

QCM : problématiques et amélioration

Jean-Marc Braibant,
Service d'évaluation en appui à la qualité (EVA)
<http://www.uclouvain.be/eva/html>

Université Paris Sud, le 6 octobre 2015



Les QCM ? Un sujet qui fait débat

Inconvénients	Avantages
On ne peut évaluer que certains objectifs	On peut balayer toute la matière pour un grand nombre d'étudiants
Difficulté pour concevoir un bon QCM	Facilité de la correction
Intervention potentielle du hasard	Objectivité de la notation



C'est quoi un QCM ?

« Une question à laquelle l'étudiant répond en opérant une sélection (au moins) parmi plusieurs solutions proposées, chacune étant jugée (par le constructeur de l'épreuve et par un consensus entre spécialistes) correcte ou incorrecte, indépendamment de l'étudiant qui doit y répondre »

D.Leclercq (1986). *La conception des questions à choix multiples*. Bruxelles : Labor.

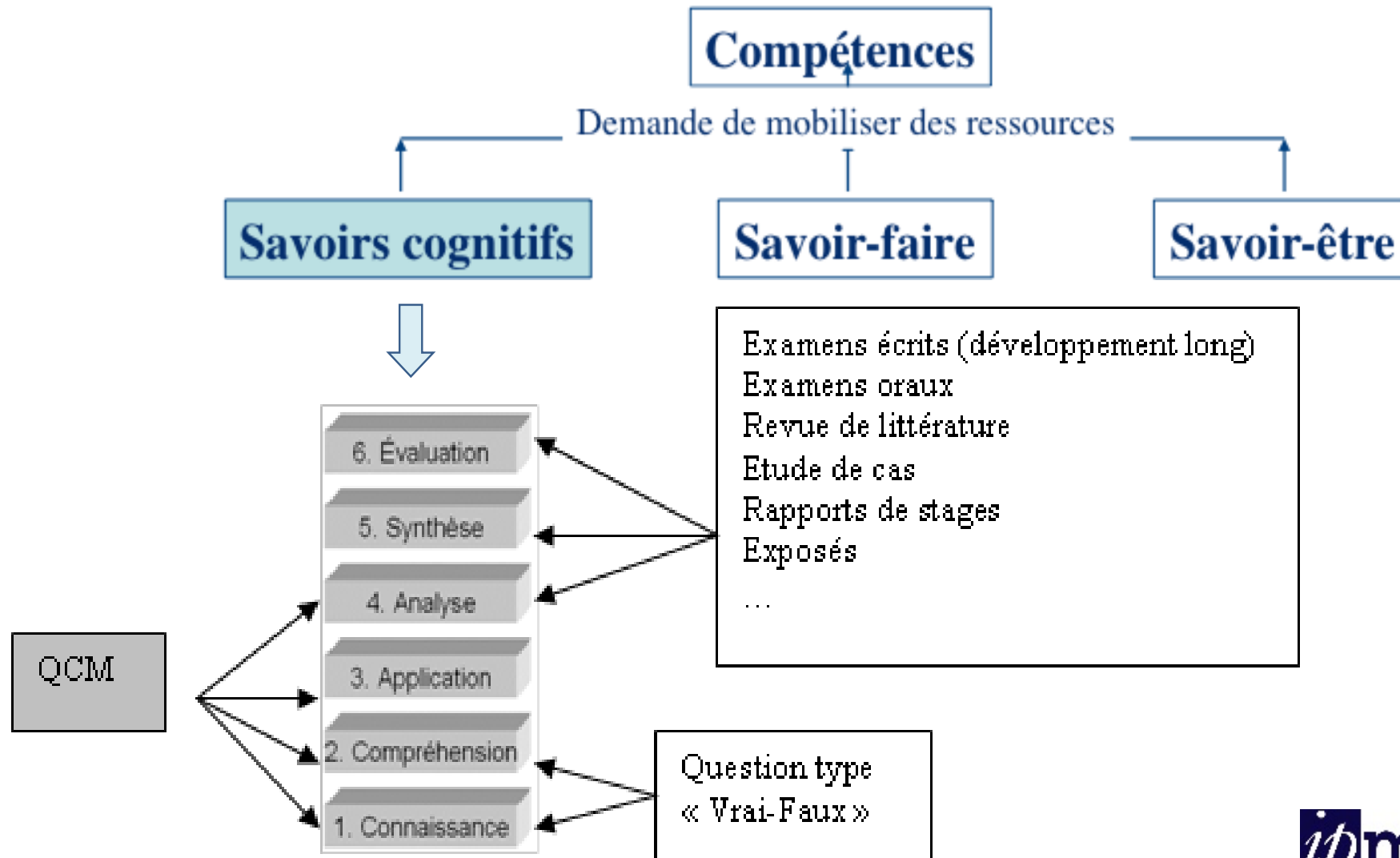


Avec un QCM, on n'évalue que certains objectifs

- ▶ **Qu'évalue-t-on ? Du décidable !**
- ▶ Une certaine image de la connaissance : l'accord indiscutable, peu de nuances
 - ▶ Dans une activité de production non spontanée
 - ▶ Dans un domaine limité du savoir : le plus plausible
 - ▶ Sans avoir accès aux démarches de l'étudiant
- ▶ On ne peut donc pas tout évaluer par QCM !
- ▶ Les QCM sont une modalité d'évaluation parmi d'autres



QCM : pour évaluer quels apprentissages ?



Difficulté pour concevoir un bon QCM

Les énoncés

- ▶ La formulation des questions doit correspondre au niveau des *acquis d'apprentissage* à évaluer : les questions portent sur des *contenus/notions essentielles*.
- ▶ L'énoncé présente *un seul* problème à solutionner
- ▶ *Eviter les termes imprécis* comme parfois, fréquemment, souvent, etc
- ▶ *Eviter de compliquer* le traitement que l'étudiant doit opérer pour comprendre l'énoncé sauf si cette compétence de compréhension veut être évaluée (exemple: concision de l'énoncé)
- ▶ Dans l'énoncé, *séparer* les informations (ex. description d'une situation) de la question
- ▶ *Eviter la forme négative* qui complique l'énoncé et ajoute un problème de logique (privilégier la forme positive sur un mode affirmatif ou interrogatif)
- ▶ L'énoncé ne demande *pas de jugement de valeur* ni une appréciation de la part de l'étudiant
- ▶ *Eviter toute formulation (indices)* qui pourrait permettre à l'étudiant de détecter la solution correcte ou d'éliminer des leurres alors qu'il ne connaît pas la matière.



Difficulté pour concevoir un bon QCM

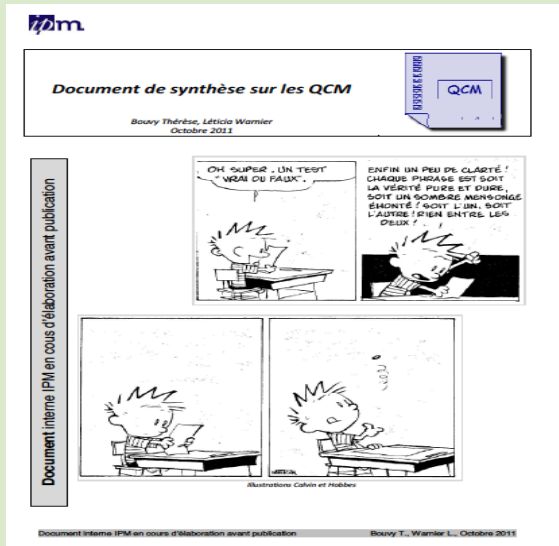
Les solutions proposées

- ▶ La (les) **solution(s)** correcte(s) est(sont) **incontestablement exacte(s)** par rapport à l'énoncé
- ▶ Les **leurres sont crédibles** mais incontestablement faux.
- ▶ Veiller à **l'homogénéité des solutions** dans leur contenu, leur forme et leur structure grammaticale. La bonne réponse n'est pas plus longue, mieux construite que les autres propositions de solutions (ce qui pourrait être un indice de bonne réponse pour l'étudiant)
- ▶ **Indépendance syntaxique** : ne pas lier les solutions les unes aux autres par des termes tels que «au contraire, de plus, ... »
- ▶ **Indépendance sémantique** : pas de propositions imbriquées ou qui s'excluent (réduisant le nombre de leurres réels)
- ▶ **Eviter des termes qui peuvent fournir des indices** comme « toujours, tous, jamais, aucun,... » (proposition probablement fausse) ou « habituellement, parfois,... » (proposition probablement vraie)
- ▶ Eviter de répéter des **termes identiques** à ceux de l'énoncé (indice de bonne réponse pour l'étudiant)
- ▶ Eviter les réponses « **tous ces choix** »
- ▶ Utiliser avec parcimonie « **aucun de ces choix** »



Ressources pour la conception d'un QCM

Vade-Mecum de l'IPM



<http://www.uclouvain.be/cps/ucl/doc/ipm/documents/vademecumQCM-V2.pdf>

Mémo de l'IPM

LES MÉMOS DE L'IPM

MÉMO NUMÉRO 29

QCM : vigilance !

Types de questions, formulation, barèmes de notation...
Comment construire un bon questionnaire à choix multiples ?

1. De quoi est composé une QCM ?
Une question à choix multiples est formée de 2 composantes principales :

- L'énoncé ou l'énoncé qui décrit le problème et/ou pose la question : phrase affirmative, interrogative, graph, situation-problème, ...
- Les solutions proposées, comprenant la (les solution(s) correcte(s) et les fautes, solutions incontestablement fautes.

Exemple :
La polymérisation est la science des
a) microbes
b) molécules
c) mollusques
d) pollutions = solution correcte

La façon de rédiger les énoncés, de rédiger et de proposer les fautes... influence la clarté de la question et la possibilité (ou non) pour l'étudiant de trouver la bonne réponse sans connaître la matière !

2. Comment concevoir un examen par QCM ?
Avant de rédiger sous question, il est important de se poser la question suivante : Quel est ce que je veux évaluer ? Il faut ainsi déterminer :

- Les objectifs d'apprentissage à évaluer, par exemple, ceux qui font appel à la mémoire, à l'interprétation de données (compréhension), à la résolution de problème (application), ...
- Les contenus / notions visés par l'évaluation par QCM.

Avant de concevoir une approche globale de l'examen par QCM, ces objectifs et ces contenus sont retranscrits dans un tableau à double entrée, le tableau de spécification (exemple ci-dessous). L'enseignant pourra ainsi identifier combien de questions mesureront chaque objectif et s'assurer que l'ensemble des questions balaye les notions ciblées à évaluer.

Contenus / notions	Objectifs d'apprentissage faisant appel à			
	La mémorisation (conséquences)	L'interprétation (compréhension)	La solution de problèmes (application)	L'analyse de cas
Niveau 1	1 question			
Niveau 2	1 question	1 question		2 questions
Niveau 3		2 questions		
Niveau 4			1 question	
...				
TOTAL	X questions	Y questions	Z questions	XX questions

Édition 2012

www.uclouvain.be/fr/cps

<http://www.uclouvain.be/99784.html>

Un grand avantage : avec un QCM, on peut poser de nombreuses questions

- ▶ Un examen est censé donner une image aussi exacte que possible de la compétence d'un étudiant sur un tout grâce à sa performance sur une partie
- ▶ Le principal défi de tout examen : concevoir des questions qui constituent un échantillon représentatif de la matière
- ▶ Concrètement, l'amélioration d'un examen passe par un meilleur choix des questions
- ▶ Avec un QCM, on peut balayer toute la matière (objectifs, contenu, difficulté)



Deux questions essentielles qui se posent pour tout examen

- L'examen constitue-t-il bien un **échantillon représentatif** de la matière ?
 - Permet-il à l'enseignant d'**inférer** correctement l'apprentissage réalisé par l'étudiant au regard des objectifs visés ?
- Les notes produites par l'examen sont-elles **fiables** ?



La situation rêvée

Le résultat obtenu par un étudiant à un examen est la mesure précise de son niveau de connaissance de la matière

Le score observé = le score vrai



Cet idéal est illusoire (même pour un QCM)

- ▶ Toute mesure est une approximation
- ▶ L'examen parfait (un test qui mesure sans erreur) n'existe donc pas
- ▶ Le score obtenu par un étudiant à un examen résulte TOUJOURS de la somme entre son score vrai et l'erreur de mesure
 - ▶ **Score observé = score vrai + erreur de mesure**



Quelles sont les sources d'erreur de mesure dans un QCM ?

Tous les facteurs qui influencent les résultats du test et qui n'ont rien à voir avec le niveau de connaissance réel de l'étudiant ...



Exemples

- Des questions inadéquates par rapport au contenu, à la difficulté, aux objectifs de la matière ...
- Des questions qui sont mal posées (ambiguïté du vocabulaire utilisé, leurre mal choisi, indices qui favorisent certains leures, etc.)
- Des erreurs d'encodage des bonnes réponses
- L'effet du hasard
- Lorsqu'il y a des pénalités, la tendance générale d'un étudiant à prendre ou non des risques
- Une limite de temps inappropriée
- Une transcription erronée des réponses par l'étudiant (dans la mauvaise case, auto-correction confuse, ...)
- Des erreurs de notation (mauvaise interprétation du correcteur ou du scanner)
- Des erreurs dans le report des notes (manipulation de fichiers)



Série 1 :(BLEU)

	Nombre de propositions formulées pour cet item (hors case absention éventuelle):	Lettre(s) de la bonne ou des bonnes réponses	Points si bonne(s) réponse(s) cochée(s)	Points si mauvaise(s) réponse(s) cochée(s)	Points si rien de coché parmi les propositions	Si usage de l'abstention : points si la case "abstention" est cochée	Quels Items contigus faut-il grouper ensemble ?	Si pertinent: quels points attribuer à chaque groupe de question ?
Item 1	3	A	1	0	0	0	Sous-total 1 1 à 80	80/100
Item 2	3	A	1	0	0	0		
Item 3	3	A	1	0	0	0		
Item 4	3	B	1	0	0	0		
Item 5	3	A	1	0	0	0		
Item 6	3	B	1	0	0	0		
Item 7	3	C	1	0	0	0		
Item 8	3	B	1	0	0	0		
Item 9	3	A	1	0	0	0		
Item 10	3	B	1	0	0	0		
Item 11	3	A	1	0	0	0		
Item 12	3	B	1	0	0	0		
Item 13	3	A	1	0	0	0		
Item 14	3	C	1	0	0	0		
Item 15	3	A	1	0	0	0		
Item 16	3	B	1	0	0	0		
Item 17	3	A	1	0	0	0		
Item 18	3	B	1	0	0	0		
Item 19	3	A	1	0	0	0		
Item 20	3	A	1	0	0	0		
Item 21	3	B	1	0	0	0		
Item 22	3	C	1	0	0	0		
Item 23	3	B	1	0	0	0		
Item 24	3	B	1	0	0	0		
Item 25	3	B	1	0	0	0		





Numéro matricule (Noma)

0

1

2

3

4

5

6

7

8

9

Série 1 2 3 4

Faculté de philosophie, arts et lettres
LROM1111
24 septembre 2015 - TestNumber 2637

Nom (Majuscules): _____

Prénom : _____

Année d'étude (Majuscules) : _____

	A	B	C
Q1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q7	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q8	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q9	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q10	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q11	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q12	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q13	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q14	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q15	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q16	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q17	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q18	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q19	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q20	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q21	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q22	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q23	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q24	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q25	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	A	B	C
Q26	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q27	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q28	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q29	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q30	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q31	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q32	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q33	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q34	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q35	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q36	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q37	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q38	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q39	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q40	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q41	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q42	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q43	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q44	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q45	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q46	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q47	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q48	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q49	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q50	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	A	B	C
Q51	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q52	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q53	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q54	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q55	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q56	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q57	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q58	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q59	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q60	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q61	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q62	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q63	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q64	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q65	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q66	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q67	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q68	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q69	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q70	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q71	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q72	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q73	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q74	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q75	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	A	B	C
Q76	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q77	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q78	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q79	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q80	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q81	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q82	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q83	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q84	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q85	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q86	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q87	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q88	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q89	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q90	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q91	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q92	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q93	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q94	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q95	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q96	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q97	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q98	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q99	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q100	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



Que faire alors ?

Minimiser l'erreur de mesure

Permettre à l'étudiant de démontrer son niveau réel de connaissance le jour J

- ▶ Des questions bien conçues qui constituent un échantillon représentatif de toute la matière étudiée
- ▶ Un nombre suffisant d'items
 - ▶ Plus il y a de questions entrant dans le calcul du score observé, plus celui-ci risque d'être précis
- ▶ Un questionnaire aussi homogène que possible
 - ▶ Eviter l'étudiant à changer de technique de réponse
 - ▶ Placer dans un même bloc les questions de même type
- ▶ Des directives claires et précises
 - ▶ Modalités de réponse : nombre de bonnes réponses attendues par question
 - ▶ Indication du barème de notation : pénalité ou non pour les effets du hasard ?
 - ▶ Possibilité ou non de corriger ses réponses
- ▶ Un nombre réduit de manipulations de fichiers



On a les moyens de contrôler tout cela !

**Le QCM est la
seule forme d'examen qui permet
une analyse a posteriori des résultats
sur base d'indices statistiques :**

les analyses d'items



Les analyses d'items

Ensemble de procédés statistiques dont le but est d'évaluer la qualité d'un instrument de mesure et des items qui le composent.



Analyses d'items : questions auxquelles on peut apporter une réponse

- ▶ Mon QCM est-il bien conçu ?
- ▶ Dois-je augmenter le nombre d'items ?
- ▶ Puis-je me permettre de les réduire ?
- ▶ L'examen est-il trop simple ou trop difficile ?
- ▶ Les scores des étudiants sont-ils fiables ?
- ▶ Les leurres sont-ils efficaces ?
- ▶ Quels items pourrais-je conserver pour un autre examen ?
- ▶ Quels sont les items à éliminer ?



Analyses d'items : questions auxquelles on peut apporter une réponse

- ▶ Mon QCM est-il bien conçu ?
- ▶ Dois-je augmenter le nombre d'items ?
- ▶ Puis-je me permettre de les réduire ?
- ▶ L'examen est-il trop simple ou trop difficile ?
- ▶ Les scores des étudiants sont-ils fiables ?
- ▶ Les leurres sont-ils efficaces ?
- ▶ Quels items pourrais-je conserver pour un autre examen ?
- ▶ Quels sont les items à éliminer ?



Exemple d'un système automatisé par lecture optique (*Contest, Consilium, 2011*)

- ▶ A l'UCL, en EVA, 3 rapports sont remis aux enseignants
 - ▶ Test analysis report
 - ▶ Item advice report
 - ▶ Item analysis report



Le test analysis report : informations générales

Entire Test		
General details		
Questions	20	
Excluded questions	0	
Participants	212	
Participants passed	76	35,85 %
Participants failed	136	64,15 %
Minimum score	-7,50	
Chance score	3,13	
Maximum score	20,00	
Minimum grade	0,00	
Maximum grade	20,00	
Pass grade	10,00	



Le test analysis report : partie statistique

Statistical details				
	Average	Stddev	Lowest	Highest
Score	22,05	4,72	10,00	32,00
P	0,70	0,21	0,13	0,93
Corrected p	0,40	0,43	-0,75	0,97
Item-rest correlation	0,14	0,08	-0,05	0,30
Grade	11,03	2,36	5,00	16,00

Cronbach's Alpha 0,50

Standard error of measurement 1,40

Estimated Alpha for similar test with norm length 40 items 0,50

Cronbach's Alpha (norm) 0,10 0,20 0,30 0,40 0,50 0,60 0,70 0,80 0,90 0,99

Required replications (K) 0,11 0,25 0,44 0,68 1,02 1,53 2,38 4,08 9,17 100,90



Deux indicateurs importants (évaluation certificative)

Est-ce que mon examen fait bien la distinction entre les bons et les moins bons étudiants ?

« Je souhaiterais que l'ensemble des questions distinguent les étudiants entre eux en ce qui concerne la dimension que j'évalue »

- ▶ **Coefficient de discrimination**
 - ▶ Item Rest Correlation (Rir)
 - ▶ Valeur > **0,20**

Est-ce que je peux faire confiance aux scores des étudiants ?

« J'espère que mes étudiants obtiendraient le même score s'ils étaient amenés à repasser le même examen »

- ▶ **Coefficient de fidélité**
 - ▶ Alpha de Cronbach
 - ▶ Valeur > **0,70**



1^{er} diagnostic rapide sur cette base

Statistical details				
	Average	Stdev	Lowest	Highest
Score	8,58	3,33	0,00	16,00
P	0,57	0,18	0,24	0,88
Corrected p	0,33	0,21	-0,02	0,75
Item-rest correlation	0,15	0,14	-0,15	0,34
Grade	8,58	3,33	0,00	16,25

Cronbach's Alpha 0,45

Standard error of measurement 1,09

Estimated Alpha for similar test with norm length 40 items 0,62

Cronbach's Alpha (norm) 0,10 0,20 0,30 0,40 0,50 0,60 0,70 0,80 0,90 0,99

Required replications (K) 0,13 0,30 0,52 0,80 1,21 1,81 2,81 4,82 10,85 119,32



Exploitation par le service EVA de ces rapports statistiques



Constitution d'une banque de données

- ▶ 331 examens
- ▶ 4 sessions
 - ▶ Janvier 2012
 - ▶ Juin 2012
 - ▶ Janvier 2013
 - ▶ Juin 2013
- ▶ 69 283 étudiants

**Métanalyses en vue d'identifier
les « bonnes pratiques »**



1^{er} constat : grande variété des pratiques

- ▶ **Nombre d'items par examen**
 - ▶ De 5 à 100
- ▶ **Nombre d'options de réponses par question**
 - ▶ De 2 (V/F) à 8
- ▶ **Type de questions**
 - ▶ Une seule réponse correcte attendue par question
 - ▶ Plusieurs réponses attendues pour une question
- ▶ **Barème de notation**
 - ▶ De 0,2 à 10 points pour une bonne réponse
 - ▶ De -0,10 à -2 points pour une mauvaise réponse
- ▶ **Adoption ou non d'un barème de pénalisation**
 - ▶ Points négatifs pour les mauvaises réponses



Quel est l'impact des ces variables sur les qualités psychométriques d'un examen ?

3 hypothèses

1. Relation entre nombre de questions et alpha de Cronbach

- ▶ *“Plus il y a de questions, plus grande sera la probabilité d'avoir une bonne fidélité”* (Laveault & Grégoire, 1997)

2. Relation entre nombre d'options de réponse et Rir

- ▶ Les Questions V/F discrimineront moins bien les étudiants entre eux que celles qui présentent au moins 3 options de réponses

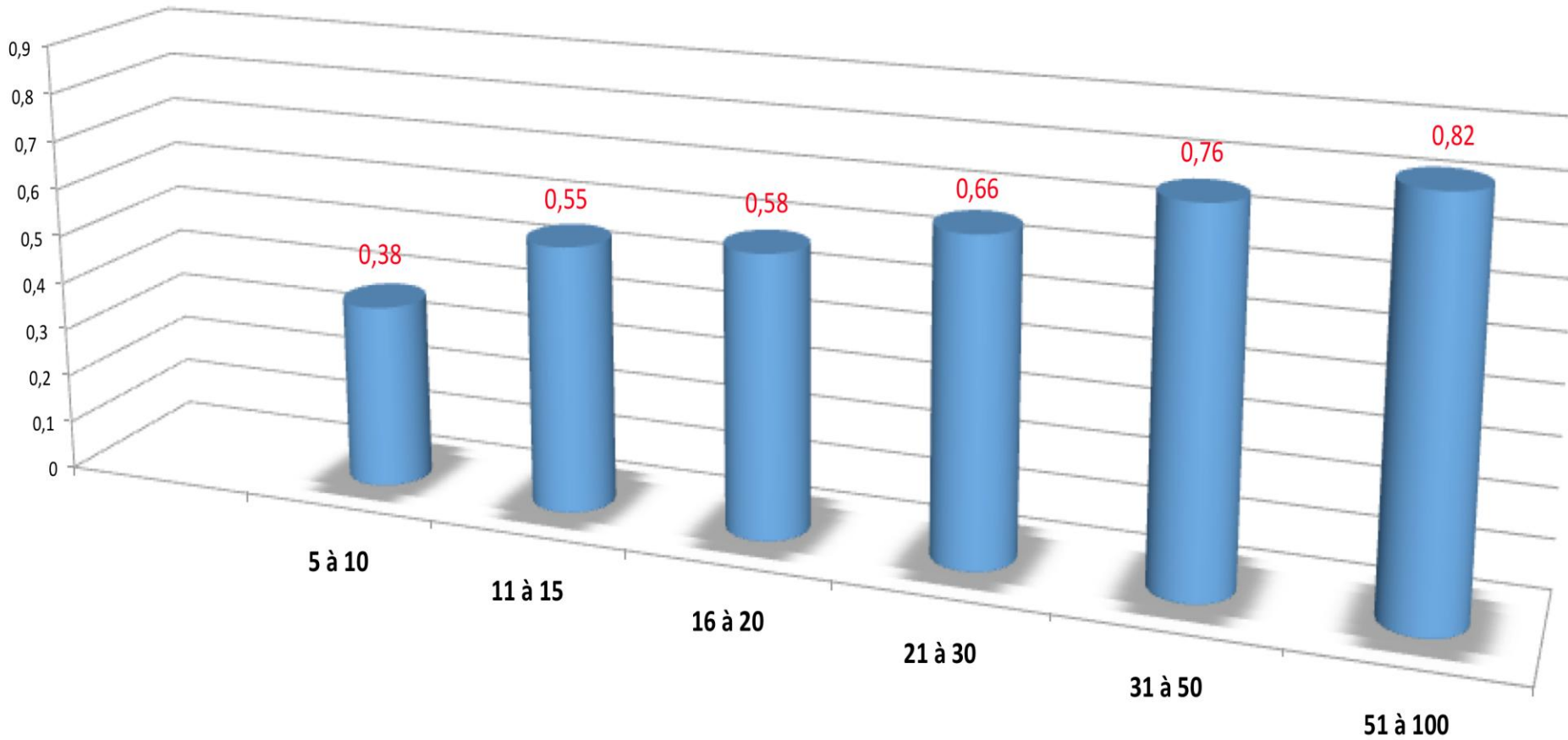
3. Relation entre pénalisation des mauvaises réponses et Rir

- ▶ Les examens avec un barème de pénalisation seront plus discriminants que ceux qui ne pénalisent pas les mauvaises réponses



Hypothèse 1

Valeur de l'Alpha cronbach en fonction du nombre de questions



Hypothèse 1

Relation entre le nombre de questions et l'alpha de Cronbach

Nombre questions par examen	N	Alpha Cronbach	Std deviation	Etendue Alpha
5 à 10	20	,38	(,16)	,05 à ,65
11 à 15	25	,55	(,17)	,01 à ,75
16 à 20	128	,58	(,12)	,28 à ,84
21 à 30	50	,66	(,12)	,32 à ,86
31 à 50	62	,76	(,08)	,46 à ,90
51 à 100	46	,82	(,10)	,52 à ,98



Hypothèse 1

Relation nombre de questions et alpha de Cronbach

Nombre questions par examen	N	Alpha Cronbach	Std deviation	Etendue Alpha
5 à 10	20	,38	(,16)	,05 à ,65
11 à 15	25	,55	(,17)	,01 à ,75
16 à 20	128	,58	(,12)	,28 à ,84
21 à 30	50	,66	(,12)	,32 à ,86
31 à 50	62	,76	(,08)	,46 à ,90
51 à 100	46	,82	(,10)	,52 à ,98



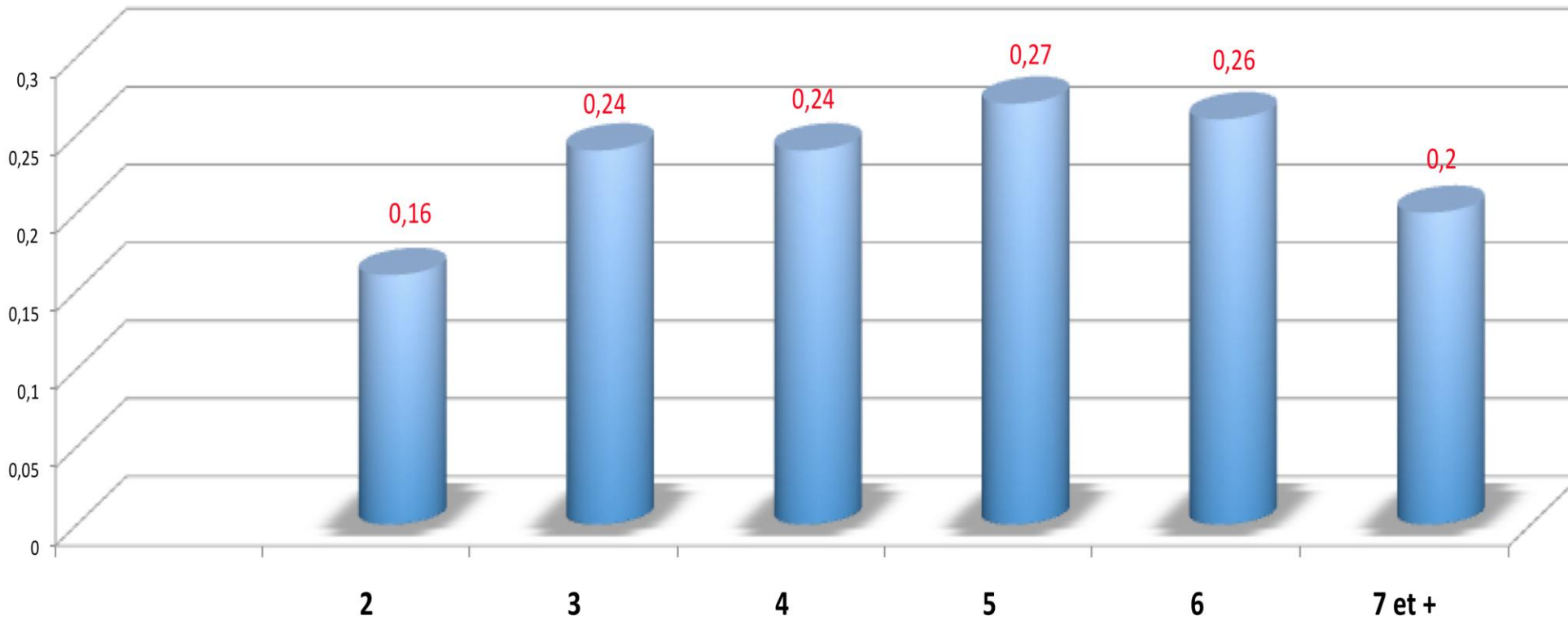
Conclusion Hypothèse 1

- ▶ Plus le nombre d'items augmente, plus grande sera la probabilité d'avoir un alpha de Cronbach satisfaisant
- ▶ Importance de poser plus de 30 questions



Hypothèse 2

Valeur de Rir en fonction du nombre d'options de réponses



Hypothèse 2

Relation entre le nombre d'options de réponses et le « rir » (n = 255)

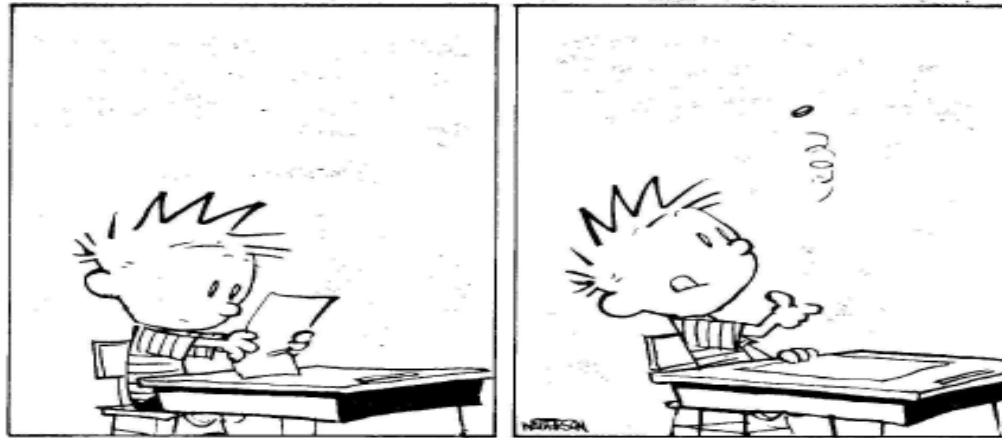
Nbre options	N	Rir	Std. deviation	Etendue Rir
2 (V/F)	45	,16	(,089)	0,01 à 0,34
3	24	,24	(,070)	0,12 à 0,44
4	109	,24	(,075)	0,07 à 0,42
5	56	,27	(,062)	0,14 à 0,42
6	11	,26	(,072)	0,14 à 0,37
7 et +	10	,20	(,045)	0,13 à 0,27

255 examens homogènes : même type de question, on ne change pas le nombre d'options par question



Conclusion Hypothèse 2

- ▶ Les V/F sont à déconseiller car ils sont associés à un indice de discrimination inférieur au seuil minimal attendu (0,20)



- ▶ On devrait proposer **au minimum 3 options de réponse** pour espérer atteindre un seuil minimal de 0,20

Hypothèse 3

Qu'en est-il de l'utilisation de barèmes de pénalisation ?

	N	RIR	Alpha Cronbach
Pénalités	247	,23	0,64
Pas de pénalités	84	,23	0,65

A l'UCL, 75 % des examens font appel à un barème de pénalisation !

Mais pénaliser les mauvaises réponses n'est pas associé à de meilleurs indicateurs psychométriques !



Conclusion hypothèse 3

Rien ne justifie le succès actuel des barèmes de pénalisation

- ▶ Ils poussent les étudiants très faibles à la prise de risque : ils n'ont rien à perdre !
- ▶ Ils incitent une partie des étudiants forts à la prudence voire à la paralysie
- ▶ Un QCM avec pénalisation peut parfois mesurer davantage les aptitudes stratégiques des étudiants à répondre que leurs connaissances
- ▶ « Les barèmes avec pénalisation pour devinette devraient être abandonnés » (Fontaine, 1982)



Quelle est la solution ?

- ▶ Pour diminuer la possibilité de réussir par hasard, offrez un « espace de réponses » suffisamment étendu
 - ▶ Jouer sur deux paramètres
 - ▶ Le nombre de questions
 - ▶ Le nombre d'options de réponses
 - ▶ Augmentez « **the size of the test** » (Grier, 1975)
 - ▶ **Size** = n (nombre de questions) x a (alternatives)



Quid des examens dans lesquels on change les paramètres en cours d'examen ?

	N	Rir	Alpha Cronbach
Examens homogènes (sauf V/F)	197	,25	,67
Autres examens	89	,24	,66



Une différence de taille : le taux de réussite des Etudiants !

	Rir	Cronbach	% réussite
Examens homogènes	,25	,67	59,8
Autres examens	,24	,66	51,4



Changer les paramètres en cours d'examen...

- ▶ N'est pas associée à une meilleure discrimination des étudiants
- ▶ N'est pas associée à une meilleure fiabilité de l'examen
- ▶ Est associée à un taux d'échecs plus élevé
 - ▶ Augmentation des absences
 - ▶ Augmentation des réponses « out of range »



En résumé

5 recommandations pour donner l'occasion à l'étudiant de démontrer son niveau réel de connaissance le jour J

- ▶ Un seul type de question
- ▶ Un nombre suffisant d'items : 40 à 60
- ▶ Un nombre suffisant d'options de réponses par question : de 3 à 5
- ▶ Pas de pénalisation des mauvaises réponses
- ▶ Un barème de notation identique pour toutes les questions



Sinon ...

- ▶ Toute autre situation est acceptable mais elle réclame
 - ▶ Une vigilance particulière
 - ▶ Ex. regrouper les types de question par bloc
 - ▶ Une préparation appropriée des étudiants
 - ▶ Information et formation des étudiants à la pratique des QCM
 - ▶ Examen à blanc

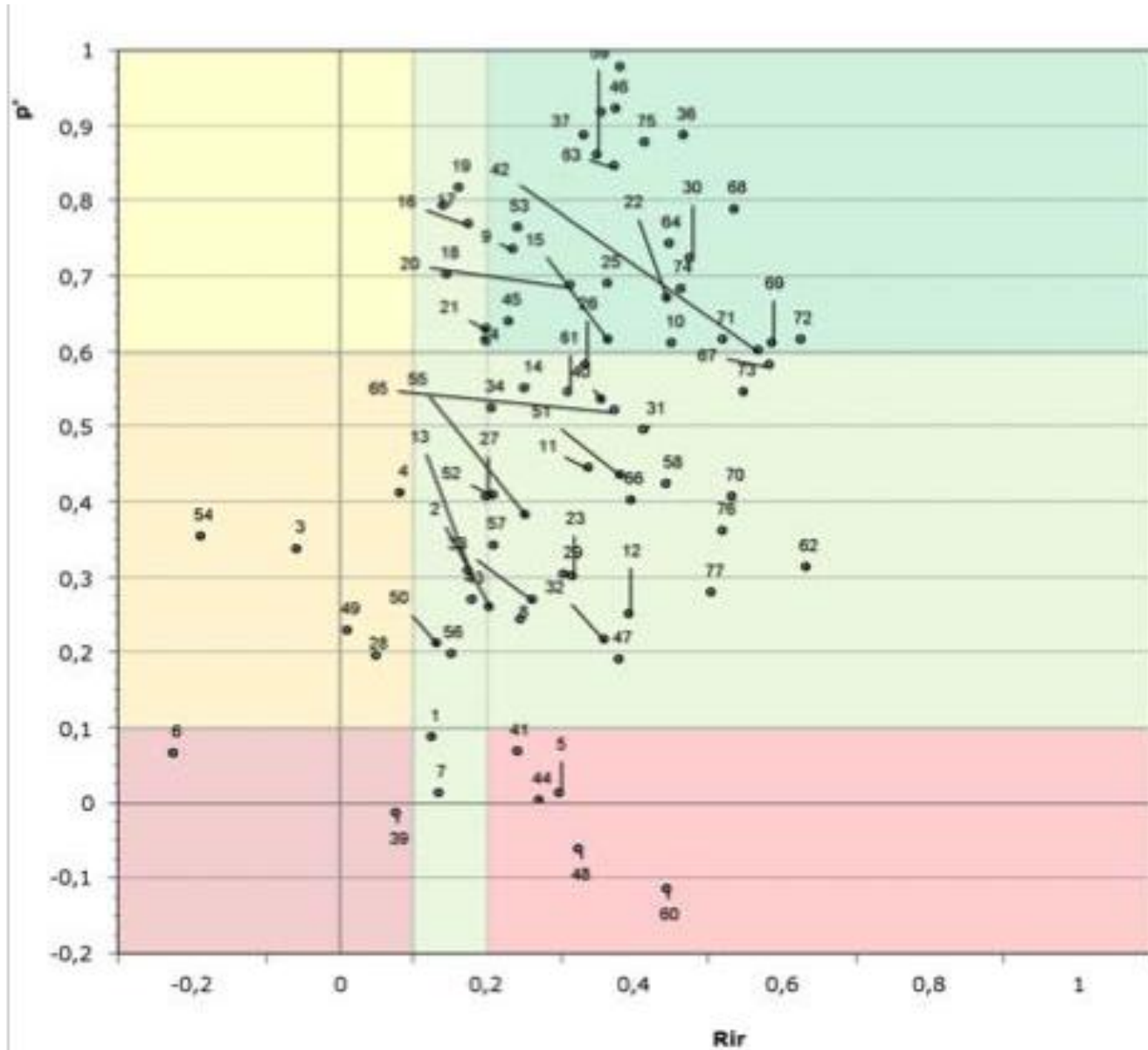


Deuxième type d'actions menées par EVA

**Diffusion des rapports d'analyses
statistiques auprès de chaque enseignant
avec les notes des étudiants**



Item advice report



Item analysis report

	Open				A		B		C		D		E		
	p	p'	Rir	Rit	f	z	f	z	f	z	f	z	f	z	
1	0,14	-0,07	0,06	0,12	14	23	-0,25	10	-1,13	281	0,07	9	-0,83	52	0,08
2	0,38	0,22	0,47	0,53	37	124	-0,12	23	-0,87	14	-0,61	51	-0,63	140	0,54
3	0,51	0,38	0,42	0,49	106	176	0,29	22	-0,61	24	-0,54	15	-0,88	46	-0,24
4	0,23	0,04	0,13	0,20	58	30	-0,64	79	0,21	22	-0,92	131	0,21	69	-0,07
5	0,21	0,01	0,37	0,42	235	42	0,08	14	-0,70	40	-0,48	24	-0,28	34	0,94
6	0,43	0,29	0,45	0,52	107	11	-0,46	60	-0,30	146	0,39	35	-0,63	30	-0,37
7	0,59	0,49	0,42	0,48	42	19	-1,05	223	0,31	52	-0,53	39	-0,62	14	0,25
8	0,72	0,65	0,30	0,36	25	4	-1,06	276	0,17	8	-0,43	38	-0,38	38	-0,67
9	0,53	0,42	0,33	0,41	62	20	-0,81	63	-0,19	40	-0,50	9	-0,67	195	0,28
10	0,46	0,33	0,25	0,33	120	35	-0,54	16	-0,57	33	-0,67	156	0,32	29	-0,02
11	0,23	0,04	-0,34	-0,27	271	18	-0,21	28	0,08	35	-0,33	31	0,25	6	0,90
12	0,55	0,43	0,44	0,50	114	46	-0,79	24	-0,86	8	-0,85	190	0,36	7	-0,69
13	0,31	0,13	0,45	0,51	127	94	0,69	70	-0,35	48	-0,48	30	-0,54	20	-0,09
14	0,53	0,41	0,52	0,58	30	199	0,46	8	-0,51	52	-0,52	45	-0,78	55	-0,47
15	0,65	0,56	0,13	0,20	106	18	-1,02	230	0,18	11	-0,99	11	-0,76	13	-0,25
16	0,31	0,13	0,35	0,42	211	29	-0,54	20	-0,34	27	-0,48	25	-0,29	77	0,55
17	0,53	0,41	0,50	0,56	80	21	-0,57	190	0,37	52	-0,60	10	-0,60	36	-0,59
18	0,27	0,09	0,23	0,30	136	79	0,32	34	-0,81	32	-0,44	35	-0,54	73	0,48
19	0,39	0,24	0,30	0,38	120	13	-0,51	49	-0,39	129	0,40	59	-0,31	19	-0,42
20	0,64	0,55	0,39	0,46	41	241	0,21	56	-0,56	18	-0,81	10	-0,53	23	-0,01
21	0,37	0,21	0,24	0,31	31	73	0,01	14	-1,02	41	-0,49	92	-0,02	138	0,26
22	0,41	0,27	0,42	0,49	164	20	-0,64	128	0,45	36	-0,79	12	-0,54	29	-0,34
23	0,37	0,22	0,34	0,41	149	25	-0,33	32	-0,46	41	-0,33	27	-0,46	115	0,43
24	0,52	0,40	0,23	0,31	41	194	0,24	63	-0,36	78	-0,18	6	-0,78	7	-0,70
25	0,54	0,43	0,38	0,45	107	56	-0,29	28	-0,64	190	0,22	5	-0,96	3	-0,90
26	0,85	0,82	0,42	0,47	7	6	-0,92	15	-1,11	28	-1,02	331	0,16	2	-0,42
27	0,71	0,64	0,01	0,08	19	16	-0,30	3	-0,46	12	-0,90	272	0,03	67	0,14
28	0,42	0,27	0,30	0,38	118	19	-0,67	82	-0,09	138	0,32	19	-0,78	13	-0,69
29	0,27	0,09	0,19	0,26	110	84	0,53	31	-0,61	13	-0,37	144	-0,14	7	-0,24
30	0,22	0,03	0,10	0,17	138	28	-0,33	58	0,28	75	-0,16	49	-0,13	41	0,28



Diffusion des rapports statistiques

► Enquête en novembre 2013 (n = 86)

- Consultation des rapports OUI : 88 %
- Données trop complexes : OUI : 63 %
- Demande d'accompagnement individuel OUI : 22 %
- Existence de cette aide OUI : 32 %
- Intention de participer à une formation OUI : 50 %
- Pas assez de temps OUI : 27 %



Actions menées suite à l'enquête

- ▶ 6 sessions de formation (n = 25)
- ▶ Rédaction et diffusion de 2 notes explicatives
 - ▶ « *Comment lire et interpréter les rapports d'analyses d'items en vue d'améliorer la qualité de vos examens ?* »
 - ▶ « *Glossaire technique des termes utilisés dans Contest* »
- ▶ Rédaction d'une note d'aide au diagnostic
 - ▶ « *Aide au diagnostic : quel regard jeter sur le rapport de statistiques ? 8 signaux qui peuvent vous aider* »

Actions menées depuis la session de juin 2014



Rapport diagnostic envoyé à chaque professeur après la remise des résultats de l'examen

▶ **4 signaux d'alerte**

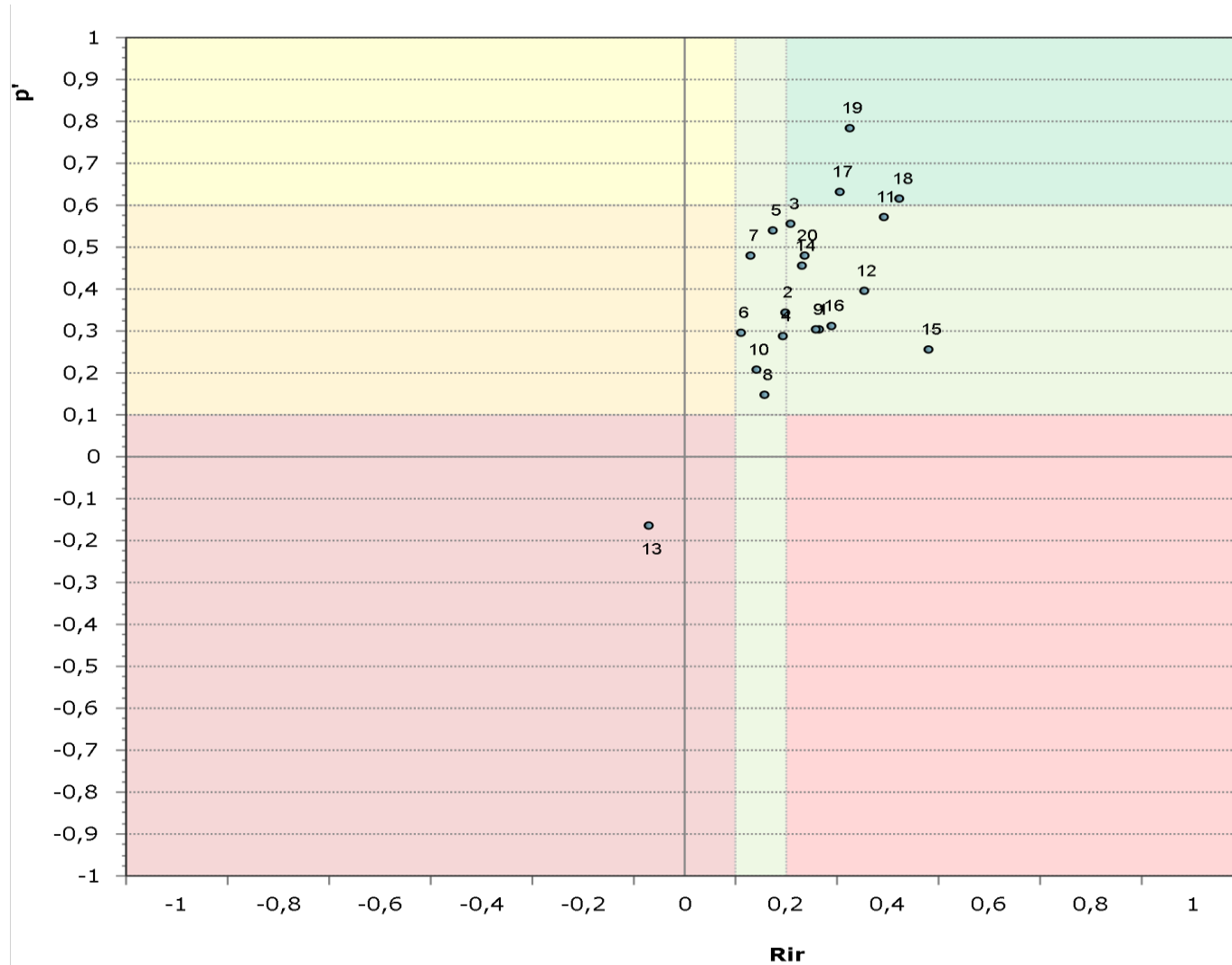
- ▶ Indice de discrimination $< 0,20$
- ▶ Alpha de Cronbach $< 0,70$
- ▶ Items ayant un indice de discrimination négatif
- ▶ Items ayant un nombre élevé d'abstentions

▶ **Si un seul de ces signaux est présent**

- ▶ Mail individuel (Feed-back personnalisé) envoyé dans les heures qui suivent la remise des résultats



Exemple d'un item à discrimination négative



Bonjour,

Suite aux résultats reçus pour l'examen XXXX, j'attire votre attention sur les points suivants :

- **L'item-rest correlation est de 0,15 et est donc inférieur à 0,20 (norme minimale attendue)**
 - Cela signifie que les questions ne permettent pas de distinguer de manière suffisamment précise les étudiants "forts" (ceux qui connaissent la matière) des étudiants plus faibles (ceux qui ne la maîtrisent pas).
- **L'alpha de Cronbach est de 0,56 et est donc inférieur à 0,70 (norme que l'on peut espérer atteindre)**
 - Cela signifie que les questions posées dans cet examen ne permettent pas de mesurer de manière fiable le niveau de connaissance de la matière vos étudiants. Supposons que l'on présente deux formes équivalentes de cet examen, on aurait beaucoup de chances d'obtenir des résultats forts différents pour un même étudiant d'une forme à l'autre. En d'autres termes, un étudiant ayant obtenu un 14 pour la forme A pourrait se retrouver avec un 8 à la forme B et vice-versa.
- **Les deux indicateurs qui précèdent sont affectés par les résultats psychométriques obtenus pour 5 questions problématiques**
 - Il s'agit des **questions 3, 4, 8, 13 et 27**. Ces questions posent problème car, paradoxalement, les étudiants qui réussissent bien votre examen ont tendance à être en échec pour ces questions et, à l'inverse, elles sont bien réussies par des étudiants qui sont plutôt "faibles" aux autres questions de l'examen.
- On observe par ailleurs qu'un **nombre conséquent d'étudiants s'abstiennent de répondre à toute une série de questions (Colonne OPEN dans l'Item analysis report)**. Un bon examen est un examen qui donne la possibilité de démontrer leur niveau réel de connaissance le jour J. Le fait que certaines questions rassemblent un nombre élevé de non réponses pose problème. Compte tenu de la pénalisation en cas de mauvaises réponse, on peut faire l'hypothèse que les étudiants ne prennent pas le risque de répondre surtout s'il s'agit de questions difficiles ou en fin de test quand ils pensent avoir accumulé suffisamment de points pour obtenir la moyenne. Dans les deux cas de figure, c'est problématique et cela soulève la question de maintenir, à l'avenir, un barème de pénalisation pour cet examen.

Si cela vous semble utile, je vous propose de me recontacter à votre meilleure convenance de manière à ce qu'on puisse regarder ensemble comment faire évoluer la qualité de ce QCM.

Bien à vous,

*Jean-Marc Braibant
Conseiller en évaluation
Administration de l'enseignement et de la formation
Service d'évaluation en appui à la qualité (EVA)*

!



Possibilité de faire des modifications avant la remise officielle des notes

- ▶ Rectifications d'erreurs dans la déclaration
 - ▶ Réponse attendue pas correctement encodée dans Contest
 - ▶ Idem pour les pénalités
- ▶ Acceptation de deux bonnes réponses au lieu d'une seule comme initialement prévu
- ▶ Retrait d'une ou plusieurs questions
 - ▶ Item trop difficile par rapport aux autres
 - ▶ Item à discrimination négative
- ▶ Allègement de la pénalité
 - ▶ - 0,5 pt au lieu de 1 pt, -0,25 au lieu de -0,5 pt, -0,20 au lieu de - 0,25



Remarques importantes

- ▶ Propositions de modification faites par EVA en faveur de l'étudiant
- ▶ Décisions qui sont toujours prises in fine par l'enseignant



Quelles sont les réactions des professeurs ?

Session	Examens EVA (N)	Rapports alertes (A)	Réactions au rapport	Modifications examens
Juin 2014	134	61 (45,5 % de N)	33 (54,1 % de A)	21 (15,7 % de N)
Janvier 2015	74	45 (60,8 % de N)	31 (68,9 % de A)	18 (24,3 % de N)
Juin 2015	111	65 (58,56 % de N)	33 (50,8 % de A)	20 (18,02 % de N)
Total	319	171	97	59



Etude d'impact réalisée pour la session de janvier 2015

17 examens (n = 4586 étudiants)	Avant	Après	Différence
Rir	0,229	0,243	+ 0,014
Cronbach	0,662	0,681	+ 0,019
Nombre de réussites	2185	2514	+ 329
% de réussite	47,65 %	54,82 %	+ 7,17 %

Contexte et enjeu :
passage de la réussite de 12 à 10 !



Etude d'impact réalisée pour la session de juin 2015

11 examens (n = 4586 étudiants)	Avant	Après	Différence
Rir	0,23	0,24	+ 0,01
Cronbach	0,76	0,77	+ 0,01
Nombre de réussites	1544	1718	+ 174
% de réussite	44,17 %	51,74 %	+ 7,57 %

Contexte et enjeu :
passage de la réussite de 12 à 10 !



En quelques mots, quelques avantages de l'analyse d'items

- ▶ Tirer profit des résultats d'un examen
- ▶ Obtenir des indications sur la précision des résultats observés
- ▶ « Valider » un examen en retirant, au besoin, les questions qui présentent des lacunes
- ▶ Améliorer la formulation des questions
- ▶ Calibrer les examens futurs
- ▶ Créer une « banque d'items »

Conclusions

- ▶ **Accueil très favorable des professeurs : cela marche !**
- ▶ **Efficacité de cette démarche proactive de diagnostic**
 - ▶ Politique concrète d'aide à la réussite
 - ▶ « Cheval de Troie » pour parler d'évaluation
- ▶ **Confirmation de 3 problématiques importantes** (*Braibant & coll, Admee 2014*)
 - ▶ Le nombre d'items par examen
 - ▶ Les pénalisations dans les QCM
 - ▶ Le pouvoir peu discriminant des QCM V/F
- ▶ **Levier important pour l'amélioration de la qualité de tous les QCM à l'UCL**



<<)



Illustration :
Calvin et Hobbes

