents copies d'examens sont analysées grâce aux logiciels *Anitem* et *SPSS*. Les pourcentages moyens des objectifs atteints en activité de connaissance, de l'analyse, de compréhension et d'évaluation sont respectivement de 61%, 52%, 37,5%, 58%, mais seulement 19% pour l'activité « application ». Le taux de réussite des items sous forme de questions ouvertes n'est que de 38,5 % contre 61,5 % pour les questions fermées et les Questions à choix Multiples (QSM). Le coefficient d'homogénéité interne est supérieur à 0.8 ($\alpha = 0,84$), ce qui signifie que l'homogénéité de l'instrument a été considérée comme satisfaisante. Selon l'indice de difficulté, sur 26 items, deux sont des extrêmes inférieurs ($P_i > 0.8$) alors que deux autres sont des extrêmes supérieurs ($P_i < 0.2$). Selon l'indice de discrimination, onze items remplissent les conditions de validité, trois ne sont pas discriminants et douze doivent être revus. La corrélation entre les taux de réussite à ces items et leurs indices de difficulté est forte ($R=0,98$). À la lumière de ces résultats, nous recommandons la nécessité d'assurer des formations continues du corps enseignant dans le domaine de mesure et d'évaluation et l'implication des experts lors d'élaboration des examens.

Mots clés : Maroc, Item, Évaluation, Validité, Examen

Introduction

L'évaluation scolaire par notation a suscité très tôt de nombreuses critiques de la part des pédagogues, une partie des ces critiques émane de la Docimologie, Science étudiant depuis le début du 20^{ème} siècle les logiques statistiques à l'œuvre dans les notes attribuées au cours de la correction des copies (Merle, 1998).

Généralement un examen, une épreuve de rendement ou un test devant servir à des fins d'évaluation certificative ou sommative doivent être préparés selon un plan bien défini. Parcourir un manuel scolaire, à la recherche d'éléments pouvant servir d'inspiration à la rédaction de questions d'examen, est un procédé depuis longtemps dépassé. Par souci de respect des «règles du jeu», les éléments de contenu et, au mieux, les objectifs d'apprentissage qui ont été l'objet d'une attention toute particulière en cours de route, doivent être considérés avec une attention toute aussi particulière dans la préparation d'un examen ou d'un ensemble d'examens pour être utilisés à des fins d'évaluation certificative (Scallon, 1996).

Ces examens doivent être validés ensuite par toute une série de données statistiques faisant état d'une analyse des items, qui contrôle que les items permettent bien de faire la différence entre les individus selon leurs compétences (paramètre de discrimination), le niveau de difficulté de chaque item (paramètre de difficulté) et la probabilité qu'un individu de niveau médiocre puisse répondre correctement à l'item (paramètre de chance, ou de "pile ou face") (Bachman, 1990).

Le test des items d'examen amène à une évaluation statistique et qualitative. Selon Hughes (2003), l'analyse statistique démontre la qualité du test (sa fiabilité) en tant qu'entité et aussi dans ses divers items (leur difficulté, leur capacité à faire la différence entre les candidats forts et faibles). D'autre part, l'analyse qualitative est effectuée afin de découvrir des erreurs d'interprétation, des réponses inattendues mais correctes, et autres signes montrant les faiblesses de tel ou tel item (Hughes, 2003).

Etant donné que la validation de l'examen contribue à poursuivre l'élaboration des tests, elle est d'une extrême importance car elle permet aux auteurs des items d'avoir un retour sur la qualité de leur travail et sur la fiabilité des résultats. C'est pourquoi, il nous paraît important, en tant que spécialistes en évaluation, d'interroger la validation de l'un des examens finaux (Sciences de la Vie et de la Terre par exemple).

Dans ce contexte, la présente étude s'articule autour des objectifs suivants :

- Analyse des performances des élèves de 3^{ème} année secondaire collégiale en Science de la Vie et de la Terre (SVT) : cet axe portera sur l'analyse descriptive des performances des élèves en sciences par domaine de contenus et par niveau taxonomique;

- Analyse des réponses aux questions de l'examen : on procédera à une analyse fine des réponses des élèves questionnés aux questions de la matière et ce afin de pouvoir apprécier le degré d'atteinte de chaque objectif. Plus précisément, il s'agit d'évaluer et d'expliquer le niveau d'apprentissage effectif des élèves de 3^{ème} année secondaire collégiale dans cette discipline

(SVT) à un moment donné durant le premier semestre (Année scolaire 2014-2015).

Materiel et methodes

Au secondaire collégial, il est à signaler que l'enseignement/apprentissage des Sciences de la Vie et de la Terre permet à l'élève d'acquérir un ensemble de connaissances scientifiques indispensables à la compréhension de son corps et de son environnement et de développer ses qualités intellectuelles, ses capacités d'observation et d'analyse, ainsi que celles d'imaginer, de créer, d'opérer des synthèses et d'avoir un esprit critique. Il permet enfin à l'élève de prendre conscience des problèmes actuels de la vie, de la santé et de l'environnement, et par là développer chez lui l'esprit de responsabilité et de citoyenneté (PNEA, 2009).

Puisque le diagnostic des apprentissages porte sur le programme scolaire officiel, nous avons d'abord collecté tous les documents pédagogiques qui s'y réfèrent, notamment la Charte Nationale d'Education et de Formation, le livre blanc, le guide des programmes et orientations pédagogiques relatif aux sciences de la vie et de la terre, les notes ministérielles ainsi que les manuels scolaires. En outre, le cadre de référence de la matière définit avec précision et d'une manière concise les modèles d'apprentissages auxquels on doit se référer lors de l'élaboration des tests de rendement scolaire. La structure de ces modèles se présente ainsi :

- Définition opérationnelle des domaines principaux d'apprentissages ;
- Détermination des objectifs finaux/compétences selon les domaines d'apprentissage ;
- Précision du degré d'importance des objectifs finaux/compétences ;
- Élaboration des tableaux de spécification
- Répartition des questions selon les domaines d'apprentissages et les niveaux d'habiletés.

Dans cette partie pratique nous avons utilisé le logiciel Anitem (Normand, 2001) pour évaluer les indices de difficulté des différentes questions. L'indice de discrimination de chaque question par rapport à la réussite globale a été calculé, et l'alpha de Cronbach permet d'apprécier la « pertinence » de chaque item par rapport aux résultats globaux. Le traitement des données et l'analyse statistique sont effectués avec le logiciel SPSS (*STATISTIC* 17.0). Les résultats sont considérés significatifs pour un degré de signification $p < 0,05$.

Plan d'échantillonnage

Le plan d'échantillonnage de base procède par étape. Dans un premier temps, nous avons sélectionné toutes les copies d'examens des élèves de 3^{ème} année secondaire collégiale durant l'année scolaire 2014

/2015 ; dont l'effectif est de 415 élèves. Puis nous avons choisi au hasard 50% de la population comme échantillon de l'étude d'évaluation. La tranche d'âge des élèves étudiés est comprise entre 14 et 18 ans. Leurs copies d'examen normalisé de SVT (unité statistique) ont été corrigées par les enseignants, ensuite nous avons fait une analyse des items (n= 26) constituant l'examen.

Resultats et discussion

Le cadre de référence énonce les paramètres qui serviront de fondement à l'évaluation des acquis des élèves de la 3^{ème} année du cycle secondaire collégial en SVT. Il résulte d'une lecture du programme du premier semestre de la 3^{ème} année du secondaire collégial et vise à assurer la congruence entre le contenu du programme et l'instrument de mesure qui servira à l'évaluation sommative dans le cadre de l'obtention du certificat du cycle secondaire collégial. Alors que le tableau de spécification fait le pont entre les objectifs/compétences visés par les programmes scolaires d'une part, et les items du test, d'autre part, compte tenu de leur représentativité en nombre et leur importance relative (Scallon, 1996). Le tableau de spécification de la matière a été élaboré par les spécialistes de la matière. En effet, il illustre les domaines ainsi que les pourcentages d'importances relatives inspirés à partir de la durée attribuée à chaque domaine selon la note ministérielle relative aux programmes (Ministère de l'Éducation Nationale, de l'Enseignement Supérieur, de la Formation des Cadres et de la Recherche Scientifique [MÉNESFCRS], 2011) et celle relative aux cadres référentiels des contrôles continus (MÉNESFCRS, 2010) (Tableau 1). Ces pourcentages nous aideront par la suite à estimer la note qui doit être réservée à chaque domaine. Un autre tableau de spécification (Tableau 2), montre cette fois les compétences visées en pourcentage selon les domaines.

Performances des élèves en SVT par domaine de contenus et par niveau taxonomique

Dans un premier temps, il est à signaler qu'en absence d'un socle de compétences et partant d'une échelle répartissant les élèves selon des niveaux taxonomiques, les performances des élèves dans cette étude représentent, en fait, le degré d'atteinte des objectifs fixés par le test. Ainsi, les compétences permettant de cerner les acquis des élèves en Science de la vie et de la terre ont été appréhendées selon les domaines suivants :

- L'éducation nutritionnelle et santé
- La respiration
- Le sang et circulation sanguine
- L'excrétion urinaire

Après correction et traitement des résultats, nous avons souligné un rendement relativement faible des élèves en SVT, notamment dans le domaine relatif au « Sang » (49 % des objectifs atteints) et les domaines relatifs à la « Respiration », où ils ont atteint 42 % des objectifs/compétences seulement, alors que ces deux domaines représentent ensemble un degré d'importance de 36 % de programme (Figure 1), le domaine relatif à l' « Urée » n'est pas abordé malgré son degré d'importance (14%). Seul le domaine relatif au domaine «Éducation nutritionnelle » qui représente 50% d'importance dans le curriculum (Figure 1), a bénéficié d'un résultat au dessus de la note moyenne (62% des objectifs atteints). Ces résultats somme toute accès faible, confirment globalement ceux des évaluations internationales. En effet, les études TIMSS (Trends in International Mathematics and Science Study), en 2007, organisées par l'IEA (International Association for the Evaluation of Educational Achievement), auxquelles le Maroc a participé, ont souligné des résultats faibles aux tests relatifs aux Sciences, administrés aux élèves de 4^{ème} année du primaire et aux élèves de 2^{ème} année du collège (PNEA, 2009). En 2011, selon les mêmes études, les résultats des élèves marocains sont restés nettement en dessous de la moyenne internationale et ce, respectivement dans les sciences et les mathématiques. Avec un score de 376 points en Sciences le Maroc reste loin de la moyenne internationale estimée à 500 points et est classé avant dernier, devant le Ghana qui a participé pour la première fois à ce type de test (Martin et Mullis, 2013).

Parmi les explications de ce rendement assez faible, on peut citer :

- L'absence des moyens didactiques appropriés à l'enseignement de l'activité scientifique;
- Méthodologie d'enseignement et outils pédagogiques insuffisants ou mal exploités;
- Nombre élevé des élèves dans la classe (Absence de groupe) ce qui entrave le déroulement des travaux pratiques ;
- Incapacité de l'élève caractérisé par de graves manques ou déficit à acquérir ou maîtriser des compétences de base ;
- Carences dans la formation continue du corps enseignant

Globalement, la moitié des élèves ont réussi les items relatifs au domaine du sang et circulation sanguine contre 62% pour les items se rapportant au domaine de l'éducation nutritionnelle et santé (Figure 1), ce dernier résultat est relativement très acceptable. En effet, cette performance peut être expliquée d'une part, par le fait que la « Digestion sous unité du domaine de l'éducation nutritionnelle et santé » est un thème très proche des intérêts des enfants et d'autre part, l'apprentissage du thème fait l'objet d'une programmation verticale et le temps imparti à ce domaine est suffisant (11H).

La comparaison entre les notes attribuées selon l'examen normalisé et celles qui devaient configurer dans le tableau de spécification, a montré que le domaine « Urée », le sous domaine « Sang » et le sous domaine « Prévention du système digestif » ne sont pas abordés lors de l'examen final (Tableau 1, Figure 2), le domaine « respiration » a bénéficié de 5 point contre seulement 3,5 qui doivent figurer dans le barème. Quant aux deux domaines restants, les normes sont plus au moins respectées (Figure 2).

En ce qui concerne la nature des acquisitions, les performances des élèves en sciences diffèrent d'un niveau taxonomique à l'autre. En effet, si les pourcentages moyens des objectifs atteints en activités de connaissance, de l'analyse, de compréhension et d'évaluation sont respectivement de 61%, 52%, 37,5% et 58% , il n'est que de 19% pour l'activité « application ». C'est donc en cette dernière que les élèves ont réalisé des performances relativement moins bonnes.

En outre, le taux de réussite des items sous forme de questions ouvertes n'est que de 38,5 % contre 61,5 % pour les questions fermées et QSM. Ce taux de réussite faible aux questions ouvertes peut avoir ses origines dans l'incapacité des apprenants à mobiliser les ressources acquises et/ou dans la non maîtrise des compétences de communication. Ceci peut être expliqué aussi par le fait que les questions ouvertes demandent des habiletés complexes (analyse, synthèse, identification des relations cause/effet ...) alors que les modes d'enseignement classique (enseignement transmissif, absence de projets pédagogiques et éducatifs ...) ne permettent pas de développer assez ces compétences. Dans tous les cas, le faible taux de réussite aux questions ouvertes est un indicateur des déficiences au niveau d'habiletés d'analyse, de synthèse et d'expression (PNEA, 2009).

Tableau 1 : Tableau de spécification représentant les domaines, les pourcentages d'importance et les points alloués à chaque domaine de la matière (SVT)

Domaines	Excrétion urinaire			Sang et circulation sanguine		Respiration		Éducation nutritionnelle et santé				Somme
Durée	3H			4 H		4 H		11 H				
Pourcentage	14%			18%		18%		50%				
Sous-domaines	Origine de l'urine	Appareil urinaire	Composants de l'urine	Sang et circulation sanguine	Sang	Prévention du système respiratoire	Respiration chez L'Homme	Prévention du système digestif	Rations alimentaires	Carences alimentaires	Alimentation, digestion et absorption	
Durée attribuée	1h	1h	1h	2h	2h	1h	3 h	1h	2h	2h	6h	22h
Pourcentage d'importance relative	4,54 %	4,54 %	4,54 %	4,54 %	9,1%	4,54 %	13,63 %	4,54 %	9,1 %	9,1 %	27,27 %	100 %
Note attribuée selon le cadre de référence	1,25	0,75	0,75	2	1,5	1	2, 5	1	2	2	5	20
Note attribuée selon l'examen normalisé	0	0	0	3,5	0	1	4	0	0,5	1,5	9,5	20

Tableau 2: Tableau de spécification selon les domaines et les compétences visées

Domaines Compétences	Excrétion urinaire			Sang et circulation sanguine		Respiration		Education nutritionnelle et santé			Somme	%	
	Origine de l'urine	Appareil urinaire	Composants de l'urine	Sang et circulation sanguine	Sang	Prévention du système respiratoire	Respiration chez L'Homme	Prévention du système digestif	Ration alimentaire	Carences alimentaires			Alimentation, digestion et absorption
Connaissance	0,25	0,25	0,5	0,75	0,75	0,25	1,25	0,5	0,75	0,75	2	8	40 %
Raisonnement scientifique	0,5	0,5	0,5	1	0,75	0,25	1,5	0,5	1	1	2,5	10	50 %
Communication écrite et graphique		0,5		0,25		0,25				1		2	10 %
Somme		3		3,5		3,5			10			20	100 %

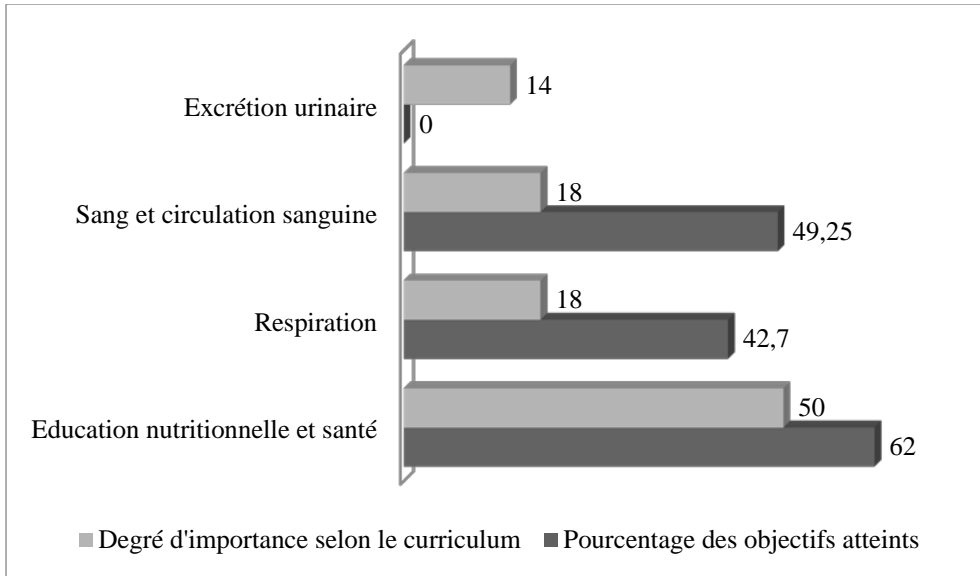


Figure 1: Degré d'importance (%) selon le curriculum et pourcentage des objectifs atteints

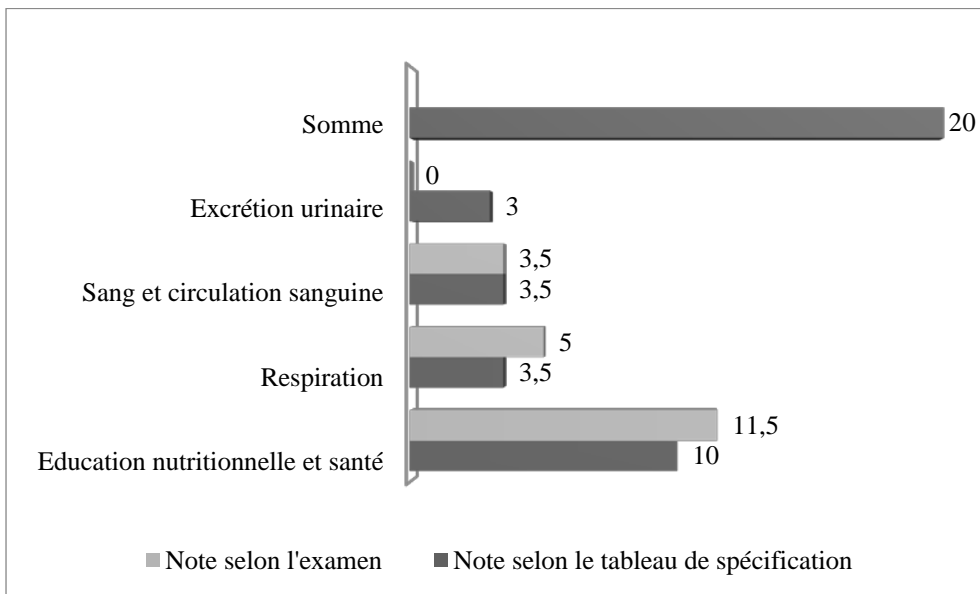


Figure 2: Notes (n /20) attribuées selon l'examen et selon le cadre de référence

Statistiques descriptives

Huit est la valeur que nous observons le plus fréquemment, la valeur qui est à « la mode » (Tableau 3). La médiane qui est égale à 9,75 se trouve au milieu de l'échantillon, elle sépare celui-ci en deux groupes de même effectif, l'un contient les notes les plus petites et l'autre les moyennes les plus grandes c'est à dire : au moins la moitié (50%) des valeurs sont inférieures

ou égales à 9,75 et au moins la moitié (50%) des valeurs est supérieure ou égales à la valeur médiane (Figure 3).

Tableau 3 : Paramètres descriptives des résultats des élèves en SVT

Effectif	Moyenne	Médiane	Mode	Ecart-type	Variance	Minimum	Maximum	Q1	Q2	Q3
200	10,10	9,75	8	3,46	11,89	3,50	16,50	7,5	9,75	12

Q1 : Quartile 1, Q2 : Quartile 2, Q3 : Quartile 3

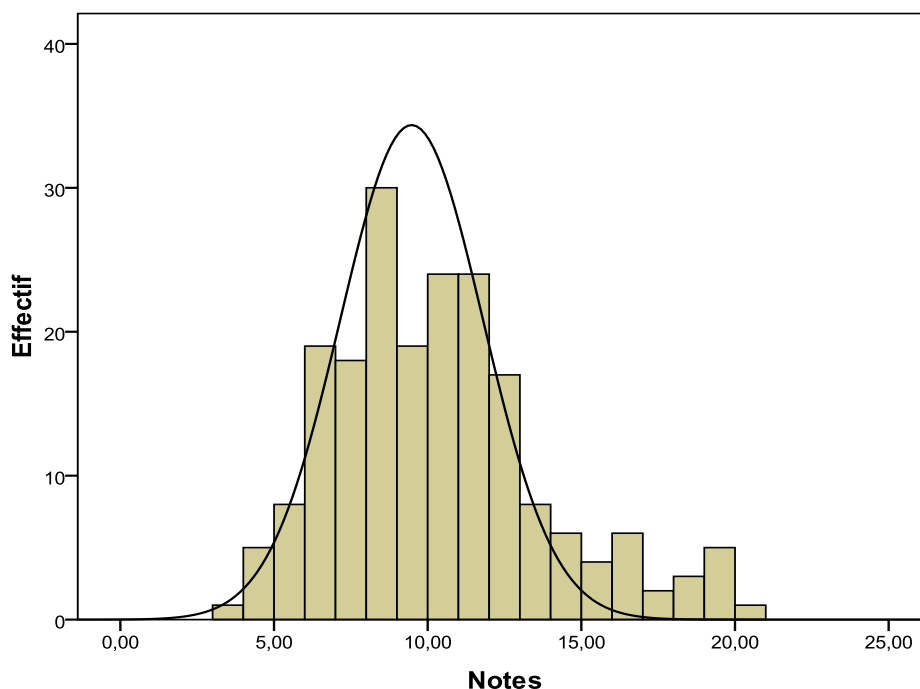


Figure 3 : Courbe illustrant la variation des notes selon l'effectif des apprenants

Au moins un quart (25%) des notes sont inférieures ou égales à 7,5. Au moins trois quarts (75%) des valeurs sont inférieures ou égales à 12. Environ la moitié des valeurs se trouvent dans l'intervalle interquartile [7,5 ; 12] (Tableau 3). La comparaison des notes des élèves selon les classes dans le même collège a montré qu'il n'y a pas une différence significative entre les apprenants ($\alpha > 0,05$). La comparaison des résultats d'une école à une autre serait non objective car les professeurs de la matière disposaient d'une grande marge de manœuvre pour compiler, gérer et noter l'examen, par conséquent, les types d'exercices variaient énormément.

Analyse des items

Coefficient d'homogénéité interne

Ce coefficient traduit le degré de « consistance interne » de l'épreuve et sert d'indice pour estimer un des aspects de la fidélité des résultats observés. Dans cette étude, il s'est avéré que le coefficient d'homogénéité interne de l'épreuve calculé selon la relation ci-dessous, est supérieur à 0.8 ($\alpha = 0,84$), ce qui signifie que l'homogénéité de l'instrument a été considérée comme satisfaisante. Alors nous pouvons postuler que les questions de l'épreuve semblent « mesurer effectivement la même caractéristique ou le même objectif » (Legendre, 1988) et contribuer par conséquent à produire des résultats fiables.

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum \sigma^2 j}{\sigma^2 t} \right)$$

où :

- k : nombre d'items
- $\sigma^2 j$: variance des scores à l'item j
- $\sigma^2 t$: variance des scores à l'ensemble du test

Dans notre cas, nous notons que quatorze items n'ont pas d'influence sur la consistance interne de l'épreuve ($\alpha = \alpha - I$), neuf ont une influence positive ($\alpha > \alpha - I$), en effet la contribution de ces items à la fidélité des résultats de l'épreuve peut ainsi être appréciée, alors que seulement trois questions influencent négativement son homogénéité interne ($\alpha < \alpha - I$) (Tableau 4).

Indice de difficulté

C'est le taux de réussite exprimé en pourcentage "décimal", il indique aussi la contribution de l'item au score total. Ainsi, en calculant le pourcentage des élèves qui réussissent l'item, on peut sélectionner les items dont la difficulté est appropriée à la situation. En effet, un item présentant un taux de réussite très bas ou très élevé est peu informatif. Selon la relation ci-dessous nous avons calculé les indices de difficulté (p_i) de toutes les questions (Figure 5)

$$\text{Avec } p_i = \frac{\text{NBR}_i}{\text{NT}}$$

NBR_i : Nombres des bonnes réponses pour la question i

NT : Nombre total des élèves

Généralement si :

$0.2 < P_i < 0.8$: On garde la question.

Pi<0.2 : La question est considérée très difficile.
Pi> 0.8 : La question est considérée très facile

D'après l'observation de l'indice de difficulté de chaque question et selon les normes requises, nous constatons que les questions 2 et 4 sont très faciles (Questions extrêmes inférieures, $P_i > 0.8$) alors que les questions 23 et 25 sont très difficiles (Questions extrêmes supérieures, $P_i < 0.2$) (Figure 4). Les items dont l'indice de difficulté P_i est supérieur à 0.8 ou inférieur 0,2 sont marqués d'une étoile (Tableau 4). Ce choix non arbitraire a pour but d'attirer l'attention du lecteur et implique que ces items posent problème. Les auteurs pourront néanmoins questionner la pertinence de ces items en rapport avec les autres items de l'épreuve; identifier un vice possible de formulation ou chercher une raison pour expliquer un taux de réussite relativement faible ou très important.

Tableau 4 : Différents paramètres de l'analyse des items par le logiciel « Anitem »

Item	Clé	Poids	Difficulté	Discrimination	Alpha-Item (α -I)	Comparaison α et α -I
Q2	2	0,5	0,95	0,25	0,84	$\alpha = \alpha$ -I
Q3	1	0,5	0,79	0,22	0,84	$\alpha = \alpha$ -I
Q4	1	0,5	0,88	0,19*	0,84	$\alpha = \alpha$ -I
Q9	3	0,5	0,63	0,27	0,84	$\alpha = \alpha$ -I
Q10	1	0,5	0,47	0,12*	0,85	$\alpha < \alpha$ -I
Q11	2	0,5	0,62	0,10*	0,85	$\alpha < \alpha$ -I
Q12	3	0,5	0,65	0,24	0,84	$\alpha = \alpha$ -I
Q13	2	0,5	0,64	0,20	0,84	$\alpha = \alpha$ -I
Q14	1	0,5	0,66	0,32	0,84	$\alpha = \alpha$ -I
Q1		1,0	0,55	0,05*	0,85	$\alpha < \alpha$ -I
Q15		0,5	0,56	0,28	0,84	$\alpha = \alpha$ -I
Q16		1,0	0,58	0,56	0,83	$\alpha > \alpha$ -I
Q17		1,0	0,53	0,51	0,83	$\alpha > \alpha$ -I
Q18		1,0	0,58	0,45	0,83	$\alpha > \alpha$ -I
Q19		1,5	0,35	0,70	0,82	$\alpha > \alpha$ -I
Q20		0,5	0,44	0,25	0,84	$\alpha = \alpha$ -I
Q21		1,0	0,51	0,53	0,83	$\alpha > \alpha$ -I
Q22		1,0	0,50	0,59	0,83	$\alpha > \alpha$ -I
Q23		1,5	0,18	0,58	0,83	$\alpha > \alpha$ -I
Q24		1,0	0,47	0,53	0,83	$\alpha > \alpha$ -I
Q25		1,0	0,19	0,64	0,83	$\alpha > \alpha$ -I
Q26		1,0	0,38	0,37	0,84	$\alpha = \alpha$ -I
Q5		0,5	0,48	0,39	0,84	$\alpha = \alpha$ -I
Q6		0,5	0,46	0,41	0,84	$\alpha = \alpha$ -I
Q7		0,5	0,44	0,37	0,84	$\alpha = \alpha$ -I
Q8		1,0	0,70	0,34	0,84	$\alpha = \alpha$ -I

Alpha de Cronbach (α) = 0,84 ; α -I : Indique la nouvelle valeur du coefficient « Alpha de Cronbach » si cet item était retiré de l'épreuve ; Qn : Numéro de question ; (*) : Question extrêmes

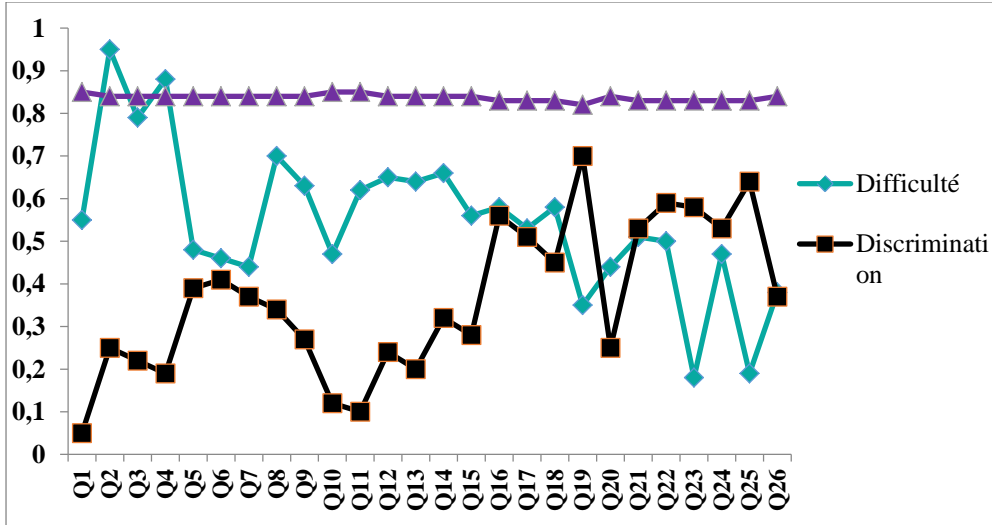


Figure 4: Différents indices des items de l'examen de SVT

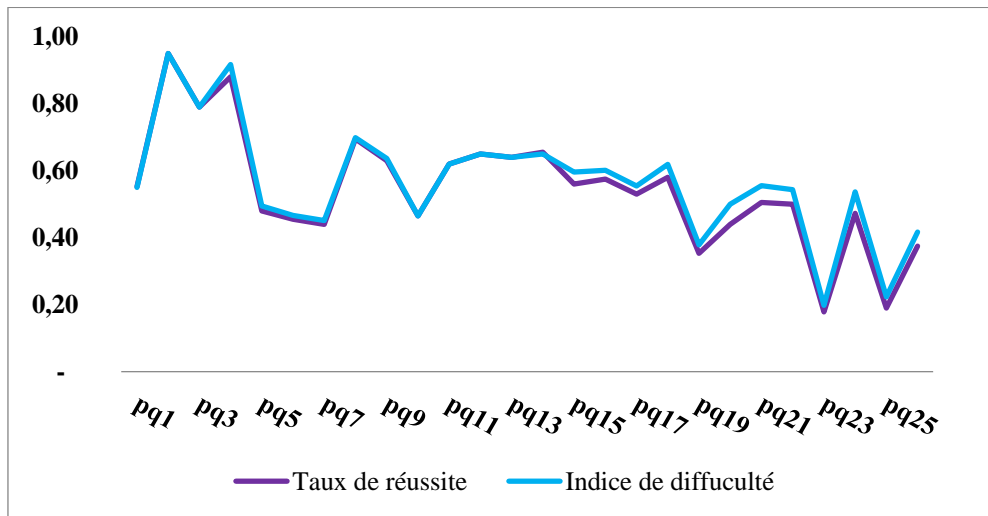


Figure 5 : Taux de réussite et indices de difficulté des items de SVT

Indice de discrimination

Ce coefficient de corrélation établit un lien entre les notes à un item donné et les notes avec le total des autres items de l'épreuve. Un item n'est utile pour évaluer les performances des élèves que s'il est significativement discriminatif, c'est-à-dire sensible aux différences d'apprentissage entre les élèves testés. Cette sensibilité est mesurée par l'indice de discrimination.

Afin de calculer l'indice de discrimination nous utilisons la relation suivante :

$$d_i = \frac{NGPE}{NE} - \frac{NGPB}{NB}$$

Avec :

NGPE : nombre de questions correctes dans le groupe supérieur.

NGPB : nombre de questions correctes dans le groupe inférieur.

NE : nombre des élèves dans le groupe supérieur.

NB : nombre d'élèves dans le groupe inférieur.

$0.2 < di < 0.4$: La question doit être revue.

$di < 0.2$: La question n'est pas retenue

$di > 0.4$: La question est retenue

Un coefficient dont la valeur se situe à 0,40 et plus indique que les sujets ayant eu des notes élevées à cet item s'avèrent avoir aussi obtenu des notes élevées aux autres items de l'épreuve c'est le cas des items Q6, Q16, Q17, Q18, Q19, Q21, Q22, Q23, Q24 et Q26 (Figure 4). Donc ces items seront retenus car elles remplissent les conditions de validité et d'objectivité. Un indice de discrimination se rapprochant de 0,00 (Q1, $di = 0,05$) (Tableau 4) implique que pareille relation n'existe pas. Par ailleurs un coefficient négatif élevé témoigne d'une relation inverse non souhaitable : les "meilleurs" sujets obtiennent une note faible à cet item, alors que les sujets qui présentent des notes faibles à cet item obtiennent de "meilleurs" notes aux autres items de l'épreuve. Les questions 2, 10 et 11 ne sont pas discriminantes car leurs indices de discrimination sont tous inférieurs à 0,20 : Elles doivent être éliminées. Quand aux questions (Q2, Q3, Q5, Q7, Q8, Q9, Q12, Q13, Q14, Q15, Q20 et Q25) dont l'indice de discrimination est inclus entre 0,2 et 0,4, elles doivent être revues.

Le taux moyen de réussite aux items est moyen : un taux moyen de réussite de 57% pour un indice moyen de difficulté de 0,54. Globalement, la corrélation entre les taux de réussite à ces items et leurs indice de difficulté est forte ($R = 0,98$) : la variation des taux de réussite à ces items est expliquée à 95% par la variation de leurs indices de difficulté.

Conclusion

L'évaluation scolaire constitue le principal socle des décisions d'orientation, il convient en effet de se poser la question de savoir dans quelle mesure elle garantit de façon effective à tous les élèves une orientation juste et pertinente.

Lorsque les indices de facilité, de discrimination ou de sensibilité ne correspondent plus à des niveaux de qualité minimale, trois raisons principales expliquent habituellement cette carence : ou bien il s'agit d'une anomalie erreur dans la formulation de la question, d'une ambiguïté dans l'exposé du problème à résoudre ou d'une ambiguïté externe, ou bien un enseignement inapproprié a pu induire en erreur les élèves les plus attentifs et ainsi causer un déséquilibre dans la répartition des échecs à l'item d'examen, ou encore des faiblesses aléatoires ont pu produire des indices

exceptionnellement inacceptables surtout si le nombre des élèves est plutôt réduit (Morissette et Laurencelle, 1993) .

Certes, ce test a pu mesurer quelques degrés de la maîtrise des ressources cognitives en SVT de la part des apprenants, Mais il n'a pas évalué le degré de maîtrise et l'habilité à les mobiliser dans des situations complexes ou peu intégrées.

Acknowledgement:

Les auteurs remercient les cadres d'orientation et l'inspecteur pédagogique de la matière, ainsi que la direction du collège et le corps enseignant de la matière SVT pour leur brillante participation.

References:

- Bachman L., (1990). *Fundamental Considerations in Language Testing*, Oxford University Press. 240
- Hughes C., (2003). *Testing for Language Teachers*, Cambridge University Press, 172 p.
- Legendre R ., (1988). *Dictionnaire actuel de l'éducation*, Larousse, Paris Montréal.
- Martin M.O., Mullis I.V.S., (2013). Eds. *TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade-implications for early learning*. Chestnut Hill, MA: Boston College.
- MÉNESFCRS., (2010). Note ministérielle N° 190, au sujet de l'encadrement et le suivi de l'exécution des contrôles continus en sciences de la vie et de la terre au secondaire collégial, 7p.
- Merle P., (1998). *Sociologie de l'évaluation scolaire*, Que sais-je n° 3278, Paris : PUF.
- Ministère de l'Éducation Nationale, de l'Enseignement Supérieur, de la Formation des Cadres et de la Recherche Scientifique [MÉNESFCRS] ., (2011). Note ministérielle N° 132, au sujet des programmes des sciences de la vie et de la terre au secondaire collégial, 3p.
- Morissette D, Laurencelle L., (1993). *Les examens de rendement scolaire*, Les presses de L'Université laval, PP 463, p 320 Editeur : Sainte-Foy, Québec : Presses de l'Université Laval, 1993, 3e éd
- Normand S., (2001). *Anitem*, Logiciel pour effectuer une analyse d'items via le chiffrier Excel, version PC et Mac, Bureau d'évaluation, Faculté de Médecine, Université de Montréal, 2001. www.medbev.umontreal.ca/docimo/DocSource/Docimologie,AnItem.
- PNEA (Programme National d'Evaluation des Acquis)., (2009) . *Conseil Supérieur de l'Enseignement, programme nationale d'évaluation des acquis*, Royaume du Maroc fascicule des Sciences, p 45 – pp : 61

Scallon G., (1996). Département d'orientation, d'administration et d'évaluation en éducation, FSÉ, Université Laval. (1996)
<http://www.fse.ulaval.ca/Gerard.Scallon/fascicules/tabspecific.pdf>