

Le Test de Concordance des Scripts

JP. Fournier

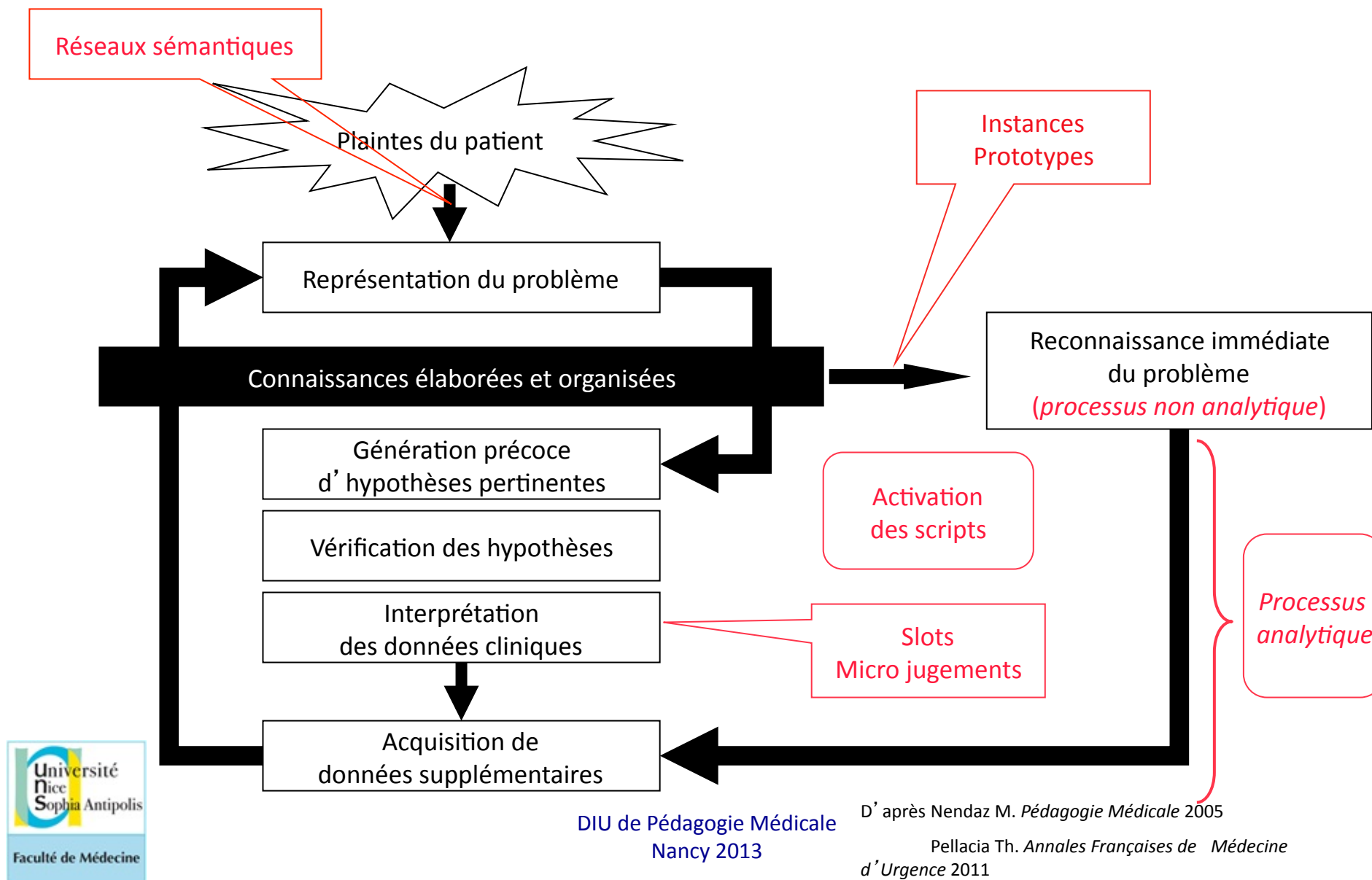
Département de Pédagogie Médicale

Faculté de Médecine de Nice Sophia Antipolis

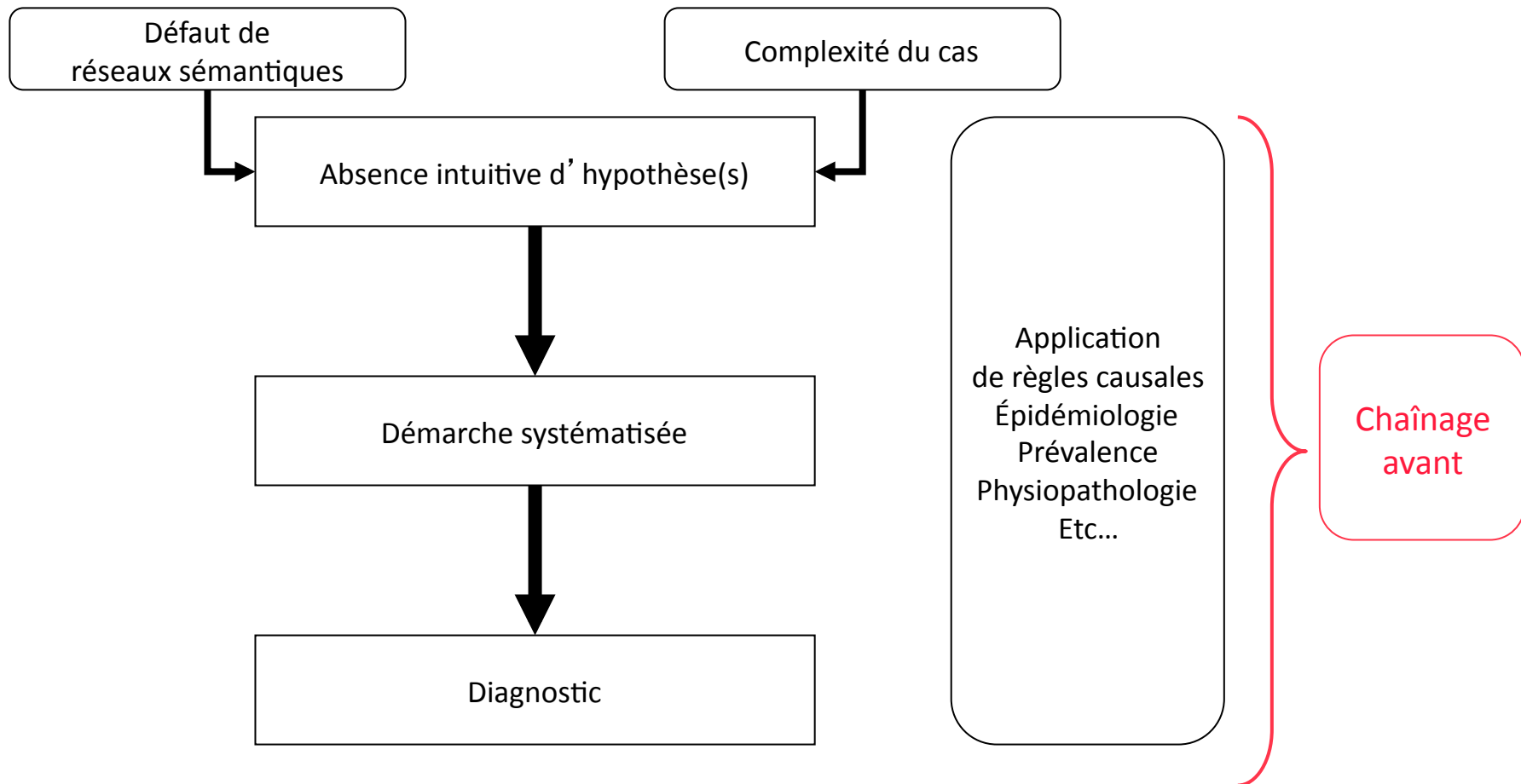


DIU de Pédagogie Médicale
Nancy 2013

Le raisonnement médical



Le raisonnement médical



Le raisonnement médical

Processus non analytique	Description
Reconnaissance de formes (<i>Pattern recognition</i>)	Identification intuitive d' une configuration caractéristique de signes (contextuels, cliniques,...) évoquant très fortement un ou des diagnostics
Reconnaissance d' exemples concrets (<i>Instances</i>)	Identification intuitive d' une situation déjà vécue, lui permettant d' évoquer un ou des diagnostics

Le raisonnement médical

Processus analytique	Description
Raisonnement hypothético déductif	Le praticien recherche consciemment, à travers interrogatoire, examen physique, examens complémentaires, à confirmer / infirmer les hypothèses envisagées
Raisonnement en chaînage avant (<i>forward reasoning</i>)	Le praticien chemine consciemment des données cliniques / paracliniques vers la solution, à travers l'application de règles causales ou conditionnelles

Quels outils pour mesurer le raisonnement clinique ?

Perspect Med Educ (2012) 1:162–171
DOI 10.1007/s40037-012-0024-1

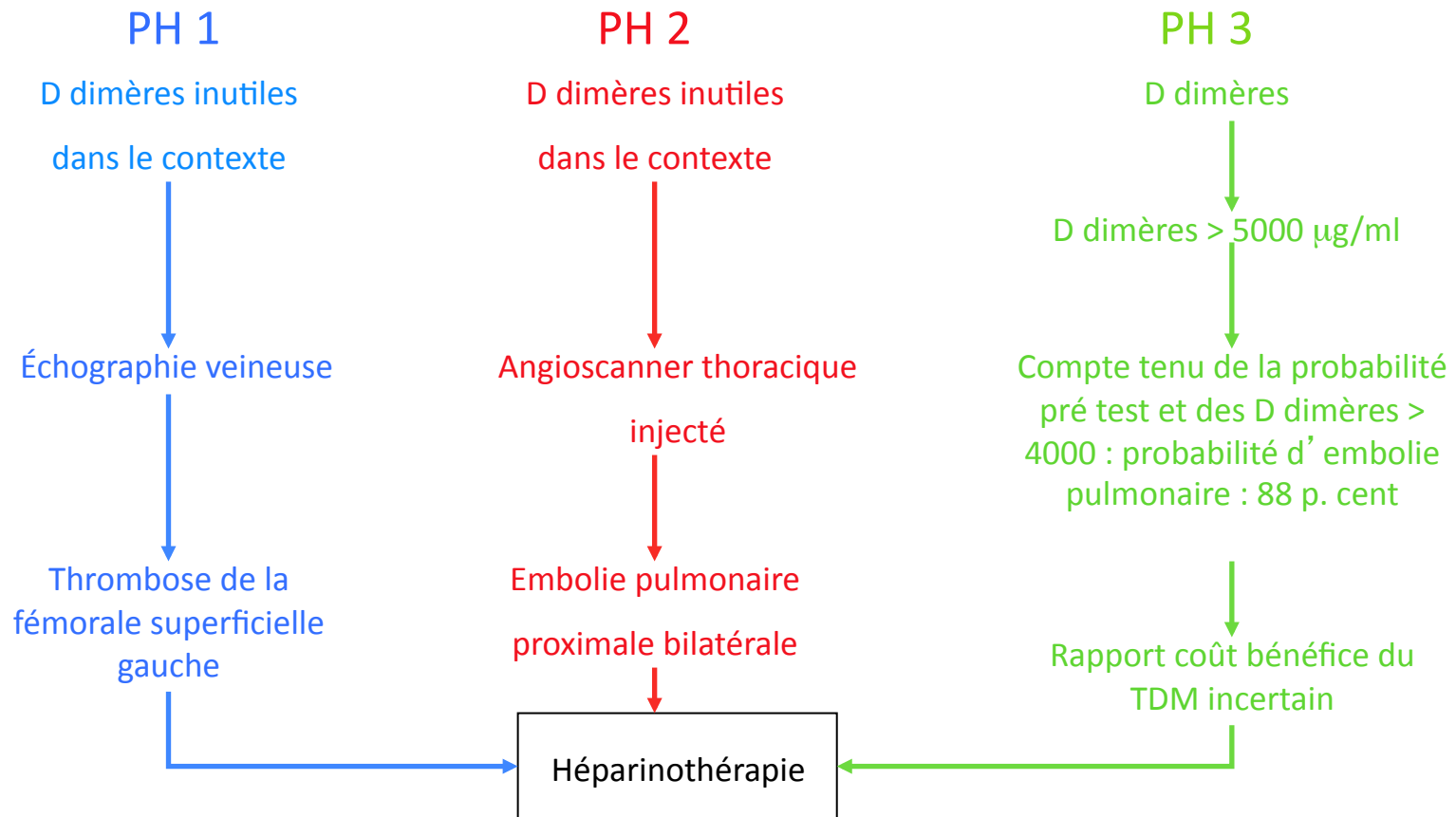
ORIGINAL ARTICLE

Preferred question types for computer-based assessment of clinical reasoning: a literature study

Lisette van Bruggen ·
Margreet Manrique-van Woudenberg ·
Emely Spierenburg · Jacqueline Vos

- ❖ Script Concordance Test ;
- ❖ Extended Matched Questions ;
- ❖ Comprehensive Integrative Puzzle ;
- ❖ Modified Essay questions / Short Answer Questions ;
- ❖ Long Menu Questions ;
- ❖ Multiple Choice Questions ;
- ❖ Multiple True False Questions ;
- ❖ Virtual Patients.

Une femme de 82 ans est adressée aux urgences par sa maison de retraite pour une douleur basi thoracique droite accompagnée de dyspnée. Elle y était hospitalisée pour prise en charge palliative d'un cancer de l'ovaire poly métastasé qui la confine au lit. Elle présente également une insuffisance cardiaque évoluée sur cardiopathie ischémique. La température est à 38,2°C. La TA est à 125 - 80 mm H, la fréquence cardiaque à 130 /min et la fréquence ventilatoire est à 30 / min. La saturation transcutanée est à 90 p. cent sous 6 l/min d'oxygène au masque facial. Il existe un épanchement pleural de faible abondance à la base droite. Elle présente un très mauvais état veineux avec des œdèmes chroniques. Vous suspectez une embolie pulmonaire



Qu' a-t-on observé ?

❖ Utilisation de 3 approches diagnostiques pertinentes dans ce contexte : recherche de phlébite par échographie, recherche directe d' embolie pulmonaire par TDM, utilisation d' une approche « non agressive » ;

Activation
des
scripts

❖ 3 PH et 3 approches, aboutissant à la même conclusion : nécessité d' une héparinothérapie pour une embolie pulmonaire ;

Variabilité
des
raisonnements

❖ Incertitude : interprétation des D-dimères chez une patiente âgée, porteuse d' un cancer, de l' ovaire ; rapport bénéfice-risque d' une TDM injectée dans ce contexte ;

Incertitude

❖ Pondération de l' apport de ces éléments sur la décision finale (micro jugements).

Pondération

Théorie des scripts

- ❖ Script : réseau de connaissances organisées (clinique, biologie,...) utilisées pour la résolution d'un problème clinique ;
- ❖ Les scripts sont élaborés dès le début des rencontres avec les patients ;
- ❖ Ils sont d'autant plus riches que les contacts sont plus fréquents ;
- ❖ Ils ne sont pas transférables ;
- ❖ Ils sont variables d'un praticien à l'autre.

Le TCS

Technique d'évaluation standardisée permettant d'évaluer le raisonnement clinique à travers l'existence et la fonctionnalité de liens (*scripts*) au sein des connaissances cliniques, dans un contexte de simulation de situations cliniques mal définies

- ❖ Technique d'évaluation *standardisée* ;
- ❖ En clinique : ne s'applique pas aux sciences fondamentales ;
- ❖ Situations mal définies : pas de consensus, pas de médecine fondée sur les preuves ;
- ❖ Mesure le raisonnement *structuré* ou *compilé* ;
- ❖ Comparaison à un panel d'experts (scores combinés - *aggregate scoring*).

Qu'est-ce qu'un TCS ?

Vignette clinique courte

Une femme de 82 ans est adressée aux urgences par sa maison de retraite pour une douleur basi thoracique droite accompagnée de dyspnée. Elle y était hospitalisée en soins palliatifs. Elle ne se déplace plus que de son lit au fauteuil. Elle présente également une insuffisance cardiaque évoluée sur cardiopathie ischémique. La température est à 38,2°C. La TA est à 125 - 80 mm Hg, la fréquence cardiaque est à 130 /min et la fréquence ventilatoire est à 30 / min. La saturation transcutanée est à 90 p. cent sous 6 l/min d'oxygène au masque facial. Il existe un épanchement pleural de faible abondance à la base droite. Elle présente un très mauvais état veineux avec des œdèmes chroniques. Vous discutez une embolie pulmonaire.

Si vous pensiez prescrire	Et qu'alors vous trouvez	L'effet sur la nécessité de demander ce test est le suivant
Un dosage des D dimères	Un cancer de l'ovaire en cours de traitement	-2 -1 0 +1 +2
Un écho doppler veineux	Un membre inférieur gauche plus volumineux que le droit	-2 -1 0 +1 +2
Un angioscanner thoracique	Une créatinine à 150 µmol/l	-2 -1 0 +1 +2
-2 : Tout à fait inutile -1 : Inutile	0 : Ni plus, ni moins utile	+1 : Plus utile +2 : Très utile

Item

Question

Pondération
(ancrage)

Hypothèses pertinentes

Informations
supplémentaires relevantes

Echelle de Likert : Mesure l'impact de l'information supplémentaire sur l'hypothèse initiale (Micro jugement)

Le TCS est-il un outil d'évaluation efficient ?

❖ Un outil valide ?

- ✓ Validité apparente : conditionnée par la rigueur de sa construction ;
- ✓ Validité de construit :
 - identification d'experts et juniors ;
 - prédictibilité des scores d'examens explorant le même domaine ;

❖ Un outil fidèle ?

- Scores alpha de 0,79 à 0,82

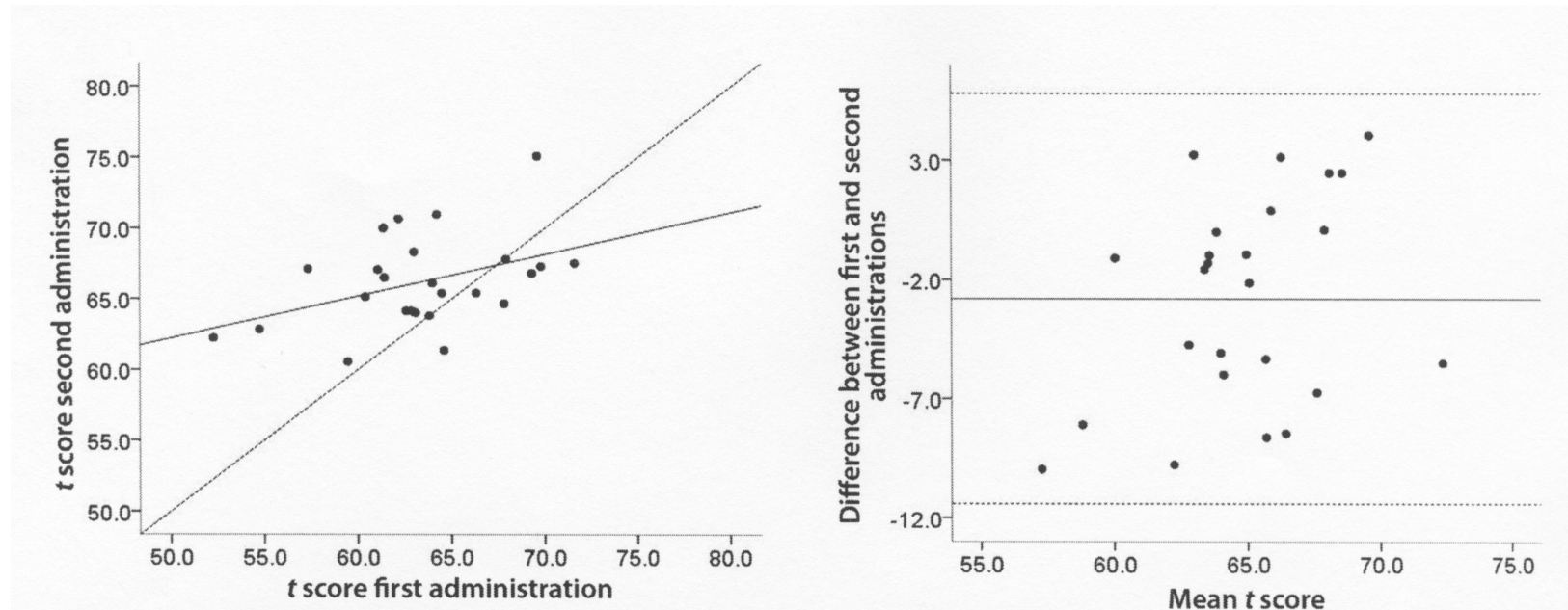
QCM : 0,76 à 0,93 (2 à 8 heures)

Patient Management Problem : 0,69 à 0,82 (4 à 8 heures)

Key Feature Case : 0,66 à 0,79 (4 à 8 heures).

Fidélité du test

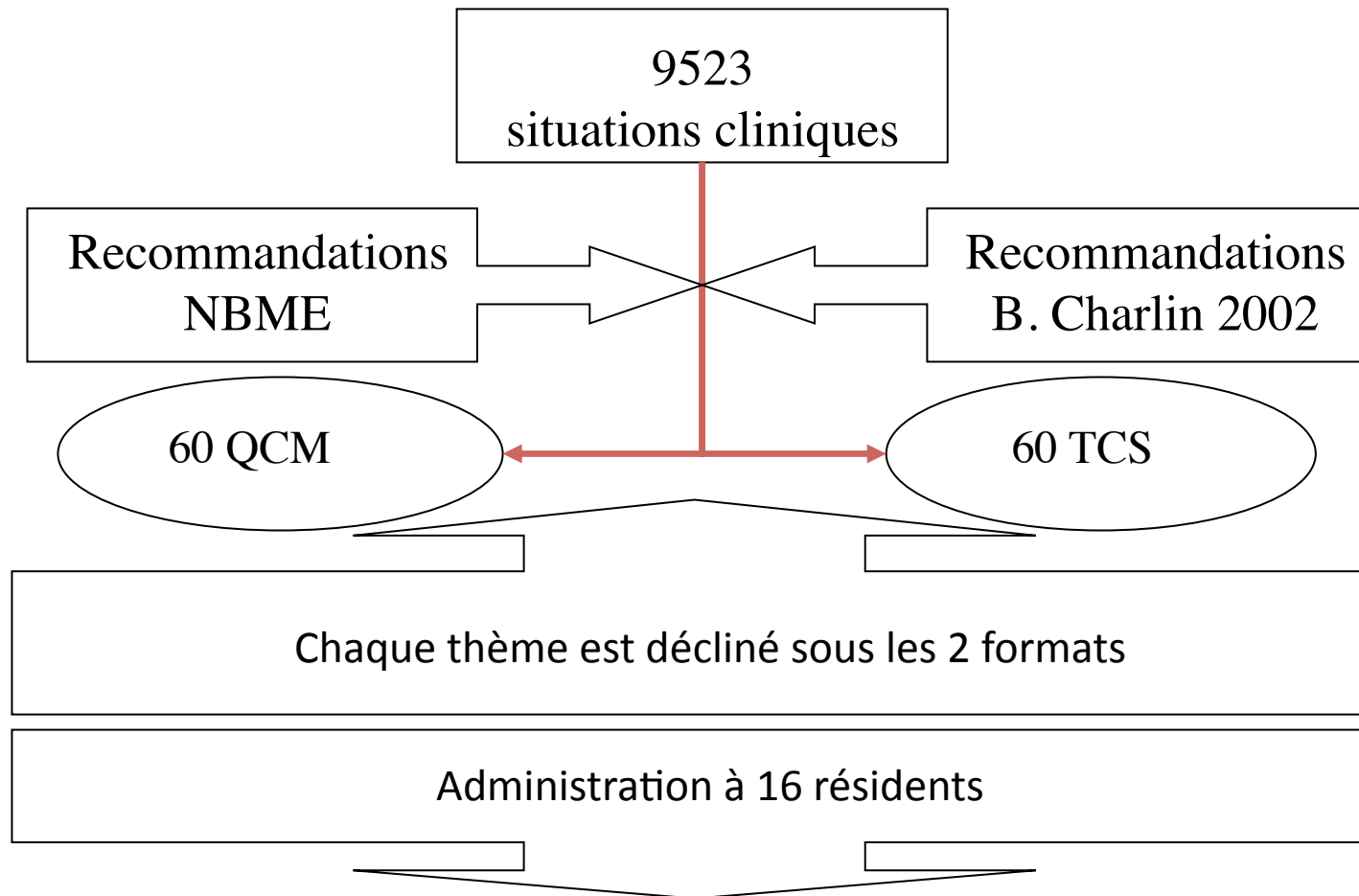
- ❖ Cohérence interne du test ;
- ❖ Mesurée par le coefficient α de Cronbach ($\geq 0,80$) ;
- ❖ Habituellement atteint avec 25 cas, 75 questions, et 60 minutes de composition.



Quelle est sa place parmi les autres outils d'évaluation ?

- ❖ Comparaison aux QCM ;
- ❖ Comparaison aux questions rédactionnelles format ECN.

TCS vs QCM à contexte riche



TCS vs QCM à contexte riche

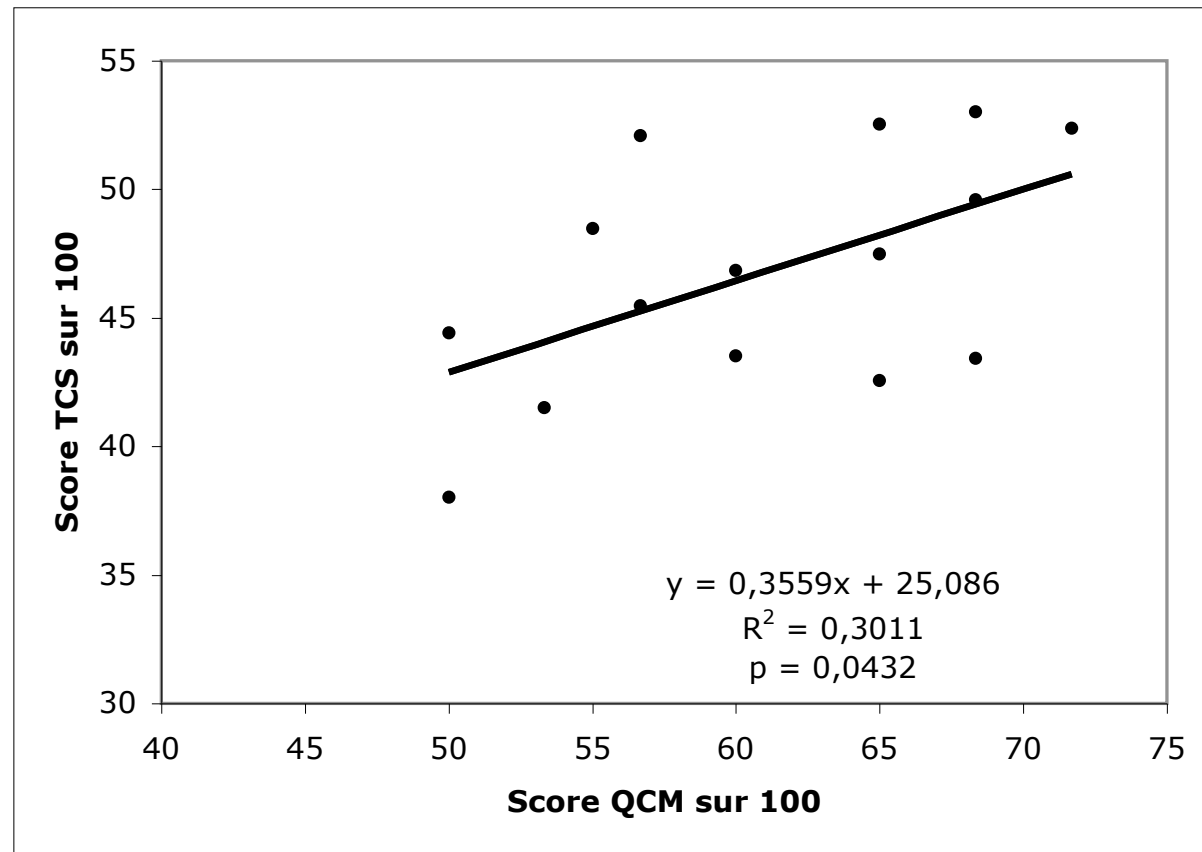
Une patiente de 17 ans consulte pour un « accident de préservatif ». Elle est au 12^{ème} jour de son cycle. Quelle est votre prescription de sortie ?

- A. Lévonorgestrel, éthinyli estradiol (Adépal[®])
- B. Lévonorgestrel (Norlévo[®])
- C. Consultation au planning familial
- D. Méthotrexate (Méthotrexate[®])
- E. Dosage des β -HCG urinaires

Une patiente consulte pour un « accident de préservatif ». Vous discutez le traitement.

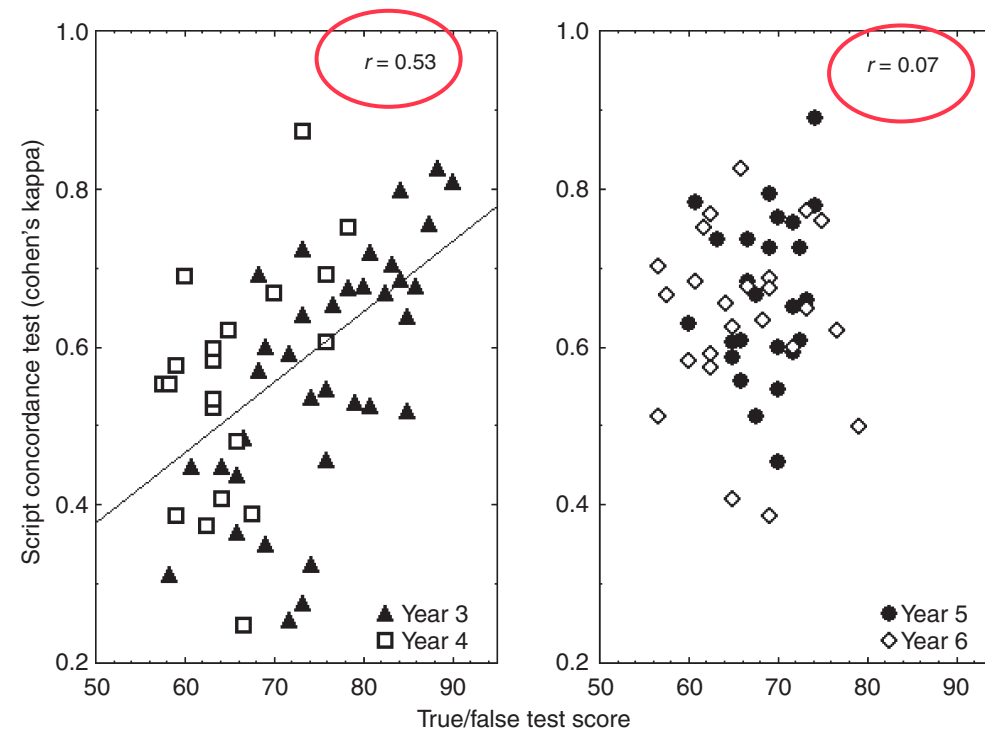
Si vous pensiez à	Et qu'alors vous trouvez	Cette nouvelle donnée rend le traitement proposé
Une prescription de lévonorgestrel (Norlévo [®])	Un « accident » survenu le 2 ^{ème} jour du cycle	-2 -1 0 +1 +2
Une prescription de lévonorgestrel (Norlévo [®])	Une patiente âgée de 15 ans	-2 -1 0 +1 +2
Une prescription de lévonorgestrel (Norlévo [®])	Des antécédents familiaux d'hypercholestérolémie	-2 -1 0 +1 +2
-2 : Totalelement contre indiqué -1 : Contre indiqué	0 : Ni plus, ni moins indiqué	+1 : Indiqué +2 : Formellement indiqué

TCS vs QCM à contexte riche

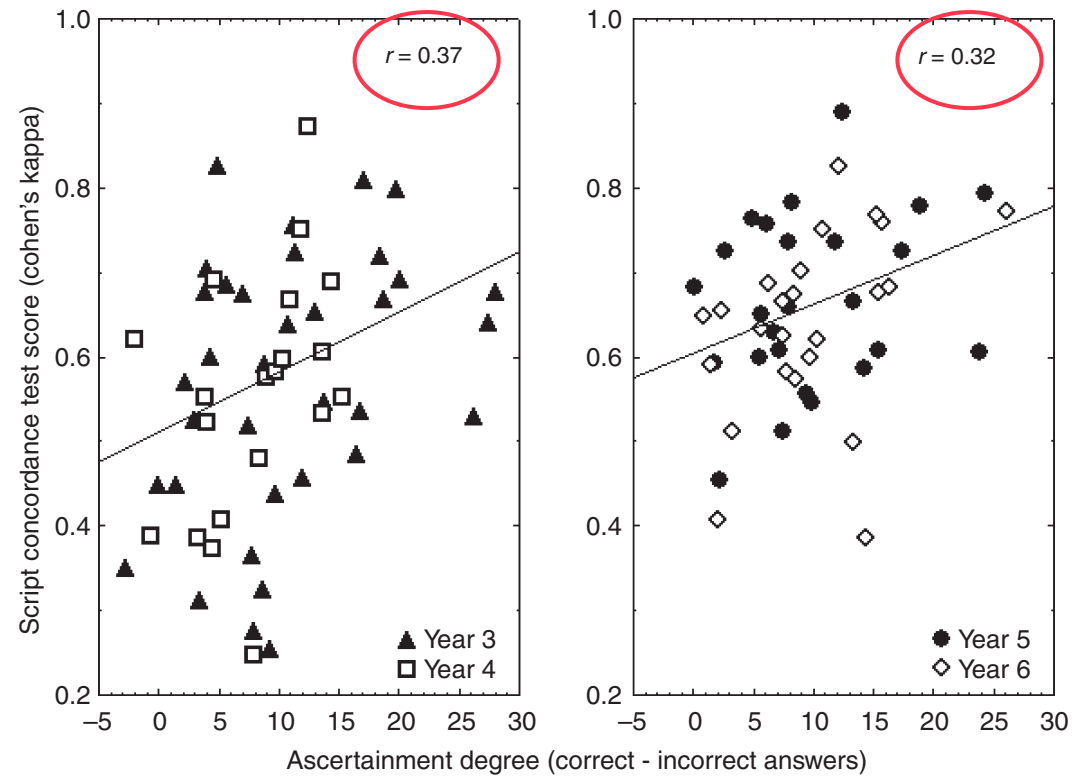


TCS vs QCM factuels

- ❖ 106 étudiants en médecine de 3^{ème} à 6^{ème} année ;
- ❖ Enseignement d'endocrinologie ;
- ❖ TCs vs QCM factuels (vrai / faux) :



TCS vs QCM factuels

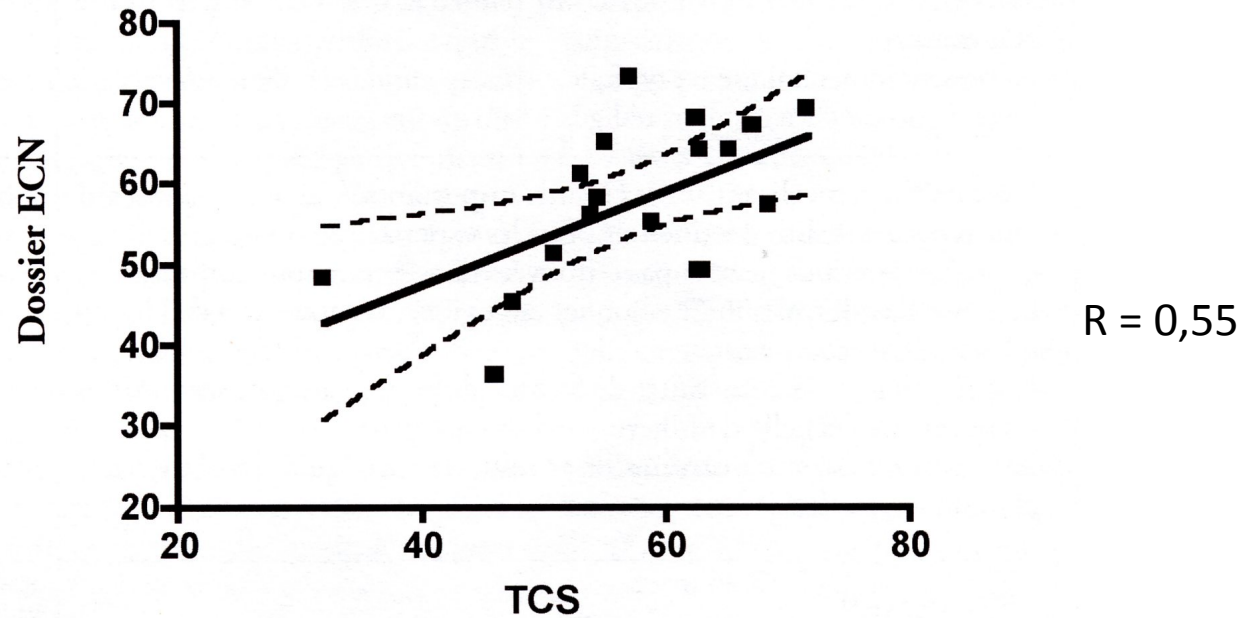


TCS vs dossier type ECN

- ❖ 18 étudiants de DCEM 2 à DCEM 4 en réanimation médicale ;
- ❖ Dossier type ECN ;
- ❖ TCS comportant 26 questions

TCS vs dossier type ECN

Corrélation entre les scores obtenus par les étudiants au test de concordance de script (TCS) (abscisses) et au dossier de « type épreuves classantes nationales » (ECN) (ordonnées)



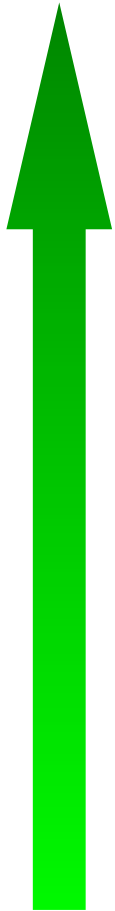
$R = 0,55$

TCS vs autres techniques d'évaluation

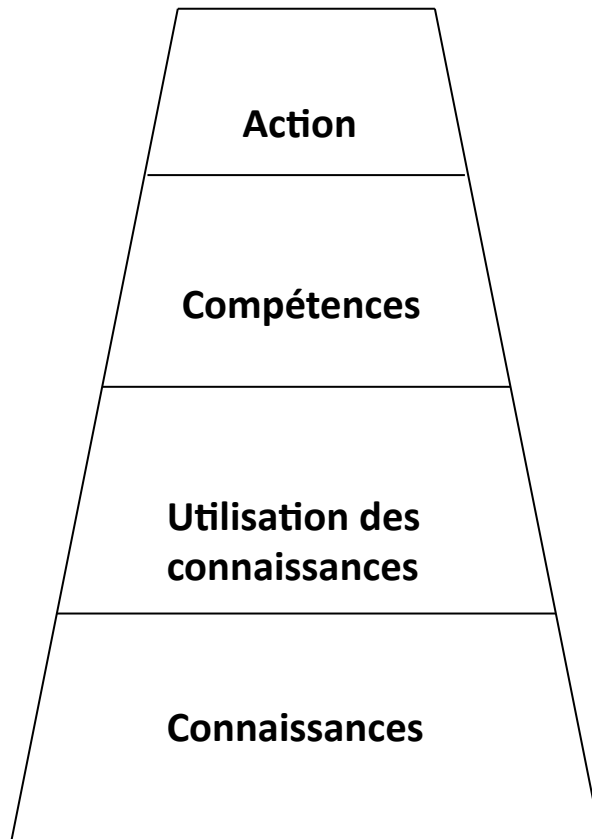
- ❖ Corrélation statistiquement significative, mais valeur intermédiaire : le TCS mesure des points communs, mais explore d'autres aspects ;
- ❖ Pas de caractère redondant entre ces différentes techniques d'évaluation.

Techniques d'évaluation courantes

Performance



Compétence



Observation directe, pairs, vidéo, audits

Examen oral, ECOS, épreuve de clinique, simulation, **TCS**

Cascades, QCM à contexte riche, **TCS**, problèmes à résolution séquentielle

QCM, QROC, questions rédactionnelles

D'après Miller *Acad. Med.* 1990

Quelle acceptabilité ?

Table 2. Satisfaction Survey Conducted Among 20 Pharmacy Students (Out of a Cohort of 68 Students) Following the SCT-Examination

Statements	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I was satisfied with this type of examination	2	1	3	6	8
I found the examination in adequacy with the course	1	1	3	9	6
I think my examination's result is linked to participation in exercises	1	2	9	7	1
I would have preferred an alternative examination	6	4	4	4	2
I would have preferred a multiple-choice examination	2	2	6	5	5
I would have preferred a Short-open question examination	9	3	2	3	3
I would have preferred a « true or false » examination	6	1	8	2	3
I would have preferred an oral examination	14	0	2	4	0
I think that SCT shows real situations of the community pharmacy	0	0	2	13	5
I think that the SCT is preparing to traineeship	1	3	5	7	4
I think that the SCT is preparing me to future practice	0	0	5	10	5
I think we should continue this type of test in the future	1	2	1	7	8



L'examen

Aspects techniques pratiques

- ❖ Table de spécification (*blueprint*) ;
- ❖ Rédaction de scénarii et questions ;
- ❖ Ancrages ;
- ❖ Panel d' experts ;
- ❖ Contrôle qualité ;
- ❖ FAQ.

Table de spécification

4 éléments :

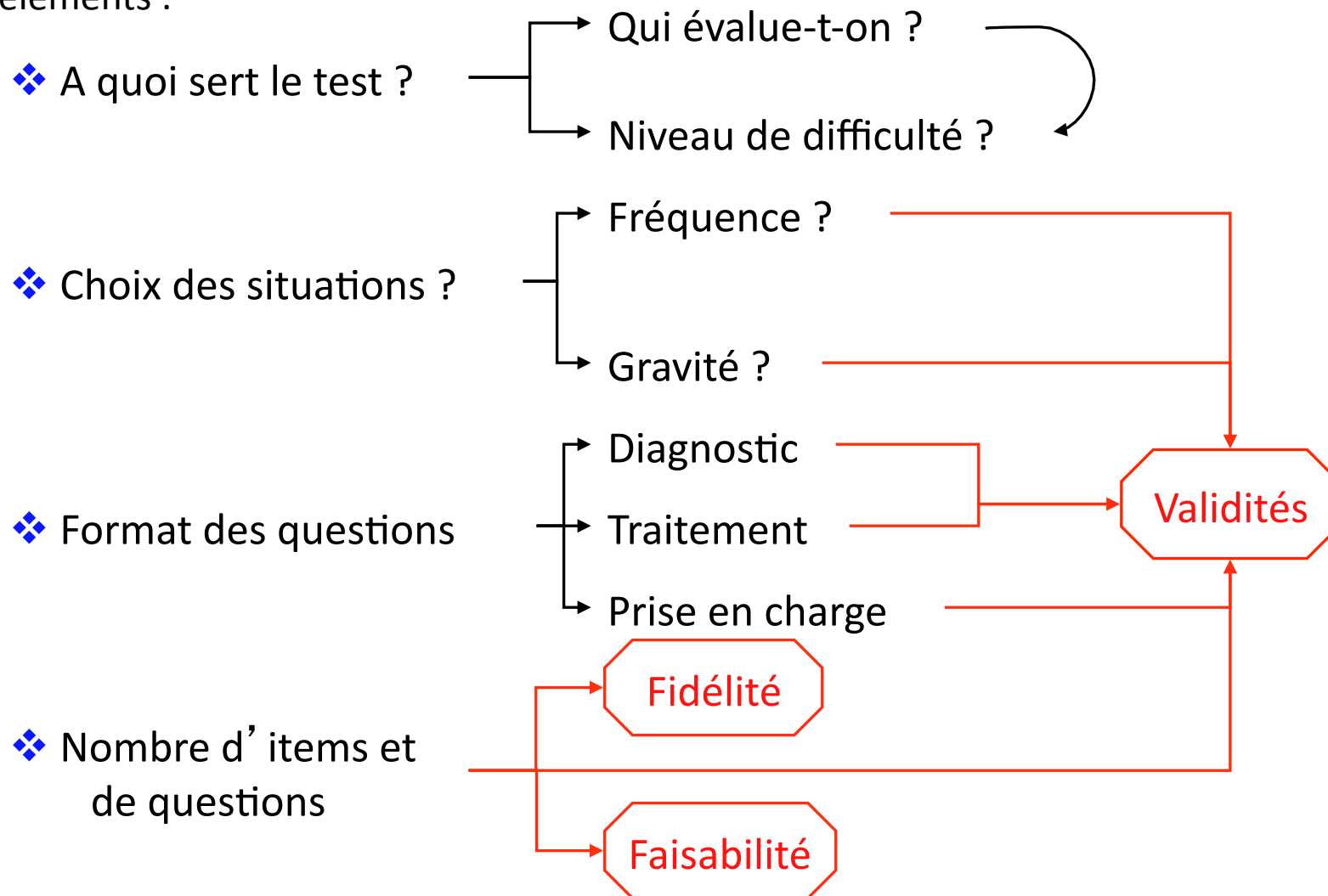


Table de spécification

	Cardio	Pneumo	Neuro	Gastro	Métabol.	Rhumato	Gériatr.	Gynéco	Psy	Trauma
Dg	X	X	X	X		X	X	X	X	X
Trt	X	X	X	X	X				X	
Prise en charge	X	X	X							

	Cardio	Pneumo	Neuro	Gastro	Métabol.	Rhumato	Gériatr.	Gynéco	Psy	Trauma
Dg	Douleur thoracique	Dyspnée aiguë	Céphalée	Douleur abdominale		Mono arthrite	AEG	GEU	Confusion	Trauma vertébral
Trt	Insuffisance cardiaque	Embolie pulmonaire	Crise comitiale	Hémorragie digestive	Deshydr.				Agitation	
Prise en charge	Synd. coronarien aigu	BPCO en poussée	AVC							

Table de spécification

En résumé :

- ❖ Architecture de l'examen ;
- ❖ Sélection des situations ;
- ❖ Sélection des hypothèses pertinentes ;
- ❖ Sélection des informations supplémentaires pertinentes ;

→ Fixe le niveau d'incertitude.

❖ 1 – 2 concepteurs
/ rédacteurs
❖ Experts

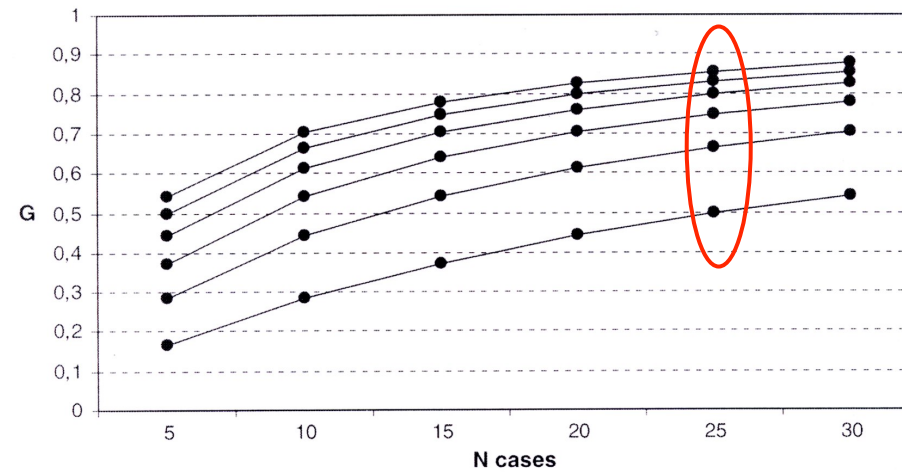
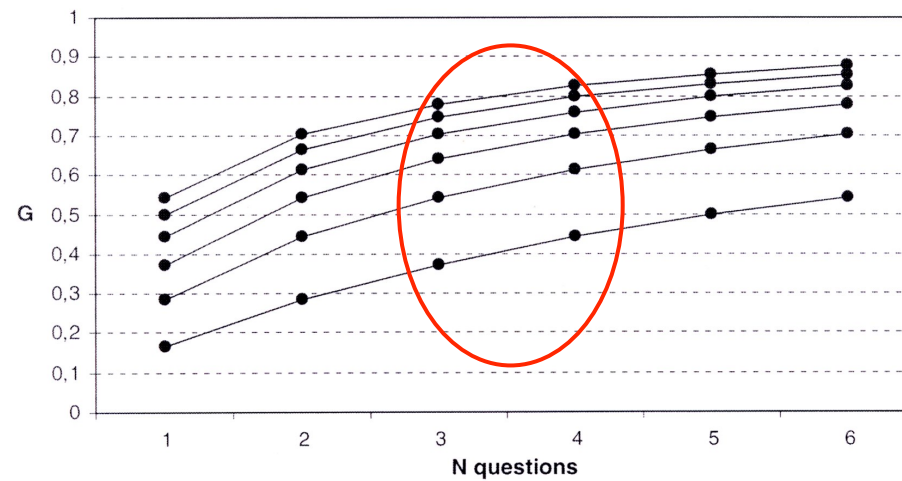
Niveau d'incertitude

- ❖ Pas de piège ;
- ❖ Pas de consensus, guidelines ;
- ❖ Plusieurs options possibles ;
- ❖ Données incomplètes.

Combien de cas et de questions ?

- ❖ Notion de spécificité de cas (*case specificity*) ;
- ❖ Nombreux cas avec 1 question ?
- ❖ 1 cas avec n questions ?
- ❖ Combien de cas /questions pour une fidélité acceptable ?

Combien de cas et de questions ?



Combien de cas et de questions ?

Table 3 D study of reliability by number of questions and cases

Number of questions per case	Number of cases	Total number of questions	Approximate testing time in minutes	G coefficient		
				Oncology	Pediatrics	Nursing
1	30	30	25	0.72	0.54	0.70
2	25	50	40	0.81	0.66	0.79
2	30	60	50	0.83	0.70	0.82
3	20	60	50	0.83	0.70	0.82
3	25	75	60	0.86	0.75	0.85
4	15	60	50	0.83	0.70	0.81
4	20	80	65	0.87	0.76	0.85
5	12	60	50	0.83	0.69	0.81
5	15	75	60	0.86	0.74	0.84

Conserver un format homogène +++++

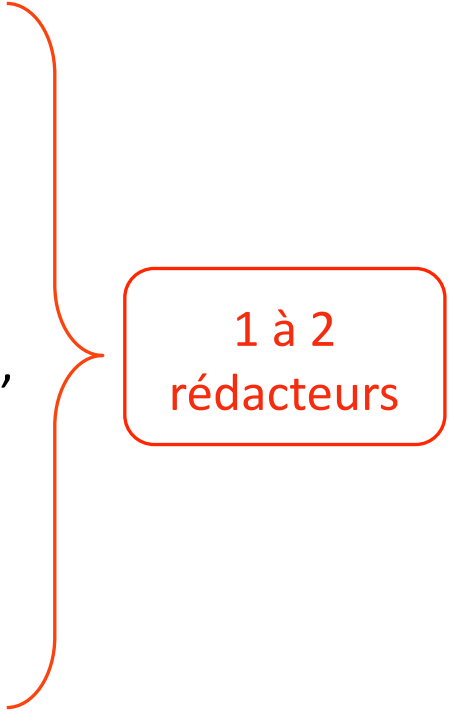
Un homme de 56 ans est adressé par les pompiers pour malaise sur la voie publique. Il a été découvert inanimé par des passants. Il reprend peu à peu conscience. En dehors d'une désorientation temporo spatiale et d'une agitation, son examen neurologique est sans particularité. Les constantes vitales sont normales. Vous planifiez sa prise en charge.

<i>Si vous pensiez prescrire :</i>	<i>Et qu'alors vous trouvez :</i>	<i>L'effet sur la nécessité de demander ce test est le suivant :</i>
Un dosage de lactates veineux	Une intervention des pompiers environ 2 heures avant votre examen	-2 -1 0 +1 +2
Un EEG	Une morsure du bord latéral de la langue	-2 -1 0 +1 +2
Une injection de 1 mg de clonazépam (Rivotril [®])	Une morsure du bord latéral de la langue	-2 -1 0 +1 +2
-1 : Inutile -2 : Tout à fait inutile	0 : Ni plus, ni moins utile	+1 : Plus utile +2 : Très utile

Rédaction de scénarii et questions

Scénario : informations nécessaires à la description du problème :

- ❖ Âge et sexe ;
- ❖ Motif de consultation ou d' admission ;
- ❖ Contexte : urgences, cabinet de consultation, hospitalisation,...
- ❖ Données cliniques,...

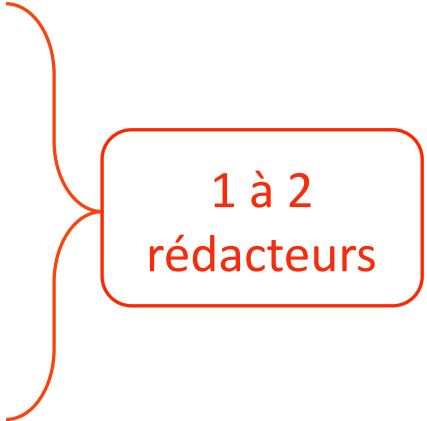


1 à 2
rédacteurs

Rédaction de scénarii et questions

Sélection des hypothèses *pertinentes* :

- ❖ Pertinentes dans le contexte (vignette clinique) ;
- ❖ Pas de piège.

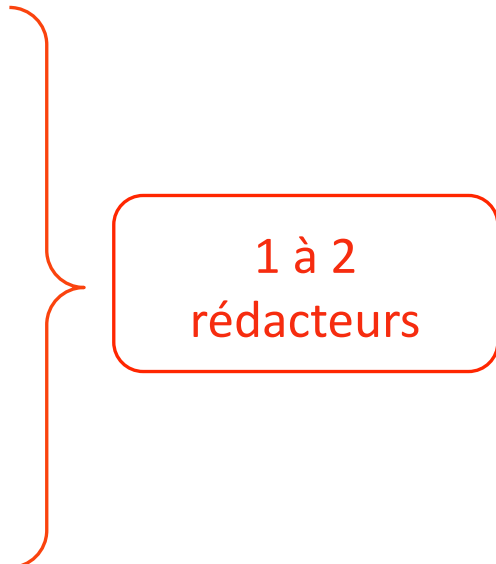


1 à 2
rédacteurs

Rédaction de scénarii et questions

Informations supplémentaires *pertinentes* :

- ❖ Donnée d'examen clinique ;
- ❖ Résultat biologique ;
- ❖ Imagerie statique ;
- ❖ Vidéo,...



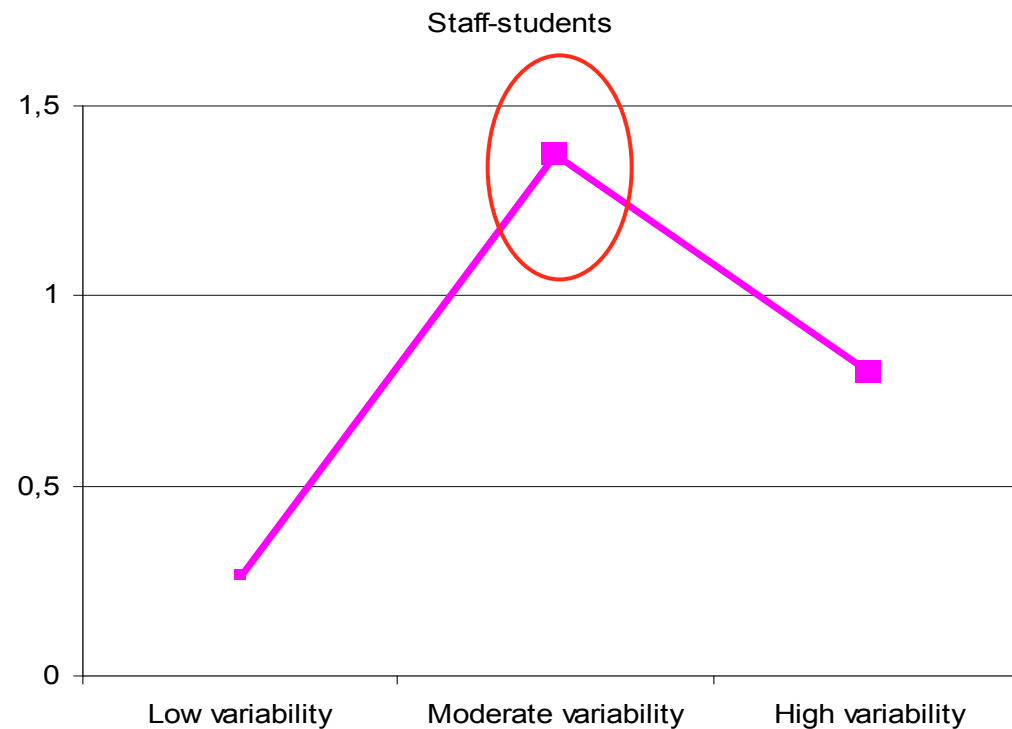
1 à 2
rédacteurs

Rédaction de scénarii et questions

Rédaction des questions et informations supplémentaires :

- ❖ Balayer l'ensemble des points de l'échelle de Likert ;
- ❖ Niveau de variabilité équilibrée ;
- ❖ Chaque question est indépendante ;
- ❖ Pas d'hypothèse ou d'information commune d'une question à l'autre ;
- ❖ Format homogène des questions ;
- ❖ Intérêt de questions à consensus ?

Impact de la variabilité des questions sur leur pouvoir discriminant



Échelle de Likert - Ancrages

Recommandations :

- ❖ Préciser la signification des différents points pour éviter toute ambiguïté, **en particulier le 0** ;
- ❖ Garder la même présentation.

-1 : Moins probable -2 : Très peu probable	0 : Ni plus, ni moins probable	+1 : Plus probable +2 : Très probable
---	-----------------------------------	--

Échelle de Likert - Ancrages

Échelles :

❖ Ancrage pour investigation (utilité) :

Si vous considériez l'utilité de l'examen suivant	Et qu'alors vous trouvez	Cette nouvelle information rend l'examen
Option d'investigation	Nouvelle information	-2 -1 0 +1 +2
-1 : Inutile -2 : Totalement inutile	0 : Ni plus, ni moins utile	+1 : Utile +2 : Très utile

N' a pas sa place ici

Prescription justifiée ici

À ne pas prescrire ici

❖ Ancrage pour investigation (risque bénéfice) :

Si vous considériez le rapport risque bénéfice de l'examen suivant	Et qu'alors vous trouvez	Cette nouvelle information rend l'examen
Option d'investigation	Nouvelle information	-2 -1 0 +1 +2
-1 : Contre indiqué -2 : Totalement contre indiqué	0 : Ni plus, ni moins indiqué	+1 : Indiqué +2 : Très indiqué

Échelle de Likert - Ancrages

Échelles :

❖ Ancrage pour traitement (utilité) :

Si vous considériez l'utilité du traitement suivant	Et qu'alors vous trouvez	Cette nouvelle information rend le traitement
Option de traitement	Nouvelle information	-2 -1 0 +1 +2
-1 : Inutile -2 : Totalement inutile	0 : Ni plus, ni moins utile	+1 : Utile +2 : Très utile

N' a pas sa place ici

Prescription justifiée ici

À ne pas prescrire ici

❖ Ancrage pour traitement (risque bénéfice) :

Si vous considériez le rapport risque bénéfice du traitement suivant	Et qu'alors vous trouvez	Cette nouvelle information rend le traitement
Option de traitement	Nouvelle information	-2 -1 0 +1 +2
-1 : Contre indiqué -2 : Totalement contre indiqué	0 : Ni plus, ni moins indiqué	+1 : Indiqué +2 : Très indiqué

Échelle de Likert - Ancrages

Échelles :

❖ Ancrage pour pronostic :

Paramètre pronostique	Information supplémentaire	Le pronostic devient
Paramètre pronostique	Nouvelle information	-2 -1 0 +1 +2
-1 : Moins favorable -2 : Pas du tout favorable (défavorable)	0 : Inchangé	+1 : Plus favorable +2 : Très favorable

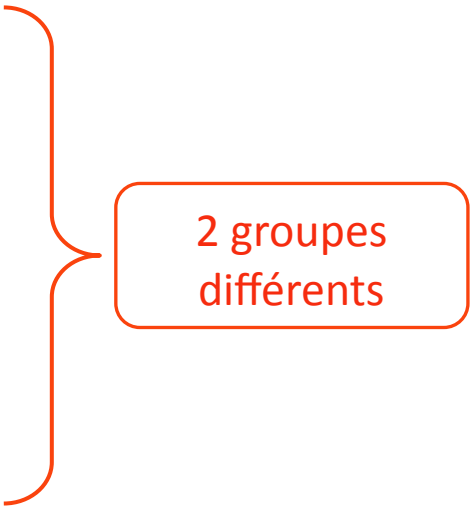
Échelle de Likert - Ancrages

Mme T. G3P1 est enceinte de 35 SA et à sa dernière consultation obstétricale on lui trouve 15/08 de TA. Le médecin lui prescrit du repos, un bilan rénal et une surveillance par SF à domicile 3 fois par semaine. L'état de la patiente est amélioré par le repos. A 37 SA, la sage-femme trouve une TA à 19/11 et fait hospitaliser Mme T. Un traitement anti-hypertenseur est administré par seringue électrique.		
À ce stade : La TA est stabilisée à 14/08	Et vous retrouvez dans le bilan que : Glycémie à jeun = 6,06 mmol/l (1,10 g/l)	Le pronostic maternel devient alors : -2 -1 0 +1 +2
À ce stade : Les biométries fœtales sont au 25^{ème} percentile.	Et vous retrouvez : Une albuminurie ++	le pronostic fœtal devient alors : -2 -1 0 +1 +2
À ce stade : L'examen clinique révèle un bassin normal.	Et vous apprenez : Mme T à 42 ans.	Le pronostic obstétrical devient alors : -2 -1 0 +1 +2
-1 : Moins favorable -2 : Pas du tout favorable (défavorable) 0 : Inchangé +1 : Plus favorable +2 : Très favorable		

Panel d' experts

Rôles :

- ❖ Préparation de l' examen ;
- ❖ Validation de scénarii et questions ;
- ❖ Établissement des scores.

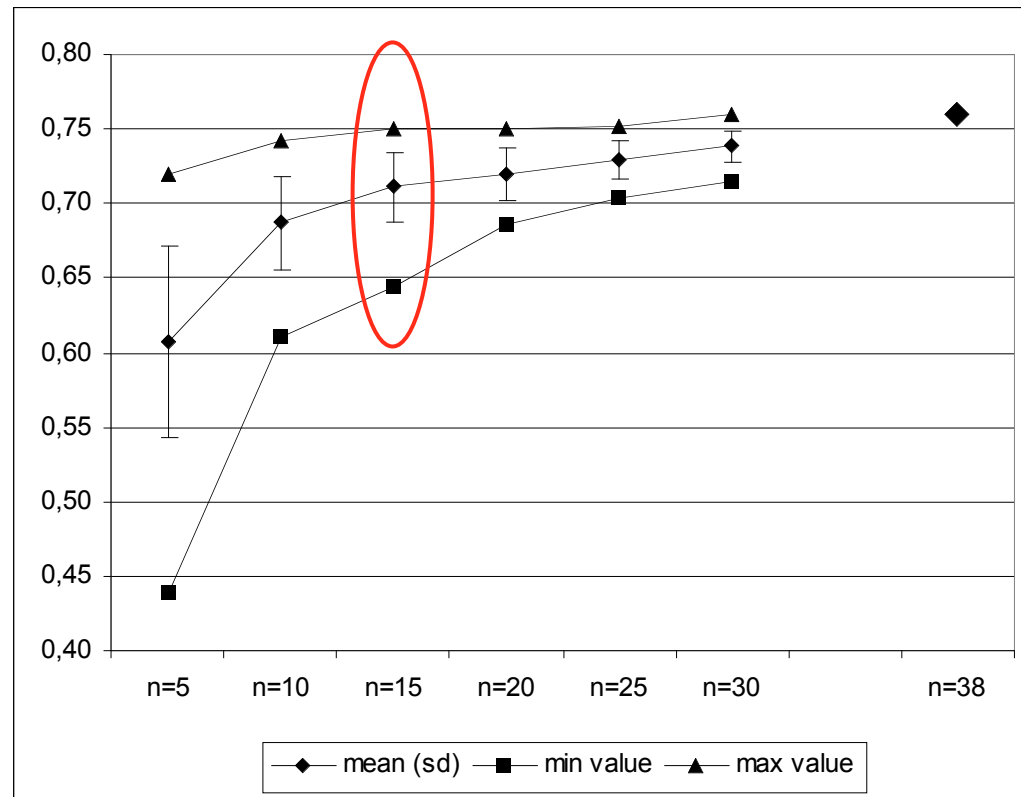


2 groupes
différents

Panel d' experts

- ❖ Qualité : experts dans le domaine clinique exploré ;
- ❖ Nombre : 10 à 20 minimum.

Fidélité moyenne en fonction de la taille du panel d'experts



Composition du panel d'expert

Enseignants vs praticiens ?

- ❖ TCS en médecine générale ;
- ❖ Comparaison des scores obtenus ;
- ❖ Panel : Généralistes enseignants vs Généralistes inscrits en FMC ;
- ❖ Concordance parfaite des scores obtenus ($R = 0,98$) ;
- ❖ Scores plus élevés avec les Généralistes inscrits en FMC ($p = NS$).

Impact de la composition du panel d'experts sur les scores en cas de TCS multi disciplinaire ?

- ❖ Duggan (2012) : panel multi spécialistes : 158 questions : Cronbach α : 0,62 (5^{ème} année) à 0,78 (6^{ème} année) ;
- ❖ Humbert (2011) : panel interniste : 75 questions : Cronbach α : 0,73 ;
- ❖ Gagnon (2011) : Pas d'impact de l'optimisation du panel sur le Cronbach α .

Impact de la composition du panel d'experts sur les scores en cas de TCS multi disciplinaire ?

1 groupe d'experts par champ (discipline) :

	Experts / champ	Items / questions	Score α
Duggan	10 – 16	51 / 158	0,62 et 0,78
Ramaekers	28 en tout	30 / 120	0,79 et 0,80
Petrucci	6	53 / 153	0,67 → 0,81

D'après

Duggan P. *BMC Med Educ* 2012

Ramaekers S. *Assessment Evaluation Higher Education* 2012

Petrucci A. *Am J Surg* 2012

Impact de la composition du panel d'experts sur les scores en cas de TCS multi disciplinaire ?

1 groupe d'experts par champ (discipline) :

	Experts	Etudiants	Items / Questions	Score α
Nouh	22	202	53 / 153	0,81 ➔ 0,89
Petrucci	6 x 5	202 + 25	53 / 153	0,61 ➔ 0,81

Établissement des scores

Correction : Établissement des scores :

❖ Choix de 10 experts :

		-2	-1	0	+1	+2
Unit 1	Question 1	7	1	0	2	0

❖ Transformation modale du choix des 10 experts :

		-2	-1	0	+1	+2
Unit 1	Question 1	1	0.14	0	0.28	0

❖ Choix des 8 résidents :

		R 1	R 2	R 3	R 4	R 5	R 6	R 7	R 8
Unit 1	Question 1	- 1	- 1	- 1	2	1	0	- 2	0

❖ Calcul des scores des 8 résidents :

		R 1	R 2	R 3	R 4	R 5	R 6	R 7	R 8
Unit 1	Question 1	0.14	0.14	0.14	0	0.28	0	1	0

Plus de signification aux scores ?

- ❖ Etablissement des scores individuels des membres du panel d'experts ;
- ❖ Transformation du score des étudiants en *t-score* :

$$t\text{-score}_{\text{étudiants}} = [(SCT_{\text{étudiants}} - \text{mean } SCT_{\text{experts}}) / SD \text{ } SCT_{\text{experts}}] \times 5 + 80$$

80 : score max habituellement rencontré chez les experts

5 : SD habituellement rencontrée chez les experts

Établissement des scores

Calculateur Excel sur le site de l' Université de Montréal :

www.cme.umontreal.ca/tcs

Au total : qui fait quoi ?

Planificateur	❖ Table de spécification ❖ Situations cliniques ❖ Hypothèses ❖ Informations supplémentaires ❖ Établissement des scores des étudiants ❖ Décision de passage
Experts 1	❖ Situations cliniques ❖ Hypothèses ❖ Informations supplémentaires
Experts 2	❖ Établissement des scores des étudiants

Contrôle qualité *a priori*

Vignette	❖ Décrit une situation problématique, même pour les experts ❖ Décrit une situation clinique adaptée au niveau de compétence des étudiants ❖ Question et contexte sont explicites ❖ Présentation clinique typique ❖ Vignette correctement rédigée	☐ Oui ☐ Non ☐ Oui ☐ Non ☐ Oui ☐ Non ☐ Oui ☐ Non ☐ Oui ☐ Non
Questions	❖ Questions focalisant sur les points importants ❖ Hypothèses pertinentes ❖ 2 options identiques non rencontrées dans 2 questions consécutives ❖ Information supplémentaire pertinente / hypothèse ❖ Ancrage explicite, en particulier la valeur neutre (0) ❖ Questions construites pour balayer les 5 points de l'échelle de Lickert ❖ Questions construites pour obtenir une variabilité moyenne	☐ Oui ☐ Non ☐ Oui ☐ Non ☐ Oui ☐ Non ☐ Oui ☐ Non ☐ Oui ☐ Non ☐ Oui ☐ Non ☐ Oui ☐ Non
Panel d'experts	❖ Entre 10 et 20 ❖ Médecins expérimentés dont la présence dans le panel est licite compte tenu du niveau de compétence des étudiants ❖ Les experts passent le test dans les mêmes conditions que les étudiants	☐ Oui ☐ Non ☐ Oui ☐ Non ☐ Oui ☐ Non

Contrôle qualité *a posteriori*

Analyse d'items :

- ❖ Indices de difficulté ;
- ❖ Indice de discrimination ;
- ❖ Fidélité globale du test (score de Cronbach) ;
- ❖ Impact de chaque item sur la fidélité du test ;
- ❖ Réponse « déviante ».

Prévoir plus de
cas pour
optimiser le
test (+ 20%)

www.cme.umontreal.ca/tcs

Que faire des experts « déviants » ?

Table 2 Comparison of optimization methods with panel of 45

	Alpha value	Correlation with original data	Effect size		
			Students-residents	Students-panel	Residents-panel
No exclusion	0.86		2.08	2.57	0.48
Removal of members	0.86	0.99	2.05	2.44	0.45
Distance-from-mode					
2 Anchors	0.86	0.99	2.06	2.53	0.47
1 Anchor	0.86	0.99	2.08	2.52	0.46
Judgment-by-experts	0.86	0.99	1.99	2.47	0.43

Table 3 Comparison of optimization methods on panels of 15 (20 samples of 15 out of 45)

	Mean alpha value (SD)	Mean correlation with original data (SD)	Mean effect size (SD)		
			Students-residents	Students-panel	Residents-panel
No exclusion	0.86 (0.01)	0.99 (0.01)	1.97(0.12)	3.44 (0.70)	1.18 (0.33)
Removal of members	0.86 (0.01)	0.98 (0.01)	2.08 (0.11)	3.07 (0.74)	0.88 (0.32)
Distance-from-mode					
2 Anchors	0.86 (0.01)	0.98 (0.01)	1.93 (0.11)	3.16 (0.75)	1.05 (0.37)
1 Anchor	0.85 (0.01)	0.98 (0.01)	2.01(0.12)	2.32 (0.17)	0.37 (0.07)
Judgment-by-experts	0.85 (0.01)	0.98 (0.01)	2.03 (0.12)	3.73 (0.53)	1.24 (0.26)

❖ Si recherche de réponses « acceptables » (validité apparente : « élimination » de réponses plutôt que d'experts ;
❖ *Distance from mode 1* ou 2 la plus pratique.

Une patiente de 86 ans est adressée aux urgences par sa maison de retraite pour « altération de l'état général ». Elle « décline » depuis 48 heures. Elle a comme principaux antécédents une démence type Alzheimer, une HTA, et un DNID. La température auriculaire est à 38,6°C. La saturation transcutanée en O₂ est à 92 p. cent (les ambulanciers l'ont transportée sous O₂). Elle est somnolente, sans signe de localisation. La pression artérielle est à 110-70 mm Hg, et la fréquence cardiaque est à 110/min. Vous recherchez l'origine de cette hyperthermie.

<i>Si vous pensiez au diagnostic suivant :</i>	<i>Et qu'alors vous trouvez :</i>	<i>L'hypothèse devient :</i>
Une pyélonéphrite aiguë	BU : leucocytes : ++, sang : ++, protéines : ++, nitrites : 0, glucose : +++, corps cétoniques : ++	-2 -1 0 +1 +2
Une pneumonie bactérienne	Un traitement par donépézil (Aricept [®]), halopéridol (Haldol [®]), méprobamate (Equanil [®]), indapamide (Fludex [®]), et glipizide (Glibénèse [®])	-2 -1 0 +1 +2
Une déshydratation globale	Pas de foyer infectieux cliniquement décelable	-2 -1 0 +1 +2
-1 : Improbable -2 : Très improbable	0 : Ni plus, ni moins probable	+1 : Plus probable +2 : Beaucoup plus probable

10 experts :

		-2	-1	0	+1	+2
Item 2	Question 1	6	3			1
	Question 2			4	5	1
	Question 3		2	5	3	

Une patiente de 86 ans est adressée aux urgences par sa maison de retraite pour « altération de l'état général ». Elle « décline » depuis 48 heures. Elle a comme principaux antécédents une démence type Alzheimer, une HTA, et un DNID. La température auriculaire est à 38,6°C. La saturation transcutanée en O₂ est à 92 p. cent (les ambulanciers l'ont transportée sous O₂). Elle est somnolente, sans signe de localisation. La pression artérielle est à 110-70 mm Hg, et la fréquence cardiaque est à 110/min. Vous recherchez l'origine de cette hyperthermie.

<i>Si vous pensiez au diagnostic suivant :</i>	<i>Et qu'alors vous trouvez :</i>	<i>L'hypothèse devient :</i>
Une pyélonéphrite aiguë	BU : leucocytes : ++, sang : ++, protéines : ++, nitrites : 0, glucose : +++, corps cétoniques : ++	-2 -1 0 +1 +2
Une pneumonie bactérienne	Un traitement par donépézil (Aricept [®]), halopéridol (Haldol [®]), méprobamate (Equanil [®]), indapamide (Fludex [®]), et glipizide (Glibénèse [®])	-2 -1 0 +1 +2
Une déshydratation globale	Pas de foyer infectieux cliniquement décelable	-2 -1 0 +1 +2
-1 : Improbable -2 : Très improbable	0 : Ni plus, ni moins probable	+1 : Plus probable +2 : Beaucoup plus probable

10 experts :

		-2	-1	0	+1	+2
Item 2	Question 1	6	3			1
	Question 2			4	5	1
	Question 3	2	1	4	2	1

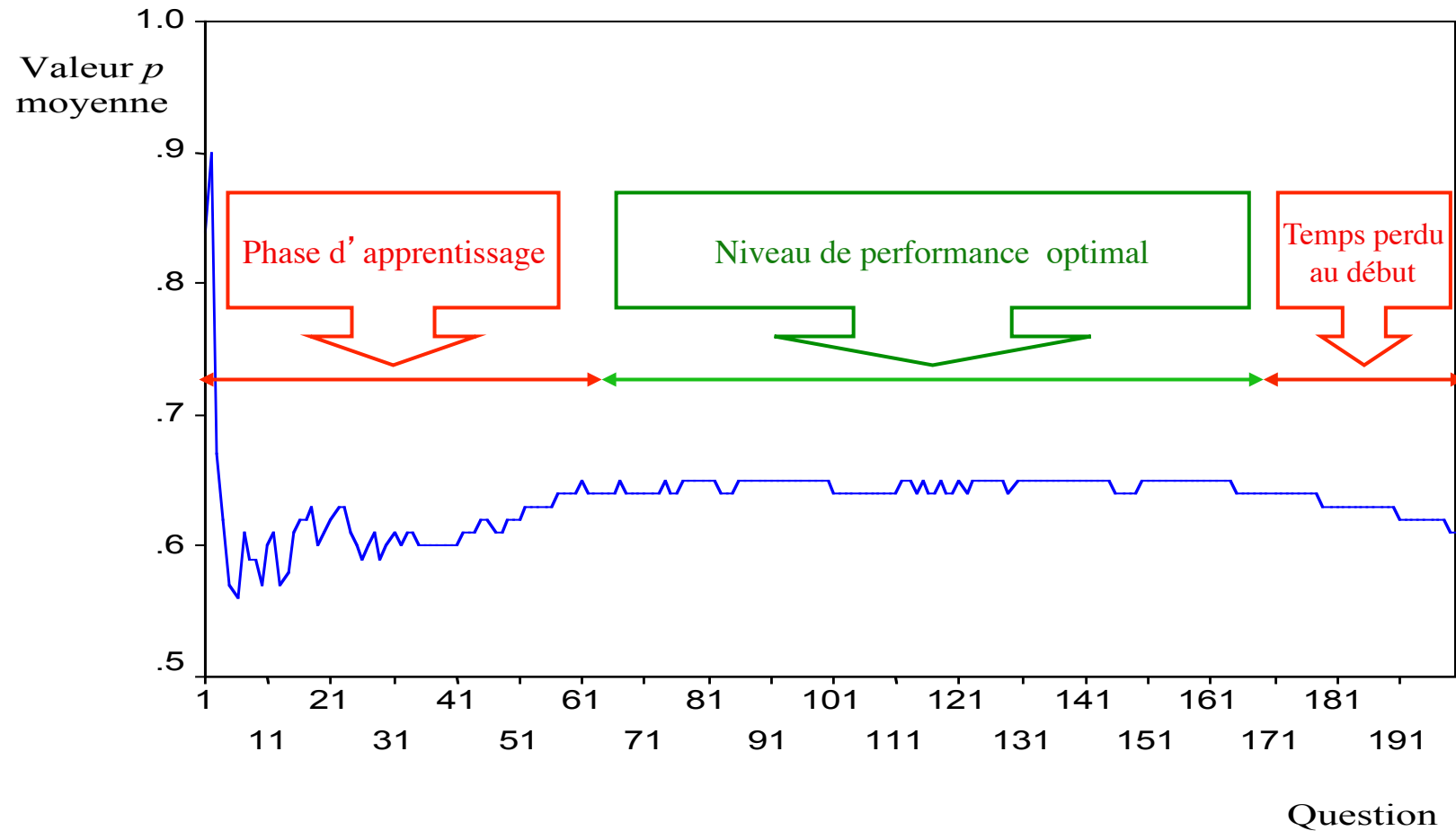
FAQ de l'ANEMF

Question	Réponse	Références
Inadapté au 2 ^{ème} cycle ?	Utilisation en 2 ^{ème} cycle : Nice (DCEM 1, DCEM 2, DCEM 3) Rouen (DCEM 3)	Sibert L. <i>BMC Med Educ</i> 2006 Sibert L <i>Med Teacher</i> 2009 Gibot S. <i>Ped Med</i> 2008 Caire F. <i>Ped Med</i> 2011 Hoff L. <i>Ped Med</i> 2010 Humbert A. <i>Med Teacher</i> 2011 Duggan P. <i>BMC Med Educ</i> 2012 Piovezan RD. <i>JAGS</i> 2012 Collard A. <i>Med Educ</i> 2009
Enseignants formés ?	Multiples ateliers : Rouen, Nice, Besançon, APNET, CEN, CGOF, Nancy, Poitiers,site SIDES...	
Impact des stages hospitaliers ?	Retour en stages hospitaliers	Newble DI. <i>Med Educ</i> 1983

FAQ de l'ANEMF

Question	Réponse	Références
Réponse au hasard ?	❖ Surévaluée ; ❖ Balayer tous les points de l'échelle de Likert ; ❖ Pas d'impact des techniques de correction	Harden RM. <i>Med Educ</i> 1976 Vanbelle S. <i>Ped Med</i> 2007 Prihoda TJ. <i>J Dent Educ</i> 2006
Discriminant ?	TCS en LCA : 1,65 % d'ex æquo (78,51 avec l'épreuve classique)	Fournier JP. SIFEM 2009
Impact de la composition du panel d'experts ?	Probablement pas en termes de fidélité ; Rangs de classement ?	Humbert A. <i>Med Teacher</i> 2011 Duggan P. <i>BMC Med Educ</i> 2012 Petrucci A. <i>Am J Surg</i> 2012 Ramaekers S. <i>Assessment and Evaluation in Higher Education</i> 2012

Finalelement l'enjeu c'est...



Merci de votre attention !

BMC Medical Informatics and Decision Making



Correspondence

Open Access

Script Concordance Tests: Guidelines for Construction

Jean Paul Fournier¹, Anne Demeester² and Bernard Charlin^{*3}

Address: ¹Département de Pédagogie Médicale, Faculté de Médecine de Nice Sophia Antipolis, France, ²École Régionale de Sages-femmes, Faculté de Médecine de Marseille, France and ³CPASS, directeur de la recherche, Faculté de Médecine, Université de Montréal, CP 6128, Succursale centre-ville, Montréal, Québec, H3C 3J7, Canada

Email: Jean Paul Fournier - Fournier.jp@chu-nice.fr; Anne Demeester - Anne.Demeester@mail.ap-hm.fr; Bernard Charlin^{*} - bernard.charlin@umontreal.ca

^{*} Corresponding author

Published: 6 May 2008

BMC Medical Informatics and Decision Making 2008, 8:18 doi:10.1186/1472-6947-8-18

This article is available from: <http://www.biomedcentral.com/1472-6947/8/18>

Received: 28 November 2007

Accepted: 6 May 2008

How to construct and implement script concordance tests: insights from a systematic review

Valérie Dory,^{1,2} Robert Gagnon,³ Dominique Vanpee^{2,4} & Bernard Charlin³