



HAL
open science

Dispositif d'évaluation pour un certificat d'aptitudes cliniques : quelles mesures pour quelles compétences ?

Olivier Catteau, Odile Beyne Rauzy, Anne Mayère, Nicolas Savy

► To cite this version:

Olivier Catteau, Odile Beyne Rauzy, Anne Mayère, Nicolas Savy. Dispositif d'évaluation pour un certificat d'aptitudes cliniques : quelles mesures pour quelles compétences ?. 9e Conference Environnements Informatiques pour l'Apprentissage Humain (EIAH 2019), Jun 2019, Paris, France. pp.157-168. hal-02435357

HAL Id: hal-02435357

<https://hal.science/hal-02435357>

Submitted on 10 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in:
<http://oatao.univ-toulouse.fr/24945>

To cite this version: Catteau, Olivier and Beyne Rauzy, Odile and Mayère, Anne and Savy, Nicolas *Dispositif d'évaluation pour un certificat d'aptitudes cliniques : quelles mesures pour quelles compétences ?* (2019) In: 9e Conference Environnements Informatiques pour l'Apprentissage Humain (EIAH 2019), 4 June 2019 - 7 June 2019 (Paris, France).

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Dispositif d'évaluation pour un certificat d'aptitudes cliniques : quelles mesures pour quelles compétences ?

Olivier Catteau¹, Odile Beyne-Rauzy², Anne Mayère³, Nicolas Savy⁴

¹ Université de Toulouse, IRIT UMR 5505, 118 route de Narbonne,
F-31062 Toulouse, France
olivier.catteau@irit.fr

² Université de Toulouse, UFR de Médecine Purpan, 37 Allées Jules Guesde,
F-31062 Toulouse Cedex 9, France
CHU de Toulouse, Médecine Interne, IUCT Oncopole, 1 avenue Irène Joliot Curie,
F-31059 Toulouse Cedex 9, France
beynerauzy.odile@iuct-oncopole.fr

³ Université de Toulouse, CERTOP UMR 5044, 115 Route de Narbonne,
F-31077 Toulouse Cedex 9, France
anne.mayere@iut-tlse3.fr

⁴ Université de Toulouse, IMT UMR 5219, 118 route de Narbonne,
F-31062 Toulouse, France
nicolas.savy@iut-tlse3.fr

Résumé. Cette communication présente les premiers résultats de l'étude de la mise en place d'un certificat d'aptitude clinique à travers la conception et la mise en œuvre d'un environnement informatique. Cet environnement informatique vise d'une part à équiper cet examen, et d'autre part à mener une démarche réflexive le concernant. Les résultats des étudiants à cet examen montrent une non redondance avec ceux des épreuves classantes typiques de ce parcours de formation. En confrontant les retours des étudiants au travers de questionnaires anonymes et l'étude du dispositif de formation lui-même, nous interrogeons l'approche qui le sous-tend concernant les compétences en général, et les compétences relationnelles en particulier.

Mots-clés. Approche par compétences, évaluation, modèles pédagogiques, collecte de données, traitement de données.

Abstract. This paper presents the first results of the study of the implementation of a clinical aptitude certificate through the design and implementation of an IT environment. This IT environment aims on the one hand to equip this examination, and on the other hand to conduct a reflexive approach concerning it. The students' results in this examination show no redundancy with those of the ranking tests typical of this training program. By comparing student feedback through anonymous questionnaires and the study of the training system itself, we question the underlying approach to skills in general and relational skills in particular.

Keywords. Competency approach, assessment, learning model, data collection, data mining.

1 Introduction

Un certificat d'aptitude clinique a été mis en place depuis 2016 à l'issue du deuxième cycle des études médicales en France (année bac+6). Ce certificat représente une innovation relative pour ce domaine, dans la mesure où il suppose de rompre avec une logique dominante de concours et d'évaluation de connaissances via des QCM, pour évaluer des compétences, en l'occurrence cliniques. Cette communication prend appui sur les premiers résultats d'une démarche initiée dans le cadre d'un programme de recherche interdisciplinaire financé par l'IRESP. A travers la conception et la mise en œuvre d'un environnement informatique, il s'agit d'une part d'équiper le dispositif d'évaluation requis pour ce certificat, d'autre part de contribuer à un retour réflexif sur le dispositif.

Nous exposerons tout d'abord le contexte d'introduction de ce nouveau certificat, ainsi que les conditions de sa mise en œuvre à l'UFR de Médecine de Toulouse Purpan. A travers l'analyse des résultats de cet examen, nous montrerons que ce certificat se distingue des autres dispositifs d'évaluation, ce qui conduit à considérer qu'il évalue des savoirs ou dispositions autres que ceux que mesurent les épreuves classantes.

Sur la base de l'analyse des résultats d'examen, qui permet d'en préciser les modalités et attendus, nous proposons une prise de recul : qu'en est-il de ce qui est ainsi évalué ? Tant les retours des étudiants via des questionnaires anonymes, que la mise en discussion des résultats intermédiaires avec ceux d'autres recherches, invitent à interroger les approches pédagogiques qui sous-tendent le dispositif, et les acceptations associées des compétences cliniques.

Enfin, nous concluons en proposant des perspectives de recherche.

2 Conditions d'élaboration et mise en œuvre d'un dispositif d'évaluation

2.1 Contexte et Méthodologie

Dans le cadre de la réforme des études médicales, la définition d'un référentiel de compétences génériques dans le Bulletin Officiel du 16 Mai 2013 a rendu nécessaire une transformation pédagogique pour mettre en place un processus d'acquisition, de développement et de certification des compétences cliniques en médecine (1er et 2ème cycle). Le certificat de compétences cliniques (CCC) a été mis en place pour en assurer la validation [12] ; 6 rôles d'ensemble ont été spécifiées, actuellement en cours de déclinaison sous forme de compétences : « Clinicien », « Communicateur », « Acteur de Santé Publique », « Scientifique », « Responsable aux plans éthique et déontologique », « Réflexif ».

Beaucoup d'auteurs se sont penchés sur la notion de compétence. Nous utiliserons dans cet article la définition de Jacques Tardif : « Un savoir-agir complexe prenant appui sur la mobilisation et la combinaison efficaces d'une variété de ressources internes et externes à l'intérieur d'une famille de situations » [15]. Cette définition souligne le caractère complexe et situé des compétences. Or l'inscription de cette démarche par compétences dans les études médicales vient en tension avec une logique de concours, qui suppose des critères objectifs et « neutres » d'évaluation pour

l'interclassement des différents candidats. Cette logique du concours a progressivement transformé les études médicales en faisant du QCM la modalité prédominante d'évaluation, et l'organisation des enseignements. Le nouveau certificat de compétences cliniques se trouve ainsi « pris en tenaille » entre l'objectif d'évaluer des savoirs complexes, interactionnels et situés, et l'exigence de critères d'évaluation objectivables et standardisables.

Nous proposons dans le cadre de cet article de questionner les logiques en tension dans un tel dispositif d'évaluation. Nous verrons qu'une forme première de résolution de ces tensions a consisté à privilégier une approche béhavioriste, focalisée sur des comportements observables et mesurables [1]. Nous poserons ensuite des jalons vers d'autres approches susceptibles de mieux évaluer les compétences mobilisées pour affronter des situations qui réclament une réflexion sur et dans l'action [2].

Pour la mise en œuvre de cet examen, un environnement informatique a été mis en place. Son rôle est de recueillir les données d'évaluations, d'automatiser et d'accélérer leurs traitements, pour faire une remontée des résultats plus rapidement. Il permet de collecter sous forme de traces brutes un ensemble d'observés, composé à ce stade essentiellement de grilles d'évaluation. Côté *front office*, le recueil de données se fait au travers d'une application web responsive permettant aux évaluateurs de saisir leurs résultats aussi bien sur ordinateur fixe, portable ou tablette, dans des réseaux hétérogènes. Chaque clic sur une case d'évaluation fait l'objet d'une trace brute envoyée au *back office*. Une fois collectées, les traces sont modélisées dans le *back office*, puis des algorithmes de calculs sont mis en place pour produire des indicateurs [4]. Les principaux indicateurs à destination des responsables de formation concernent la réussite à l'examen, le taux de satisfaction des enseignants, le taux de satisfaction des étudiants, le taux de difficulté perçue, mais également d'autres indicateurs qui prennent appui sur une analyse statistique et docimologique des grilles d'évaluation et des résultats associés [3]. Une des finalités est de vérifier la pertinence du dispositif d'évaluation et de l'améliorer pour les futures sessions : mesures de position et de dispersion des différents résultats, analyse en composantes principales [6], corrélation avec d'autres méthodes d'évaluation [8] telles que les épreuves théoriques de 6^{ème} année et les épreuves classantes nationales (ECN). Les principaux indicateurs à destination des étudiants relèvent de la rétro-information : réussite à l'examen, commentaires des évaluateurs, évaluation et commentaires des acteurs (patients standardisés). Lorsque des indicateurs nécessitent des calculs statistiques (par exemple, coefficient de corrélation), le *back office* sollicite l'application R version 3.5.0 (2018-04-23).

Le certificat de compétences cliniques n'a pas à ce stade finalité à participer à l'interclassement des étudiants mais à vérifier la maîtrise de compétences associées à la pratique clinique. Il ne s'agit pas de reproduire la logique « concours » du PACES₂ ou des ECN en favorisant des critères discriminants ; la priorité est donnée à l'homogénéité des grilles d'évaluation.

L'environnement informatique constitué permet par ailleurs, distinctement, de recueillir de façon anonyme les retours tant des enseignants que des étudiants. La

1 Patient standardisé : personne formée à simuler l'histoire d'un vrai patient et à reproduire systématiquement les signes cliniques, la personnalité, le langage corporel et les réactions émotionnelles qui auront été définis d'entrée de jeu (<http://cis-ge.ch/simulation/patient-standardise/definition/>)

2 PACES : Première Année Commune des Etudes de Santé

réflexivité ainsi proposée aux acteurs est un appui important pour questionner plus avant les compétences attendues et évaluées.

2.2 Expérimentation

A l'UFR de Médecine de Toulouse Purpan, le CCC a donné lieu à une nouvelle modalité d'évaluation sous forme d'un Examen Clinique Objectif Structuré (ECOS) à destination des étudiants de 6^{ème} année.

L'examen a concerné 141 étudiants en Janvier 2016 et 134 étudiants en Février 2017 qui sont passés successivement sur 7 stations (ECG³, Dermatologie, Thérapeutique, Annonce, Interrogatoire, Examen clinique, Synthèse). Ces stations formalisent des étapes standards de la pratique clinique. Il s'agit en particulier d'évaluer la capacité à mettre en œuvre un examen clinique, et à formuler des hypothèses de prise en charge. Un lien est établi entre plusieurs stations pour envisager un même cas sous différents angles.

L'ensemble des données collectées et analysées est résumé dans le tableau 1 ci-dessous.

Tableau 1. Effectifs collectés et analysés de 2016 à 2018

| | 2016 | 2017 | 2018 |
|--|------|------|------|
| Station ECG | 141 | 134 | |
| Station ECG/Thérapeutique | | | 135 |
| Station Dermatologie | 141 | 134 | 135 |
| Station Thérapeutique | 141 | 134 | |
| Station Annonce | 141 | 134 | 135 |
| Station Annonce (retour des acteurs) | | 134 | 101 |
| Station Interrogatoire | 141 | 134 | 135 |
| Station Interrogatoire (retour des acteurs) | | 125 | 94 |
| Station Clinique | 141 | 134 | 135 |
| Station Examen complémentaire | | | 135 |
| Station Synthèse | 141 | 134 | 135 |
| Questionnaire post-évaluation A destination des étudiants | | 134 | 128 |
| Questionnaire post-évaluation A destination des évaluateurs | | 30 | |
| Résultats aux épreuves théoriques de 6 ^{ème} année | 131 | 128 | 124 |
| Résultats des ECN | 137 | 128 | 123 |

Fusion des stations ECG et Thérapeutique

Création de la station examen complémentaire

L'environnement informatique n'a été mis en place qu'en 2017 mais une analyse statistique a pu être effectuée a posteriori sur les résultats de 2016. La logistique de l'examen, organisé sur 4 demi-journées, est assez lourde avec environ 25 examinateurs

3 ECG : Electrocardiogramme

mobilisés, certains participant à plusieurs demi-journées. Chaque étudiant doit passer 7 minutes par station. Le changement de station se fait au coup de sifflet. Chaque station dispose de plusieurs « vignettes » ou sujets d'examens pour éviter la communication entre les étudiants qui ont terminé leur parcours et ceux qui ne sont pas encore passés.

L'examen a concerné 135 étudiants en Février 2018 avec une modification des stations : les thématiques ECG et Thérapeutique ont été regroupées en 1 seule station, une nouvelle station « Examen complémentaire » a été introduite à la suite de l'examen clinique (ce dernier utilisant un mannequin). C'est une autre caractéristique de ce dispositif que d'être réflexif et de supposer un environnement informatique qui équipe cette réflexivité et qui soit aisément évolutif tout en assurant des formes de comparabilité des résultats.

La communication médecin / patient est plus particulièrement considérée au cours de deux stations faisant intervenir des comédiens jouant le rôle de patients standardisés (Annonce et Interrogatoire).

Sur chaque station les examinateurs disposent d'une grille d'évaluation informatisée. Les grilles des stations ECG, Dermatologie et Synthèse sont composées de 10 à 20 critères évalués entre 1 et 6 points. Les autres grilles sont composées de 10 à 30 éléments pour lesquels l'évaluateur indique s'ils ont été effectués complètement, partiellement ou pas du tout. Le rythme de passage contraint le temps passé par les enseignants à l'évaluation. Il suppose un environnement informatique très robuste et soulève des défis quant à sa conception et son évolution pour combiner efficacité et pertinence.

Un travail exploratoire a été mené par les responsables pédagogiques pour tenter de mettre en correspondance d'une part les éléments d'évaluation et d'autre part les rôles définis dans l'annexe du BO pour décrire les compétences génériques dont l'acquisition est nécessaire pour l'obtention du certificat : Clinicien, Communicateur, Acteur de Santé Publique, Scientifique, Responsable aux plans éthique et déontologique, Réflexif [12].

En fin de parcours, l'avis des étudiants et des enseignants a été collecté sur une base volontaire et anonyme au travers de questionnaires qui leur étaient dédiés. Cette base d'information est centrale pour la recherche ici menée dans la mesure où elle permet de recueillir des avis indépendants du contexte de l'examen. Ce dernier est en effet achevé au moment où il est proposé aux étudiants de prendre quelques instants pour répondre, ce que la plupart d'entre eux font bien volontiers (100% de participation en 2017, 95% en 2018) ce moment constituant une forme de sas de sortie d'examen.

2.3 Résultats

L'environnement informatique a permis de diffuser aux candidats les résultats à l'examen en 2017 et 2018 beaucoup plus rapidement qu'en 2016 en évitant un traitement manuel et potentiellement source d'erreurs.

Les questionnaires post-évaluation démontrent un intérêt pour la démarche, tant de la part des étudiants que des enseignants-examinateurs. En 2017, le taux de satisfaction du processus d'évaluation pour les enseignants était de 100%, il était de 92% pour les étudiants (98% en 2018). A 76% les étudiants ont trouvé que l'examen était adapté par rapport à la durée et au niveau demandé (92% en 2018). En 2017, la station dite «

d'annonce » - au cours de laquelle une pathologie doit être annoncée au patient standardisé - était perçue comme étant la plus difficile par les étudiants (64% de taux de difficulté). Les compétences mises en jeu dans la relation soignant-patient, domaine encore peu enseigné au cours des études médicales, sont plus particulièrement mises à l'épreuve dans de tels contextes. La situation a évolué en 2018, la station « annonce » ayant été mieux préparée avec la mise en place d'un enseignement dédié à la relation médecin/patient ; par ailleurs les situations d'annonce ont été simplifiées, en portant sur des annonces d'examen et non directement de pathologie. Le taux de difficulté perçue par les étudiants a alors diminué à 44% alors que la station examen clinique sur mannequin a été perçue comme étant la plus difficile (49% de taux de difficulté en 2018, 44% en 2017). Des éléments d'explication peuvent tenir aux limites techniques des mannequins et au manque d'entraînement des étudiants dans ce contexte.

Le retour sur évaluation était très attendu par les étudiants. Les résultats de chaque station ont été communiqués par groupe de critères. Exemples : Savoir Etre, Diagnostic, Prise en Charge, Raisonnement... Les commentaires saisis par les évaluateurs étaient également communiqués. De plus, sur les 2 stations où intervenaient des acteurs jouant le rôle de patient standardisé (annonce et interrogatoire), une évaluation formative du candidat leur était demandée. Le résultat de cette évaluation était communiqué comme *feedback* aux étudiants mais ne rentrait pas dans le calcul de la note finale pour l'attribution du certificat de compétences cliniques. Cette évaluation pourrait toutefois devenir sommative à l'avenir. En 2017, la corrélation entre l'évaluation formative des acteurs et l'évaluation sommative des enseignants sur le savoir être des étudiants n'était que de 0.39 pour l'annonce ($p=3.4e-6$) et de 0.31 pour l'interrogatoire ($p=2.427e-4$). Il y a là un élément intéressant à investiguer plus avant pour expliciter les attendus d'une part des acteurs jouant les patients standardisés, qui ont été formés au *feedback*, et d'autre part les enseignants.

L'examen n'avait pas de visée sélective, et cela s'est confirmé par la réussite de la quasi-totalité des étudiants, seuls 1 à 2 étudiants passant chaque année à la seconde session.

Une analyse statistique plus fine des résultats, résumée dans le tableau 2, montre que chaque station a son utilité dans le dispositif. Les résultats sont calculés de la manière suivante : le score de chaque station est obtenu en faisant la moyenne pondérée de tous les éléments évalués lors du passage de l'étudiant sur la station. Les résultats sont présentés sur un barème de 20 points par station. Un élément effectué « complètement » rapporte 1 point, un élément effectué « partiellement » un demi-point et un élément non effectué ne rapporte pas de point. Le score global est obtenu sur un barème de 100 points en faisant la moyenne pondérée des 7 stations. Prenons l'exemple de la station Dermato en 2016, le coefficient de corrélation est calculé entre (1) les résultats de cette station et (2) la somme des résultats des 6 autres stations. La normalité des résultats a été vérifiée. Pour chaque calcul de coefficient de corrélation, la méthode suivante a été appliquée : (1) vérification que l'allure de l'histogramme des variables est sous forme de cloche, (2) vérification que l'allure du diagramme quantile - quantile est sous forme de droite de pente égale à 1, et (3) élimination des valeurs isolées (inférieures à $Q1 - 1.5(Q3-Q1)$ et supérieures à $Q3 + 1.5(Q3-Q1)$).

Les coefficients de corrélations calculés entre les résultats des grilles d'une station donnée et les résultats des grilles des autres stations, quand ils sont significatifs, restent

positifs et faibles (entre 0.17 et 0.24 en 2016 ; entre 0.22 et 0.36 en 2017 ; entre 0.18 et 0.29 en 2018) justifiant ainsi de l'intérêt de chaque station.

Tableau 2. Analyse statistique des résultats des différentes stations

| 2016 | | | | | | | | | | |
|----------|------|------|------|------|------|------|------|------------------|-------------|---------|
| Station | Min | Q1 | Med | Moy | Q3 | Max | ET | Cor | p | |
| Global | 49.5 | 68.0 | 72.8 | 72.3 | 77.0 | 88.2 | 6.89 | | | |
| ECG | 8.0 | 14.0 | 16.0 | 15.2 | 18.0 | 20.0 | 2.88 | non significatif | | |
| Dermato | 2.0 | 11.0 | 15.0 | 13.9 | 17.0 | 20.0 | 3.99 | 0.17 | [0.00;0.33] | 0.048 |
| Thérapeu | 6.0 | 13.0 | 15.0 | 14.6 | 17.0 | 20.0 | 3.26 | 0.19 | [0.02;0.34] | 0.029 |
| Annonce | 0.0 | 12.0 | 15.0 | 14.1 | 17.0 | 20.0 | 3.80 | non significatif | | |
| Interro | 5.0 | 11.5 | 13.0 | 13.0 | 14.5 | 18.2 | 2.11 | 0.24 | [0.08;0.39] | 0.0039 |
| Clinique | 7.5 | 14.0 | 15.5 | 15.1 | 17.0 | 20.0 | 2.43 | non significatif | | |
| Synthèse | 8.0 | 13.0 | 16.0 | 15.3 | 18.0 | 20.0 | 3.16 | 0.18 | [0.01;0.33] | 0.037 |
| 2017 | | | | | | | | | | |
| Station | Min | Q1 | Med | Moy | Q3 | Max | ET | Cor | p | |
| Global | 49.5 | 68.6 | 73.0 | 72.5 | 77.0 | 85.2 | 6.89 | | | |
| ECG | 6.5 | 13.0 | 16.5 | 15.6 | 18.0 | 20.0 | 3.34 | 0.28 | [0.11;0.43] | 0.0014 |
| Dermato | 6.0 | 12.5 | 14.5 | 14.3 | 16.5 | 19.5 | 3.10 | 0.36 | [0.21;0.51] | 2.1e-05 |
| Thérapeu | 5.0 | 12.0 | 14.0 | 13.7 | 15.5 | 19.5 | 2.66 | 0.26 | [0.10;0.42] | 0.0024 |
| Annonce | 5.5 | 13.5 | 15.0 | 15.0 | 16.9 | 19.5 | 2.44 | non significatif | | |
| Interro | 9.8 | 13.6 | 14.9 | 14.8 | 16.2 | 18.8 | 2.03 | non significatif | | |
| Clinique | 8.0 | 12.5 | 14.0 | 13.9 | 15.5 | 17.5 | 1.95 | 0.35 | [0.20;0.50] | 3e-05 |
| Synthèse | 4.0 | 13.0 | 15.0 | 14.5 | 17.0 | 20.0 | 3.11 | 0.22 | [0.05;0.38] | 0.011 |
| 2018 | | | | | | | | | | |
| Station | Min | Q1 | Med | Moy | Q3 | Max | ET | Cor | p | |
| Global | 48.5 | 67.6 | 72.5 | 71.6 | 76.1 | 89.5 | 6.99 | | | |
| ECGThe | 4.0 | 11.5 | 14.5 | 14.0 | 17.0 | 19.0 | 3.32 | non significatif | | |
| Dermato | 4.0 | 11.5 | 13.5 | 13.5 | 15.5 | 19.0 | 3.04 | non significatif | | |
| Annonce | 9.5 | 15.0 | 17.0 | 16.4 | 18.0 | 20.0 | 2.39 | 0.29 | [0.12;0.44] | 0.00078 |
| Interro | 8.8 | 13.5 | 15.0 | 14.8 | 16.2 | 19.0 | 1.96 | non significatif | | |
| Clinique | 4.5 | 11.8 | 13.5 | 13.3 | 15.0 | 17.5 | 2.61 | non significatif | | |
| Compl | 2.0 | 12.0 | 14.0 | 13.7 | 16.0 | 20.0 | 3.77 | 0.18 | [0.01;0.34] | 0.04 |
| Synthèse | 6.5 | 13.0 | 15.0 | 14.8 | 17.0 | 19.0 | 2.70 | 0.29 | [0.12;0.44] | 0.00094 |

La faible dispersion des résultats montre également l'homogénéité de ces derniers. Une analyse en composantes principales montre une chute d'inertie faible (39% de l'information sur les 2 premières dimensions en 2016, 47.21% en 2017 et 31.88% en 2018, cf. Figure 1). Là encore, on retrouve une homogénéité, il n'y a pas de caractère discriminant entre les stations.

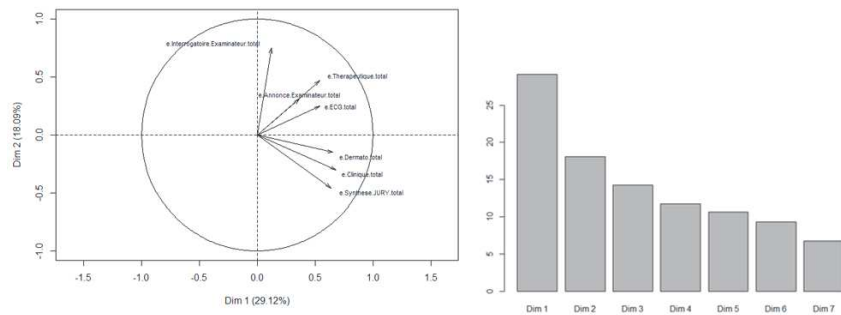


Fig. 1. Cercle des corrélations et diagramme de chute d'inertie issues de l'analyse en composantes principales des données de 2017.

Il n'y a pas non plus de redondance avec les résultats aux épreuves théoriques. En effet, comme le montre l'exemple de la figure 2, le coefficient de corrélation entre les résultats à l'ECOS et les résultats de l'UE d'enseignements théoriques de 6^{ème} année est également significatif, positif et faible (0.40 [0.25;0.53] en 2016 ; 0.39 [0.23;0.53] en 2017, 0.25 [0.08;0.41] en 2018) justifiant de l'intérêt de l'ECOS par rapport aux modalités d'évaluation classiques existantes. La modification des stations en 2018 apporte une tendance à la diminution de la corrélation.

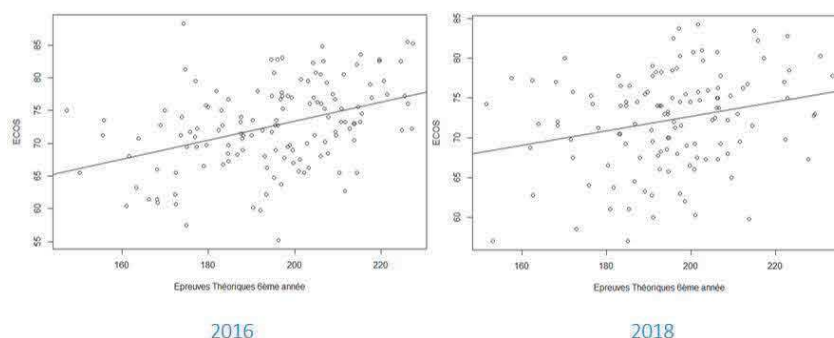


Fig. 2. Nuage de points de l'ECOS et des épreuves théoriques de 6^{ème} année en 2016 et 2018

Enfin, il n'y a pas de redondance avec les ECN. En effet, le coefficient de corrélation entre les résultats à l'ECOS et les résultats des épreuves classantes nationales est lui aussi significatif, positif et faible (0.35 [0.19;0.49] en 2016, 0.43 [0.27;0.56] en 2017, 0.35 [0.19;0.50] en 2018). Ainsi l'examen ECOS présente une contribution spécifique, qui peut être associée à la mise en situation.

Tous ces résultats mettent en évidence l'intérêt de la démarche en ce que l'évaluation qui en résulte n'est pas redondante au regard d'autres dispositifs d'évaluation. Nous allons maintenant prendre un peu de recul au regard de ce qui peut être appréhendé comme une évaluation priorisant des critères d'évaluation objectivables et standardisables, dans une approche à dominante behavioriste.

3 Analyse critique

L'annexe du BO décrit les compétences génériques requises en les articulant autour de 7 rôles. Mais il ne donne pas d'indication sur la façon dont ces rôles et plus globalement les compétences doivent être évaluées, ici tout au long des 7 stations. Quelles conceptions sont alors retenues des attendus et des modalités de les atteindre ? En termes de ressenti, 87% des enseignants estiment que la formation actuelle permet aux étudiants d'acquérir les compétences requises. Les étudiants sont beaucoup plus circonspects quand on les interroge sur la maîtrise de chacun des 7 rôles dans le cadre du questionnaire anonyme soumis après achèvement du parcours des stations d'évaluation : les taux de maîtrise déclarée varient de 47% pour le rôle « d'acteur de santé publique » à 55% pour celui de « coopérateur ». On retrouve cette tendance en

2018 : les taux de maîtrise déclarée varient de 50% pour le rôle « d'acteur de santé publique » à 57% pour celui de « coopérateur ». Les étudiants concernés n'avaient découvert que tardivement ces différents rôles tels qu'énoncés, auxquels ils n'avaient été sensibilisés que quelques semaines avant l'examen ; ces réponses traduisent également une difficulté de compréhension des attendus, tant les termes sont vastes et les définitions étendues.

Selon Rey, un référentiel (tel celui décrit dans le BO pour les compétences génériques) devient parole de commandement, mais ne garantit pas la faisabilité, surtout quand on aborde le caractère civique et social [13]. Rey interroge la culture du résultat susceptible d'y être associé : il y a souvent, dans nombre d'approches par compétences, confusion entre la compétence et le résultat d'une action qui s'apparente alors à une performance. A titre d'exemple, le rôle de communicateur requiert de la part de l'étudiant qu'il soit à même « d'établir une communication axée sur le patient, par le biais de la prise de décisions partagées et d'interactions efficaces fondées sur l'éthique et l'empathie ». Or il reste difficile de formuler des critères d'évaluation au-delà des formules de politesse et de civilité usuelles. Le BO laisse la mise en œuvre du référentiel sous la responsabilité des enseignants et des étudiants et fait quasi-silence sur le processus d'engendrement de ces compétences et de leur évaluation ; seules les modalités « d'évaluations en situation authentique et des auto-évaluations » sont évoquées.

Hébrard dénonce quant à lui le fait de privilégier l'apprentissage par l'action en situation au détriment des connaissances, des concepts relevant des disciplines [5]. Dans la formation d'infirmier, il constate que les grilles d'évaluations des compétences sont souvent hétérogènes et sont constituées d'une liste de preuves à vérifier ; les évaluateurs font alors face à la difficulté de cocher telle ou telle case. De plus, les dimensions affectives, conflictuelles, la question du pouvoir, les enjeux identitaires sont occultés. Dans les études médicales, les connaissances issues des sciences sociales sont en bonne part ignorées, l'enseignement étant essentiellement le fait de praticiens. Or l'approche par compétences, si elle est ramenée à une liste d'activités, ne permet pas de prendre en compte ce qui est identifié comme une caractéristique clé du travail de thérapeute, à savoir la tension intrinsèque entre d'une part un mouvement nécessaire d'objectivation du patient pour interpréter les signes éventuels de la maladie, et d'autre part une capacité à « se déprendre de sa position d'objectivation et se placer avec le patient dans une relation de vivant à vivant et de sujet à sujet » [9].

Dès lors il apparaît qu'une opération qui paraît aussi simple que de définir des listes de critères d'évaluation est en fait sous-tendue par une conception plus générique de ce qu'est en l'occurrence ici un thérapeute. Une approche par compétences ramenée à une liste d'attitudes peut s'avérer très proche d'un néo-paternalisme usuel, précisément parce qu'elle n'interroge pas les dispositions requises pour être à même de mener ce double mouvement d'objectivation et de construction d'un échange effectif avec le patient.

Dans la définition de la compétence par Jacques Tardif, figure la notion de « famille de situations ». Cela interroge à deux niveaux par rapport à l'évaluation proposée : la répétition et la complexité. D'une part, cela sous-entend que la compétence ne peut être acquise que si elle a pu être mobilisée dans plusieurs situations. Quel crédit peut-on

alors donner à une certification qui ne repose que sur une situation unique (celle de l'examen) ? Une réflexion est actuellement en cours pour étendre le dispositif d'évaluation (formative et sommative) aux situations de stage de la 2^{ème} à la 5^{ème} année de formation. D'autre part, dans une famille de situations, il y en aura qui seront plus complexes à traiter que d'autres. On se retrouve alors face à un paradoxe : dans un souci d'homogénéité de l'examen, les « vignettes » sont construites sur des situations simples et courantes du point de vue des médecins. C'est aussi une volonté d'éviter de confronter l'étudiant à des situations de 3^{ème} cycle.

Roegiers souligne l'importance d'établir une articulation entre le référentiel métier, le référentiel de compétences et le référentiel de formation initiale [14]. Le Boterf ajoute la nécessité d'élaborer préalablement un référentiel d'évaluation [7]. On sent bien que la volonté ministérielle d'introduire l'approche par compétences dans les études de médecine n'en est qu'au début de la démarche. La mise en œuvre de l'ECOS l'est également. Les critères d'évaluation ont été définis de manière empirique. La priorité donnée à l'opérationnel pour faire face aux défis de l'ingénierie pédagogique a conduit à mettre en partie de côté la réflexion sur les conceptions des relations soignants-patients et leurs modes d'enseignement.

Il est en cela utile d'interroger la posture behavioriste de l'approche par compétences. Comme le souligne Boutin, le behaviorisme est adapté pour la maîtrise de comportements observables, objectivés. Dans cette approche, « le comportement visé doit être répété jusqu'à ce qu'il devienne automatique » [2]. Ainsi en est-il lors d'examens cliniques de l'identification de certains signes cliniques. Leung démontre les dérives d'une approche qui peut dans son principe de répétitivité étendue à l'évaluation tendre à l'absurdité [10] : il donne l'exemple d'un étudiant qui doit montrer au moins 4 fois qu'une compétence est mise en œuvre au cours de 7 situations filmées différentes. Meirieu dénonce l'atomisation des apprentissages globaux en une multitude de petits actes dont la recombinaison peut être difficile [11].

D'autres stratégies ne sont pas à écarter : le cognitivisme peut être utile à la résolution de problèmes tandis que le constructivisme aide à affronter des situations qui réclament une réflexion sur et dans l'action [2]. Là encore il apparaît que les orientations retenues ont des liens avec l'approche retenue de l'activité thérapeutique : si les compétences cliniques reposent non seulement sur la connaissance des pathologies et des moyens d'investigation et de traitement, mais de façon imbriquée, sur « la compréhension fine de ce que peut signifier le fait d'être malade pour un individu et son entourage, ainsi que la connaissance des enjeux sociaux et institutionnels de cette expérience » [9], alors les situations de relation soignant-patient sont fondamentalement des situations qui réclament une réflexion sur et dans l'action, et une réflexion menée non sur le patient mais avec lui.

La démarche des ECOS est également répandue dans les pays anglo-saxons avec ses checklists et ses patients standardisés. Elle apporte certes l'individualisation de l'apprentissage mais revient selon certains auteurs à démontrer la performance de l'étudiant, ce potentiellement au détriment du processus sous-jacent [10]. L'approche peut être perçue comme superficielle et parfois source de démotivation. Elle encourage les étudiants à faire ce qu'on attend d'eux plutôt que de prendre le risque de développer une distance critique. A cela s'ajoute le risque pour les examinateurs de définir des

règles minimalistes de validation. L'autre risque souligné par Leung, la réduction du contenu pédagogique, est particulièrement sensible dans le contexte d'études médicales où les étudiants désertent les cours magistraux et où le primat donné aux QCM fait que toute évaluation doit se conformer aux modalités d'un QCM, quand bien même les connaissances en cause sont complexes et relèvent d'un paradigme différent de celui de l'*evidence-based medicine*, ce qui est le cas des connaissances issues des sciences sociales.

4 Conclusion et perspectives

Cette communication rend compte des premiers résultats d'une démarche innovante en termes de formation aux compétences cliniques de futurs médecins et de leur évaluation. L'environnement informatique constitué permet d'une part d'équiper l'ingénierie pédagogique, et d'autre part, distinctement, de recueillir de façon anonyme les retours tant des enseignants que des étudiants.

L'étude des résultats de cet examen de compétences cliniques montre que cet examen n'est pas redondant au regard des examens évaluant les compétences plus « théoriques » sur les maladies, les techniques diagnostiques et thérapeutiques, et les sciences dites fondamentales associées. En cela cet examen présente une contribution spécifique, dont la pleine signification reste à expliciter.

L'analyse critique proposée en complément permet d'ouvrir des pistes sur les questions que soulève cet examen. L'observation du dispositif en train de se faire permet d'interroger les paradigmes en présence. Il apparaît que, compte tenu des impératifs de calendrier, et de conceptions implicites sur les relations soignants-patients, l'approche behavioriste ait été dominante dans les premiers temps de ce dispositif.

En perspectives, la question est celle de dispositif complémentaire ou renouvelé de formation et d'évaluation qui fasse place à une approche constructiviste susceptible d'aider à affronter des situations qui réclament une réflexion sur et dans l'action [2]. Dans cette perspective, la refonte des stages et de leur évaluation pourrait être l'occasion d'une réflexion sur la mise en place de démarches pédagogique et d'évaluation accordant plus de place à cette réflexion sur et dans l'action, non seulement sur mais avec les patients. Il s'agit également de réfléchir à un dispositif permettant de mesurer la progression des compétences acquises au fil du cursus, voire au-delà si une articulation peut être établie avec la formation d'internat voire la formation continue des praticiens.

Dans ce cadre, l'EIAH doit permettre de mesurer la progression des compétences acquises en cours de formation et de contribuer à la réingénierie pédagogique, tout en étant lui-même objet d'enquête et de réflexion dans la perspective interdisciplinaire retenue.

Remerciements. Cette recherche est financée grâce au soutien de l'Institut de Recherche en Santé Publique, IReSP-17-HSR-16. Elle a bénéficié de l'aide des partenaires financeurs de l'IReSP dans le cadre de l'appel à projets général 2017 Volet Services de Santé.

J.C. Basson (CRESCO EA 7419*), E. Coeurdevey (CERTOP UMR 5044*), E. Fournales (CERTOP UMR 5044*), N. Haschar-Noé (CRESCO EA 7419*), P. Marrast (CERTOP UMR 5044*), T. Lang (INSERM UMR 1027**), M. Kelly-Irving (INSERM UMR 1027*), S. Lamy (INSERM UMR 1027*), L. Bonneville (GRICO**) et S. Grosjean (GRICO**), participent également à cette recherche dans le cadre d'une collaboration menée au sein de l'IFERISS, Institut Fédératif d'Etudes et de Recherches Interdisciplinaires Santé Société.

* Université Toulouse 3

** Université d'Ottawa

Références

1. Boutin G. : L'approche par compétences en éducation : un amalgame paradigmatique. *Connexions* 2004/1 n° 81 (2004) 25-41
2. Boutin G. : Le béhaviorisme et le constructivisme ou la guerre des paradigmes. *Québec français* 119 (2000) 37-40
3. Dessus P. : Analyser les items d'un questionnaire. <http://espe-rted-reflexpro.u-ga.fr/docs/scied-cours-qcm/fr/latest/tuto-analyse-items.html>, (2018)
4. Djouad T. : Ingénierie des indicateurs d'activités à partir de traces modélisées pour un environnement informatique d'apprentissage humain. Thèse Université Claude Bernard Lyon 1 (2013)
5. Hébrard P. : L'humanité comme compétence ? Une zone d'ombre dans la professionnalisation aux métiers de l'interaction avec autrui. *Les sciences de l'éducation – Pour l'Ere nouvelle* 2011/2 Vol.44 (2011) 103-121
6. Husson F., Lê S., Pagès J. : *Analyse de données avec R*. 2^{ème} édition, Presses Universitaires de Rennes (2016)
7. Le Boterf G. : *Ingénierie et évaluation des compétences*. 6^{ème} Edition, Editions d'Organisation, Eyrolles (2011)
8. Lecoutre J.P. : *Statistique et probabilités*. 6^{ème} édition, Dunod (2016)
9. Lefève C., Mino J. C. : Former de vrais thérapeutes. La place des sciences humaines et sociales dans les études de médecine, *revue Etudes*, n°4142, Février 2011, p. 187-198.
10. Leung W.C. : Competency based medical training : review. *British Medical Journal* Volume 325 (2002) 693-696
11. Meirieu P. : Si la compétence n'existait pas, il faudrait l'inventer... <https://www.meirieu.com/ARTICLES/SUR%20LES%20COMPETENCES.pdf> (2005)
12. Ministère de l'enseignement supérieur, de la Recherche et de l'Innovation : *Etudes Médicales, Régime des études en vue du premier et du deuxième cycle*, Bulletin Officiel n°20 du 16 Mai (2013)
13. Rey B. : La notion de compétence : usages et enjeux. *Le français d'aujourd'hui* 2015/4, n° 191 (2015) 15-24
14. Roegiers, X. : *Des curricula pour la formation professionnelle initiale : la Pédagogie de l'intégration comme cadre de réflexion et d'action pour l'enseignement technique et professionnel*. Bruxelles : éd. De Boeck Université (2010)
15. Tardif J. : *L'évaluation des compétences : Documenter le parcours de développement*. Montréal, Canada, Chenelière Education (2006)