



Recyclage de questions à choix multiples d'épreuves certificatives de physique en items de tests formatifs en ligne

Pierre-Xavier Marique (Didactique de la Physique – ULg)

Jean-François Van de Poël (IFRES – ULg)

Maryse Hoebeke (Didactique de la Physique – ULg)



Contexte

PUBLIC : 1Bac Médecine et Dentisterie ULg (~600 et.)

MATIERE : Physique (50h TH, 8h TP, 20h TD)

Environ **200 items** conçus chaque année

METHODE D'EVALUATION : QCM

MAIS étudiants peu (pas) entraînés aux QCM

→ **nécessité d'un outil d'entraînement !**

L'outil : le simulateur d'examens

Outil d'entraînement = **Simulateur d'examens**

Les questions du simulateur sont :

- d'anciennes questions d'examens (retravaillées/améliorées si besoin)
- classées par matière et par niveau de difficulté
- tirées aléatoirement lors de chaque test

Simulateur d'examen

3 paramètres

MATIERE

DUREE

DIFFICULTE

Simulateur d'examen

3 paramètres

MATIERE →

DUREE

DIFFICULTE

5 grands thèmes

60'

Optique 60'

Electricité 60'

Mécanique 60'

Fluides 60'

Imagerie 60'

Simulateur d'examen

3 paramètres

MATIERE

DUREE →

DIFFICULTE



Tests introductifs

Questions de base et de prérequis.

→ Débloque les simulations d'examens de la matière correspondante.



Simulateur d'examens

Questions des examens des années antérieures

Simulateur d'examen

3 paramètres

DUREE

MATIERE

DIFFICULTE

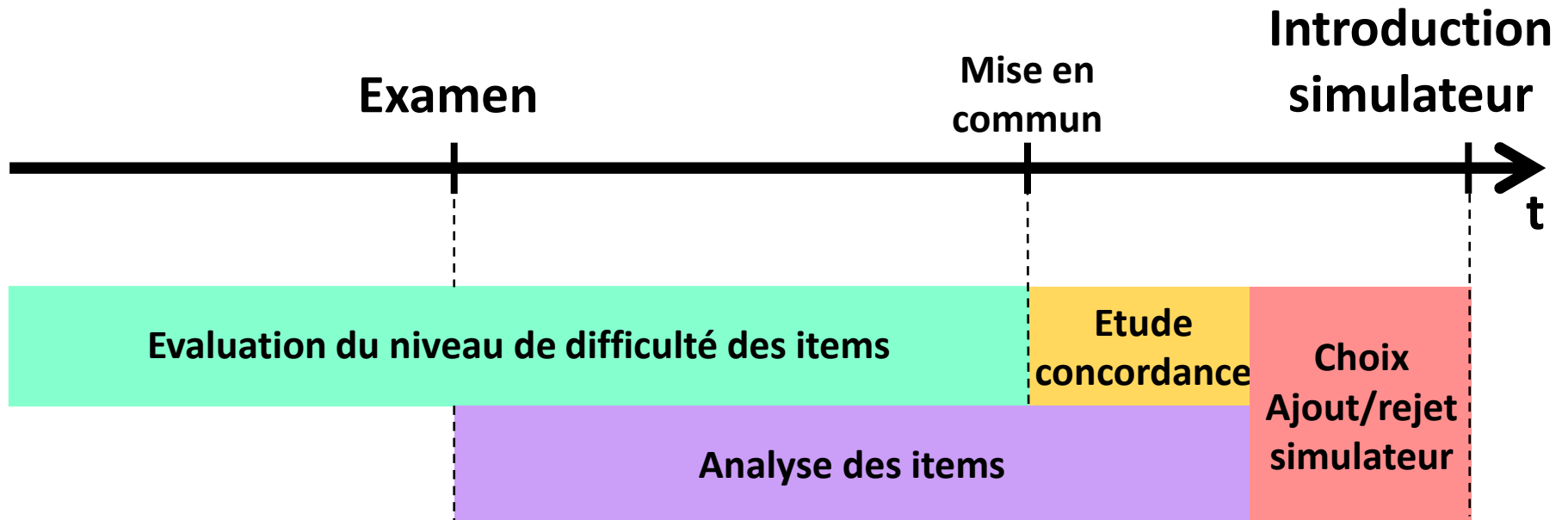
Test niveau A ———> Niveau le plus facile

Test niveau B

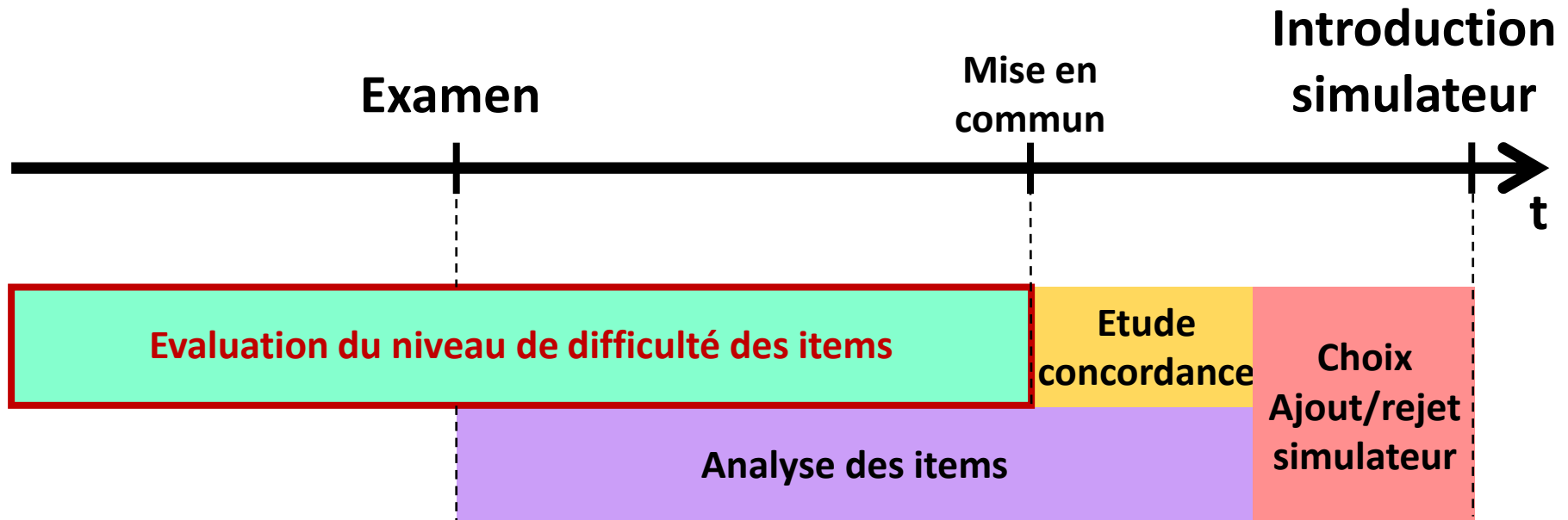
Test niveau C ———> Niveau le plus difficile

Dépend du niveau de difficulté des questions !

Recyclage des questions



Recyclage des questions



Niveau de difficulté

Test niveau A —————> Niveau le plus facile

Test niveau B

Test niveau C —————> Niveau le plus difficile

Le niveau de difficulté des items fixe le niveau des tests

- 4 niveaux de difficultés
- 4 critères d'évaluation
- Évalués par les membres de l'équipe pédagogique

Niveau de difficulté

4 critères :

- ***Abstraction***
- ***Réflexion***
- ***Mixité***
- ***Mathématiques***

Chaque évaluateur attribue une note de 1 à 3 à chaque critère.

La somme indique le ***niveau de la question.***

Niveau de difficulté

4 critères :

- **Abstraction**
- *Réflexion*
- *Mixité*
- *Mathématiques*

Exemple – niveau 1 :

Lorsqu'une artère a son diamètre multiplié par 2 au niveau d'un anévrisme, sa résistance à l'écoulement est :

1. **divisée par 16.**
2. divisée par 4.
3. inchangée car sa vitesse a quadruplé.
4. multipliée par 4.
5. multipliée par 16.

Illustration :

1. Enoncé clair, données, explicites, ...
2. Enoncé complexe, quelques données implicites, ...
3. Enoncé très complexe, relation à trouver, ...

Niveau de difficulté

4 critères :

- **Abstraction**
- *Réflexion*
- *Mixité*
- *Mathématiques*

Illustration :

1. Énoncé clair, données, explicites, ...
2. Énoncé complexe, quelques données implicites, ...
3. Énoncé très complexe, relation à trouver, ...

Exemple – niveau 3 :

En scintigraphie, on utilise des « fantômes » pour le contrôle d'uniformité d'image. Un fantôme est un cylindre rempli d'eau à laquelle on mélange un liquide radioactif. On peut alors scanner ce cylindre radioactif et vérifier si l'image obtenue est bien uniforme.

Dans un service de médecine nucléaire, on prépare un fantôme pour avoir 300 MBq de Tc-99m ($T_{1/2} = 6$ heures) au moment du scan avec une gamma-caméra. Malheureusement, on perd l'accès à la caméra que l'on pourra récupérer plus tard. Le deuxième scan est prévu 90 minutes après le moment auquel le premier scan était prévu. Quelle est l'activité à réinjecter dans le fantôme 10 min avant ce deuxième scan, pour retrouver 300 MBq au moment du deuxième scan ?

Niveau de difficulté

4 critères :

- *Abstraction*
- ***Réflexion***
- *Mixité*
- *Mathématiques*

Illustration :

1. 1 formule, 1 étape
2. Au moins 2 étapes
3. Nombreux passages d'une formule à une autre

Exemple – niveau 1 :

Un faisceau d'électrons se propage horizontalement dans un oscilloscope avant de passer entre deux plaques horizontales chargées. Ces plaques créent ainsi un champ électrique uniforme. La plaque supérieure est chargée positivement, la plaque inférieure négativement.

1. Le faisceau d'électrons ne sera pas dévié et continue donc en ligne droite horizontalement.
2. Le faisceau d'électrons suit une trajectoire circulaire orientée vers le haut.
3. Le faisceau d'électrons suit une trajectoire parabolique orientée vers le bas.
4. Le faisceau d'électrons suit une trajectoire rectiligne orientée obliquement vers le haut.
5. Le faisceau d'électrons suit une trajectoire hyperbolique orientée vers le bas.

6. Aucune

Niveau de difficulté

4 critères :

- *Abstraction*
- ***Réflexion***
- *Mixité*
- *Mathématiques*

Illustration :

1. 1 formule, 1 étape
2. Au moins 2 étapes
3. Nombreux passages d'une formule à une autre

Exemple – niveau 3 :

Un morceau de pin pèse 45 N dans l'air. Un bloc de plomb pèse 130 N quand il est immergé dans l'eau. Attaché l'un à l'autre, ils pèsent 85 N dans l'eau. Quelle est la masse volumique du bois ? ($\rho_{\text{plomb}} = 1100 \text{ kg/m}^3$)

1. 500 g/dm^3
2. $5000 \cdot 10^{-4} \text{ kg/l}$
3. $5 \cdot 10^5 \text{ g/m}^3$
4. $0,5 \cdot 10^3 \text{ g/l}$
5. $5 \cdot 10^{-1} \text{ g/cm}^3$
6. Aucune de ces propositions (1 à 5) n'est correcte
7. **Toutes ces propositions (1 à 5) sont correctes**

Niveau de difficulté

4 critères :

- *Abstraction*
- *Réflexion*
- ***Mixité***
- *Mathématiques*

Illustration :

1. 1 chapitre
2. 2 chapitres ou plusieurs éléments d'1 chapitre
3. Plus de 2 chapitres

Exemple – niveau 1 :

On estime la température de la flamme d'une bougie à 1577 °C. Quelle est la longueur d'onde principalement émise par cette bougie ?

1. $1,113 \cdot 10^{-6}$ m
2. $1,284 \cdot 10^{-6}$ m
3. **$1,566 \cdot 10^{-6}$ m**
4. $1,729 \cdot 10^{-6}$ m
5. $1,838 \cdot 10^{-6}$ m

Niveau de difficulté

4 critères :

- *Abstraction*
- *Réflexion*
- ***Mixité***
- *Mathématiques*

Illustration :

1. 1 chapitre
2. 2 chapitres ou plusieurs éléments d'1 chapitre
3. Plus de 2 chapitres

Exemple – niveau 2 :

Une antilope se trouve à une distance de 25 m d'une lentille convergente de focale 50 cm. La lentille forme une image de l'animal. L'antilope se met alors à courir en direction de la lentille à la vitesse constante de 5 m/s. Déterminez le moment auquel l'image réelle formée se trouvera à 60 cm de la lentille.

1. Il n'y a aucun moment auquel l'image formée se trouve à 60 cm de la lentille (on trouve en effet un temps négatif).
2. 25,20 s après le démarrage.
3. 4,4 s après le démarrage.
4. 12,4 s après le démarrage.
5. Je n'ai pas assez d'éléments pour répondre.

Niveau de difficulté

4 critères :

- *Abstraction*
- *Réflexion*
- *Mixité*
- ***Mathématiques***

Illustration :

1. 1 formule simple
2. Plusieurs formules simples ou 1 relation complexe
3. « Truc » à connaître, calcul spécifique, ...

Exemple – niveau 1 :

On estime la température de la flamme d'une bougie à 1577 °C. Quelle est la longueur d'onde principalement émise par cette bougie ?

1,113.10⁻⁶ m

1,284.10⁻⁶ m

1,566.10⁻⁶ m

1,729.10⁻⁶ m

1,838.10⁻⁶ m

Niveau de difficulté

4 critères :

- *Abstraction*
- *Réflexion*
- *Mixité*
- *Mathématiques*

Somme → Niveau de la question

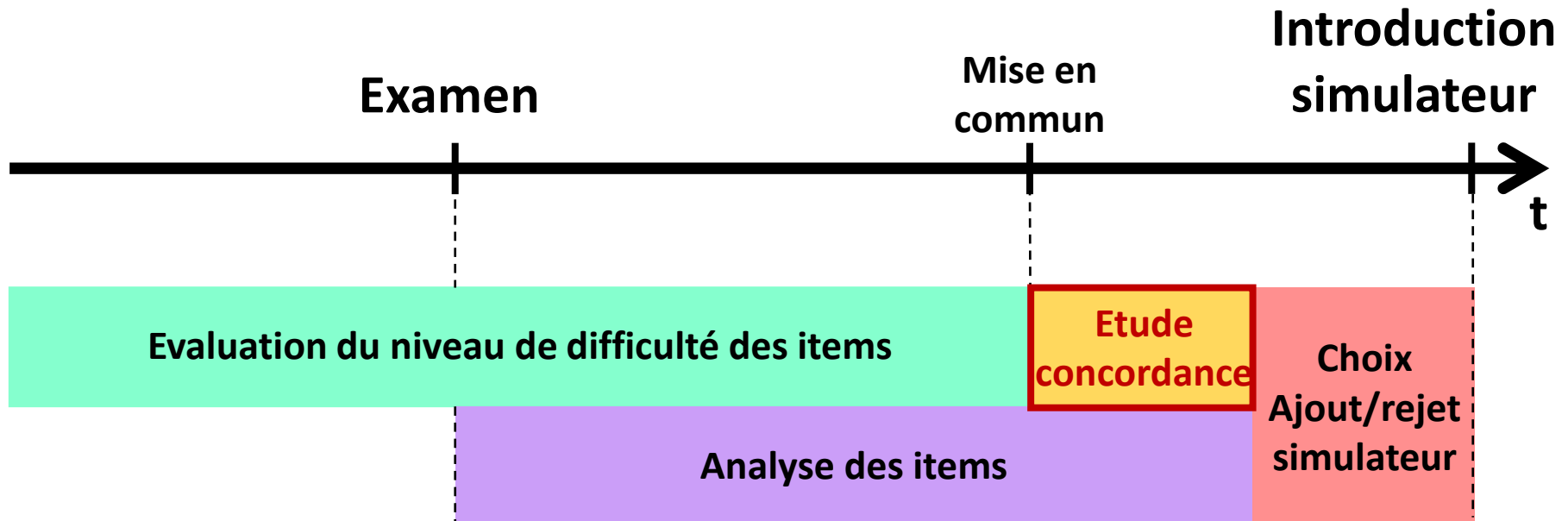
[4;5] → niveau 1

[6;7] → niveau 2

[8;9] → niveau 3

[10;12] → niveau 4

Recyclage des questions



Etude de la concordance inter-codeurs

Par item :

- minimum 3 évaluateurs
- niveau de difficulté de 1 à 4
- évaluation subjective

➔ mesure de la concordance : **Kappa de Fleiss**

Kappa de Fleiss κ ≡ Mesure statistique évaluant la concordance lors de l'assignation qualitative d'objets au sein de catégories par un certain nombre d'observateurs

Etude de la concordance inter-codeurs

Kappa de Fleiss κ \equiv Mesure statistique évaluant la concordance lors de l'assignation qualitative d'objets au sein de catégories par un certain nombre d'observateurs

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

$$\bar{P} = \frac{1}{Nn(n-1)} \left(\sum_{i=1}^N \sum_{j=1}^k n_{ij}^2 - Nn \right)$$

$$\bar{P}_e = \sum_{j=1}^k \left(\frac{1}{Nn} \sum_{i=1}^N n_{ij} \right)^2$$

où :

n = nombre d'évaluateurs (observateurs)

N = nombre de questions (sujets)

k = nombre de niveau de difficulté (catégories)

κ	Interprétation
< 0	Pauvre concordance
0,01 \rightarrow 0,2	Faible concordance
0,21 \rightarrow 0,4	Légère concordance
0,41 \rightarrow 0,6	Concordance moyenne
0,61 \rightarrow 0,8	Concordance importante
0,81 \rightarrow 1	Concordance presque parfaite

Etude de la concordance inter-codeurs

Kappa de Fleiss κ \equiv Mesure statistique évaluant la concordance lors de l'assignation qualitative d'objets au sein de catégories par un certain nombre d'observateurs

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

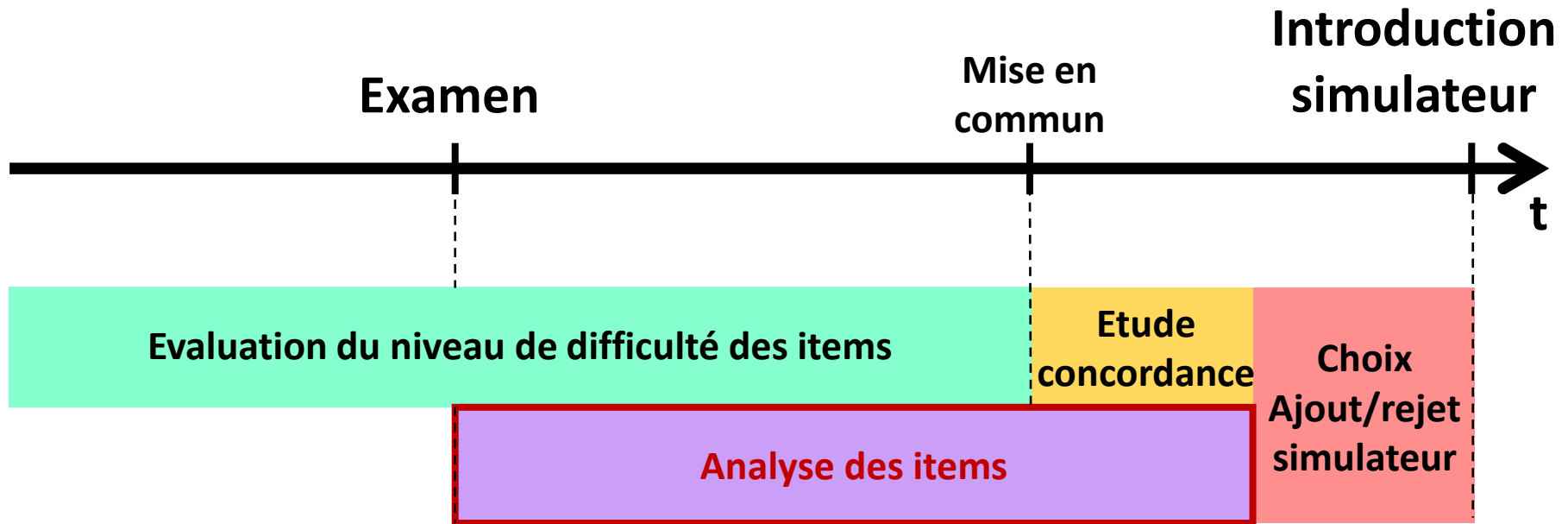
Examens janvier 2013 \rightarrow mai 2015 :

n	Nb Eval	3
N	Nb sujet (Q)	345
k	Nb cat (niv)	4

$$\kappa = 0,398$$

κ	Interprétation
< 0	Pauvre concordance
0,01 \rightarrow 0,2	Faible concordance
0,21 \rightarrow 0,4	Légère concordance
0,41 \rightarrow 0,6	Concordance moyenne
0,61 \rightarrow 0,8	Concordance importante
0,81 \rightarrow 1	Concordance presque parfaite

Recyclage des questions



Analyse des items

Après chaque examen,

analyse de chaque item à l'aide de \neq indices :

- Indice de facilité P (et indice corrigé P')
- Indice D de discrimination des 2 groupes extrêmes
- Indice B de discrimination au seuil de maîtrise

Analyse des items

Indice de facilité P (et indice corrigé P')

Indice p :

$$P = \frac{\sum_{s=1}^{ns} x_{qs}}{ns} \quad P \in [0 ; 1]$$

avec :

s = l'indice des sujets

x_{qs} = la note à la question q obtenue par le sujet s (ici 0 ou 1)

ns = le nombre total de sujets ayant répondu à la question q

Indice corrigé p' :

$$P' = P - \left(\frac{1 - P}{M - 1} \right) \quad P' \in [0 ; 1]$$

avec :

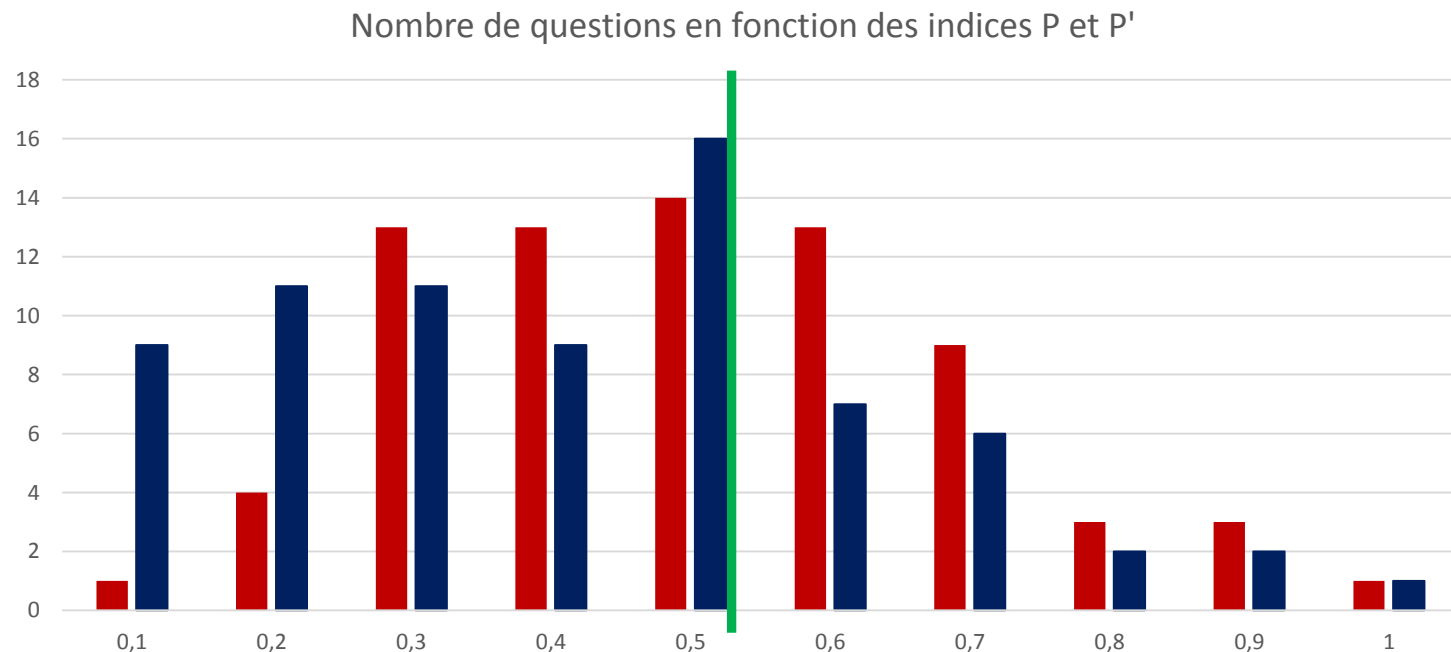
M = nombre de propositions contenues dans la QCM

(Laveault & Gregoire, 1997)

Analyse des items

Indice de facilité P (et indice corrigé P')

Exemple : examen de janvier 2013 (partie 1)



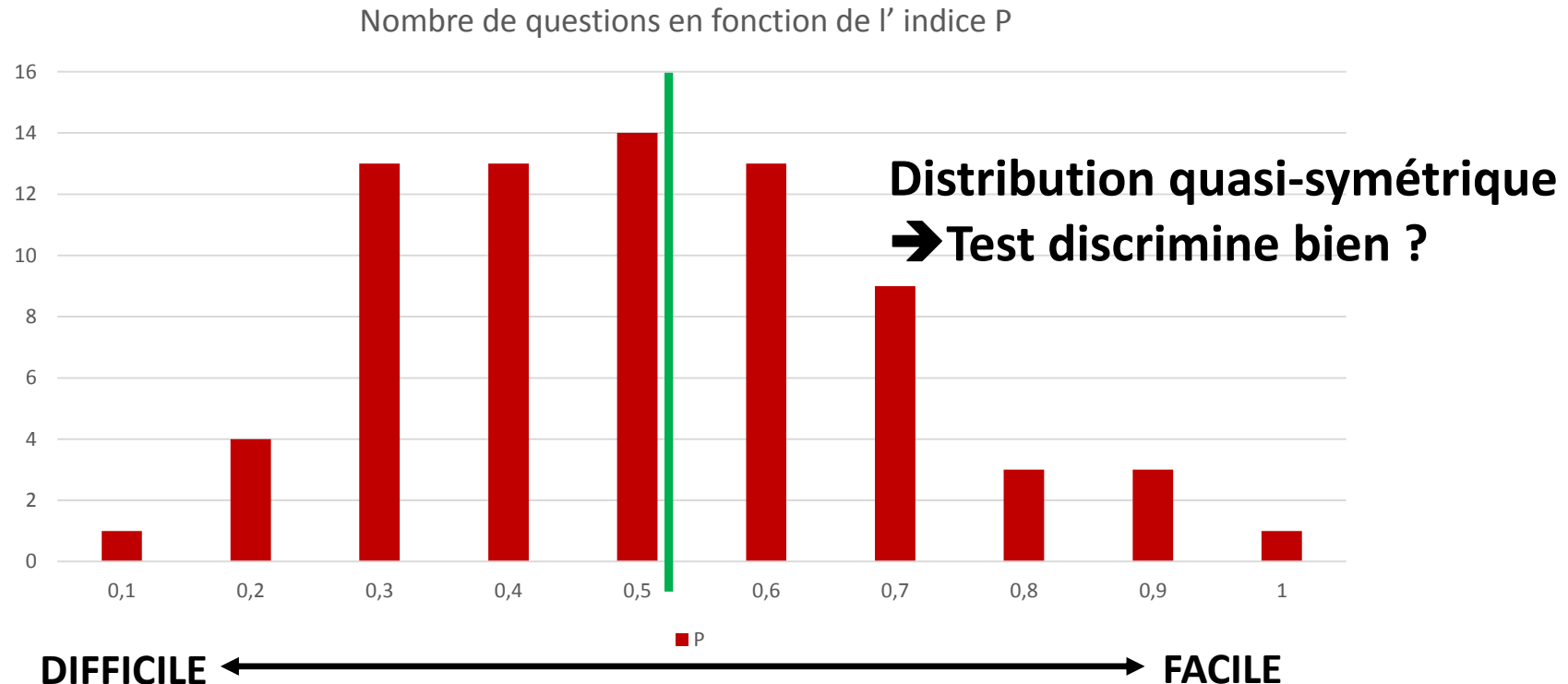
Indice P' tient compte de l'effet du hasard ■ P ■ P'

MAIS P' trop sévère car très peu plausible que le sujet réponde véritablement au hasard
Connaissance partielle → Elimination de discriminant(s)

Analyse des items

Indice de facilité P (et indice corrigé P')

Exemple : examen de janvier 2013 (partie 1)



Analyse des items

En incluant une échelle de 1 à 4 à partir de l'indice de facilité P :

$P \in [0 ; 0,25] \rightarrow 4$

$P \in]0,25 ; 0,5] \rightarrow 3$

$P \in]0,5 ; 0,75] \rightarrow 2$

$P \in]0,75 ; 1] \rightarrow 1$

**Comment évolue la concordance ?
(Kappa de Fleiss)**

Examen de janvier 2013 (74 questions)

3 évaluateurs (équipe pédagogique) : $\kappa = 0,478$

« 4 évaluateurs » (équipe + échelle P) : $\kappa = 0,253$

Analyse des items

Indices de discrimination :

Analyse des items

Indices de discrimination :

Principe : Mesure de la différence entre l'indice de facilité P d'un item obtenu par un groupe « fort » et par un groupe « faible »

Différence de définition des groupes « fort » et « faible » pour les indices D et B

	Groupe « Fort »	Groupe « faible »
Indice B	Seuil de maîtrise atteint	Seuil de maîtrise non atteint
Indice D	27 % meilleurs au test	27 % plus faibles au test

Analyse des items

Indices de discrimination :

Indice de discrimination B (seuil de maîtrise) :

$$B = p_{M+} - p_{M-}$$

avec :

p_{M+} : proportion de sujets qui ont atteint le seuil de maîtrise pour l'ensemble des questions du test et qui ont répondu correctement à l'item

p_{M-} : proportion de sujets qui n'ont pas atteint le seuil de maîtrise et qui ont répondu correctement à l'item.

Analyse des items

Indices de discrimination :

Indice de discrimination B (seuil de maîtrise) :

$$B = p_{M+} - p_{M-}$$

$$B \in [-1 ; 1]$$

Interprétation :

- Plus B est grand, mieux l'item discrimine
- Si $B = 0$: Pas de discrimination
- Si $B < 0$: Situation paradoxale

Exemple : examen de physique de janvier 2013 en 1^{ère} année de médecine à l'ULg :

74 questions (items) au total

73 items discriminent ($B > 0$)

1 item présente une anomalie ($B = -0,006$)

Analyse des items

Indices de discrimination :

Indice de discrimination D (groupes extrêmes) :

Indice D = Différence entre l'indice de facilité P de l'item obtenu par :

- 1 groupe constitué des 27% d'étudiants ayant les meilleurs scores sur l'ensemble du test (Groupe Supérieur [GS])
- 1 groupe constitué des 27% d'étudiants ayant les moins bons scores sur l'ensemble du test (Groupe Inférieur [GI])

$$D = p_+ - p_-$$

avec :

p_+ : la proportion de sujets du groupe GS ayant répondu correctement à l'item

p_- : la proportion de sujets du groupe GI ayant répondu correctement à l'item.

(Findley, 1956 ; Gilles, 2002 ; Kelley, 1939)

Analyse des items

Indices de discrimination :

Indice de discrimination D (groupes extrêmes) :

$$D = p_+ - p_-$$

$$D \in [-1 ; 1]$$

Interprétation :

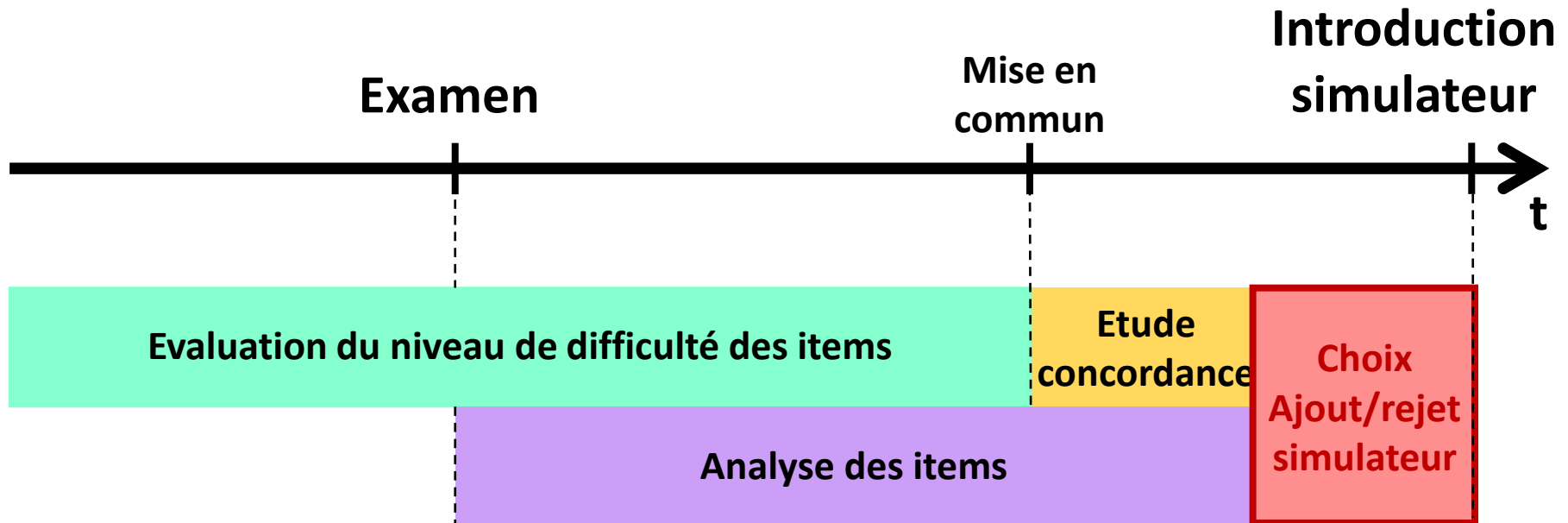
- Plus D est grand, mieux l'item discrimine
- Si $D = 0$: Pas de discrimination
- Si $D < 0$: Situation paradoxale

Ebel (1965) propose des valeurs repères pour interpréter l'indice D :

<u>valeur D</u>	<u>Item discrimine</u>
$D \geq 0,4$	très bien
$0,3 \leq D \leq 0,39$	bien
$0,2 \leq D \leq 0,29$	peu
$0,1 \leq D \leq 0,19$	<i>item à améliorer</i>
$D \leq 0,1$	<i>item inutile</i>

Janvier 2013 (74 Q)	N	%
discrimine très bien	40	54,05
discrimine bien	14	18,92
discrimine peu	11	14,86
à améliorer	4	5,41
Inutile	5	6,76

Recyclage des questions



Recyclage des questions

Insertion des questions dans le simulateur :

Pour l'instant :

- Niveau de difficulté attribué uniquement en fonction de l'évaluation effectuée par l'équipe pédagogique
- Question améliorée en fonction des indices B et D

A l'avenir :

- Niveau de difficulté attribué uniquement en fonction de l'évaluation effectuée par l'équipe pédagogique ET en fonction de l'indice de facilité P
- Question améliorée en fonction des indices B et D

Simulateur d'examen

3 paramètres

DUREE

50 % de questions de niveau 1
50 % de questions de niveau 2

MATIERE

Objectif : Obtenir 60 %

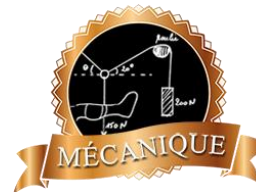
DIFFICULTE

Récompense : Médaille de bronze

Test niveau A

Test niveau B

Test niveau C



Simulateur d'examen

3 paramètres

DUREE

MATIERE

DIFFICULTE

Test niveau A

Test niveau B

Test niveau C

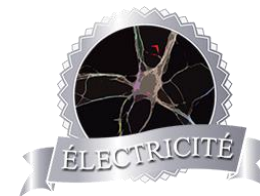
20 % de questions de niveau 1

40 % de questions de niveau 2

40 % de questions de niveau 3

Objectif : Obtenir 60 %

Récompense : Médaille d'argent



Simulateur d'examen

3 paramètres

DUREE

MATIERE

DIFFICULTE

Test niveau A

Test niveau B

Test niveau C

10 % de questions de niveau 1

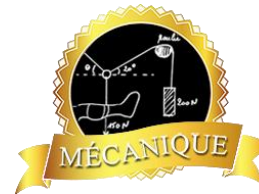
20 % de questions de niveau 2

40 % de questions de niveau 3

30 % de questions de niveau 4

Objectif : Obtenir 60 %

Récompense : Médaille d'or



Simulateur d'examen

A la fin du quadrimestre :

Test formatif (n°2) en présentiel

Conditions d'accès :

Avoir une médaille d'or pour chaque thème (sauf imagerie)

→ **Pass personnel** à télécharger

Déc 2015 : N = 150 (26,83 %)

Déc 2016 : N = 264 (45,36 %)

Feedback :

Général : Juste après le test

Personnel : Quelques jours plus tard

(informations générales et thème par thème)



Références

- BRENNAN, R.-L., *A generalized upper-lower item discrimination index*, Educational and psychological measurement, 32, 289-303, 1972
- EBEL, R.L., *Confidence—Weighting and Test Reliability*, Journal of Educational Measurement, 2, 49-57 B., 195
- FINDLEY, W.G., *A rationale for evaluation of item discrimination statistics*, Educational and Psychological Measurement, 16, 175-180, 1956
- FLEISS, J.-L., *Statistical methods for rates and proportions*, 2nd ed., John Wiley, New York, 1981
- GILLES, J.-L., *Qualité spectrale des tests standardisés universitaires*, Thèse de doctorat, Université de Liège, 2002
- KELLEY, T.L., *Selection of upper and lower groups for the validation of test items*, Journal of Educational Psychology, 30, 17-24, 1939
- LAVEAULT, D., GREGOIRE, J. *Introduction aux théories des tests en sciences humaines*. Paris, Bruxelles : De Boeck Université & Larcier s.a., 1997
- LECLERCQ, D. *La conception des questions à choix multiple*, Bruxelles, Labor, 1986
- LECLERCQ, D. *Qualité des questions et signification des scores avec application aux QCM*, Bruxelles, Labor, 1987
- MARIQUE, P.-X., HOEBEKE, M. *Plate-forme interactive au service des grandes populations d'étudiants suivant un cours de Physique*, Actes de la Conférence TICE 2014, Béziers, France, 2014
- MARIQUE, P.-X., VAN DE POËL, J.-F., HOEBEKE, M. *Quel outil d'entraînement pour des étudiants en médecine évalués par QCM en physique ?*, Actes du Colloque ADMEE 2015, Lisbonne, Portugal, 2015
- ROEDIGER, H. L., BUTLER, A. C. *The critical role of retrieval practice in long-term retention*, Trends in Cognitive Sciences, volume 15, n°1, pp. 20–27, 2011. <http://dx.doi.org/10.1016/j.tics.2010.09.003>
- ROEDIGER, H. L., KARPIPCKE, J. D., *The power of testing memory: Basic research and implications for educational practice*, Perspectives on Psychological Science, volume 1, n°3, pp. 181–210, 2006. <http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x>
- VIAU, R. *La motivation en contexte scolaire*, St-Laurent, Éditions du Renouveau pédagogique, 1994
- https://www.uclouvain.be/cps/ucl/doc/edef/documents/EVA_QCM_version3.pdf, consulté le 24/09/2016
- ADMEE 2017 (Dijon) : Marique et al. ; Application des principes de la classe inversée aux travaux pratiques de physique



Université
de Liège



*Merci pour votre
attention !*

Contact :

Pierre-Xavier Marique

Didactique de la Physique
Université de Liège (Belgique)

pxmarique@ulg.ac.be