

Docimologie

**Dr. Ch. Kohler,
Pr. M. Braun, Pr. F. Kohler**

Introduction :

- ▶ La **docimologie** (grec *dokime* signifiant épreuve)
 - ▶ = science et pratique du contrôle des connaissances. (Henri PIERON)
 - ▶ = étude de la qualité et de la validité des différents systèmes de notation du contrôle des connaissances
 - ▶ = étude des évaluations
 - ▶ = étude de l'attribution des notes par un correcteur.
-



Docimologie

- ▶ Rechercher les facteurs qui entrent en jeu dans l'évaluation d'un travail — écrit, oral, production matérielle... — indépendamment de la **valeur intrinsèque de ce travail ou de l'apprenant**
-

Les évaluations

Quand faire une évaluation ?

- ▶ Au **début** d'un enseignement :
 - ▶ Pour situer les apprenants
 - ▶ Au **terme** d'un enseignement :
 - ▶ Pour évaluer les performances des étudiants
 - ▶ Pour évaluer les performances de l'enseignement
 - ▶ Pour **sélectionner**
 - ▶ Concours ou recrutement
-



Pourquoi faire une évaluation ?

➤ Evaluation formative

- En cours de formation
- Mettre en évidence les difficultés dans le processus d'apprentissage

Auto-évaluation

➤ Evaluation sommative

- Fin de la période de formation
- Certifier que l'apprenant a atteint un certain niveau de formation
- Mesurer les effets de la formation

Examen

➤ Evaluation pronostique

- Prévoir l'adaptation de l'apprenant, la réussite de l'apprenant dans une formation à venir ou un poste

➤ *Entretien d'embauche*



Attention !!!

- ▶ Lien entre contenu enseignement et modalités des épreuves
 - ▶ Constituent les **objectifs de fait** d'un enseignement ou d'une formation
 - ▶ => **Nécessité d'une forte cohérence entre les objectifs annoncés et les épreuves.**
 - S'il y a discordance, pour les étudiants réussir les épreuves devient l'objectif unique.
 - Exemples : PACES, ECN
-



Modalités évaluation

- ▶ Déterminent les modalités d'apprentissage
 - ▶ Les étudiants adoptent les stratégies qu'ils estiment les plus efficaces et les plus efficaces pour réussir
 - Lorsque les épreuves et l'enseignement ne correspondent pas aux besoins, l'apprentissage s'effectue en marge de la formation ; l'étudiant devient autodidacte
 - ▶ Un fait : Un minimum d'énergie est mis en œuvre pour satisfaire aux examens
-



Exemple d'inadéquation

- ▶ Entre **objectif** du 2^e cycle des études médicales (apprentissage de séméiologie clinique) et **réalité** (apprentissage des items de l'ECN)
- ▶ **Étude : évaluation du savoir-faire en séméiologie clinique des étudiants en fin de deuxième cycle des études médicales ***
 - ▶ **Résultats :**
 - ▶ **Savoir-faire global médiocre** alors que l'étude est faite dans deux facultés qui sont très bien classées à l'ECN
 - ▶ **Savoir-faire individuel** indépendant du rang de classement à l'ECN

* O. Steichen et al. Rev Med Int 2015; 36 : 312-318

Que pensez vous de ces évaluations

- Évaluation de l'apprentissage de la séméiologie digestive par un examen au lit du malade
- Évaluation de l'observation au microscope de coupes histologiques par des QCM
- Evaluation du savoir être avec des patients par des QROC
- Evaluation de la maîtrise de la prise de tension artérielle sur un patient simulé



Examens et Concours

▶ Examens :

▶ Tous les étudiants

- ▶ Ayant la **capacité de suivre** la formation peuvent la suivre
- ▶ Ayant suivi la formation et **ayant acquis les connaissances / les compétences** valident la formation

▶ Concours :

- ▶ Nombre de candidats admis **fixé à l'avance**,
 - ▶ le concours fait une **sélection**
 - ▶ Pour certains concours, examinateurs $\neq$ enseignants
 - ▶ => Incertitude sur les réponses attendues
 - ▶ Classer les étudiants au mieux
-



Examens et Concours :

Une Distinction souvent arbitraire

▶ Examens Classants

▶ Deux rôles

- ▶ Valider une formation
- ▶ Etablir une sélection par le classement

▶ Concours déguisés en examens et vice versa

- ▶ Formation sans concours d'entrée mais avec des débouchés limités
 - ▶ Sélection par étapes au travers d'examens difficiles
- ▶ Concours avec un nombre de places supérieur au niveau minimal requis
 - ▶ Devient un examen sans sélection



Fiabilité des corrections ?

Présentation de diverses expériences

Voici trois QROC posés lors d'un examen avec leur grille de correction.
Faites la correction.



Dispersion importante entre les correcteurs

- ▶ 1930. Pr. Laugier. Multi-correction de copies d'agrégation d'histoire
 - ▶ **166** Copies corrigées par **2** professeurs indépendants en aveugle.
 - ▶ Moyenne des notes du premier supérieure de 2 points à celle du second
 - ▶ Candidat classé avant dernier par le premier, classé second par l'autre
 - ▶ 21 copies avec une note ≤ 5 pour le premier notées entre 2 et 14 par le second
 - ▶ **Moitié des candidats reçus par le premier, refusés par le second**
-



La double correction : Une illusion (1)

- ▶ La double correction est illusoire (Laugier et Weinberg)
 - ▶ il faudrait, pour obtenir une note "exacte" (Une note "exacte" étant une moyenne de notes telle que l'adjonction d'une autre note ne modifie pas sensiblement cette moyenne) .
 - ▶ - 127 correcteurs en philosophie
 - ▶ - 78 en composition française
 - ▶ - 28 en anglais - 19 en version latine
 - ▶ - 16 en physique
 - ▶ - 13 en mathématiques
-



La double correction : Une illusion (2)

- ▶ Pas de meilleur résultat quand un correcteur reprend ses copies plus tard !!!.
 - ▶ Un professeur de physiologie de la Faculté des sciences accepta **37 copies** - dactylographiées et anonymes - qu'il avait corrigées **trois ans et demi auparavant**.
 - ▶ Dans **7 cas** seulement, il remit la même note au même devoir.
 - ▶ Dans les 30 autres cas, il y eut des divergences comprises **entre 1 et 10 points**.
-



Il n'est pas nécessaire d'être enseignant pour corriger....

- ▶ « On demanda à **une bachelière**, intelligente mais **ignorant tout de la question traitée**, de noter à son tour des compositions de physiologie, après les avoir lues une fois pour se faire une idée du sujet.
 - ▶ Ses notes eurent une corrélation de 0,51 (semblable à celle entre les enseignants correcteurs) avec celles attribuées par les professeurs compétents.
 - ▶ **La bachelière ne se trouvait pas plus en désaccord avec les spécialistes que ceux-ci entre eux** ». (Science et Vie, 1968, n° 610).
-



Expériences de Rosenthal (1)

- ▶ **Trois groupes** d'expérimentateurs devaient étudier **le comportement de rats** et les noter, au cours d'un programme comprenant des techniques classiques (apprentissage du labyrinthe et boîte de Skinner).
-

Expériences de Rosenthal (2)

- ▶ Informations différentes à chaque groupe :
 - ▶ à la moitié des expérimentateurs : les rats à étudier ont été spécialement choisis pour se comporter brillamment,
 - ▶ à l'autre moitié : les rats ont été spécialement choisis pour être stupides
 - ▶ les rats ont été choisis au hasard.
-

Expériences de Rosenthal (3)

- ▶ Bien meilleur apprentissage pour le groupe des rats jugés brillants que pour le groupe des rats jugés stupides.
 - ▶ L'évaluation est fortement teintée par la personnalité de l'évaluateur.
 - ▶ L'enseignant s'évalue autant qu'il évalue ses étudiants!
-



Pourquoi des discordances dans les corrections?

Quelles sont les causes possibles de discordances dans les corrections ?



Discordances repérées dans les pratiques d'évaluation

- ▶ 1. Effet de fatigue ou d'ennui
 - ▶ Peut engendrer laxisme ou sur-sévérité
- ▶ 2. Effet de contagion ou de contamination des copies
 - ▶ Les notes attribuées successivement aux différents aspects d'un même travail s'influencent mutuellement.



Discordances repérées dans les pratiques d'évaluation

▶ 3. Effet de halo

- ▶ Le professeur, influencé par des caractéristiques de présentation (soin, écriture, orthographe) surestime ou sous-estime la note.

▶ 4. Effet de l'ordre de correction des copies

- ▶ Devant un nouveau travail ou un nouveau candidat à évaluer, un juge se laisser influencer par la qualité du candidat précédent. Un travail moyen paraîtra bon s'il suit un travail médiocre.
-



Discordances repérées dans les pratiques d'évaluation

▶ 5. Effet d'ancrage ou de relativisation

- ▶ Plutôt que de juger intrinsèquement d'un travail, les professeurs jugent ce dernier en fonction des travaux dans lesquels il est inséré.

▶ 6. Effet de stéréotypie

- ▶ Le professeur maintient un jugement immuable sur la performance d'un élève, quelles que soient ses variations effectives

▶ 7. Effet de tendance centrale

- ▶ Par crainte de surévaluer ou de sous-évaluer un élève, le professeur groupe ses appréciations vers le centre de l'échelle.
-



Discordances repérées dans les pratiques d'évaluation

▶ 8. Effet de flou

- ▶ Les objectifs poursuivis et les critères de notation ne sont pas toujours définis avec précision

▶ 9. Effet de trop grande indulgence et de trop grande sévérité

- ▶ Certains juges sont systématiquement trop indulgents ou trop sévères dans toutes leurs évaluations
-



Influences du correcteur (2014)

- ▶ Si **apprenant connu** :
 - ▶ Si bon élève -> meilleure note
 - ▶ Sexe = filles mieux corrigée
 - ▶ Aspect physique
- ▶ Si **copie anonyme** :
 - ▶ Ordre de correction : 1^{er} tiers noté de façon plus indulgente
 - ▶ Niveau copie précédente :
 - ▶ Bonne copie, suivante note plus sévère
 - ▶ Hasard : variation pour la même copie corrigée deux fois

Intérêt d'une grille de correction

Plusieurs expériences

La grille (barème de correction)

- ▶ La plus grande partie des épreuves d'un examen fait intervenir un barème plus ou moins précis selon la matière.
 - ▶ Ce barème a pour but d'uniformiser la codification des appréciations.
 - ▶ **Mais ce barème que les correcteurs sont "priés de respecter scrupuleusement" est-il un instrument fiable ?**
 - ▶ Plusieurs équipes de chercheurs se sont penchés sur ce problème.
-



La grille : un exemple

- ▶ Influence de la grille sur la note à un dossier de rhumatologie pour le concours d'internat
- ▶ Dossier de rhumatologie envoyé à plusieurs rhumatologues
 - ▶ 1^{ère} grille : items retenus uniquement s'ils sont proposés par tous les enseignants
 - ▶ 2^{ème} grille : tous les items proposés (même par un seul enseignant) sont pris en compte
 - Nombre de points donné à chaque item correspond au nombre d'enseignants ayant noté l'item

Enoncé du dossier

- ▶ **Femme de 71 ans avec douleur lombaire depuis trois semaines**
 - ▶ Intensité douloureuse progressive, irradiant depuis 10 j vers la fesse D et face postérieure cuisse D
 - ▶ Douleur exacerbée au mouvement
 - ▶ **ATCD :**
 - ▶ Hystérectomie à 47 ans
 - ▶ Ablation de varices
 - ▶ Cystopexie pour prolapsus il y a 6 semaines
 - ▶ **Examen :**
 - ▶ raideur vertébrale majeure
 - ▶ Douleur à l'ébranlement des dernières vertèbres
 - ▶ Examen neuro normal
 - ▶ Examen général normal
-



Q1 : Quels diagnostics pouvez vous évoquer ? Argumentez

- ▶ **Grille n° 1**
 - ▶ Spondylodiscite à pyogène
 - ▶ Tassement vertébral : métastase et/ou Myélome
- ▶ **Grille n° 2**
 - ▶ Spondylodiscite
 - ▶ À pyogène
 - ▶ Tuberculeuse
 - ▶ Inflammatoire dans le cadre d'une spondylarthropathie ou d'une chondrocalcinose ou arthrite septique interapophysaire postérieure
 - ▶ Tassement vertébral : métastase et/ou Myélome
 - ▶ Tassement vertébral bénin
 - ▶ Lombosciatique discale
 - ▶ Fracture du sacrum
 - ▶ Sacriilélite infectieuse
 - ▶ Lombalgie d'origine articulaire arthrosique
 - ▶ Anévrisme de l'aorte compliqué
 - ▶ neurinome



Q2 : Quels examens complémentaires en première intention ?

▶ Grille n° 1 :

- ▶ Rachis lombaire face et profil
- ▶ VS
- ▶ CRP

▶ Grille n° 2 :

- ▶ Rachis lombaire face et profil
- ▶ VS
- ▶ CRP
- ▶ Hémocultures
- ▶ Radiographie de bassin
- ▶ IRM d'emblée
- ▶ Calcémie
- ▶ Electrophorèse des protéines
- ▶ Échocardiographie cardiaque
- ▶ ECBU



Q3 : Il s'agit d'une spondylodiscite. Quels seront les principes du traitement ?

▶ Grille n°1

- ▶ Immobilisation
 - ▶ Repos au lit
 - ▶ Corset
- ▶ Antibiothérapie double
 - ▶ IV dans un premier temps
 - ▶ Per os

▶ Grille n°2

- ▶ Immobilisation
 - ▶ Repos au lit
 - ▶ Corset
- ▶ Antibiothérapie double
 - ▶ IV dans un premier temps
 - ▶ Per os
- ▶ Hospitalisation
- ▶ Rééducation et kinésithérapie d'entretien
- ▶ Prévention de phlébite et d'escarre
- ▶ Antalgiques
- ▶ Bilan de contrôle (radio et VS/CRP)



Résultats

- ▶ Reproductibilité des deux méthodes
 - ▶ Coefficient de corrélation interclasse : 0,92
 - ▶ Corrélation de Spearman : 0,97 ($p < 0,0001$)
 - ▶ Meilleures notes avec grille n°1
 - ▶ Majoration des notes des étudiants qui écrivent beaucoup (à tort ou à raison) avec grille n°2
 - ▶ Corrélation entre les deux mesures très satisfaisante
 - Classement des étudiants peu modifié
-

En conclusion

► Résultats :

- Grille simple (n°1) **plus facile** à corriger
- **Même pertinence** avec les deux grilles
- Mais **éventail de note plus faible** avec grille n°1 -> plus d'ex aequo

► Améliorations proposées :

☐ Questions fermées

- ☐ « donnez le diagnostic le plus probable »

☐ Utilisation des points négatifs

- ☐ Réponses aberrantes
- ☐ Réponses avec trop d'items avec réponses fermées
 - Liste de diagnostics quand il est demandé LE diagnostic le plus probable

☐ Éviter les listes complètes qui comportent à tout coup la bonne réponse

Copies de mathématiques :

- ▶ **L'I. R. E. M* de Grenoble** a fait corriger 22 copies de mathématiques du B.E.P.C. par 10 professeurs
- ▶ **Cinq d'entre eux l'ont fait avec barème, les autres sans barème.**
- ▶ **Résultats**
 - ▶ Les utilisateurs du barème ont corrigé plus sévèrement
 - ▶ L'écart entre les notes extrêmes est moindre quand on tient compte du barème
 - ▶ Le barème ne supprime pas la dispersion des notes.
- ▶ Le barème se limite à un rôle vaguement modérateur

** Institut de Recherche sur l'Enseignement des Mathématiques*



Même expérimentation

- ▶ **En 1975**, l'IREM de Toulouse entreprend une expérience analogue de **multi-correction**.
- ▶ Echantillon de 6 copies de mathématiques (niveau BEPC)
- ▶ 64 correcteurs,
- ▶ Barème, très précis, sur 40 points.

Les résultats confirment ceux de l'enquête précédente,

- ▶ **La dispersion des notes atteint près de 20 points.**
-



Autre expérimentation

- ▶ **Le Groupe de Recherche de Montauban dirigé par Cransac et Dauvisis 1975.**
 - ▶ 15 heures de travail pour se mettre d'accord et élaborer un barème très précis (sur 120 points) ramené à une note entre 0 et 20.
 - ▶ **Malgré cet effort pour une copie la note varie de 4 à 13 sur 20,**
 - ▶ Chaque correcteur concerné a justifié sa notation et il a été impossible de donner raison à l'un ou à l'autre.
 - ▶ Certains ont donné tour à tour raison à l'un puis à l'autre.
 - ▶ D'autre part, le barème devient illusoire quand dans "une réponse peu claire, l'un voit un bon raisonnement, l'autre un raisonnement faux
-



Suite ...

- ▶ => **Le barème imposé est appliqué en fonction de la personnalité des correcteurs.**
 - ▶ Certains l'appliquent à la lettre en faisant fi de leurs sentiments et de leur expérience pédagogique,
 - ▶ d'autres ne peuvent ou ne veulent le faire et modulent le dit barème."
-



De l'importance de la formulation du sujet

- ▶ **A la question** : « On achète deux pains, on donne dix francs, la marchande rend une pièce de cinq francs et une de un franc ; quel est le prix d'un pain ? » on obtient **93,44 %** de réponses exactes dans une population d'élèves de troisième.
- ▶ **A la question** : « Résoudre dans $\mathbb{R}$: $2x+6=10$ », le pourcentage de réponses exactes tombe à **81,49 %**.
- ▶ **A la question** : « Résoudre dans $\mathbb{R}$: $2000x+6000=10000$. » on n'obtient plus que **60,34 %** de bonnes réponses.

Groupe de Recherche d'Epernay de l'I.R.E.M. de Reims 1975



Qualités de l'évaluateur

- ▶ Trois questions que se posent les enseignants face à une copie :
 - ▶ Est-ce que dans un mois je considérerai excellent un devoir que je considère excellent aujourd'hui ?
 - ▶ Sera-t-il jugé excellent par un collègue ?
 - ▶ Reproductibilité - Fidélité
 - ▶ Est-ce que je mesure correctement ce que je cherche à mesurer ?
 - ▶ Validité
 - ▶ Est-ce que mes notes varient assez finement en fonction des différences de qualité des travaux à évaluer ?
 - ▶ Sensibilité



Les différents types d'épreuves

Plusieurs possibilités

- ▶ Questions rédactionnelles
 - ▶ Questions conduisant à une réponse ouverte et courte (Q.R.O.C.)
 - ▶ Questions à choix multiples de réponse (Q.C.M.)
 - ▶ Exercice à trous
 - ▶ Mise en correspondance et ordonnancement
 - ▶ Annotation d'image
 - ▶ Epreuves de travaux pratiques
 - ▶ Epreuves orales
 - ▶ Mises en situations fictives (simulation) ou réelles
 - ▶ Examen Clinique Objectif Structuré (E.C.O.S.)
 - ▶ Tests de concordance de script
 - ▶ Port folio
-



Quel que soit le type d'épreuve

- ▶ 3 phases
 - ▶ 1° Conception du sujet
 - ▶ Choix de la méthode (question ouverte, QCM,...)
 - ▶ Rédaction des questions
 - ▶ 2° Correction « brute »
 - ▶ 3° « Post » traitement docimologique et rôle du jury
-

Pour chacune des situations proposées, trouver le mode d'évaluation le mieux adapté



Questions rédactionnelles

- ▶ **Objectifs :**

- ▶ Montrer que les étudiants sont « capables » :
 - ▶ de lire et de comprendre la question
 - ▶ d'analyser,
 - ▶ d'argumenter,
 - ▶ de faire preuve des capacités de synthèse,
 - ▶ d'écrire un document structuré

- ▶ Développer l'esprit critique

- ▶ Mais :

- ▶ Limite l'étendue de l'investigation des connaissances des étudiants
 - ▶ Difficulté de la correction
-



QROC

- ▶ Question rédactionnelle à réponse ouverte et courte
- ▶ Avantage :
 - ▶ Explore les connaissances des candidats sur de nombreux sujets d'un enseignement
- ▶ Inconvénients
 - ▶ Correction manuelle
 - ▶ Difficultés de la grille
 - ▶ Liste de mots clés...
- ▶ Cas clinique de l'ECN



Question à Choix Multiple

▶ Question fermée

- ▶ Un énoncé
- ▶ Plusieurs propositions
 - ▶ Dont une est juste, les autres fausses : QCM Simple (qcm)
 - ▶ Dont plusieurs sont justes : QCM Multiple (multiselection)
- ▶ Le nombre de propositions est variables en général de 3 à 5
- ▶ Présentation
 - ▶ Forme habituelle : on coche les propositions justes
 - ▶ Forme vrai/faux : on coche pour chaque proposition si elle est vraie ou fausse
 - ☐ Cette façon permet de prendre ne compte la notion « Je ne sais pas »



QCM

- ▶ Finalité initiale : tester les connaissances en balayant largement le programme
 - ▶ Mais dans le cadre de l'épreuve
 - ▶ On peut faire des QCM « intelligents » qui peuvent tester un raisonnement
 - ▶ QCM en cascade....
 - ▶ Application : très nombreuses dans tous les domaines
 - ▶ Aéronautique : Epreuve théorique des examens de pilote du pilote privé au pilote de ligne
 - ▶ Permis de conduire
 - ▶ Médecine : PACES
-



QCM

► Avantages :

- Explore les connaissances des candidats sur de nombreux sujets d'un enseignement
- Correction :
 - Manuelle : transparent posé sur chaque copie
 - Informatique avec lecteur :
 - ☐ Grille optique
 - ☐ Boitier individuel



QCM

► Inconvénients

- Difficulté de rédaction
- Réponses au hasard
- Perversité du raisonnement :
 - chercher les propositions fausses
 - Récurrence statistique (pour un enseignant les assertions vraies sont souvent au même endroit ou caractérisées par des éléments comme la longueur)



QCM : Correction

- ▶ QCM présentation traditionnelle
 - ▶ QCM « simple » :
 - ▶ L'étudiant a coché uniquement la seule réponse exacte, il a les points attribués à la QCM sinon il a 0
 - ▶ QCM « à réponse multiple »
 - ▶ **Conformité stricte** : l'étudiant a coché toutes les propositions exactes, il a les points attribués à la QCM sinon il a 0



QCM : Correction

► QCM « à réponse multiple »

► Conformité partielle :

- ☐ L'étudiant a coché une partie des propositions exactes et aucune des propositions fausses, il a une partie des points attribués à la QCM, sinon il a 0
 - ☐ *Par exemple Une QCM valorisé 6 points comporte 5 propositions dont 3 sont exactes, L'étudiant n'a coché aucune proposition fausse mais*
 - ☐ *Il a coché une seule proposition juste : il a 2 points, Il coché 2 des 3 propositions justes, il a 4 points, il a coché les 3 propositions juste, il a 6 points*
 - ☐ *Barème linéaire*
-



QCM : Correction

► QCM « à réponse multiple »

☐ **D'autres barèmes** peuvent être utilisés :

☐ Barème non linéaire : Avec l'exemple cité, si l'étudiant à coché les 3 propositions exactes, il a 6 points, 2 propositions parmi les 3 il a 2 points et 0 si il n'a coché qu'une proposition

☐ **Conformité à un cadre** (profil) de réponse

☐ on attribue un certain nombre de points pour des profils de réponses combinant des propositions exactes et fausses,



QCM avec présentation Vrai/Faux

- ▶ Possibilité de prendre en compte « je ne sais pas »
- ▶ Exemple
 - ▶ Réponse attendue A : **V**, B : **F**, C : **V**, D : **F**, E : **F**
 - ▶ Réponse de l'étudiant A : **V**, B : -, C : **V**, D : **F**, E : **F**
 - ▶ L'étudiant n'a pas répondu à la proposition B
 - (à noter que ceci n'était pas repérable avec la forme traditionnelle dans laquelle il aurait tous les points)
 - ▶ Conformité stricte : il a 0



QCM avec présentation Vrai/Faux

- ▶ Possibilité de prendre en compte « je ne sais pas »

▶ Exemple

▶ Réponse attendue A : **V**, B : **F**, C : **V**, D : **F**, E : **F**

▶ Réponse de l'étudiant A : **V**, B : -, C : **V**, D : **F**, E : **F**

▶ Conformité partielle :

- ☐ Les points attribués vont dépendre de la prise en compte ou non de non réponse en B :
- ☐ Si on considère que la non réponse correspond à une réponse inexacte la non réponse est assimilée à F et en barème linéaire, l'étudiant a 4/5 des points
- ☐ Si on considère que la non réponse correspond à une réponse exacte la non réponse est assimilée à V et en barème linéaire, l'étudiant a 5/5 des points (cf forme traditionnelle)
- ☐ Si on considère que la non réponse correspond à une réponse à valoriser de manière arbitraire, la non réponse est valorisée arbitrairement et en barème linéaire, l'étudiant a cette des points entre 4/5 et 5/5

Exercices à trous

- ▶ Il s'agit de compléter un texte :
 - ▶ « Dans la rougeole, l'érythème est de type(1)
l'éruption (2) de peau saine. La personne (2)
prurit . La rougeole est due à à »
 - ▶ (1) morbiliforme; (2) laisse des intervalles ; (3) ne présente pas
de; (4) un virus à ARN
- ▶ Avantages
 - ▶ Large couverture des connaissances
 - ▶ Guidage de l'étudiant (précision)
- ▶ Inconvénients :
 - ▶ Correction manuelle
 - ▶ Difficulté de correction (synonymes, périphrases....)



Mise en correspondance, ordonnancement

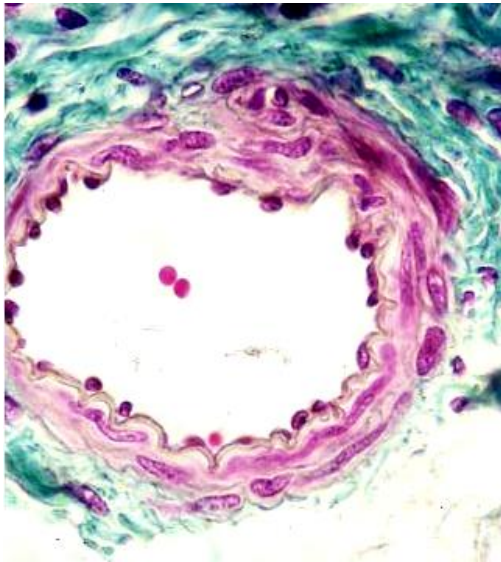
► Exemple :

- « Ordonner les différentes étapes de traitement d'un biopsie à son arrivée au laboratoire en vue de son étude anatomopathologie :
 - Montre un savoir faire
 - Fixation, inclusion, coupe, coloration
- Associer à chacun des médicaments son utilisation.....



Annotation d'image ou de schéma

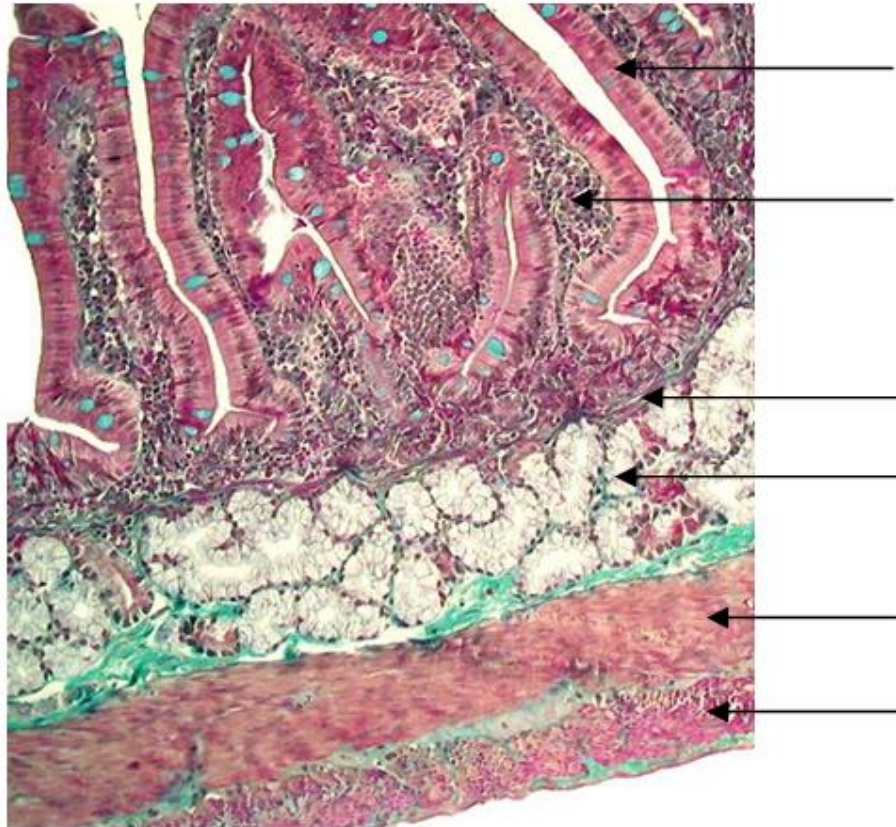
- ▶ Exemple sans orientation pour l'étudiant
- ▶ Annoter les différentes structures de cette photographie



Annotation de photographies

► Avec aide pour l'étudiant

1. Annotez la micrographie ci dessous



De quel organe s'agit-il ?

Epreuves de travaux pratiques

- ▶ L'étudiant est mis en situation (observation au microscope par exemple).
 - ▶ Il a besoin de connaissances théoriques
 - ▶ Il doit être capable d'observer
 - ▶ Il doit être capable de faire une synthèse de son observation
- ▶ Avantages
 - ▶ Permet d'analyser une démarche diagnostique transposable à d'autres situations
- ▶ Inconvénients
 - ▶ Prend du temps
 - ▶ subjectivité



Epreuves orales

- ▶ Avec temps de préparation
 - ▶ Permet d'apprécier
 - ▶ Les capacités de synthèse sur un sujet donné
 - ▶ Les capacités de s'exprimer
 - ▶ Inconvénients
 - ▶ Sujet très limité
 - ▶ Grande subjectivité



Mises en situations fictives (simulation) ou réelles

- ▶ Simulation de raisonnement
 - ▶ Cas clinique évolutif
- ▶ Simulation de gestes
 - ▶ Mannequin du plus simple au plus complexe
- ▶ Simulation de comportement de groupe
 - ▶ Mise en situation d'une équipe



Examen Clinique Objectif Structuré (E.C.O.S.) Harden 1975

- ▶ **L'Examen Clinique Objectif Structuré** : examen à stations multiples, avec patients réels ou simulés
 - ▶ Évalue les habiletés, les attitudes ou les aptitudes, ainsi que les aspects cognitifs
 - ▶ Outil d'évaluation formative et normative.
 - ▶ Réputé pertinent pour sa fiabilité et sa validité lors de l'évaluation clinique.
 - ▶ Chaque compétence est découpée en tâches facilement évaluable
-



Examen Clinique Objectif Structuré (E.C.O.S.) Harden 1975

- ▶ Environ 20 postes d'une durée de quatre à cinq minutes
 - ▶ Deux types: répondre à des questions ou exécuter des tâches.
 - ▶ Histoire du patient, examen physique, interprétations des examens para-cliniques, traitement, prévention, communication et éducation du patient
 - ▶ L'étudiant évolue de poste en poste,
 - ▶ L'évaluation porte sur la connaissance (le savoir), sur l'habileté technique (savoir-faire), sur le comportement (savoir être).
 - ▶ L'évaluateur ne fait qu'observer, il ne doit pas communiquer avec l'étudiant.
 - ▶ Grille d'items afin d'élaborer un score donc une note..
-



ECOS

► Avantages

- Proche de la réalité
 - Plusieurs évaluateurs impliqués
 - Standardisation
 - Validité et fiabilité de ce type d'examen
 - Multiples compétences évaluées
 - "Globalité" de l'évaluation
 - Évaluation longitudinale
-



ECOS

► Inconvénients

- Lourdeur d'organisation et coûts
 - Difficulté de réalisation avec de vrais patients
 - Situation clinique artificielle
 - Difficulté de "jouer le rôle"
 - Certains signes cliniques sont impossibles à simuler
 - Facteur stress (cf oraux)
 - Difficultés d'élaborer une grille
 - Difficile pour l'évaluateur de ne pas commenter
-



Tests de concordance de script (TCS)

- ▶ Le test de concordance de script (TCS) est une technique d'évaluation du raisonnement clinique en contexte d'incertitude par une simulation de situations
 - ▶ Ils apprécient l'organisation des connaissances et non uniquement les connaissances factuelles.
 - ▶ Ils s'appuient sur la théorie hypothético-déductive du raisonnement clinique et sur la théorie des scripts, elle même issue de la psychologie cognitive
 - ▶ Ils permettent de discriminer des niveaux d'expérience
-



Comment se concrétise un TCS

- ▶ Une courte mise en situation « vignette » puis des items
- ▶ Un item

Si vous pensez à..	Et qu'alors vous trouvez ...	L'effet sur l'hypothèse est ...
Une hypothèse	Nouvelle information	-2 □ -1 □ 0 □ +1 □ +2 □
[...]	[...]	[...]

La grille de réponse est :

- -2 : Pratiquement éliminé
- -1 : Moins Probable
- 0 : Non influencée
- +1 : Plus probable
- +2 : Pratiquement certaine

Exemple

Un item :
correspond à la façon dont le clinicien mobilise ses connaissances pour traiter et résoudre le problème.

Une vignette : une situation clinique problématique même pour un expert, mais du niveau d'étudiants à évaluer

La nouvelle donnée
permet de tester la force qui l'unit à l'option, dans le contexte précis de la vignette

Scénario clinique :

Une jeune femme de 22 ans, accompagnée de son mari, consulte pour des métrorragies abondantes faites de sang rouge et douleurs abdominales importantes. Elle ne prend aucune contraception depuis 6 mois car elle désire un enfant. Elle déclare avoir eu ses règles il y a 6 semaines.

Item #1

Si vous pensez à :

Une grossesse extra-utérine

... et que vous trouvez à l'échographie :

Une image évoquant un caillot intra-utérin

L'effet de cette nouvelle donnée sur votre hypothèse diagnostique sera (cocher votre réponse) :

- ☐ -2 Elle rend l'hypothèse beaucoup moins probable
- ☐ -1 Elle rend l'hypothèse moins probable
- ☐ 0 La nouvelle donnée n'aura aucun effet sur l'hypothèse diagnostique.
- ☒ +1 Elle rend l'hypothèse plus probable
- ☐ +2 Elle rend l'hypothèse beaucoup plus probable

La méthode de correction permet de mesurer et de comparer chacun des jugements fait par le candidat avec celui d'un panel de cliniciens expérimentés

Les ancrages de l'échelle de Likert correspondent à la réalité du raisonnement clinique : une seule donnée permet rarement d'affirmer qu'il ne peut s'agir que de ce diagnostic, mais mis dans une vignette, une donnée peut rendre pratiquement certain ce diagnostic

Score des experts

- ▶ Panel de 10 à 20 experts
- ▶ Exemple de résultats

Une jeune femme de 24 ans vient consulter pour un écoulement vaginal apparu depuis 1 semaine. Elle a un nouveau partenaire depuis 3 mois et est inquiète de la possibilité d'une maladie sexuellement transmissible.

Si vous pensez à :	Et qu'alors vous trouvez :	L'hypothèse diagnostique est :				
Mycose	Elle a eu une MST il y a 4 ans	-2	-1	0	+1	+2
Répartition des réponses des experts		0	2	7	1	0

Extrait de : Evaluation du raisonnement clinique. Le TCS. Sibert Louis.
<http://www.univ-rouen.fr/med/breeze/LStcs/index.htm>

Score de l'étudiant

- La réponse modale (la plus fréquente des experts) vaut 1 point
les autres réponses sont rapportées à la valeur modale

Si vous pensez à :	Et qu'alors vous trouvez :	L'hypothèse diagnostique est :				
Mycose	Elle a eu une MST il y a 4 ans	-2	-1	0	+1	+2
Répartition des réponses des experts		0	2	7	1	0

• Tous les items : score maximum de 1

0 : $7/7 = 1\text{pt}$ -1 : $2/7 = 0,29\text{pt}$ +1 : $1/7 = 0,13\text{pt}$

TCS

► **Avantages :**

- place les candidats à évaluer dans un contexte professionnel authentique.
 - confronte les candidats à une situation problématique,
 - sonde leur raisonnement
 - donne un score en comparant leurs réponses à celle d'un panel de référence.
 - Correction automatisable
 - TCS en ligne (département de médecine générale de Nancy)
 - Grille optique
-



TCS

► Inconvénients

- Difficulté de rédaction
- Choix du panel d'experts
 - Sensibilité au panel : exemple pour un TCS de pédiatrie : experts pédiatres et/ou experts généralistes



Portfolio

- ▶ **Dossier personnel** contenant les acquis de formation et les acquis de l'expérience d'une personne
 - ▶ Permet une reconnaissance.
 - ▶ « Book » des artistes
 - ▶ **Outil dynamique** qui permet de suivre l'évolution de la progression d'un élève dans ses apprentissages.
 - ▶ Mais pas un dossier contenant un tas de documents sans liens entre eux ni un recueil de productions, sans traces et sans réflexion.
 - ▶ Permet de se situer par rapport à un référentiel
-



Portfolio

► Buts :

- Suivre le cheminement de l'apprenant
 - Mieux comprendre son processus d'apprentissage Apprendre à l'étudiant à s'autoévaluer
 - L'aider à prendre conscience de ses apprentissages
 - Faire la preuve de l'atteinte du niveau de développement d'une compétence ou de plusieurs compétences
-



Portfolio

- ▶ **Caractéristiques du portfolio d'un point de vue de la pédagogie :**
 - ▶ Une actualisation permanente avec la possibilité de consigner l'historique du processus d'apprentissage.
 - ▶ Une ouverture sur le monde avec un accès de partout facilitant des interventions tierces (ePortfolio) .
 - ▶ Une structure composite qui oblige l'apprenant à définir et créer ses propres repères et à structurer son espace de travail.
 - ▶ Un support éditorial qui démultiplie les possibilités d'expression, l'occasion de faire jouer des connexions multiples propres à générer une démarche heuristique (ePortfolio).
-



Extrait du portfolio du DES de Santé Publique

A vu faire A fait avec aide A fait seul

Santé, déterminants, système de santé

a fait une activité dans laquelle il a pu montrer sa connaissance du système de santé français			
a réalisé une activité qui l'a amené à maîtriser les systèmes de santé des pays en voie de développement			
a réalisé une activité requérant la connaissance, la maîtrise et/ou l'interprétation des déterminants de l'état de santé des populations humaines			
a réalisé une activité impliquant l'utilisation des modèles définition et dimension de la santé			
a réalisé un travail requérant une source d'information utile en santé publique (PubMed exclus) utilisée en communication et recherche clinique et épidémiologique			

Recherche documentaire

a su sélectionner ou a participé activement à l'identification des mots clefs signifiants indispensables à la recherche des données nécessaires à la mise en place d'un projet (de santé publique)			
a réalisé une recherche bibliographique à partir des bases de données bibliographiques disponibles et utiles dans le champ de la thématique d'un projet (de santé publique).			
a su constituer une base de données bibliographiques utilisable pour la conduite d'un projet (de santé publique).			
a réalisé ou a participé activement lors de son stage, à une revue critique de la littérature sur un thème donné			

Communication

être capable de lire un article en anglais et de restituer son contenu			
être capable de communiquer à l'oral de façon adaptée par rapport au public visé			
être capable de rédiger un article rapportant une recherche (en anglais et en français)			

Informatique

maîtriser les compétences du C2i niveau 1			
maîtriser les compétences du C2i niveau 2			
maîtriser la conception d'une base de données relationnelles simple			
maîtriser un logiciel de statistiques généraliste			

Management

savoir conduire et animer une réunion de projet			
savoir énoncer les principes de la gestion d'équipe			
savoir construire la justification d'un budget			

Traitement des notes

Post traitement docimologique

Traitement global de l'épreuve

- ▶ La correction amène une note brute mais plusieurs types de traitement docimologique sont souvent nécessaires :
 - ▶ Par exemple dans une épreuve de 50 QCM, chaque QCM est notée 0 ou 1, la note brute théorique varie entre 0 et 50.
 - ▶ L'épreuve est notée entre 0 et 20.
-



- ▶ Comment ramener les notes brutes dans la fourchette imposée
- ▶ (de 0 à 20) ?

Traitement global de l'épreuve 1

- ▶ La note finale est une **simple règle de trois** : on ramène 0 à 0 et 50 à 20
 - ▶ Exemple : note brute = 27 / 50
 - ▶ Note sur 20 = $27/50 \times 20 = 10,8 / 20$
-

Traitement global de l'épreuve 2

- ▶ La note finale est telle que la **note brute minimale est ramenée à 0 et 50 à 20**
 - ▶ Soit le minimum = 10 ; maxi = 45 (18/20)
 - ▶ -> transformation linéaire $y = ax + b$
 - ▶ $0 = 10a + b$
 - ▶ $18 = 45a + b$
 - ▶ Si note brute = 27/50
 - ▶ Note pondérée = **8,74 /20**
-



Traitement global de l'épreuve 3

- ▶ La note finale est telle que la **note brute minimale est ramenée à 0 et la note brute maximale à 20**
 - ▶ Note brute mini 10 ; Note brute maxi 45
 - ▶ $0 = 10a + b$
 - ▶ $20 = 45a + b$
 - ▶ Si note brute = 27/50
 - ▶ Note pondérée = **9,71 / 20**
-



Avantages / inconvénient TTT 2 ou 3

- ▶ Permet de lutter contre le hasard
 - ▶ Remplissage d'une grille optique au hasard = 6 à 7 / 20
- ▶ Inconvénients
 - ▶ Étalement important si note la plus basse « haute »



Traitement question par question

- ▶ On peut vouloir pondérer les notes brutes :
 - ▶ De manière **a priori** :
 - ▶ Exemple dans une épreuve par QCM, les réponses fausses à certains QCM sont très pénalisantes (QCM qui tuent)
 - ▶ **A posteriori**
 - ▶ En tenant compte de la **fraction de réussite** à une question :
 - **Pénaliser fortement** l'étudiant qui ne sait pas ce que tout le monde sait
 - **Valoriser faiblement** celui qui connaît ce que la majorité des étudiants ignore.

QCM : Post traitement docimologique

- ▶ Chaque QCM peut avoir une pondération
 - ▶ A priori décidée par l'enseignant lors de la création du sujet
 - ▶ A posteriori calculée à partir de l'ensemble des réponses des étudiant par prise ne compte de la fraction de réussite :
 - ▶ Exemple une QCM qui a été répondue de façon exacte (FR) par 90% des étudiants sera mieux valorisée qu'une question avec une FR de 10% : On pénalise les étudiants qui ne savent pas ce que « tout » le monde sait.
- ▶ On doit ramener les notes brutes à une note finale dans la fourchette imposée
 - ▶ Transformation linéaire mais
 - ▶ Utilisation du minimum théorique et du maximum théorique
 - ▶ Utilisation du minimum observé et du maximum observé
 - ▶ Utilisation du minimum observé et du maximum théorique



L'effet jury :

- ▶ Dans une épreuve type ECN notée sur 120 points, 30 copies ont été confiées par tirage au sort à 3 binômes faisant une double correction. Chaque binôme constitue un « jury ».
-

L'effet Jury : Notes brutes

Etudiant	Jury	Note Initiale	Etudiant	Jury	Note Initiale	Etudiant	Jury	Note Initiale
1	1	39	11	2	68	21	3	120
2	1	8	12	2	71	22	3	120
3	1	87	13	2	79	23	3	108
4	1	10	14	2	88	24	3	104
5	1	40	15	2	96	25	3	88
6	1	90	16	2	102	26	3	88
7	1	45	17	2	104	27	3	72
8	1	34	18	2	108	28	3	62
9	1	47	19	2	115	29	3	44
10	1	60	20	2	120	30	3	28
Minimum		8			68			28
Maximum		90			120			120
Moyenne		46,00			95,10			83,40
Ecart type		27,455			18,058			31,553
CV		60%			19%			38%
Pour les 30 étudiants								
Minimum		8						
Maximum		120						
Moyenne		74,83						
Ecart type		33,13						
CV		0,44						

A vous :

Devant ces résultats que faites vous lors de la délibération ?

Comment réaliser une harmonisation entre les jurys ?

Quelles seront les conséquences de cette harmonisation ?

- ▶ En termes de classement (influence sur un concours)
- ▶ En termes de validation de crédits (avoir la moyenne à l'épreuve)



L'harmonisation selon la loi normale

► Hypothèse :

- La différence entre les jury ne s'explique pas par une différence de niveau des étudiants : les copies ont été réparties au hasard (Attention à l'anonymat reposant sur l'ordre alphabétique des noms des candidats par exemple)
 - La différence entre les jurys est liée à une sévérité différente de notation
-



L'harmonisation selon la loi normale

► Principe :

- On transforme la note brute de chaque jury en note centrée réduite $\Rightarrow$ Tous les jurys ont des notes ayant comme moyenne 0 et écart type 1
 - Toutes les notes centrées réduites font l'objet d'une transformation inverse en utilisant la moyenne générale et l'écart type général
 - Les notes ainsi obtenues sont ramenées dans la fourchette de notation exigée
-



Effet de l'Harmonisation

► Sur la validation des crédits

- **15 étudiants** sur 30 valident leurs crédits avec ou sans harmonisation
 - **6 étudiants** valident leurs crédits sans harmonisation mais ne les valident pas avec harmonisation
 - **2 étudiants** ne valident pas sans harmonisation mais valident leurs crédits avec harmonisation
-



Effet de l'Harmonisation

► Sur le classement

- Le classement est très influencé par l'harmonisation
 - Seulement 9 étudiants varient au plus de 2 places
 - 4 étudiants varient de plus de 9 places avec un maximum de 13
 - 7 étudiants sont classés dans les 10 premiers avec ou sans harmonisation
-



Bibliographie :

- ▶ André QUINTON. Docimologie. Mémoire de DIU de pédagogie
 - ▶ Abernot, Y. (1996). *Les méthodes d'évaluation scolaire*. Paris, Dunod.
 - ▶ Bacher, F. (1973). La docimologie. In M. Reuchlin (Ed.), *Traité de psychologie appliquée*, Vol. 6, Paris PUF.
 - ▶ Bloom, B.S. (1979). Caractéristiques individuelles et apprentissages scolaires. Editions Labor / Fernand Nathan.
 - ▶ Bonboir (1972). *La docimologie*. Paris, PUF.
 - ▶ Ketele (1982). Docimologie. Introduction aux concepts et aux pratiques, Louvain-la-Neuve, Cabay.
 - ▶ Landsheere (1980). *Evaluation continue et examens : précis de docimologie*. Bruxelles, Editions Labor.
-



Effet de l'harmonisation :

Et.	Jury	N.Ini	≥60	Place	N.Fin	≥60	Pl. H.	Et.	Jury	N.Ini	≥60	Place	N.Fin	≥60	Pl. H.
1	1	39		26	58,05		19	16	2	102	Oui	9	79,30	Oui	12
2	1	8		30	20,40		28	17	2	104	Oui	7	82,99	Oui	11
3	1	87	Oui	15	116,36	Oui	2	18	2	108	Oui	5	90,38	Oui	8
4	1	10		29	22,82		26	19	2	115	Oui	4	103,30	Oui	6
5	1	40		25	59,27		18	20	2	120	Oui	1	112,54	Oui	3
6	1	90	Oui	11	120,00	Oui	1	21	3	120	Oui	1	105,24	Oui	4
7	1	45		23	65,34	Oui	17	22	3	120	Oui	1	105,24	Oui	4
8	1	34		27	51,98		22	23	3	108	Oui	5	92,55	Oui	7
9	1	47		22	67,77	Oui	16	24	3	104	Oui	7	88,33	Oui	9
10	1	60	Oui	21	83,56	Oui	10	25	3	88	Oui	12	71,42	Oui	13
11	2	68	Oui	19	16,51		29	26	3	88	Oui	12	71,42	Oui	13
12	2	71	Oui	18	22,05		27	27	3	72	Oui	17	54,50		20
13	2	79	Oui	16	36,82		24	28	3	62	Oui	20	43,94		23
14	2	88	Oui	12	53,44		21	29	3	44		24	24,91		25
15	2	96	Oui	10	68,22	Oui	15	30	3	28		28	8,00		30

