

Intégration du degré de certitude dans l'évaluation des connaissances des étudiants en médecine d'urgence

Use of the Degree of Certainty in the Assessment of the Emergency Medical Students' Knowledges

C.H. Houzé-Cerfon · D. Lauque · S. Charpentier

Reçu le 26 juillet 2016 ; accepté le 28 octobre 2016
© SFMU et Lavoisier SAS 2016

Résumé En médecine d'urgence, les connaissances doivent être maîtrisées avec un niveau élevé de certitude afin d'être mobilisables rapidement en situation critique. Le degré de certitude qu'estiment les étudiants dans leurs connaissances peut être mesuré au cours de la formation par un questionnaire « spectral » qui distingue les connaissances utilisables (sûres et correctes), les connaissances dangereuses (sûres et incorrectes), et les connaissances non utilisables car incertaines.

Objectif : Nous avons étudié le niveau de certitude des connaissances lors de l'évaluation formative d'un enseignement sur la lecture de l'électrocardiogramme (ECG).

Méthode : Vingt-neuf médecins inscrits au DESC de médecine d'urgence à l'université Paul Sabatier de Toulouse étaient évalués avant et après un atelier de formation sur les tracés ECG de syndrome coronarien aigu. Les étudiants répondaient à dix questions à choix multiples en notant leur niveau de certitude. La distribution spectrale des réponses en fonction du niveau de certitude était calculée pour l'ensemble des réponses et par QCM. Les analyses bivariées des variables quantitatives et qualitatives étaient réalisées respectivement sur échantillon apparié par le test de rang de Wilcoxon et par le test de Mac Nemar.

Résultats : La médiane des notes augmentait de 6/20 (intervalle interquartile 6-8) à 12/20 (10-14) avant et après l'atelier, respectivement ($p < 0,001$). Le pourcentage de connaissances utilisables augmentait de 24 % à 55 %, celui de réponses incertaines diminuait de 42 % à 7 % et celui de réponses dangereuses passait de 34 % à 38 % après la séance, respectivement. Les connaissances étaient améliorées mais également le niveau de confiance des apprenants

dans ces connaissances ; une part importante des notions dans lesquelles les apprenants avaient confiance restait néanmoins erronée.

Conclusion : Le niveau de certitude peut être mesuré lors de l'évaluation formative en médecine d'urgence. Il permet aux étudiants de s'assurer que leur niveau de confiance s'améliore et de cibler les connaissances à revoir prioritairement.

Mots clés Certitude · Métacognition · Évaluation · Prise de décision

Abstract In emergency medicine, the knowledge must be mastered with a high degree of certainty to be mobilized quickly in critical situations. The degree of certainty taken by the students in their knowledge can be measured during training by a "spectral" questionnaire which separates usable knowledge (safe and correct), dangerous knowledge (safe and incorrect), and the knowledge not usable because uncertain.

Objective: The aim of this study is to evaluate the interest of degrees of certainty in learning ECG in emergency medicine.

Methods: Twenty nine students in emergency medicine were evaluated before and after a training workshop on the ECG of acute coronary syndrome. Students answered 10 questions with their degree of certainty. The spectral distribution of the answers of the confidence level was calculated for all the answers and questions. The medians were compared with a Wilcoxon test and the percentages with a Mac Nemar test.

Results: The median score increased from 6/20 (interquartile range (IQR) 6-8) to 12/20 (IQR 10-14) before and after the learning, respectively ($p < 0,001$). The percentage of knowledge usable increased from 24% to 55%, and uncertain answers decreased from 42% to 7% and that of dangerous answers increased from 34% to 38% after the session, respectively.

Conclusion: The degree of certainty can be measured during the assessment in emergency medicine. He allows the

C.H. Houzé-Cerfon (✉) · D. Lauque · S. Charpentier
Pôle de médecine d'urgence, CHU Toulouse,
F-31059 Toulouse cedex 9
e-mail : houze-cerfon.ch@chu-toulouse.fr

D. Lauque · S. Charpentier
Université Paul Sabatier, F-31059 Toulouse cedex 9

students to make sure that their reliable level improves and to target the knowledge to be firstly seen again.

Keywords Certainty · Metacognition · Assessment · Decision making

Introduction

Les étudiants en médecine sont évalués régulièrement au cours de leur apprentissage pour s'assurer que les savoirs sont acquis, interprétés correctement et qu'ils permettent de prendre les décisions appropriées pour soigner les malades [1]. L'évaluation ne doit toutefois pas se limiter à l'attribution d'une note validant un examen. Elle a également un rôle formateur en indiquant aux étudiants les connaissances acquises et en facilitant le recentrage de leur processus d'apprentissage vers les savoirs non maîtrisés [2]. L'évaluation dite formative est selon Perrenoud [3] une évaluation qui aide l'élève à apprendre et à se développer, autrement dit qui participe à la régulation des apprentissages et du développement dans le sens du projet éducatif.

Le degré de certitude des réponses des étudiants devrait être pris en compte car des réponses pourtant correctes peuvent correspondre à des connaissances incertaines qui limiteront la prise de décision ou à des connaissances erronées considérées à tort comme exactes et certaines [4]. Le niveau de certitude permet de distinguer les connaissances sûres et correctes (connaissances assurées) qui sont utilisables, celles qui sont sûres et incorrectes (ignorance ignorée) potentiellement dangereuses et celles qui sont incertaines (correcte ou incorrecte) et qui nécessiteront une vérification externe pour être utilisées. Cette dernière catégorie regroupe la connaissance partielle (incertaine et correcte) et l'ignorance reconnue (incertaines et incorrectes) [4,5]. Cette démarche paraît essentielle dans l'apprentissage de métiers où les risques vitaux sont évidents, tels qu'en médecine d'urgence. Le niveau de certitude des connaissances doit y être élevé pour savoir prendre rapidement les bonnes décisions.

L'étudiant peut ainsi réaliser une démarche métacognitive [6] en prenant conscience que des réponses exactes peuvent correspondre à des connaissances potentiellement inutilisables ou nuisibles [7]. Leclercq et Poumay donnent de la métacognition une définition proche de celle de l'apprentissage autorégulé (*Self Regulated Learning*) : « Un ensemble d'opérations (jugements, analyses, régulations) menées à différents moments (pré- per- et post-) de situations (d'apprentissage ou d'évaluation) et portant sur deux types d'objets (ses propres processus ou ses propres productions) ». Les degrés de certitude constituent le volet « jugements » de l'autorégulation. Il s'agit ensuite pour chaque étudiant de passer à l'opération d'analyse afin d'identifier ces connaissances incertaines et ces

connaissances erronées. Une fois les causes comprises, l'étudiant pourra réguler ces apprentissages (réapprentissage, désapprentissage de ce qui est faux, consolidation, etc.).

Un questionnaire « spectral » permet de mesurer les degrés de certitude accompagnant les réponses à des questions [8]. Ce questionnaire a été testé chez des étudiants [9,10] en utilisant des questions à choix multiples (QCM) lors de pré-tests et post-tests. Ce questionnaire a également été utilisé dans l'éducation thérapeutique des patients diabétiques [11,12]. L'objectif de cette étude était d'observer l'amélioration des connaissances et de la confiance dans ces connaissances lors de l'évaluation formative d'un enseignement sur la lecture de l'électrocardiogramme (ECG).

Méthodes

Trente-trois internes et assistants, inscrits en 1^{re} et 2^e année du DESC de médecine d'urgence à l'université Paul Sabatier de Toulouse (Midi-Pyrénées), participaient à un atelier sur la lecture de l'ECG. Quatre ne répondaient pas au questionnaire pré- ou post-test et étaient exclus de l'analyse. Les résultats des 29 autres étudiants sont ici présentés. Tous avaient une expérience pratique d'au moins un an en médecine d'urgence et avaient déjà suivi pendant leurs études un enseignement sur l'analyse de l'ECG.

L'enseignement sur la lecture de l'ECG était réalisé pendant un séminaire de trois jours sur les urgences cardiologiques, respiratoires et neurologiques. Il comprenait un rappel magistral d'une heure sur l'interprétation de l'ECG normal puis un atelier de 2h30 sur l'interprétation d'ECG présentant des troubles de la repolarisation évocateurs d'ischémie myocardique. L'objectif de la formation était de savoir reconnaître les tracés de syndromes coronariens aigus, en particulier ceux avec sus-décalages du segment ST, et les diagnostics différentiels. Les étudiants interprétaient collectivement 50 tracés en suivant une grille de lecture systématique, en présence de l'enseignant qui commentait les réponses et expliquait les ECG présentés.

Avant le début de l'enseignement, le pré- et post-test et la mesure du degré de certitude étaient expliqués aux étudiants pendant 20 minutes environ. Le test comprenait dix QCM comportant chacun cinq propositions correctes ou incorrectes. Les trois premiers QCM présentaient des tracés ECG normaux, les sept suivants des troubles de la repolarisation. L'ensemble des QCM étaient réalisés selon les règles proposées par Guilbert dans son Guide pédagogique pour les professionnels de santé, OMS [13]. Le temps imparti pour le pré- ou le post-test était de 15 minutes. Les étudiants répondaient à chaque QCM et indiquaient le degré de certitude de leur réponse en utilisant une échelle à six niveaux allant de 0 % (doute maximal) à 100 % (certitude totale). Cette échelle était limitée à six niveaux car des niveaux plus nombreux

n'ajoutent pas d'élément supplémentaire pour classer les informations cognitives dans les domaines de l'attention et de la perception en particulier sur les bas niveaux de certitude [14,15].

Après le post-test, l'enseignant présentait les réponses aux étudiants, puis chaque étudiant corrigeait lui-même en séance ses QCM du pré- et post-test en utilisant une grille. Les QCM étaient corrects si les réponses à toutes ses propositions étaient correctes, incorrects lorsqu'au moins une des réponses était incorrecte. La distribution spectrale des réponses était calculée pour chaque QCM en fonction de son exactitude et de son degré de certitude. Les réponses dont le degré de certitude était ≥ 60 % étaient considérées comme liées à des connaissances utilisables si elles étaient correctes ou comme liées à des connaissances dangereuses si elles étaient incorrectes. Celles dont le degré de certitude était < 60 % étaient jugées difficilement utilisables. Ces qualificatifs étaient choisis pour que les étudiants comprennent l'impact potentiellement délétère de connaissances erronées ou peu sûres sur leur pratique clinique. L'enseignant ne disposait pas pour son enseignement des résultats du pré-test qui était corrigé avec le post-test en fin de séance.

Ultérieurement, après vérification des résultats par l'enseignant, la note moyenne de chaque étudiant au pré- et post-test était calculée en attribuant deux points aux réponses exactes et zéro point aux réponses inexactes. Des points dits de métacognition liés au degré global de certitude des réponses étaient ensuite ajoutés à ces notes [4,10,11] : un point de confiance était ajouté à la moyenne lorsque le degré de certitude moyen avec les réponses correctes (CMRC) était ≥ 60 %, un point de prudence lorsque le degré de certitude moyen avec les réponses incorrectes (CMRI) était < 60 %. Un point supplémentaire de discrimination était ajouté lorsque la différence entre CMRC et CMRI était > 20 %.

Une information sur le protocole de recherche était donnée aux étudiants et leur consentement obtenu pour l'utilisation des données anonymisées.

Méthodologie statistique

Les variables quantitatives étaient décrites en médiane et intervalle interquartile (IIQ). Les variables qualitatives

étaient décrites en termes d'effectif et de pourcentage. L'hypothèse testée était l'augmentation des réponses correctes et certaines après la formation. Les notes des étudiants étaient comparées par un test de rang de Wilcoxon (série appariée). Les niveaux de connaissances en fonction des degrés de certitude étaient comparés par le test de Mac Nemar (série appariée). Un $p < 0,05$ était considéré comme significatif. Les analyses statistiques étaient réalisées avec le logiciel STATA 12[®].

Résultats

Les résultats globaux du pré- et post-test sont présentés dans le Tableau 1. La médiane des réponses correctes des 29 étudiants aux QCM du pré-test était 6/20 (IIQ 6-8) et correspondait à 97 QCM corrects parmi un total de 290. La note des réponses correctes était majorée d'un point de prudence pour 11 étudiants, d'un point de discrimination pour 9 et d'un point de confiance pour 21 étudiants. La médiane pondérée par le degré de certitude augmentait à 8/20 (IIQ 7-9).

Après l'enseignement, les notes des réponses correctes augmentaient significativement par rapport au pré-test ($p < 0,001$) avec une médiane à 12/20 (IIQ 10-14) et correspondait à 165 réponses correctes parmi 290. Ainsi, la progression la plus notable concernait le point de confiance qui concernait la totalité des étudiants après le test. La note moyenne des 29 étudiants, pondérée par ces points de bonification, augmentait significativement comparativement à celle du pré-test ($p < 0,001$) avec une médiane de 13/20 (11-16).

La représentation spectrale des résultats visualise la répartition des réponses correctes (P) ou incorrectes (M) aux 290 QCM en fonction de leur niveau de certitude (Fig. 1). Au pré-test, les réponses incertaines, correctes ou incorrectes, dont le niveau de certitude était inférieur à 60 % étaient les plus nombreuses (121 soit 42 %). Quarante-deux réponses incorrectes (34 %) étaient considérées comme dangereuses car énoncées avec un niveau de certitude supérieur ou égal à 60 %. Seulement 70 (24 %) réponses correctes présentaient un niveau de certitude supérieur ou égal à 60 % et étaient considérées comme utilisables. Après l'enseignement, leur nombre augmentait à 161 (55 %). Celui des

Tableau 1 Résultats des notes des 29 étudiants au pré- et post-test, sans et avec les points de métacognition.

	Note médiane (IIQ)	Points de confiance	Points de prudence	Points de discrimination	Note pondérée médiane (IIQ)
Nombre d'étudiants					
Pré-test	6 (6-8)	21	11	9	8 (7-9)
Post-test	12 (10-14)*	29	2	7	13 (11-16)*

* $p < 0,001$; IIQ : intervalle interquartile.

réponses non utilisables diminuait à 20 (7 %), mais le nombre de réponses dites dangereuses augmentait légèrement à 109 (38 %). La proportion de réponses correctes parmi celles qui bénéficiaient d'un haut niveau de certitude augmentait de 41 % lors du prétest à 60 % après la formation.

L'analyse des niveaux de certitude des réponses à chaque QCM est présentée dans le Tableau 2. L'enseignement permettait d'améliorer les connaissances utilisables et de diminuer celles inutilisables pour la quasi-totalité des QCM (Tableau 2). À l'inverse, les réponses dangereuses aux

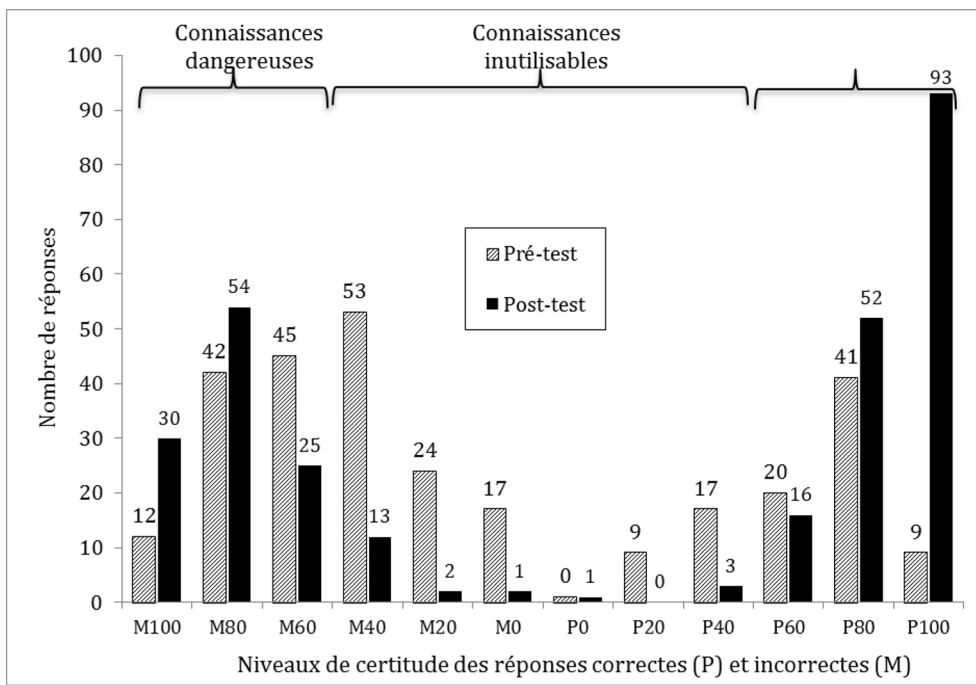


Fig. 1 Distribution spectrale du degré de certitude des réponses au pré-test et au post-test. Les pourcentages de certitude précédés de la lettre P correspondent aux réponses correctes (hémispectre droit), ceux précédés de la lettre M correspondent aux réponses incorrectes (hémispectre gauche). Par exemple, le nombre de réponses correctes avec un niveau de certitude de 80 % était de 41 au pré-test puis 52 au post-test

	Réponses					
	Dangereuses		Inutilisables		Utilisables	
	Pré-test	Post-test	Pré-test	Post-test	Pré-test	Post-test
Q1	13	9	4	0	12	20
Q2	12	20	17	5	0	4
Q3	3	3	25	0	1	26
Q4	7	15	17	1	5	13
Q5	12	12	16	2	1	15
Q6	17	16	5	2	7	11
Q7	8	8	10	1	11	20
Q8	5	3	11	2	13	24
Q9	8	17	12	6	9	10
Q10	14	6	4	1	11	22
Total	99 (34 %)	109 (38 %)	121 (42 %)	20 (7 %)	70 (24 %)	161 (55 %)
	p<0,001		p<0,001		p<0,001	
Test de Mac Nemar ; Q : question.						

QCM 2, 4 et 9 augmentaient lors du post-test et concernaient plus de la moitié des étudiants (respectivement 20, 15 et 17 étudiants).

Discussion

Nous avons étudié l'impact d'un atelier de formation à la lecture de l'ECG sur l'amélioration des connaissances d'étudiants du DESC de médecine d'urgence en intégrant le degré de certitude des réponses. Le niveau global des connaissances et de la confiance liée à ces connaissances progressait nettement après la formation. La formation augmentait largement le nombre de réponses exactes et sûres, correspondant à des connaissances qualifiées d'utilisables.

La mesure du niveau de certitude des connaissances est un élément important dans l'apprentissage. La confiance doit être élevée pour que les connaissances soient utilisées dans la pratique réelle [8]. L'intégration du niveau de certitude dans les évaluations permet aux étudiants de construire un apprentissage basé sur une démarche réflexive et autonome. Ils prennent ainsi conscience non seulement des savoirs acquis mais également du niveau de confiance qu'ils ont dans ces savoirs [11]. Les résultats de l'évaluation leur permettent de savoir lesquels sont acquis avec un niveau de confiance élevé, et lesquels devront être améliorées prioritairement après la formation.

Cette démarche paraît adaptée à la médecine d'urgence car les décisions rapides et fréquentes nécessitent un niveau de confiance élevé dans les connaissances mobilisées. Une anomalie évocatrice d'un syndrome coronarien aigu sur l'ECG doit être reconnue rapidement car elle permet de débiter sans tarder, dans un contexte clinique évocateur, une prise en charge adaptée et d'améliorer le pronostic. L'objectif de la formation était d'acquérir une expertise sur la lecture des troubles de la repolarisation, en insistant sur le sus-décalage du segment ST et l'ischémie myocardique. Les étudiants qui participaient à cet enseignement étaient susceptibles d'être confrontés à ces situations cliniques. Ils pouvaient ainsi mieux prendre conscience que des connaissances considérées à tort comme acquises étaient en réalité incertaines voire potentiellement dangereuses, même au terme de la séance d'enseignement. Cela doit leur permettre de mettre en place, avec l'aide des enseignants, des stratégies correctives telles que la prise de décision partagée en pratique clinique. La confiance dans les connaissances pourrait être étudiée dans d'autres urgences que la douleur thoracique, telles que l'arrêt cardiorespiratoire, au cours duquel les décisions sont trop urgentes pour permettre de les vérifier.

Pour l'enseignant, l'amélioration significative des notes permettait de considérer la séance comme formatrice et adaptée à l'objectif. La prise en compte des niveaux de certitude confirmait cette impression favorable puisque de nom-

breuses réponses incertaines devenaient certaines et utilisables après l'enseignement. Ce constat devait toutefois être nuancé par le fait que le nombre de réponses incorrectes que les étudiants considéraient à tort comme correctes avec un niveau élevé de certitude augmentait après l'enseignement. Elles concernaient surtout trois questions (QCM 2, 4 et 9) et plus de la moitié des étudiants à chacune de ces trois questions. Ces réponses qualifiées de potentiellement dangereuses auraient dû disparaître au profit des réponses utilisables si l'objectif de l'apprentissage avait été pleinement atteint. [4] Le QCM 2 reprenait les règles d'interprétation d'une repolarisation normale. Tous les items n'ont pas été assimilés par les étudiants. Pour les QCM 4 et 9, les étudiants ont tous interprété correctement le trouble de la repolarisation comme un syndrome coronarien aigu (SCA) mais ne l'ont pas relié au territoire coronarien suspecté lésé. L'objectif était surtout qu'ils diagnostiquent le SCA mais les étudiants avaient des connaissances erronées sur les correspondances entre les territoires électrocardiographiques et les lésions coronariennes suspectées. Le doute constaté évoque des possibilités importantes de réapprentissage et de remédiation ultérieure [12]. Sur le plan docimologique, le fait d'avoir choisi une réponse binaire aux QCM à choix multiples (tous les items justes vs un item faux) ne permet pas d'avoir la valeur réelle de l'étudiant au regard de l'objectif d'apprentissage attendu. Dans ces cas, l'enseignant doit revoir la pondération des QCM et leur construction afin qu'ils correspondent aux objectifs de l'enseignement. Il doit aussi revoir son enseignement en ciblant les objectifs qui n'ont pas été acquis, en particulier auprès des étudiants dont les connaissances erronées persistaient à la fin de l'enseignement (ici sur la correspondance entre territoire électrocardiographique et lésion coronarienne suspecté).

L'ajout de points de confiance, de prudence et de discrimination aux notes permettait d'intégrer la dimension métacognitive à l'évaluation. Les points de confiance augmentaient largement, alors que ceux de prudence diminuaient, confirmant l'acquisition de connaissances sûres. Une pondération positive des réponses utilisables et négative des réponses dangereuses pourrait être ajoutée dans le cadre d'une évaluation normative des étudiants. Ce type d'évaluation peut être réalisé dans des grands groupes d'étudiants car elle nécessite peu de ressources humaines et matérielles.

Notre travail présente plusieurs limites. L'évaluation du niveau de certitude des connaissances est peu répandue. Elle a été utilisée dans l'éducation thérapeutique mais peu ou pas dans l'apprentissage de la médecine, à notre connaissance [9]. Notre travail mériterait d'être testé dans d'autres situations cliniques et par d'autres enseignants. La phase socle du prochain DES de médecine d'urgence prévoit une évaluation des connaissances par QCM et pourrait permettre de tester plus largement la dimension métacognitive. Une autre limite tient au choix arbitraire des seuils entre les connaissances

utilisables, non utilisables et dangereuses ont été fixés à 60 %. Le seuil des connaissances certaines aurait pu être placé plus proche de 100 % en raison de l'exigence de sécurité concernant des détroites potentiellement vitales. Le niveau de certitude des réponses exactes lors du post-test était cependant très majoritairement supérieur à 80 %. Enfin, les étudiants n'étaient pas interrogés sur leur satisfaction concernant cette méthode originale d'évaluation. Ils notaient toutefois que cet atelier avait été l'un des plus intéressants du séminaire et en soulignaient son interactivité.

Conclusion

En conclusion, la mesure du niveau de certitude peut être utilisée lors de l'évaluation formative en médecine d'urgence. Elle permet aux étudiants de s'assurer que leur niveau de confiance dans leurs connaissances s'améliore et de cibler celles à revoir prioritairement. Cette méthode d'évaluation pourrait être diffusée si ces résultats sont confirmés dans d'autres situations cliniques et par d'autres enseignants.

Liens d'intérêts : Les auteurs déclarent ne pas avoir de lien d'intérêt.

Références

- Jouquan J (2002) L'évaluation des apprentissages des étudiants en formation initiale. *Pédagogie Med* 3:38–52
- Legendre R (1993) Dictionnaire actuel de l'éducation. Guérin-ESKA, Ed. Montréal-Paris
- Perrenoud P (1999) L'évaluation des élèves. De la fabrication de l'excellence à la régulation des apprentissages. Porto Alegre (Brésil), Artmed Editora Ed. De Boeck
- Hunt D (1993) Human self-assessment: theory and application to learning and testing. In: Leclercq D, Bruno J. *Item Banking: Interactive Testing and Self-Assessment*. NATO ASI Series, Springer Ed, Berlin pp 177-89
- Shuford EH (1993) Pursuit of the fallacy: resurrecting the penalty. In: Leclercq D, Bruno J. *Item Banking: Interactive Testing and Self-Assessment*. NATO ASI Series. Springer Ed, Berlin, pp 76-98
- Lafortune L (2007) Pour guider la métacognition. Collection éducation - intervention. Presse de l'université du Québec, Ed. Québec, p 11
- Gombert JE (1990) Le développement métalinguistique. PUF, Paris, p 27
- Leclercq D, Poumay M (2008) La métacognition. In: Leclercq D. *Psychologie éducationnelle de l'adolescent et du jeune adulte*. Éditions de l'Université de Liège, Ed. Liège, pp 203-63
- Delangue C, Wiel E, Ammirati C, Assez N (2011) Évaluation préliminaire de l'impact d'un dispositif de formation aux gestes et soins d'urgence. *Pédagogie Med* 12:103–19
- Leclercq D (2003) Diagnostic cognitif et métacognitif au seuil de l'université. Le projet Mohican mené par les 9 universités de la Communauté Française Wallonie Bruxelles. Éditions de l'Université de Liège, Ed. Liège
- De Finetti B (1965) Methods for discriminating levels of partial knowledge concerning a test item. *Br J Math Stat Psy* 18:87–123
- Bruttomesso D, Leclercq D, Gagnayre R, et al (2003) The use of degrees of certainty to evaluate knowledge. *Patient Educ Couns* 51:29–37
- Guilbert JJ (1981) Guide pédagogique pour les personnels de santé. Édition révisée OMS. Genève
- Leclercq D, Poumay M (2003) La connaissance partielle chez l'apprenant : pourquoi et comment la mesurer. *Educ Ther Patient* 1:S201–12
- Miller GA (1956) The magical number seven, plus or minus two. *Psy Rev* 63:81–97