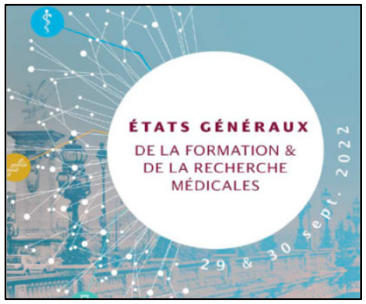


Tout ce que vous avez voulu savoir sur la docimologie des TCS,…en osant le demander

L. Sibert (Rouen), JP. Fournier (Nice)

Les TCS ne sont appréciés ni des étudiants…

Students' feedbacks	*N (%)*
SCT are adapted to graduate medical students	9 (10%)
SCT are adapted to post-graduate medical students	7 (8%)
SCT are adapted to doctors for continuing medical education	4 (4.5%)
The principle of SCT is excellent	1 (1%)
SCT are discriminant	1 (1%)
SCT are interesting in theory but not in practice	21 (24%)
Too high variability between experts 'responses	27 (30.5%)
SCT are too ambiguous, not clear enough	22 (24%)
SCT are not adapted to graduate medical students	13 (15%)
Inadequacy is felt between obtained scores and skills / knowledge	9 (10%)
insufficient students preparation	8 (7%)
SCT are useless	6 (6.5%)
SCT are confusing	4 (4.5%)
SCT are too subjective	4 (4.5%)
Frustrating because no possibility to justify one's answer	3 (3.5%)
The principle of SCT is bad	3 (3.5%)
SCT are too difficult	3 (3.5%)
Inadequacy is felt between SCT experts answers and national referential about the subject	2 (2%)
Lack of detailed correction	1 (1%)
SCT are not discriminant	1 (1%)

D'après Kün-Darbois JD. BMC Med Educ 2021

…Ni des enseignants

Teachers' feedbacks	*N (%)*
SCT are adapted to post-graduate medical students	1 (9%)
SCT need to be developed	1 (9%)
SCT are useful only once knowledge is acquired	1 (9%)
Difficult to write questions	3 (27%)
SCT are not satisfying	2 (18%)
Too high variability between experts 'responses	1 (9%)
SCT are too ambiguous, not clear enough	1 (9%)
SCT are not adapted to graduate medical students	1 (9%)
Difficult to recruit a sufficient number of experts	1 (9%)
SCT are confusing	1 (9%)
SCT prevent students from good medical reasoning	1 (9%)

D'après Kün-Darbois JD. BMC Med Educ 2021

A juste titre !

Conceptuellement, les TCS ont au moins 3 limites intrinsèques qui limitent leur utilisation à grande échelle pour des examens d'enjeux importants (EDN) :

- Validité de contenu et établissement des scores
- Stabilité des scores (*reliability*)
- Processus de réponse

D'après Lineberry M. *Med Educ* 2013

Etablissement des scores

- ❖ *Agregate scoring* :
- ✓ Base conceptuelle du TCS
- ✓ Remplacer *l'agregate scoring* par des réponses consensuelles :
  - Perte de discrimination
  - *Intermediate state*
- ❖ Evaluation du raisonnement clinique des étudiants en contexte d'incertitude :
- ✓ Objectif très important de la formation des étudiants
- ✓ Défaut de gestion de l'incertitude et conséquences

- ❖ Comment y remédier ?
- ✓ Construction des TCS
- ✓ Composition du panel

- ✓ Elimination des réponses « déviantes » ?

D'après

Lubarsky S. *Med Educ* 2011
Williams RG. *Acad Med* 2011
Fournier JP. *BMC Med Inform and Decis Making* 2008
Lubarsky S. *Med Teacher* 2013
Sibert L., Fournier JP. Epreuves TCS – Le guide méthodologique, Maloine, Paris , 2015
Gagnon R. *Adv in Health Sci Educ Theory Pract* 2011
Simpkin AL. *N Engl J Med* 2016
Cooke A. *Acad Med* 2017

Stabilité des scores

Schématiquement, 3 façon de l'appréhender :

- Estimation de la cohérence interne de l'épreuve : KR20 ou score de Cronbach
- Mesure des composantes de la variance des scores (généralisabilité)
- Test-retest

D'après Bertrand R, Blais JG. Modèles de mesures. L'apport de la théorie de réponse aux items. Presses Universitaires du Québec, Sainte Foy, 2004
Tavakol M. *Med Teacher* 2011
Downing SM. *Med Educ* 2004
Schmitt N. *Psychol Assess* 1996
Cortina JM. *Appl Psychol* 1993

TCS et scores de fidélité

- Cronbach :

0,70 à 0,90 (KFP : 0,45 à 0,95)

- Généralisabilité :

G :0,554 à 0,88 (KFP : 0,579 à 0,83 et 0,45 à 0,73)

- Test-retest :

Rares publications
Effectifs limités
Thématiques variées
Méthodologie différente (wash-out, tests utilisés : corrélation, concordance-corrélation, test Kappa, taille de l'effet)
KFP : taille de l'effet : 0,48

- Groupe fixe d'étudiants et plusieurs groupes de panelistes : scores constants

D'après
Lubarsky S. *Med Educ* 2011
Hrynchak P. *Med Educ* 2014
Brailowsky C. *Med Educ* 2001
Gagnon R. *Adv in Health Sci Educ* 2009
Ramaekers S. *Ass Eval Higher Educ* 2010
Huwendiek S. *Med Teacher* 2017
Holloway R. *Med Educ* 2004
Bland AC. *Acad Med* 2005
Park AJ. *Am J Obstet Gynecol* 2010
Giguère A. *Pat Educ Counsel* 2012
Dufour S. *JVME* 2012
Nickendei C. *Med Teacher* 2009
Gagnon R. *Adv in Health Sci Educ Theory Pract* 2011

Processus de réponse

❖ 3 limites objectives :

✓ Choix de la valeur neutre
✓ Choix des valeurs extrêmes (-2 et +2)
✓ Réponses au hasard

❖ Mais des solutions disponibles :

✓ Limitation de la valeur neutre comme valeur modale
✓ Valeurs modales sur les valeurs extrêmes
✓ Ancrage « générique »
✓ Expliciter la signification de la valeur neutre
✓ Instructions aux étudiants et panelistes

D'après Bland AC. *Acad Med* 2005
Lubarsky S. *Perspect Med Educ* 2018
Lubarsky S. *Med Teacher* 2013
See KC. *Med Educ* 2014
Van den Broeck M. *Perspect Med Educ* 2012
Gawad N. *Med Educ* 2021

Et dans la vraie vie ?

Test TCS nationaux 2021

- 2 épreuves administrées en ligne en avril-mai 2021 avec 1 *wash-out* de 3 semaines minimum
- Etudiants de DFGSM 3 sur la base du volontariat
- Sémiologie dans 3 domaines : Cardiologie, Pneumologie et Hépato-Gastro-Entérologie
- Construction selon les recommandations établies par 2 enseignants
- Valeurs neutres limitées
- Valeurs extrêmes systématiquement retenues
- 2 épreuves construites selon la même table de spécification (répartition des 3 disciplines, thématiques, images, biologie)
- 10 vignettes (30 qiuestions) communes aux 2 tests
- Instructions diffusées aux étudiants et aux panelistes :
  - ✓ Signification de la valeur « 0 »
  - ✓ Toutes les valeurs sont possibles, notamment « -2 » et « +2 »
  - ✓ Les réponses au hasard génèrent des scores plus faibles
- 2 groupes de panelistes :
  - ✓ Panélistes spécialistes d'organe
  - ✓ Panélistes « polyvalents »
- Temps de composition relevé

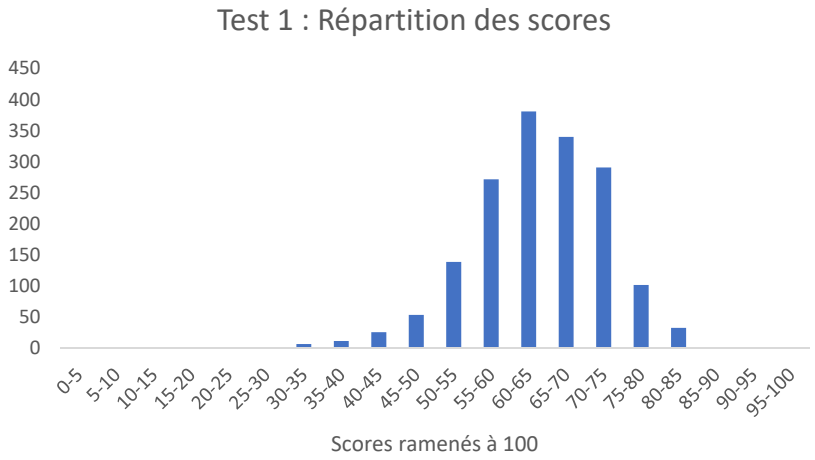
Tests nationaux 2021

- 2 épreuves de 28 vignettes et 84 questions chacune
- 9 vignettes de Cardiologie et d'HGE et 10 de Pneumologie
- Analyse uniquement des données des étudiants ayant répondu à toutes les questions

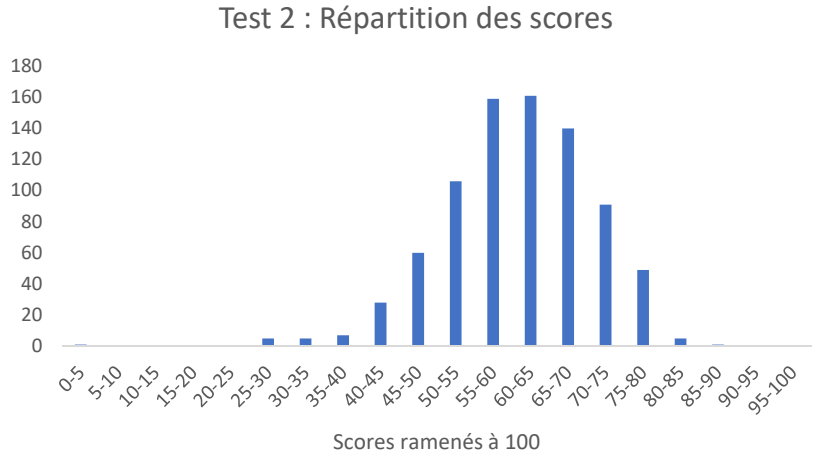
Résultats

- Test 1 et 2 : aucune question avec un impact négatif sur la fidélité du test : toutes les vignettes et questions ont été conservées
- Temps moyen de composition : 50,03 ± 25,17 minutes
- Test 1 : 1858 étudiants, 1661 analysables
- Test 2 : 1122 étudiants, 819 analysables
- Vignettes et questions communes (test-retest) : 427 analysables
- Test 1 : 42 ± 3 panelistes par question
- Test 2 : 37 ± 3 panelistes par question

Résultats



Score moyen :	60,71 ± 9,38
Cronbach :	0,7911
Index de facilité :	0,44 ± 0,13
Index de discrimination :	0,22 ± 0,13



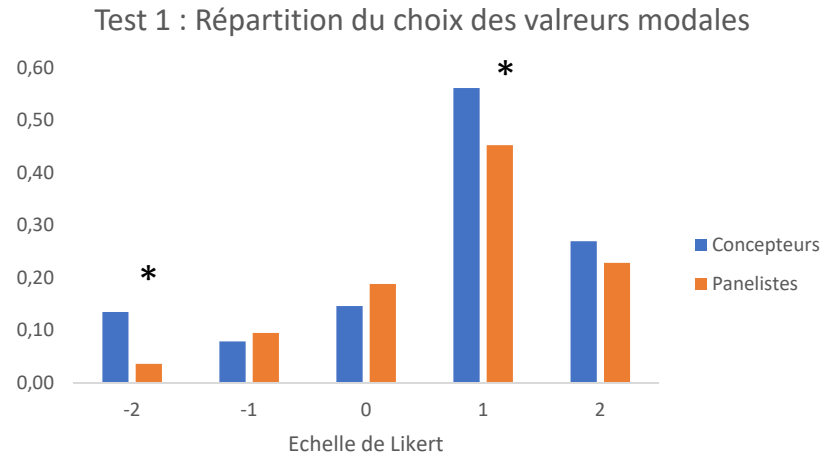
Score moyen :	60,78 ± 9,95
Cronbach :	0,8424
Index de facilité :	0,44 ± 0,14
Index de discrimination :	0,26 ± 0,18

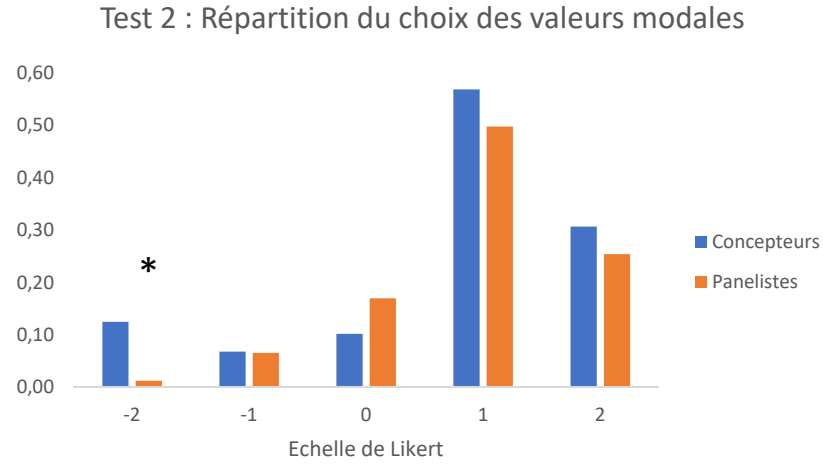
Résultats

- Très peu de différence de choix

Très peu de difficulté à comprendre ce qu'attendaient les concepteurs

- Plus faible proportion de choix des valeurs extrêmes.

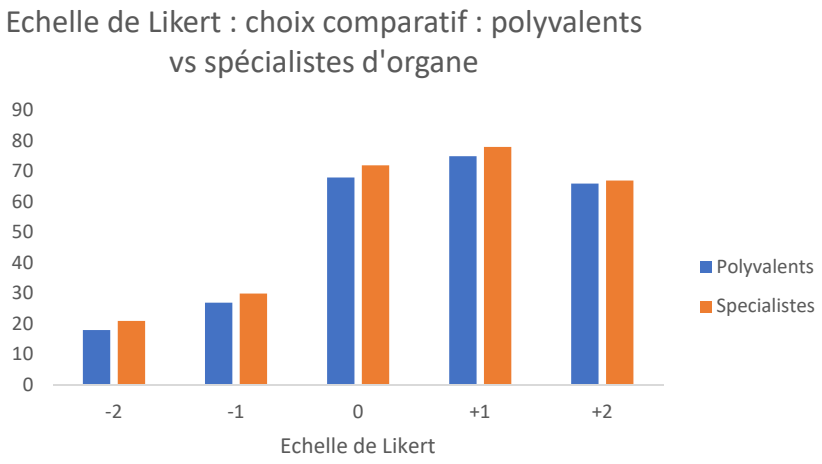


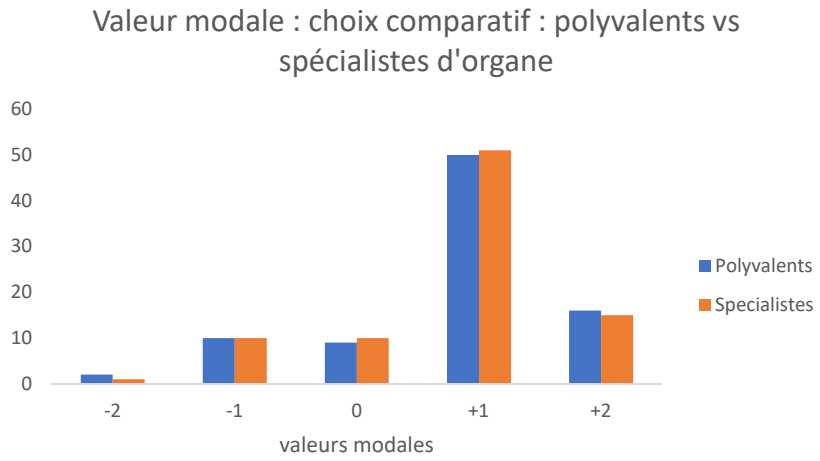


* p < 0.05

Polyvalents *vs* spécialistes d'organe

Test 1





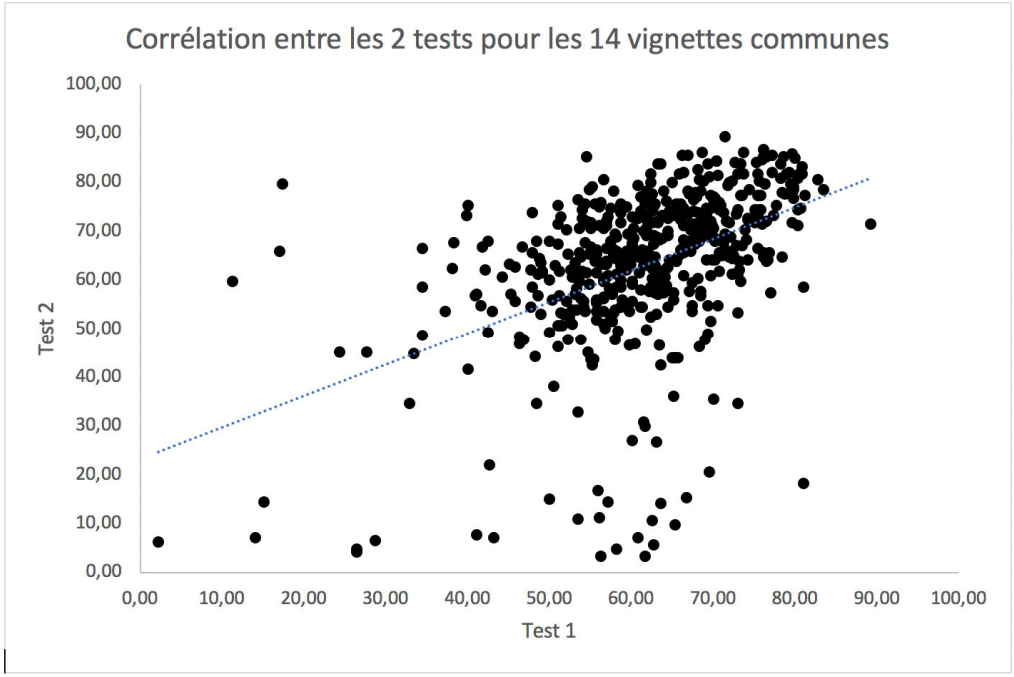
Scores attribués par les 2 groupes de panélistes

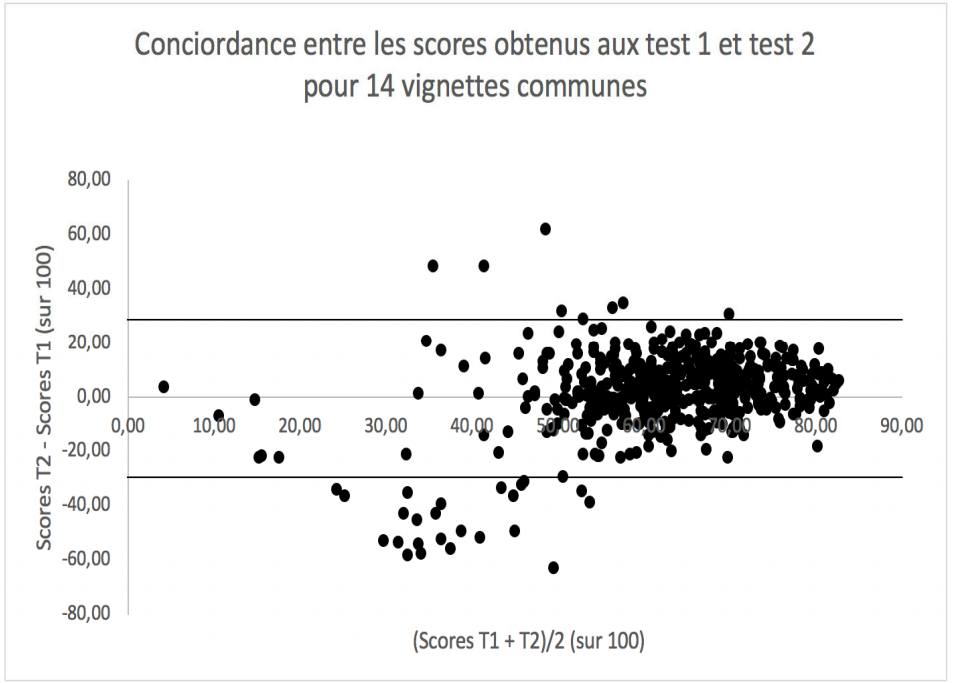
	Test 1		Test 2	
	Polyvalents	Spécialistes	Polyvalents	Spécialistes
Total	61,94 ± 8,38	64,53 ± 8,54	63,27 ± 8,93	65,96 ± 8,99
Cardiologie	60,08 ± 9,61	62,20 ± 9,88	60,96± 10,63	62,48 ± 10,15
HGE	60,10 ± 10,26	66,35 ± 9,93	62,84 ± 9,86	69,08 ± 10,11
Pneumologie	65,27 ± 10,46	65,00 ± 10,61	65,74 ± 10,96	66,28 ± 11,06

- Toutes les différences sont significative (sauf Pneumologie au test 1) avec des corrélations 2 à 2 très significatives
- Pas d'impact sur les taux de réussite et la proportion d'ex-aequos

Test-retest

Scores moyens : 63,78 ± 10,05 *vs* 63,76 ± 11, 15, p = NS





Résultats

- Impact du choix systématique de la valeur neutre(« 0 »)
- Choix moyen : 12 ± 6
- Choix moyen + 1SD : ≥ 19
- Choix moyen + 2 SD : ≥ 25

	moyenne + 1 SD	Moyenne + 2 SD	p
Etudiants (%)	189 (0,11)	68 (0. ,4)	
Choix	24 ± 6	30 ± 9	
Score si moyenne + DS (/ 100)	57,92 ± 10,97	55. ,1 ± 11,50	NS
Score si < moyenne + DS (/ 100)	61,06 ± 9,09	60,87 ± 9,25	NS
p	0,0021	0,0002	

Conclusion

- C’est réalisable (mais ancienne plateforme…)
- TCS utilisables chez des étudiants en début de parcours
- Très peu de différences entre le choix des panelistes et les valeurs attendues par les concepteurs, sur les valeurs extrêmes
- Pour des tests polyvalents les scores attribués par les panelistes polyvalents et spécialistes d’organes sont très significativement corrélés
- Pour des tests polyvalents, les panelistes spécialistes d’organe donnent des scores plus élevés, mais sans impact
- D’un point de vue logistique, les panelistes spécialistes d’organe constituent un meilleur choix (contraintes de temps, collèges, etc…)
- La stratégie « 0 » n’est pas rentable
- La préparation du test est fondamentale, vignettes, questions, instructions.