

UNIVERSITE PARIS 13

**UFR Santé, Médecine, Biologie Humaine
Léonard de Vinci**

Département de Pédagogie des Sciences de la Santé

**Mémoire de Maîtrise Universitaire
de Pédagogie des Sciences de la Santé**

**EVALUATION
DU RAISONNEMENT CLINIQUE
DES ETUDIANTS SAGES-FEMMES
PAR LE TEST
DE CONCORDANCE DE SCRIPT**

Présenté par

Anne Demeester

Née le 17 mars 1962

Soutenu le

29 novembre 2004

Directeur de mémoire : Professeur Rémi Gagnayre

Année 2004

Je remercie les personnes qui m'ont accompagnée tout au long de cette formation en pédagogie des sciences de la santé :

Annie-Claire Cottu, directrice de l'Ecole de Sages-femmes de Marseille, pour m'avoir guidée vers le Département de Pédagogie des Sciences de la Santé à Bobigny.

Le Professeur Jean-François d'Ivernois, directeur du Département.

Le Professeur Rémi Gagnayre, pour ses enseignements, ses conseils de lectures, ses encouragements et surtout ses commentaires sur les travaux rendus qui m'ont chaque fois incitée à une plus grande réflexion.

Claire Marchand, maître de conférence, pour la qualité de ses interventions.

Anne-Marie, Isabelle, Véronique, François, mes compagnons de formation, pour les productions en groupe et les discussions passionnées.

Florence et Anne-Marie pour leur accueil au Département et pour la gestion à distance des petits problèmes administratifs.

L'Assistance Publique des Hôpitaux de Marseille et l'ANFH, pour la prise en charge des frais de formation.

Je remercie les personnes qui ont contribué, peu ou prou, de près ou de très loin (au Québec), à la réalisation du travail de recherche présenté ici :

Le Professeur Rémi Gagnayre (Bobigny) pour avoir accepté de diriger ce mémoire.

Le Professeur Bernard Charlin (Montréal), concepteur du TCS, pour l'intérêt porté à ce travail, pour la communication de nombreuses ressources documentaires et bibliographiques, et pour sa particulière bienveillance à mon égard.

Le Professeur Marc Gamerre et Annie-Claire Cottu (Marseille), directeurs de l'école de sages-femmes de Marseille, pour m'avoir témoigné leur confiance en me permettant de développer le TCS auprès des étudiants.

Le Professeur Jean-Louis Bernard (Marseille), pour son expertise pédagogique, ses précieux conseils, son soutien et son enthousiasme.

Le Professeur Xavier Thirion (Marseille), pour ses compétences et son engagement dans l'exploitation statistique des données.

Le Docteur Robert Gagnon (Montréal) pour son écoute attentive, son expertise en méthodologie et son aide à l'interprétation des résultats.

Driss Kazitani (Montréal), pour son assistance technique dans l'utilisation du correcteur en ligne, sa disponibilité et son dévouement.

Jean-Marc Dubois (Bordeaux), pour avoir très aimablement offert de reviser les calculs de données et pour sa réflexion pertinente autour du TCS.

Caroline Cardinale, Catherine Dutertre, Brigitte Massari, Nicole Nguyen, Isabelle Patot et Isabelle Guyonnet, pour leur participation active au recrutement sur chaque site des sages-femmes qui ont participé à l'étude.

Mes collègues sages-femmes enseignantes pour avoir participé au recueil de situations authentiques, au réajustement des vignettes TCS et à l'organisation du passage du test par les étudiants.

Chantal, Joëlle, Sandrine, Virginie et Anna Rita, pour leur soutien et leur amitié.

Marc et Bertrand pour leur amour et leur soutien moral.

Merci à toutes les sages-femmes qui ont accepté de participer à cette étude :

Annie Allègre, Muriel Athlani, Valérie Béchadergue, Pascale Bernard, Brigitte Berthelot, Christèle Bienvenu, Fabienne Billon, Sandie Boisis, Céline Cabane, Laurence Caillol, Caroline Cardinale, Myriam Céas, Fabienne Clauss, Lydie Coston, Christine Croutelle, Elisabeth Darrieutort, Sylvie Davin, Catherine Dutertre, Danièle El Ouertani, Ghislaine Emeric, Françoise Foucaut, Sophie Garel, Christelle Gauthier, Hélène Guiolet, Isabelle Guyonet, Marion Grimangne, Lila Halil, Monique Hislen, Laurence Joseph, Marie-Josée Jourlain, Danièle Kmieckowiak, Nicole Lecomte, Aude Lambert, Pascale Lopez, Martine Mahé, Sophie Mazabrard, Nicole Mercier, Isabelle Moreau, Isabelle Patot, Laurence Peda, Aude Petetin, Annie Petiteau, Laurence Pigeon, Violaine Pin, Sylvie Renard, Chantal Roux, Sylvie Salmon, Nadine Sénatore, Françoise Serri, Myriam Simonin, Fella Tahri, Christine Ter Akopoff, Anne Tissot, Marlène Vanuxem, Valérie Vaugirard, Sandrine Viala.

Merci à toutes les étudiantes sages-femmes qui ont été impliquées dans cette étude :

Caroline Allio, Sara Amrane, Damaris Astier, Johanne Audren, Virginie Bailleul, Cécile Barety, Amandine Barnoin, Anne Binguy, Caroline Bogaty, Marianne Cattani, Heïdi Charras, Eva Chaurial, Marion Chauvet, Chloé Chuzeville, Emilie Colotte, Sylvie Cormenier, Claire Domini, Isabelle Filippi, Géraldine Flacher, Sophie Fulachier, Julia Galasso, Alexandra Garcia, Laurence Gentil, Sandy Grima, Stéphanie Grimer, Gwénola Hervi, Noémie Hot, Elsa Kohler, Séverine Lanteri, Leslie Lathus, Sophie Laurencin, Anna-Gaëlle Laurent, Céline Leconte, Hélène Le Guillan, Noémie Lenriot, Nathalie Lesieur, Christine Lopes, Béatrice Maître, Youssra Mansouri, Sabrina Monchaud, Marion Mottino, Fanny Muraglia, Claire Nicolas, Laetitia Nivaggioni, Alexandre Perrin, Emilie Pommier de Santi, Anne-Laure Pons, Anne-Cécile Pratési, Sabrina Ricard, Julie Ronot-Blanc, Nathalie Rossi, Clémence Roubaud, Isabelle Rouras, Marie Rouzaud, Céline Saez Gonzales, Karine Sanchez, Lenaïg Serazin, Marion Suponter, Laetitia Vidal, Marie-Cerise Virey, Audrey Vuola.

SOMMAIRE

PROLOGUE

De la pratique clinique au raisonnement p.6

PRESENTATION DU TRAVAIL DE RECHERCHE

Introduction p.11

Contexte p.12

Finalité p.15

Hypothèses et objectifs de recherche p.16

Matériels et méthode p.17

Résultats p.24

Limites de l'étude p.27

Discussion p.29

Conclusion p.33

Références p.34

EPILOGUE p.37

BIBLIOGRAPHIE p.41

ANNEXES p.45

PROLOGUE

De la pratique clinique au raisonnement

Je suis une ancienne élève de l'Ecole de Sages-femmes de Marseille où je travaille à présent. J'ai obtenu mon Diplôme d'Etat il y a 20 ans, en 1984.

J'ai commencé ma carrière par des remplacements qui m'ont permis de découvrir le secteur public hors CHU et le secteur privé, en métropole et outre-mer .

En 1987 j'ai travaillé 9 mois pour l'organisation non gouvernementale Médecins Sans Frontières comme sage-femme bénévole dans un Camp de réfugiés Erythréens au Nord-Est du Soudan. L'isolement géographique, la barrière de la langue et des conditions de vie et d'exercice précaires m'ont confronté à des expériences humaines difficiles et à une pratique médicale complexe. J'ai du gérer seule et avec peu de moyens, des situations qui en France seraient souvent sorties du cadre des compétences légalement conférées à la sage-femme.

C'est dans ce contexte très particulier qu'est né mon goût pour la formation dans le domaine de la santé. De manière intuitive j'ai imaginé et mis en œuvre ma première mission de formation ayant pour objectif d'améliorer la pratique des accoucheuses traditionnelles du Camp. Ce fut pour moi l'occasion d'une formidable prise de conscience : très peu de moyens pouvaient suffire à mes actions pédagogiques et les bénéfices étaient rapidement observables.

Quelques mois après mon retour de mission, je m'installais comme sage-femme libérale à Marseille pour créer le premier service de surveillance à domicile, centré sur le dépistage, le suivi médical et l'accompagnement des grossesses pathologiques, en collaboration avec des obstétriciens. Malgré le rapide succès de cette activité et la richesse des contacts humains qu'elle générait, j'avais un autre projet en tête : enseigner.

J'ai donc préparé le concours d'entrée à l'Ecole des Cadres Sages-femmes de Dijon, et personnellement financé une année de formation dans l'unique Institution

française préparant au Certificat Cadre indispensable pour accéder aux fonctions d'enseignante en école de sages-femmes.

Pendant cette formation mes choix de lecture et de stages se sont préférentiellement orientés vers la pédagogie et j'ai choisi comme sujet de mémoire : *La formation des sages-femmes enseignantes de la CEE, proposition d'objectifs communs* . Ce travail m'a initié à la démarche de recherche et m'a permis d'obtenir mon Certificat Cadre en juin 1993 avec les félicitations du jury.

En 1995 j'ai été sollicitée pour occuper un poste de sage-femme enseignante à l'Ecole de Sages-femmes de Marseille, et c'est dans cette Institution que je travaille depuis 9 ans maintenant.

Pendant les 6 premières années je me suis investie dans une mission exclusivement ciblée sur l'encadrement et l'évaluation des étudiants en salle d'accouchement. Je fut rapidement confrontée aux difficultés de l'évaluation : construction des grilles, sélection des critères de jugement, pondération, etc.... J'ai alors pris conscience de la nécessité d'acquérir d'autres connaissances et de disposer de nouveaux outils pédagogiques .

C'est dans ce but que je me suis inscrite en Maîtrise Universitaire de Pédagogie des Sciences de la Santé (MUPSS) au Département de Pédagogie des Sciences de la Santé (DPSS) de la Faculté de Médecine de Bobigny. Tout au long de cette formation j'ai apprécié la souplesse de son organisation qui malgré l'éloignement la rendait compatible avec mon activité professionnelle et ma vie privée. J'ai aussi apprécié la qualité des enseignements reçus, la disponibilité et l'ouverture d'esprit des formateurs et par-dessus tout, le respect de mon propre rythme d'apprentissage. Au DPSS, on ne *suit* pas une formation, on la *vit*. C'est certainement cette nuance qui donne à la MUPSS toute son originalité, sa spécificité et son efficacité. Dès la première année, la participation active aux séminaires, les nombreux travaux personnels à rendre et les lectures conseillées, m'ont incité et m'ont aidé à mettre en œuvre plusieurs actions pédagogiques innovantes à l'école de sages-femmes et dans les différentes formations extérieures où j'interviens également comme enseignante.

La validation finale de ce cursus par l'obtention de la maîtrise nécessitait mon implication dans un travail de recherche en pédagogie de la santé.

C'est la lecture d'un article publié par Bernard Charlin (2001) dans la revue d'expression francophone *Pédagogie Médicale* qui m'a « soufflé » l'idée de travailler sur *L'évaluation des processus de raisonnement clinique en contexte d'incertitude*. Cet article décrivait un outil d'évaluation appelé Test de Concordance de Script (TCS), dont la finalité n'était pas de juger le résultat du raisonnement clinique dans la résolution d'un problème de santé complexe, mais de juger du processus même de ce raisonnement en le comparant à celui d'experts professionnels. Le TCS m'est apparu comme un outil pertinent et innovant, susceptible d'élargir le champ de l'évaluation clinique en école de sages-femmes. En choisissant ce sujet, j'ai pu concilier deux domaines qui m'intéressent plus particulièrement : la démarche clinique qui est au cœur de toute pratique médicale, et l'évaluation qui est un thème central des pratiques pédagogiques.

J'ai élaboré un protocole de recherche qui a été validé au DPSS en novembre 2003 par le Pr. Rémi Gagnayre, directeur de ce mémoire. Parallèlement j'ai construit et développé un TCS spécifique de notre spécialité qui a la particularité de se situer au carrefour entre trois disciplines : l'obstétrique, la gynécologie et la néonatalogie.

Ce travail a pour objet de situer cette recherche dans le contexte où elle a eu lieu, de présenter les objectifs fixés et les moyens et méthodes mis en œuvre pour les atteindre, de présenter les résultats de l'étude et ses limites, et d'alimenter une discussion sur la pertinence d'introduire le TCS dans une école de sages-femmes : analyse des résultats, difficultés, intérêt du test, et perspectives.

Conformément aux directives reçues, ce mémoire est rédigé sous la forme d'un article qui sera traduit et soumis à la revue anglophone d'éducation médicale *Medical Teacher*.

Tout au long de l'article des notes renvoient le lecteur aux annexes qui viennent compléter et enrichir le texte.

Avertissement

La profession de sage-femme est accessible aux candidats des deux sexes depuis 1985. L'étude qui suit n'a concerné que des femmes, les quelques professionnels ou étudiants masculins ne répondant pas aux critères d'inclusion définis, et ceci sans aucune intention de discrimination. Cependant le genre masculin a souvent été utilisé dans un souci d'allègement de la syntaxe.

PRESENTATION DU TRAVAIL DE RECHERCHE

**Evaluation du raisonnement clinique
des étudiants sages-femmes
par le Test de Concordance de Script**

Introduction

Cet article présente une étude dont le but est d'évaluer les processus de raisonnement clinique des étudiants sages-femmes en fin de cursus par l'introduction d'un Test de Concordance de Script ou TCS (Charlin et al. 2002-b).

Les sages-femmes utilisent quotidiennement **le raisonnement clinique médical** indispensable à la prise en charge de femmes enceintes, de parturientes, d'accouchées et de nouveau-nés. **Le raisonnement clinique médical**, forme particulière de raisonnement fait de processus de pensée et de prise de décision, est caractéristique de l'expertise professionnelle en sciences de la santé (Barrows & Tamblyn 1980, Kassirer 1983, Kassirer 1989, Guilbert 1992, Paolaggi & Coste 2001, Charlin et al. 2003). Comme tous les professionnels de santé, les sages-femmes sont confrontées dans leur pratique à deux types de problèmes médicaux (Guilbert 1992) : a) le plus couramment, il s'agit de problèmes simples où l'application de règles ou de procédures suffit ; b) dans d'autres cas, il peut s'agir de problèmes médicaux en contexte d'incertitude, c'est à dire pour lesquelles une solution n'est pas d'emblée évidente mais nécessite le recours à des opérations intellectuelles successives, de sélection et d'élimination de données qui permettront de définir une conduite à tenir adaptée.

Dans les sciences de la santé, le raisonnement le plus couramment utilisé est de type hypothético-déductif, c'est à dire qu'il va de l'hypothèse vers les faits. Il peut arriver que le clinicien ne trouve pas d'hypothèse, il passe alors à un raisonnement de type inductif qui à l'inverse, part des faits pour aboutir à une l'hypothèse. Cette étude propose d'explorer la forme hypothético-déductive du raisonnement développé par les sages-femmes.

Le concept de raisonnement clinique en situation complexe s'appuie à la fois sur la **théorie hypothético-déductive du raisonnement clinique**, et sur la **théorie des scripts** issue de la psychologie cognitive. Les praticiens disposent de connaissances organisées et reliées entre elles par des réseaux ou liens (Elstein et al.1990, Schmidt et al. 1990, Charlin et al. 2000-b) appelés ici scripts.

Le degré d'organisation des connaissances et la qualité des réseaux qui les relie sont deux éléments essentiels du savoir agir en situation clinique (Bernard 2001). Selon Bordage, les cliniciens les plus performants sont ceux qui possèdent de

nombreux réseaux de connaissances cliniques, qualifiés de réseaux « élaborés » et « compilés » (Bordage & Lemieux 1991, Bordage 1994).

L'élaboration de scripts puis leur activation devient possible pour les étudiants dès leur confrontation avec des cas cliniques authentiques (Schmidt et al. 1990), c'est à dire pendant les stages hospitaliers.

Contexte

- *Evaluation du raisonnement clinique à l'école de sages-femmes*

En France, la profession de sage-femme est inscrite au Code de la Santé Publique en tant que profession médicale ¹. Les textes qui réglementent son exercice, confère à la sage-femme une grande autonomie dans l'accompagnement de la naissance physiologique et dans la prévention des risques (nombreuses possibilités d'investigations complémentaires). Face à la pathologie suspectée ou déclarée, elle doit informer immédiatement l'obstétricien et collaborer avec lui, mais lorsqu'elle se trouve confrontée à l'urgence, elle a le devoir d'agir en attendant son arrivée ou celle du pédiatre (pratique de toutes les manœuvres obstétricales et de la réanimation néonatale) ².

La formation de sage-femme est dispensée dans l'une des 35 écoles de sages-femmes de métropole et outre-mer ³. Les étudiants sages-femmes sont recrutés à l'issue de la première année validée du Premier Cycle des Etudes Médicales, PCEM1 ⁴. Ils poursuivent ensuite 4 ans d'études (deux phases de deux ans chacune). Un peu plus de 60 % du temps total de formation est consacré à l'apprentissage en milieu clinique auprès de professionnels qui gèrent seuls ou en équipe des problèmes médicaux des plus simples aux plus complexes. L'apprentissage des différents temps de la démarche clinique se fait au cours des stages et lors des encadrements individuels formatifs organisés et assurés par les sages-femmes enseignantes qui se déplacent sur les terrains de stage.

¹ Code de la Santé Publique, titre 1^{er} - Livre IV

² Code de Déontologie de la Sage-Femme

³ Arrêté du 11/12/2001 fixant le programme des études de sage-femme

⁴ Loi 2002-73, art.70

En conformité avec les dispositions législatives en vigueur, les études sont validées par des épreuves organisées en cours d'année (contrôle continu) et en fin de phase⁵. Le tableau 1 résume et présente les types d'épreuves utilisées pour l'évaluation sommative en école de sages-femmes. Il s'agit quasi exclusivement d'évaluation interne, sous la responsabilité des enseignants de l'école.

Type d'épreuves	Nature	Année d'étude	Durée	Contenu	Choix du sujet ou choix du patient par	Évalué par
Epreuves Ecrites : question de cours, CROQ, QCM, question de synthèse, ou cas clinique	Contrôle continu Théorique (8 à 10 /an)	1-2-3-4	1 h	Chaque matière enseignée	Le professeur qui a fait le cours	Le professeur qui a fait le cours
	Examen Diplôme d'Etat	2 4	De 1h à 3 h	Matières principales : Gynécologie Obstétrique Pédiatrie Sciences humaines et sociales Droit et Législation	Le professeur responsable des enseignements dans la discipline concernée	Le professeur responsable des enseignements dans la discipline concernée
Epreuves cliniques : Recueil de données + examen clinique + raisonnement et présentation orale de la conduite à tenir	Contrôle continu	1-2-3-4	1 h	Sur femme enceinte, accouchée, ou nouveau né.	Les sages-femmes enseignantes de l'école + les sages-femmes du terrain où à lieu l'évaluation	Les sages-femmes enseignantes de l'école
	Examen Diplôme d'Etat	2 4	1 h	Sur femme enceinte, accouchée, ou nouveau né.	2 sages-femmes : 1 enseignante + 1 professionnelle	1 sage-femme enseignante + 1 obstétricien ou 1 pédiatre
Epreuves Orales : Préparation et présentation d'une question	Diplôme d'Etat	4	15 mn	Obstétrique Législation professionnelle	Tirage au sort d'une question parmi une liste connue des étudiants	1 obstétricien + 1 sage-femme enseignante 2 sages-femmes : 1 enseignante + 1 professionnelle
Mémoire : Recherche et rédaction. <i>Soutenance</i>	Diplôme d'Etat	3 –4 4	 20 mn	En rapport avec la profession de sage-femme	L'étudiant	<i>Un Jury composé de sages-femmes et médecins</i>

Tableau 1 : Epreuves organisées pour l'évaluation sommative des étudiants à l'école de sages-femmes de Marseille

Ces épreuves se complètent et visent l'évaluation de connaissances et de compétences variées. Chacune présente des avantages et des inconvénients

⁵ Arrêté du 11 décembre 2001 relatif au contrôle des connaissances et des aptitudes des étudiants sages-femmes et à l'organisation des examens – Journal Officiel 303 du 30 décembre 2001 page 21456.

(Jouquan 2002) notamment en matière d'évaluation du raisonnement clinique (Charlin et al. 2003).

Les épreuves orales reflètent bien la qualité du raisonnement de l'étudiant mais elles sont discutables en terme de fidélité car l'étudiant peut être intimidé ou stressé, et l'évaluateur indulgent ou sévère.

Les sujets proposés à l'écrit renseignent souvent sur les capacités rédactionnelles de l'étudiant et ses aptitudes à mémoriser des connaissances, plus rarement sur ses compétences à gérer un problème clinique (étude de cas) éventuellement en situation de stress.

Les épreuves cliniques organisées tout au long du curriculum offrent l'avantage d'observer l'étudiant face à une situation complexe authentique présentée par une patiente ou un nouveau-né et renseignent sur les niveaux acquis en terme de savoir, de savoir-faire, de savoir-être et effectivement de savoir-raisonner car elles sollicitent systématiquement la forme hypothético-déductive de raisonnement de l'étudiant. Cependant ce sont des épreuves longues (1 heure et plus par étudiant) qui ne permettent pas d'évaluer chacun des apprenants sur plus de 4 situations cliniques différentes par an. De plus, il est reconnu que ce type d'épreuve présente des limites de fidélité, de spécificité de contenu et d'équité entre les candidats : les patients ne sont pas standardisés et les cas proposés à chaque étudiant présentent rarement un niveau de réflexion ou de résolution équivalents. Comme lors des épreuves orales, l'étudiant peut se trouver en difficulté face à l'évaluateur.

- *Le Test de Concordance de Script*

Aucune épreuve n'étant actuellement prévue pour explorer les processus de raisonnement clinique des étudiants sages-femmes en contexte d'incertitude, il est proposé ici d'introduire le Test de Concordance de Script (TCS) conçu par Charlin (1998-2004). L'originalité du TCS est de noter et comparer les réponses données par l'étudiant en fonctions des réponses de professionnels expérimentés, d'où le terme de concordance. Contrairement aux épreuves qui permettent d'évaluer le résultat de leur raisonnement, comme les questions à choix multiple (QCM) par exemple, le TCS permet d'évaluer les processus de raisonnement nécessaires à la résolution de problèmes cliniques « mal définis », encore appelés problèmes « complexes » ou « en contexte d'incertitude ».

Le TCS est un instrument encore récent puisque les premières publications sur le sujet ont six ans (Charlin 1998-a&b). Depuis, plusieurs responsables de formation et enseignants chercheurs (Belgique, Canada, France, USA, Thaïlande) ont montré l'intérêt du TCS en formation médicale continue (Charlin 1999, Labelle 2003) et initiale. Des travaux de recherche touchant à des disciplines médicales variées figurent dans la littérature. C'est le cas en urologie (Sibert et al.2001-2002), éthique médicale (Llorca et al. 2003), radiologie (Charlin et al. 1998-a, Brazeau-Lamontagne et al. 2003), gynécologie (Charlin et al. 2002-a), médecine d'urgence (Birlouez 2001), neurochirurgie (Caire et al. 2004-a&b), rééducation fonctionnelle (Cohen et al. 2003), pharmacie (Khonputsa et al. 2003). Ces recherches ont contribué à mettre en évidence les caractéristiques psychométriques du TCS.

La validité de construit a été montrée dans plusieurs études : le test discrimine significativement le niveau d'expérience entre médecins de niveaux d'expérience différents : externes, internes, chefs de clinique et séniors (Charlin et al. 1998-a&b, Sibert et al. 2001). **La validité prédictive** du test montre que les étudiants qui organisent bien leurs connaissances à un stade de leur formation continuent à le faire ultérieurement (Brailovsky et al. 2001). **La stabilité des scores en fonction des cultures** a été mise en évidence par Sibert et al. (2002) dans une étude menée en urologie auprès de médecins français et canadiens anglophones : bien que présentant une plus grande concordance dans la culture d'origine, le TCS conserve ses capacités de discrimination dans l'autre culture.

L'ensemble des études publiées jusqu'ici sur le TCS le décrivent comme **un instrument valide pour analyser les processus de raisonnement clinique en contexte d'incertitude.**

A notre connaissance aucun TCS n'a jusqu'ici été développé pour des sages-femmes, alors qu'en pratique ces dernières utilisent systématiquement le raisonnement hypothético-déductif et gèrent régulièrement des situations complexes présentées par les patientes ou les nouveau-nés.

Finalité

Dans le cadre d'un processus d'amélioration global de notre système de formation et sous réserve que le TCS s'avère être un outil pertinent pour des sages-femmes, **la**

finalité de cette étude serait d'introduire le TCS à l'école de sages-femmes, comme une évaluation complémentaire, objective et standardisée des processus de raisonnement clinique des étudiants, dans la résolution de problèmes cliniques authentiques. Cette mesure devrait apporter plus de cohérence à l'ensemble du dispositif d'évaluation de la formation initiale sage-femme. L'introduction du TCS pourrait également développer la prise de conscience par les étudiants de leurs propres processus de réflexion en contexte d'incertitude, et faciliter le repérage des éventuelles difficultés à raisonner présentées par certains d'entre eux.

Hypothèses et objectifs de recherche

Le TCS discrimine-t-il le niveau de performance entre étudiants et sages-femmes experts comme cela a été montré pour les médecins ? Le TCS pourrait-il s'intégrer au dispositif de formation et d'évaluation des écoles de sages-femmes ? Et si oui, comment ? A l'issue de leur formation les sages-femmes développent-elles un raisonnement proche de celui développé par les professionnelles expérimentées ?

Deux hypothèses ont été émises.

Première hypothèse : le TCS discrimine clairement le niveau de performance entre étudiants sages-femmes et sages-femmes expérimentées.

Ceci revient à montrer la validité de l'outil dans ce contexte professionnel particulier puisque cela n'a jamais été fait jusqu'ici. Les objectifs sont : « *construire un TCS adapté aux sages-femmes hospitalières* », puisque c'est dans un environnement essentiellement hospitalier (secteurs publics et privés confondus) que pratiquent les sages-femmes et que sont formés les étudiants sages-femmes en France, « *développer ce TCS auprès d'étudiants sages-femmes de troisième et quatrième années* » et « *rechercher une différence de résultat significative entre les étudiants et les sages-femmes expérimentées* ».

Seconde hypothèse : il existe une corrélation significative mais non absolue entre les résultats obtenus par les étudiants sages-femmes aux épreuves habituelles et au TCS.

Ceci revient à définir le lien qui existe entre les résultats obtenus par les étudiants au TCS et aux autres épreuves. Les objectifs sont : « *comparer les résultats entre*

*épreuves cliniques traditionnellement organisées et TCS », « comparer les résultats entre étape spécifique de raisonnement de ces épreuves cliniques et TCS »
« comparer les résultats obtenus entre moyenne générale théorique et TCS ».*

Matériels et méthode

- *Cadre de la recherche*

Cette recherche-action évaluative et prospective, a été menée dans le cadre de la Maîtrise Universitaire de Pédagogie des Sciences de la Santé, de l'UFR Santé, Médecine et Biologie Humaine Léonard de Vinci de l'Université de Paris 13 à Bobigny.

- *Date et lieu*

L'étude s'est déroulée dans le courant du premier semestre 2004 à l'Ecole Régionale de Sages-femmes de Marseille, relevant conjointement de l'Assistance Publique des Hôpitaux de Marseille et de la Faculté de Médecine (secteur nord), Université de la Méditerranée.

- *Population*

La population étudiée est représentée par 3 groupes de sages-femmes de niveaux d'expérience différents : 1) 50 sages-femmes hospitalières en activité dans des établissements publics ou privés ont été réparties de manière aléatoire en deux sous-groupes : a) 20 ont été tirées au sort pour constituer le panel de référence nécessaire pour établir les scores des trois groupes suivants, b) les 30 restantes ont constitué un groupe témoin servant à mesurer l'écart entre étudiantes et praticiennes expérimentées, 2) 29 étudiantes sages-femmes de 4^{ème} année (ESF4) qui seront diplômées quelques semaines plus tard ; 3) 35 étudiantes de 3^{ème} année (ESF3) qui seront autorisées à effectuer des replacements en tant que sage-femme dès la validation de cette troisième année ⁶.

Toutes les personnes concernées par l'étude présentent donc un niveau de compétence officiellement reconnu et défini. Selon une définition graduée de

⁶ Circulaire du 9 juin 2004 du ministère de la Santé et de la Protection Sociale

l'expertise proposée par Schmidt et al. (1990, in Guilbert 1992), les étudiantes peuvent être considérées comme des sages-femmes de niveau intermédiaire 1 en troisième année et niveau intermédiaire 2 en quatrième année. Selon ces mêmes auteurs les professionnelles expertes sont celles qui « *ont une utilisation efficiente de l'ensemble structuré de connaissances ou d'expériences antérieures* ». Selon Benner (1984) dont les travaux en sciences de la santé sont fondés sur le modèle de différenciation des étapes du cheminement vers l'expertise de Dreyfus & Dreyfus (Benner 1984, Frenette-Leclerc 1992), les professionnelles seraient expertes au bout de 5 ans de pratique continue dans les mêmes fonctions. Ainsi les sages-femmes qui ont participé à cette étude devaient toutes justifier de 5 ans de service au moins dans une maternité publique ou privée accueillant régulièrement les étudiants sages-femmes en stage. En collaboration avec leur encadrement, une liste des praticiennes susceptibles de répondre à ces critères avait été établie en décembre 2003. Cent trente lettres expliquant les objectifs de l'étude ont ainsi été nominativement adressées de décembre 2003 à janvier 2004 afin de solliciter une participation volontaire et bénévole. 43 % des personnes sollicitées ont accepté de participer à l'étude. Le recrutement a pris fin le 28 février 2004.

A l'issue d'une décision collégiale au sein de l'Institution, le test a été présenté aux étudiants comme une épreuve d'obstétrique obligatoire, mais dont chacun pourrait choisir d'intégrer ou non la note obtenue à sa moyenne de contrôle continu en obstétrique. Cette décision ayant pour double objectif d'obtenir le maximum de participants pour l'étude tout en ne pénalisant aucun d'entre eux.

Une réunion a été organisée pour informer les étudiants. Il a été annoncé à tous les participants que les résultats de la recherche leur seraient communiqués.

- *Présentation du Test de Concordance de Script*

Le test se présente sous la forme d'une série de « vignettes ». Chaque vignette (figure 1) décrit une courte situation clinique (scénario clinique) qui comporte des données incomplètes, contradictoires, ou éventuellement ambiguës⁷. Cette situation est suivie d'un ou plusieurs items. Chaque item se compose lui même d'une proposition cohérente – une hypothèse diagnostique par exemple – puis, d'une nouvelle donnée.

⁷ voir aussi schéma en Annexe I

Scénario clinique : Une jeune femme de 22 ans, accompagnée de son mari, consulte pour abondantes métrorragies de sang rouge avec douleurs abdominales. Elle ne prend aucune contraception depuis 6 mois car ils désirent un enfant. Elle déclare avoir eu ses règles il y a 6 semaines.	
Si vous pensez à : Une grossesse extra utérine	Et que vous trouvez à l'échographie : Une image évoquant la présence de caillots intra-utérins
Cette nouvelle donnée rend votre hypothèse diagnostique : -2 quasi improbable -1 moins probable 0 ni plus ni moins probable +1 plus probable +2 quasi certaine	
Entourez votre réponse : -2 -1 0 +1 +2	

Figure 1 : Exemple d'une vignette du test

La personne qui passe le test doit se prononcer sur l'effet de la nouvelle donnée sur l'hypothèse en choisissant l'un des 5 niveaux d'ancrage possible sur une échelle de Likert. Le sens de lecture des vignettes est intuitif : de haut en bas et de gauche à droite. Les vignettes présentent des situations cliniques authentiques pouvant respectivement relever des dimensions diagnostique, pronostique, investigations complémentaires et thérapeutique. Le pronostic est considéré pour de nombreux auteurs comme étant un diagnostic évolutif (Le Maout 1999). En obstétrique le pronostic revêt une importance particulière, c'est pourquoi il a été choisi de faire apparaître dans ce test plusieurs vignettes de dimension pronostique. La terminologie utilisée pour la déclinaison sur les 5 niveaux de l'échelle diffère légèrement d'une dimension à l'autre (tableau 2).

Niveau sur L'échelle	Dimension Pronostique	Dimension Diagnostique	Dimension Investigations	Dimension Thérapeutique
	<i>Cette nouvelle donnée rend le pronostic maternel* :</i>	<i>Cette nouvelle donnée rend l'hypothèse diagnostique :</i>	<i>Cette nouvelle donnée Rend l'examen proposé :</i>	<i>Cette nouvelle donnée rend Le traitement** proposé :</i>
+2	Beaucoup plus péjoratif	Quasi certaine	Indispensable	Indispensable
+1	Plus péjoratif	Plus probable	Plus utile, plus pertinent	Plus utile, plus pertinent
0	Inchangé	Ni plus ni moins probable	Ni plus ni moins utile	Ni plus ni moins utile
-1	Plus favorable	Moins probable	Moins utile, moins pertinent	Peu utile, peu pertinent
-2	Beaucoup plus favorable	Quasi improbable	Inutile	Contre-indiqué

* ou obstétrical, ou fœtal, ou néonatal selon le cas

** ou geste

Tableau 2 : Déclinaison proposée pour chaque niveau de l'échelle de Likert, en fonction de la dimension explorée (diagnostique, pronostique, investigations, thérapeutique)

- *Méthodologie*

1) Construction du test : de juin à novembre 2003

Afin de recueillir les situations cliniques réelles, une grille de recueil de données a été élaborée et soumise à Bernard Charlin qui l'a validée (figure 2)⁸.

Recueil des données cliniques significatives :
Hypothèses diagnostiques les plus probables (4 maximum) :
Investigations complémentaires pertinentes (4 maximum) :
Conduites thérapeutiques adaptées : traitements, soins, surveillance (les 4 premières) :
Facteurs susceptibles d'influencer le pronostic (les plus importants) :

Figure 2 : Grille de recueil du matériel nécessaire à la construction de TCS
Situations cliniques avec incertitude

Avec le concours de trois autres sages-femmes enseignantes, des situations médicales réelles ont été recueillies jour après jour en maternité. A partir de ce matériel 84 vignettes ont été construites de façon à couvrir aussi largement que possible les domaines de compétences de la sage-femme hospitalière (tableau 3)⁹. Cette table de spécification garantie la validité de contenu du test (Messick 1995) dans le contexte où il est développé : en secteur d'hospitalier.

⁸ fiche grandeur réelle en Annexe II

⁹ table succinctement présentée ici, reprise en détail en Annexe III

Domaine De compétences	Dimensions explorées en ce domaine	Nombres de vignettes cliniques portant sur les connaissances en ce domaine
<i>Surveillance clinique et para-clinique de la grossesse</i>	Diagnostic/Pronostic Investigation Thérapeutique	43
<i>Surveillance clinique et para-clinique du travail</i>	Diagnostic/Pronostic Thérapeutique	12
<i>Pratique de l'accouchement</i>	Diagnostic/Pronostic Thérapeutique	6
<i>Pratique des manœuvres obstétricales</i>	Pronostic Thérapeutique	6
<i>Pratique des techniques obstétricales</i>	Diagnostic Thérapeutique	7
<i>Surveillance clinique et para-clinique du nouveau-né</i>	Diagnostic/Pronostic Investigation Thérapeutique	9
<i>Surveillance clinique et para-clinique des suites de couches</i>	Diagnostic Investigation Thérapeutique	14
<i>Gynécologie</i>	Diagnostic Investigation Thérapeutique	12

NB : Chaque vignette clinique peut concerner plusieurs domaines de compétences

Tableau 3 : Table de spécification des domaines de compétences de la sage-femme hospitalière pris en compte dans ce TCS, dimensions explorées et nombre d'items

A ce stade, un groupe pilote de 3 jeunes diplômées en juin 2003 a été réuni pour tester et réajuster les vignettes. Ces 3 sages-femmes ont passé le test en temps limité, en présence d'un observateur et avec pour consigne de pointer les difficultés rencontrées : difficulté de compréhension de l'énoncé, incohérence des propositions, niveau de résolution inadapté, ou au contraire réponse d'emblée trop évidente. Le temps de passage a pu ainsi être confirmé et tous les commentaires ont été notés. A l'issue de ce test 3 items ont été éliminés et 15 ont été modifiés.

Les 81 vignettes restantes ont été soumises au contrôle qualité d'un groupe d'experts constitué de 2 sages-femmes cliniciennes expertes, 3 sages-femmes enseignantes et la directrice de l'école. Chaque item a été analysé à l'aide de la grille de contrôle élaborée par Bernard Charlin¹⁰. 11 items jugés non pertinents ont été éliminés, fixant à 70 le nombre d'items définitifs pour un test de 60 minutes environ, le temps de réflexion étant estimé à 45 secondes environ par vignette.

¹⁰ voir Annexe IV

2) Passage du test : mars, avril 2004

64 étudiantes (35 ESF3 et 29 ESF 4), et 53 sages-femmes ont passé le test administré sous forme papier & crayon. Chaque participant a reçu un livret de 31 pages comportant : une page d'identification du participant¹¹, 29 pages de vignettes¹², et un court questionnaire d'opinion visant à évaluer l'acceptabilité du test¹³.

Deux sessions ont été organisées à la faculté pour le passage du test par les étudiants. Concernant les professionnelles, il était au départ prévu de leur faire choisir une date parmi 5 ou 6 proposées et de venir passer le test à la faculté. En réalité, il s'est avéré impossible de les réunir par plus de 2 ou 3 et de les faire se déplacer. A l'exception des unités de lieu et de date, les professionnelles ont passé le test dans des conditions semblables à celles des étudiants.

Trois sages-femmes ont abandonné en cours d'épreuve : une pour ne pas en avoir compris le but, une pour être trop souvent en désaccord avec les pratiques du service, et une pour ne pas être arrivée à se concentrer suffisamment. Ceci a ramené le nombre final de participants à 50 sages-femmes également réparties entre les secteurs d'hospitalisation privé et public et à 64 étudiantes.

3) Saisie des données : juin 2004

Les livrets ont été classés, numérotés et codés afin de respecter l'anonymat des participants. Les questionnaires d'opinion sur l'acceptabilité du test ont été regroupés pour être exploités séparément.

Un tirage au sort parmi les 50 professionnelles a permis de définir les 20 membres du panel de référence recommandés par Gagnon et al. (2003) à l'issue d'une recherche sur le nombre d'experts nécessaires pour constituer le panel. Comme les participantes étaient également réparties entre les secteurs d'hospitalisation privé (25) et public (25), il a été décidé de respecter la même proportion dans le panel : 10 sages-femmes du privé et 10 du public¹⁴. Le nombre d'années d'exercice de ces sages-femmes varie de 7 à 32 ans.

¹¹ voir Annexe V

¹² le livret de vignettes TCS fait l'objet d'un « tiré à part » et est exclusivement communiqué aux membres du jury de ce mémoire, pour des raisons évidentes de confidentialité (réutilisation du test)

¹³ voir Annexe VI

¹⁴ voir Annexe VII

4) Traitement et analyse des données : juillet, août 2004.

- *Traitement des données sur le site TCS.*

Toutes les données ont été saisies en ligne sur le site TCS : <http://www.cme.umontreal.ca/tcs>. Le correcteur du site a permis d'établir les clés de correction à partir des réponses de chaque membre du panel de référence et de noter les participants au TCS selon la méthode des scores combinés ou *aggregate scoring method* (in Charlin 2003 : *Norcini 1987*, in Charlin 2004-a : *Norcini 1990*).

Comparée à la méthode par consensus, la méthode des scores combinés est mieux adaptée à l'évaluation en contexte d'incertitude et discrimine mieux les participants selon leur niveau d'expérience (Charlin et al. 2002-a), Elle a pour principe de prendre en compte l'avis de tous les experts et de définir une clé de correction item par item, servant à noter chaque étudiant au prorata du nombre d'expert(s) ayant le même choix de réponse que lui (tableau 4).

Echelle de réponse	- 2	-1	0	+1	+2
Nombre de membre du panel pour ce choix de réponse	1	0	3	5	11
Clé de correction : Score attribué	0.9	0.00	0.27	0.45	1

Tableau 4 : Exemple de clé de correction pour l'item n° 2 de ce TCS
Source : calculé par le correcteur en ligne du site TCS, Université de Montréal

- *Analyse statistique*

Les résultats obtenus sur le site TCS ainsi que les résultats obtenus aux contrôles continus théoriques et cliniques à l'école ont été saisis sur Excel, puis les analyses statistiques ont été réalisées à l'aide du logiciel SPSS.

Résultats

- *Vérification de la première hypothèse :*

Les résultats au TCS sont présentés au tableau 5 .

	ESF3 N = 35	ESF4 N = 29	SF N = 30	TOTAL N = 94
Niveau	Débutantes		Expertes	
Moyenne / 100	68	69,5	71,5	69,5
Ecart Type	6.49	4.05	4.52	5.29
Minimum*	50	60	62	50
Maximum*	78	77	81	81
Coefficient α	0.69	0.28	0.34	0.52

* Le correcteur applique automatiquement une règle d'arrondi des notes

Tableau 5 : Statistiques globales et évaluation du TCS

Chiffres donnés par le correcteur en ligne du site TCS

La moyenne des scores sur 100 est très proche d'un groupe à l'autre. Les étudiants de troisième année (ESF3 - niveau intermédiaire 1) obtiennent au test la moyenne la plus basse (68) et les professionnelles la moyenne la plus élevée (71,5). Le score le plus élevé au test (81/100) est obtenu par une sage-femme expérimentée, le score le plus bas (50/100) est obtenu par une étudiante de troisième année. L'écart type est plus faible pour les étudiants de quatrième année (ESF4 -niveau intermédiaire 2) qui arrivent au terme de leur formation.

Le coefficient α renseigne sur la cohérence interne du test (Cronbach 1951). On estime que le test est pertinent quand le coefficient est $>$ à 0,70. Sur l'ensemble de la population testée, ce TCS présente un coefficient α de Cronbach assez moyen (0.52) qui s'avère optimal pour les troisième années (0.69), et médiocre pour les quatrième années (0.28) et les professionnelles (0.34).

Pour les 3 groupes, les notes présentent une fréquence de distribution normale (courbe de Gauss) , ce qui a permis d'appliquer le test-t de Student et de vérifier que l'écart constaté entre étudiantes et sages-femmes expérimentées n'était pas purement lié au hasard. On trouve $p = 0,04$. Il existe donc une différence

statistiquement significative entre celles qui sont proches d'accéder à la pratique (ESF3 + ESF4) et celles qui pratiquent depuis 5 ans au moins.

→ **La première hypothèse est confirmée : le TCS discrimine significativement le niveau de performance entre étudiantes sages-femmes de niveau intermédiaire et sages-femmes de plus de 5 ans d'expérience.**

- *Vérification de la deuxième hypothèse*

Les notes obtenues par les étudiants aux épreuves théoriques et cliniques traditionnellement utilisées¹⁵ ont été comparées aux notes obtenues au TCS.

Parmi les ESF 3 qui obtiennent les notes les plus élevées au TCS (> 15/ 20) présentées au tableau 6, deux ont effectivement d'excellents résultats par ailleurs (entre 16 et 17). Ceci pourrait laisser penser que les « bons » élèves ont de bons résultats au TCS (cas n° 1 et cas n° 2). Or une étudiante ayant de bons résultats en théorie et clinique a eu tout juste la moyenne au TCS (cas n° 3) , et une étudiante ayant des résultats moyens, notamment en clinique, a bien réussi au TCS (cas n° 4).

	Code ESF	Moyenne Contrôle continu théorique	Moyenne épreuves cliniques	Moyenne raisonnement clinique	Note TCS
Cas n° 1	34VMESF3	16,33	16,33	16,22	15,2
Cas n° 2	35VAESF3	16,89	16,75	17,33	15,6
Cas n° 3	04BAESF3	15,32	16,17	16,00	10
Cas n° 4	26PAESF3	13,33	12,50	11,33	15,2

Tableau 6 : comparaison entre résultat au TCS et aux autres épreuves – ESF3

L'étude des notes obtenues par les ESF4 amène aux mêmes constats bien que le groupe affiche globalement des résultats plus homogènes aussi bien pour le TCS que pour les autres épreuves.

Il semble difficile à priori de trouver un lien entre le niveau moyen de l'étudiant et son résultat au TCS. Afin de le démontrer sur le plan statistique, une recherche de

¹⁵ voir Annexe VII

corrélation intra-classe a été faite. Les coefficients retrouvés sont présentés aux tableaux 7 et 8. Dans les deux groupes ESF3 et ESF4, il existe une corrélation modérée, c'est à dire significative mais non absolue, entre les résultats au contrôles théoriques et aux épreuves cliniques.

	Moyenne des épreuves cliniques	Moyenne du « raisonnement clinique » des épreuves cliniques	Note TCS
Moyenne des contrôles continus	0, 41 C modérée	0, 32 C modérée	0,02 C absente
Moyenne des épreuves cliniques		0,84 C forte	- 0,07 C absente
Moyenne du « raisonnement clinique » des épreuves cliniques			- 0,06 C absente

C = corrélation

Tableau 7 : Présentation du coefficient de corrélation intra classe
Etudiants sages-femmes 3^{ème} année (ESF3)

Concernant les ESF3 il n'existe aucune corrélation entre le résultat obtenu au TCS et les résultats aux contrôles continus cliniques et théoriques.

	Moyenne des épreuves cliniques	Moyenne du « raisonnement clinique » des épreuves cliniques	Note TCS
Moyenne des contrôles continus	0, 40 C modérée	0, 52 C modérée	0,10 C très faible Non significative.
Moyenne des épreuves cliniques		0,83 C forte	0,24 C faible Non significative
Moyenne du « raisonnement clinique » des épreuves cliniques			0,30 C faible Non significative

C = corrélation

Tableau 8 : Présentation du coefficient de corrélation intra classe
Etudiants sages-femmes 4^{ème} année (ESF4)

Concernant les ESF4 il existe une corrélation faible et non significative entre le résultat obtenu au TCS et les résultats aux contrôles continus théoriques et épreuves cliniques.

La forte corrélation établie entre raisonnement clinique et épreuve clinique ne présente pas d'intérêt dans cette étude. Elle s'explique par le fait que les points attribués pour le raisonnement ont été prélevés sur la note globale de la clinique comprenant l'évaluation de l'interrogatoire, l'examen clinique, et le raisonnement clinique (diagnostic, pronostic et conduite à tenir proposée).

→ La deuxième hypothèse n'est donc pas confirmée : il n'existe pas de corrélation significative entre les résultats des étudiants au TCS et aux autres épreuves.

Limites de l'étude

- *La cohérence interne du test n'est pas optimale.*

Depuis la construction de ce test entre juin et novembre 2003, les chercheurs et les méthodologistes ont mis en évidence que décliner 3 items par situation clinique rendait le test plus homogène et augmentait le coefficient de corrélation interne du TCS (α de Cronbach). En appliquant cette règle dans l'élaboration du TCS, nous aurions probablement obtenu une plus grande fiabilité de mesure de l'instrument.

Par ailleurs le test comprend certainement plusieurs items indésirables, c'est à dire ne répondant pas parfaitement aux critères de qualité du test et ce malgré la vérification de chaque vignette par un jury d'experts. Certaines vignettes mériteraient d'être reprises dans leur formulation et notamment quant à l'hypothèse proposée qui, selon le retour de certains évalués, n'est pas forcément la première qui leur vient à l'esprit. Une analyse d'items permettrait de dépister les items indésirables. En supprimant ces derniers, le test gagnerait en cohérence interne et en pertinence.

- *Le choix des experts est discutable*

Pour constituer le groupe panel de référence dont les réponses ont permis d'établir les clés de correction pour chaque item, deux méthodes ont été envisagées. La

première était de sélectionner 20 sages-femmes réputées comme étant d'excellentes professionnelles par leurs paires et leur hiérarchie médicale ; méthode comparable à celle utilisée dans d'autres études TCS (Sibert et al. 2001, Charlin et al. 2002-a, Llorca et al. 2003, Caire et al. 2004a-b). La seconde était de procéder à un tirage au sort entre les professionnelles participant à cette étude, afin de créer 2 groupes de sages-femmes d'expertise équivalente : le panel de référence et le groupe témoin pour servir la première hypothèse. Nous avons retenu cette deuxième option car l'objectif de cette étude n'était pas d'évaluer des professionnelles expérimentées par rapport à un groupe de « super » expertes, mais bien de mesurer l'écart entre des futurs professionnels tout près de rentrer en exercice et des professionnels ayant 5 ans d'expérience au moins. Un biais de sélection peut être évoqué ici puisqu'il est fortement probable qu'en modifiant le panel de référence (par un second tirage au sort par exemple), les résultats du test soient différents et que le coefficient de corrélation interne du test soit modifié. Néanmoins ce second échantillon serait aussi un reflet de la réalité. Une analyse plus précise du profil de chaque expert pour permettre un éventuel rééquilibrage entre les groupes pourrait être envisagée.

- *La recherche de corrélation s'appuie sur des données peu fiables.*

Comme nous l'avons déjà évoqué, les épreuves cliniques que passent actuellement les étudiants offrent peu de garantie de fiabilité (limites psychométriques). Cette étude utilise les notes de cliniques et de raisonnement clinique attribuées lors de ces épreuves. Un biais d'information (voire de mesure) peut être évoqué à ce niveau.

- *Ce test est probablement lié au contexte dans lequel il a été construit.*

Ce test a été construit et développé dans le contexte propre à la formation et à l'exercice des sages-femmes hospitalières à Marseille. Comme cela été fait pour vérifier la stabilité d'un TCS d'urologie entre 2 cultures (Sibert et al. 2002), il serait intéressant de vérifier la stabilité de ce test entre 2 écoles d'obstétrique différentes.

- *L'étude concerne des groupes de faible effectif.*

Les conclusions seront donc émises avec réserve.

Discussion

- *Analyse à partir des résultats*

Comme cela a été montré dans d'autres études et notamment celle de Sibert et al. (2001), le TCS développé ici est un instrument de mesure qui discrimine significativement le niveau d'expertise entre sages-femmes de niveau intermédiaire (ESF3+ESF4) et sages-femmes exerçant depuis plus de 5 ans. Cependant il n'existe pas de différence statistiquement significative entre ESF3 seules et professionnelles, ESF4 seules et professionnelles, ni entre ESF3 et ESF4.

Comme le laisse supposer le faible écart entre les moyennes obtenues au TCS par les 3 groupes, ceci tendrait à confirmer que notre dispositif de formation actuel forme des futures sages-femmes proches des professionnelles dans la résolution de problèmes cliniques en contexte d'incertitude. Ce constat vient conforter la légitimité pour les étudiants d'effectuer des gardes rémunérées de sage-femme dès la fin de la troisième année validée.

Le coefficient α de Cronbach qui permet d'estimer la fiabilité du test est très variable d'un groupe à l'autre. Plusieurs facteurs rendent compte de cette variabilité : la capacité de raisonnement des évalués (plus ou moins élevée d'un groupe à l'autre), la présence d'items indésirables (qui pourraient être déterminés par un processus d'analyse d'items et écartés), mais aussi l'hétérogénéité du groupe panel. Cette étude montre que c'est pour les étudiantes de troisième année que ce TCS est le plus fiable et le plus pertinent puisque c'est pour ce groupe que l'on obtient le meilleur coefficient α de Cronbach, comparable aux coefficients retrouvés dans les autres études portant sur des TCS (Charlin et al. 2004).

Le principe du TCS est d'explorer l'organisation des connaissances et la qualité des liens qui relient ces connaissances entre elles chez un individu, par rapport à un groupe de référence. Mais le coefficient α est calculé sur un ensemble aléatoire d'individualités, rapporté au groupe de référence. Le groupe des ESF3 obtient un coefficient α plus élevé au TCS car les étudiants ont apporté des réponses plus cohérentes. Le groupe des ESF3 semble avoir disposé de connaissances mieux organisées et d'un raisonnement clinique plus performant que les ESF4 alors qu'ils ont moins d'expérience clinique et moins de connaissances théoriques au moment

du test. Ce « déficit » a-t-il incité les étudiants à mobiliser une plus grande réflexion pour résoudre les cas cliniques présentés et finalement mieux réussir le test ?

En troisième année les étudiants ont l'opportunité de faire des stages très diversifiés, dans des maternités de la région, éloignées du Centre Hospitalier Universitaire. C'est pour eux l'occasion de rencontrer des situations très variées et abordées de manière moins académique qu'au CHU. La troisième année est toujours une période d'enrichissement professionnel et de forte progression pour les étudiants.

Les étudiants de quatrième année reviennent souvent vers le CHU, réputé de haut niveau technologique. Ils sont particulièrement préoccupés par la préparation du Diplôme d'Etat et notamment l'épreuve clinique. Notre pratique d'enseignant nous permet de constater qu'à ce stade les étudiants modélisent énormément leur pratique sur les conduites protocolaires. L'obstétrique est une spécialité où les protocoles sont nombreux et ont pour objectif d'appliquer des conduites à tenir consensuelles. Lors du passage d'une épreuve clinique les étudiants sont souvent très intimidés pour développer leur propre raisonnement clinique et argumenter leur conduite à tenir face aux experts qui les évaluent. Ils cherchent à résoudre leur cas clinique en s'appuyant sur une conduite protocolaire qui a de bonnes chances d'être approuvée et appréciée par l'évaluateur. L'enquête d'opinion visant à apprécier l'acceptabilité du TCS a révélé que plusieurs professionnelles travaillant en CHU avait été gênées pour répondre au test. L'une d'entre elle a écrit le commentaire suivant : *« je ne savais pas si je devais répondre ce que je pense personnellement en tant que praticienne responsable, ou si je devais répondre ce que l'on nous impose de faire avec les protocoles »*. Ce témoignage montre l'influence des protocoles sur les pratiques. Nous avons peut-être là une autre piste pour expliquer le faible coefficient α retrouvé au test pour les ESF4.

Ces éléments de réflexion ne permettent néanmoins pas de déterminer si ce sont les individus formant le groupe d'étudiants de troisième année qui raisonnent particulièrement bien ou si il existe des composantes de la formation susceptibles d'inhiber les capacités de raisonnement clinique au cours de la quatrième année. Il serait intéressant de faire repasser ce TCS en avril prochain aux nouveaux ESF3 et aux ESF3 devenus ESF4. Ces derniers auront augmenté leurs connaissances et surtout le nombre de situations cliniques rencontrées : ils devraient logiquement augmenter leur performance au TCS. Si les scores se trouvaient diminués, nous

serions amenés à réfléchir en équipe sur l'approche curriculaire de la clinique et du raisonnement clinique.

- *Difficultés rencontrées*

Des difficultés se sont présentées et ont été surmontées à chacune des étapes de l'élaboration et du développement du test. La construction des vignettes a représenté une charge de travail très importante pour recueillir les situations cliniques authentiques et pour monter les vignettes. Le passage du test par les professionnelles a été plus compliqué que prévu et a exigé de nombreux déplacements sur 5 maternités marseillaises. La saisie des données a été particulièrement fastidieuse (70 réponses multipliées par 114 livrets = 7980 entrées). L'exploitation statistique des données a été soumise à la disponibilité des personnes compétentes.

- *Intérêt du TCS pour la formation des sages-femmes*

Plusieurs éléments viennent favorablement appuyer l'introduction d'un TCS à l'école de sages-femmes. a) Le TCS place les étudiants face à de nombreuses situations cliniques complexes, authentiques et variées, en un temps très court. C'est un bénéfice important en matière d'évaluation car les passages de cliniques individuelles comme nous les concevons actuellement sont particulièrement « chronophages » : pour une promotion de 35 étudiants de 4^{ème} année nous ne parvenons pas à assurer plus de 4 évaluations sommatives avant la clinique du Diplôme d'Etat. b) Le TCS présente l'originalité d'évaluer les étudiants en tenant compte de l'avis de sages-femmes en exercice qui possèdent les compétences attendues. C'est donc une reconnaissance des professionnels au travers de l'évaluation et l'implication de ces derniers dans la formation. c) Le test explore les processus de raisonnement hypothético-déductif des étudiants, actuellement peu et mal explorés par les autres formes d'évaluation utilisées (absence de corrélation significative entre TCS et autres épreuves). Le TCS est donc un instrument d'évaluation complémentaire aux autres. d) Le TCS permet de repérer aussi bien les étudiants qui selon Bordage (1994) possèdent des réseaux de connaissances « élaborés » voire « compilés », que les étudiants qui ont des difficultés à raisonner c'est à dire au plan cognitif, des connaissances peu ou mal organisées, des réseaux « réduits » ou « dispersés ». e) Le TCS présente un niveau de qualité

de mesure rarement égalé en matière d'évaluation : validité de construit, fiabilité (pour les ESF3 surtout), objectivité de notation et équité entre les candidats. f) Le TCS a été globalement bien accepté par l'ensemble des étudiants. L'exploitation partielle du questionnaire d'opinion rempli à la fin du test montre que : 94 % des étudiants (60/64) ont trouvé ce test intéressant pour leur formation. Les raisons évoquées sont les suivantes (commentaire libre) : « *mise en situation, réflexion sur la pratique obstétricale, sur les gestes systématiques et les protocoles ; révision, balayage des connaissances, synthèse, mise en évidence des lacunes ; prendre une décision, prendre conscience de l'étendue de nos compétences ; remise en question, doute, différence entre théorie et pratique* ». Bien que 84 % des étudiants (53/64) déclarent avoir très fortement mobilisé leur concentration, la durée du test a été jugée acceptable par 87,5 % d'entre eux (56/64).

Tous ces éléments permettent de confirmer que le TCS est un outil d'évaluation sommative complémentaire et pertinent dans le curriculum des études de sage-femme.

- *Perspectives*

Deux principaux axes d'actions à mener peuvent être définis. Le premier s'oriente vers un réajustement de notre dispositif de formation consécutif aux résultats de cette étude, le second vers la poursuite des recherches sur le TCS pour les sages-femmes et autres professionnels de santé.

Concernant la formation plusieurs actions seraient pertinentes. a) Expliquer le raisonnement clinique médical (définition, composantes, processus,...) aux étudiants dès le début du curriculum. b) Développer des activités didactiques (ateliers et travaux de groupes) au travers desquels l'étudiant pourra se confronter à ses pairs ou à des experts professionnels dans la résolution de problèmes cliniques, prendre conscience de ses propres processus de raisonnement clinique et développer sa métacognition. c) Proposer des mini TCS formatifs aux étudiants (10 à 20 cas) afin de les habituer à cette forme d'évaluation et pour les entraîner à raisonner dans des situations diverses et variées. d) Organiser une évaluation sommative par TCS en fin de troisième année et dépister les étudiants en difficulté dans le raisonnement clinique et ayant besoin d'une remédiation personnalisée.

Concernant la recherche sur l'approche de concordance de script, plusieurs pistes méritent d'être explorées. a) Refaire passer ce TCS cette année au groupe ESF3 (devenu ESF4) pour vérifier si les étudiants ont amélioré leurs scores. b) Procéder à une analyse d'items visant à éliminer les items indésirables du test pour en augmenter l'homogénéité et la cohérence interne (α de Cronbach augmenté). c) Remplacer certains membres du panel de référence pour rendre le TCS plus discriminant et donc améliorer sa validité de construit (test t de Student significatif). d) Poursuivre la collaboration avec les enseignants chercheurs et les équipes qui mettent en place des TCS pour affiner les méthodes de notation du TCS et les outils électroniques : correcteur en ligne pour les enseignants, auto évaluation en ligne pour étudiants et professionnels (FMC¹⁶), promotion du TCS auprès des responsables de formation d'autres écoles de sages-femmes, mise en commun de vignettes par la création de banques d'items par disciplines médicales.

Conclusion

La première hypothèse de recherche est confirmée : le TCS développé ici discrimine significativement le niveau de performance entre sages-femmes de niveau intermédiaire et sages-femmes de 5 ans d'expérience au moins. La seconde hypothèse est infirmée : il n'existe aucune corrélation entre le niveau moyen des étudiants sages-femmes aux épreuves habituelles et les résultats obtenus au TCS.

Les limites de l'étude ont été détaillées. L'analyse des résultats révèle que les futures sages-femmes sont proches des professionnelles quant à la résolution de problème en contexte d'incertitude mais elle soulève des questions sur l'approche que nous avons de la clinique et du raisonnement clinique dans notre contexte particulier : l'organisation actuelle des enseignements théoriques et pratiques offre-t-elle aux étudiants les meilleures chances d'élaborer des réseaux de connaissances qui en feront des professionnels performants ? Les stages proposés (temps et lieux) incitent-ils les étudiants à véritablement développer leurs processus de raisonnement clinique ? Ne focalisons nous pas trop (étudiants et enseignants) sur l'épreuve clinique du diplôme d'état ?

¹⁶ Formation Médicale Continue

Ces questions nous incitent à poursuivre cette recherche en renouvelant le passage du test en troisième et quatrième année dès le printemps prochain.

Cette étude a permis de montrer que le TCS élaboré ici demande à être amélioré pour augmenter sa fiabilité mais peut déjà être considéré comme un instrument valide, pertinent et complémentaire dans l'évaluation sommative des compétences cliniques des sages-femmes en formation initiale. Son introduction augmenterait la cohérence globale du dispositif d'évaluation puisqu'il évalue une dimension jusqu'ici peu ou mal explorée : le raisonnement clinique des étudiants sages-femmes.

Le rôle des formateurs est d'anticiper les modifications de comportements d'apprentissage des apprenants qui seront directement liées à l'introduction du TCS. Ces modifications contribueront à former les sages-femmes de demain : plus conscientes de la légitimité et de la qualité de leur raisonnement clinique, et donc plus présentes dans les prises de décisions au sein de l'équipe médicale pluridisciplinaire.

Le TCS est un outil récent qui va progressivement s'affiner au travers des travaux de recherche qui le concernent et par l'évolution des connaissances en sciences cognitives dont il est directement issu. L'approche de concordance de script commence à s'étendre des sciences médicales aux autres sciences de la santé. Elle pourra probablement intéresser d'autres domaines professionnels dès lors qu'un processus de raisonnement en contexte d'incertitude précède la prise de décision.

Références

BARROWS H.S. & TAMBLYN R.M., 1980, Problem based learning. An approach to medical education, New York, Springer.

BENNER P., 1984, *From novice to expert, excellence and power in clinical nursing practice*, Addison-Wesley Publishing Company Inc., édition française 2003, *De novice à expert, excellence en soins infirmiers*, Paris, Masson.

BERNARD J-L., REYES P., 2001, Apprendre en médecine, *Pédagogie Médicale*, 2, 163-169 & 235-241

BORDAGE G., LEMIEUX M., 1991, Semantic structures and diagnostic thinking of experts and novices, *Academic Medicine*, 71, 55-61

BORDAGE G., 1994, Elaborated Knowledge : A key to successful Diagnostic Thinking, *Academic Medicine*, 69, 883-885.

- BOURLIEZ C., 2001, Utilisation d'un outil pédagogique original d'évaluation du raisonnement clinique en médecine d'urgence : le test de concordance de script, Thèse pour le diplôme d'état de docteur en médecine, Faculté de médecine de Bobigny Léonard de Vinci, Université Paris 13.
- BRAILOVSKY C., CHARLIN B., BEAUSOLEIL S., COTE S., VAN DER VLEUTEN C., 2001, Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency : an experimental study on the script concordance test, *Medical Education*, 35, 430-436.
- BRAZEAU-LAMONTAGNE L., CHARLIN B., GAGNON R., SAMSON L., VAN DER VLEUTEN C., 2003, Measurement of perception and interpretation skills along radiology training : utility of the script concordance approach, *Medical Teacher*, 2004,26 : 326-332.
- CAIRE F., SOL J-C., MOREAU J-J., ISIDORI P., CHARLIN B., 2004, Auto-évaluation des internes en neurochirurgie par Test de Concordance de Script (TCS) : processus d'élaboration des tests, *Neurochirurgie*, 50, 66-72.
- CAIRE F., SOL J-C., CHARLIN B., ISIDORI P., MOREAU J-J., 2004, Le Test de Concordance de Script (TCS) comme outil d'évaluation formative des internes en neurochirurgie : implantation du test sur Internet à l'échelle nationale, *Pédagogie Médicale*, 5, 87-94.
- CHARLIN B., BRAILOVSKY C.A., BRAZEAU-LAMONTAGNE L., SAMSON L., LEDUC C., 1998-a, Script Questionnaires : their use for assessment of diagnostic knowledge in radiology, *Medical Teacher*, 20, 567-571.
- CHARLIN B., BRAILOVSKY C.A., LEDUC C., BLOUIN D., 1998-b, The Diagnostic Script Questionnaire : A new tool to assess a specific dimension of clinical competence, *Advances in Health Sciences Education*, 20, 51-58.
- CHARLIN B., 1999, L'apport des tests de concordance de scripts en formation médicale continue, *L'organisateur d'EMC*, 12, 6-9.
- CHARLIN B., ROY L., BRAILOVSKY C., VAN DER VLEUTEN C., 2000-a, The Script Concordance Test, a tool to assess the reflective clinician, *Teaching and Learning in Medicine*, 12, 189-195.
- CHARLIN B., TARDIF J., BOSHUIZEN H., 2000-b, Scripts and medical diagnosis knowledge : theory and applications for clinical reasoning instruction and research, *Academic Medicine*, 75, 182-190.
- CHARLIN B., DESAULNIERS M., GAGNON R., BLOUIN D., VAN DER VLEUTEN C., 2002-a, Comparison of an aggregate scoring method with a consensus scoring method in a measure of clinical reasoning capacity, *Teaching and learning in medicine* 14, 150-156.
- CHARLIN B., GAGNON R., SIBERT L., VAN DER VLEUTEN C., 2002-b, Le test de concordance de script, un instrument d'évaluation du raisonnement clinique, *Pédagogie Médicale*, 3, 135-144.
- CHARLIN B., BORDAGE G., VAN DER VLEUTEN C., 2003, L'évaluation du raisonnement clinique, *Pédagogie Médicale*, 4, 42-52.
- CHARLIN B., VAN DER VLEUTEN C., 2004-a, Standardised assessment of reasoning in contexts of uncertainty : the script concordance approach. *Evaluation and the Health professions*, 27, 3, 304-319.
- CHARLIN B., GAGNON R., 2004-b, The effect of variability of answers among criterion experts to detect expertise with a SCT, *Academic Medicine*, submitted.
- CRONBACH L.J., 1951, Coefficient alpha and the internal structure of tests, *Psychometrika*, 16, 3, 297-334.
- ELSTEIN A.S., SHULMAN L.S., SPRAFKA S.A. (1990), Medical problem solving, a ten years retrospective, *Evaluation and the Health Professions*, 13, 5-36.

- FRENETTE-LECLERC C-A., 1992, Sur la route de l'expertise, *Nursing Québec*, 12, 1, 48-54.
- GAGNON R., CHARLIN B., COLETTI M., SAUVE E., VAN DER VLEUTEN C., 2003, Assessment in context of uncertainty : how many experts are needed on the Script Concordance Test panel of reference ? *Medical Education*, submitted.
- GUILBERT J-J., 1992, *Comment raisonnent les médecins, réflexions sur la formation médicale*, Genève, Editions médecine et hygiène.
- JOUQUAN J., 2002, L'évaluation des apprentissages des étudiants en formation initiale médicale, *Pédagogie Médicale*, 3, 38-51.
- KASSIRER J.P., 1983, Teaching Clinical Medicine by Iterative Hypothesis Testing, *The new England Journal of Medicine*, 309, 921-923.
- KASSIRER J.P., 1989, Diagnostic Reasoning, *Annals of Internal Medicine*, 110, 893-900.
- KHONPUTSA P., BESINQUE K., FISHER D., GONG W.C., 2003, The Pharmacy Script Concordance Test : a pilot study assessing clinical pharmacy competence in Thailand, *Medical Teacher*, submitted.
- LABELLE M., GAGNON R., THIVERGE R., LAPRISE R., SAINTE-MARIE L-G., CHARLIN B., 2003, Formation continue en petits groupes sur l'ostéoporose : comparaison d'un atelier basé sur le test de concordance de scripts (TCS) et d'un atelier classique, *Pédagogie Médicale*, 4, 145-153.
- LE MAOUT G., 1999, *Le raisonnement médical en gynécologie obstétrique*, Paris, Maloine.
- LLORCA G., ROY P., RICHE B., 2003, Evaluation de résolution de problèmes mal définis en éthique clinique : variation des scores selon les méthodes de correction et les caractéristiques des jurys. *Pédagogie Médicale*, 2, 80-87.
- MESSICK S., 1995, Standards of validity and the validity of standards in performance assessment, *Educational Measurement*, 5-8.
- PAOLAGGI J-B, COSTE, 2001, *Le raisonnement médical de la science à la pratique clinique*, Paris, Estem.
- SCHMIDT H.G., NORMAN G.R., BOSHUIZEN H., 1990, A cognitive perspective on medical expertise : theory and implication, *Academic Medicine*, 65, 611-21.
- SIBERT L., CHARLIN B., GAGNON R., CORCOS J., KHALAF A., GRISE P. 2001, Evaluation du raisonnement clinique en urologie, *Progrès en Urologie*, 11, 1213-1219.
- SIBERT L., CHARLIN B., CORCOS J., LECHEVALLIER J., GRISE P., 2002, Assessment of clinical reasoning competence in urology with the Script Concordance Test : an exploratory study on two sites from different countries, *European Urology*, 41, 227-233.

EPILOGUE

Les bénéfices retirés de cette expérience sont à la fois institutionnels et personnels. Sur le plan de l'Institution, un nouvel outil d'évaluation a été construit, en collaboration avec des membres de l'équipe de sages-femmes enseignantes et des sages-femmes exerçant sur les terrains d'accueil des étudiants.

Les soixante-dix vignettes élaborées constituent une base de situations cliniques authentiques en contexte d'incertitude. A cette base pourront s'ajouter à l'avenir d'autres situations recueillies à l'occasion des encadrements formatifs ou des évaluations cliniques sur le terrain. La diversité des situations rencontrées au fil du temps permettra au test de mieux correspondre à la table de spécification dont quelques points n'ont pu être abordés dans ce test du fait de vignettes éliminées lors du contrôle qualité par le jury d'experts (anémie, hémorragie de la délivrance, ...).

En sollicitant les professionnels, ce test a permis de les reconnaître en tant que référents dans l'évaluation qui est habituellement « l'affaire » de l'école et de quelques cadres sages-femmes qui y sont ponctuellement associées. Cette démarche a été globalement estimée comme originale et novatrice par les sages-femmes praticiennes qui l'ont manifesté par leur participation volontaire et bénévole et par leurs commentaires pertinents (questionnaire d'opinion).

Le fait que l'école envisage d'évaluer le processus de raisonnements mis en œuvre par les étudiantes a été perçu par les sages-femmes comme la reconnaissance de l'existence de ce raisonnement, et par là même une reconnaissance du caractère médical de leurs compétences et du haut niveau de leurs responsabilités.

L'analyse des résultats apporte à l'ensemble de l'équipe quelques points de réflexion sur notre dispositif de formation. Ceux-ci pourront être repris après que le test ait été réadministré aux étudiants de troisième et quatrième année. En fonction de cette deuxième série de résultats, nous serons en mesure de juger des réajustements nécessaires concernant l'approche de la clinique et du raisonnement clinique dans notre Institution.

Construire et développer ce test m'a demandé un investissement personnel important. J'avais heureusement prévu un délai suffisamment long entre la soumission de mon sujet de recherche et sa validation. Ceci m'a permis de ne jamais

être dans l'urgence en poursuivant par ailleurs de nombreuses autres activités. L'élaboration des items m'a semblée particulièrement fastidieuse, attachant sans doute trop d'importance à la description de chaque scénario clinique et à des détails de mise en forme.

Concernant la rédaction des vignettes, la déclinaison sur l'échelle de Likert et la présentation (que j'ai personnellement allégée par rapport au modèle initial) : j'ai travaillé lentement et de manière relativement isolée. Je n'ai sollicité l'aide de mes collègues que très ponctuellement : au début pour recueillir quelques cas cliniques, pour élaborer ensuite les quelques vignettes de néonatalogie, et finalement à un stade très avancé du travail pour le contrôle qualité de l'ensemble des vignettes. Le recrutement des professionnelles m'a contraint à de nombreux déplacements sur le terrain pour obtenir leur participation, expliquer le déroulement du test, veiller au respect de ses modalités et récupérer les livrets.

Cette expérience m'a fait prendre conscience que la construction d'un TCS est un travail important nécessitant l'engagement de plusieurs personnes. Une plus grande implication de notre équipe aurait certainement permis au test de gagner en qualité, mais il s'agissait au départ d'un projet de recherche personnelle et non d'un projet institutionnel.

La forte implication personnelle nécessaire à ce travail fut pour moi très enrichissante. Elle m'a permis d'accroître mes connaissances en éducation médicale en général (psychologie cognitive, processus de raisonnement clinique, processus d'élaboration des connaissances, ...) et d'acquérir une bonne connaissance du TCS en particulier . Le recrutement des sages-femmes m'a incité à venir les rencontrer sur leur lieux d'exercice, ce que je suis moins fréquemment amenée à faire depuis que je m'occupe des étudiants de première année. Ce fut un bon exercice de communication. J'ai également pu échanger (via Internet) avec plusieurs personnes travaillant sur les TCS et apporter ma modeste contribution dans l'approche de concordance de script, plus particulièrement sur la mise au point de quelques outils annexes (grille de recueil des situations cliniques, format des vignettes, déclinaison sur l'échelle, ...) et sur l'amélioration du site TCS dont je figure parmi les premiers utilisateurs.

Au delà de la problématique purement liée à l'approche de concordance de script, ce travail m'oriente plus particulièrement vers deux thèmes de réflexion personnelle :

- Expérience et expertise : Le cumul d'expérience suffit-il pour accéder à l'expertise ? Qui sont les experts ? Quels critères d'expertise retenir ?
- Les conduites protocolaires : Quelle influence (ou quels « effets secondaires ») les protocoles ont-ils sur la pratique des cliniciens ? Ne confond-on pas protocole et procédure ?

Comme je m'y suis engagée, je vais maintenant communiquer les résultats de ce travail à tous les participants qui en ont manifesté le désir.

Les résultats de l'étude m'encouragent à poursuivre cette recherche en renouvelant le passage du test sur les promotions suivantes, mais aussi à l'approfondir en procédant à une analyse d'items d'une part et à des modifications du panel de référence d'autre part. Parallèlement à ce travail, dans le cadre du mémoire de fin de séminaire « Apprentissage et Tutorat » je suis actuellement en train d'analyser un entretien d'explicitation réalisé auprès d'une étudiante venant de passer un TCS.

J'envisage également d'exploiter les questionnaires d'opinion remplis par les étudiants et les professionnelles afin de définir et comparer l'acceptabilité du test dans les deux groupes.

Malgré les quelques difficultés rencontrées et surmontées, mon enthousiasme reste entier. Comparativement au stade où j'ai découvert le TCS au travers d'un article, je dispose aujourd'hui de plusieurs arguments permettant de le présenter comme un outil d'évaluation particulièrement intéressant dans la seconde phase de la formation initiale sage-femme et certainement aussi en formation continue. C'est pourquoi dans un deuxième temps (après l'analyse d'items), je souhaiterais promouvoir le TCS auprès de mes collègues enseignantes d'autres écoles de sages-femmes : faire découvrir le TCS et l'apport qu'il peut représenter dans nos dispositifs d'évaluation. La finalité de cette démarche serait de créer une banque d'items d'obstétrique, à la disposition des écoles qui voudraient l'utiliser et l'enrichir par la création et la mise en commun de nouvelles vignettes. Ce projet pourrait se réaliser dans la création prochaine d'une « plateforme » et d'un site de formation en ligne, sur l'initiative de quelques enseignants chercheurs de facultés de médecine francophones qui m'ont proposé de collaborer avec eux.

Ce travail se trouve à une période charnière de mes propres apprentissages puisqu'il arrive à la fin d'un cycle de formation en pédagogie des sciences de la santé et au commencement d'un troisième cycle en sciences de l'éducation.

Poursuivre mon parcours par l'acquisition de nouvelles connaissances et par mon engagement dans la recherche en éducation médicale contribuera à m'épanouir professionnellement dans un domaine qui me passionne.

BIBLIOGRAPHIE

BARROWS H.S. & TAMBLYN R.M., 1980, Problem based learning. An approach to medical education, New York, Springer.

BENNER P., 1984, *From novice to expert, excellence and power in clinical nursing practice*, Addison-Wesley Publishing Company Inc., édition française 2003, *De novice à expert, excellence en soins infirmiers*, Paris, Masson.

BERNARD J-L., REYES P., 2001, Apprendre en médecine, *Pédagogie Médicale*, 2, 163-169 & 235-241

BORDAGE G., LEMIEUX M., 1991, Semantic structures and diagnostic thinking of experts and novices, *Academic Medicine*, 71, 55-61

BORDAGE G., 1994, Elaborated Knowledge : A key to successful Diagnostic Thinking, *Academic Medicine*, 69, 883-885.

BOURLIEZ C., 2001, Utilisation d'un outil pédagogique original d'évaluation du raisonnement clinique en médecine d'urgence : le test de concordance de script, Thèse pour le diplôme d'état de docteur en médecine, Faculté de médecine de Bobigny Léonard de Vinci, Université Paris 13.

BRAILOVSKY C., CHARLIN B., BEAUSOLEIL S., COTE S., VAN DER VLEUTEN C., 2001, Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency : an experimental study on the script concordance test, *Medical Education*, 35, 430-436.

BRAZEAU-LAMONTAGNE L., CHARLIN B., GAGNON R., SAMSON L., VAN DER VLEUTEN C., 2003, Measurement of perception and interpretation skills along radiology training : utility of the script concordance approach, *Medical Teacher*, 2004,26 : 326-332.

BRUN D., HARRICHE N., CROZET C., d'IVERNOIS J-F, GAGNAYRE R., 2003, *Guide pour la réalisation de la bibliographie, des tableaux et des figures dans un travail de recherche*, UFR SMBH Léonard de Vinci, Bobigny.

CAIRE F., SOL J-C., MOREAU J-J., ISIDORI P., CHARLIN B., 2004, Auto-évaluation des internes en neurochirurgie par Test de Concordance de Script (TCS) : processus d'élaboration des tests, *Neurochirurgie*, 50, 66-72.

CAIRE F., SOL J-C., CHARLIN B., ISIDORI P., MOREAU J-J., 2004, Le Test de Concordance de Script (TCS) comme outil d'évaluation formative des internes en neurochirurgie : implantation du test sur Internet à l'échelle nationale, *Pédagogie Médicale*, 5, 87-94.

CHAMBERLAND M., HIVON R., TARDIF J, BEDARD, 2001, Evolution du raisonnement clinique au cours d'un stage d'externat : une étude exploratoire, *Pédagogie Médicale*, 2, 9-17.

CHARLIN B., BRAILOVSKY C.A., BRAZEAU-LAMONTAGNE L., SAMSON L., LEDUC C., 1998-a, Script Questionnaires : their use for assessment of diagnostic knowledge in radiology, *Medical Teacher*, 20, 567-571.

CHARLIN B., BRAILOVSKY C.A., LEDUC C., BLOUIN D., 1998-b, The Diagnostic Script Questionnaire : A new tool to assess a specific dimension of clinical competence, *Advances in Health Sciences Education*, 20, 51-58.

CHARLIN B., 1999, L'apport des tests de concordance de scripts en formation médicale continue, *L'organisateur d'EMC*, 12, 6-9.

CHARLIN B., ROY L., BRAILOVSKY C., VAN DER VLEUTEN C., 2000-a, The Script Concordance Test, a tool to assess the reflective clinician, *Teaching and Learning in Medicine*, 12, 189-195.

CHARLIN B., TARDIF J., BOSHUIZEN H., 2000-b, Scripts and medical diagnosis knowledge : theory and applications for clinical reasoning instruction and research, *Academic Medicine*, 75, 182-190.

CHARLIN B., DESAULNIERS M., GAGNON R., BLOUIN D., VAN DER VLEUTEN C., 2002-a, Comparison of an aggregate scoring method with a consensus scoring method in a measure of clinical reasoning capacity, *Teaching and learning in medicine* 14, 150-156.

CHARLIN B., GAGNON R., SIBERT L., VAN DER VLEUTEN C., 2002-b, Le test de concordance de script, un instrument d'évaluation du raisonnement clinique, *Pédagogie Médicale*, 3, 135-144.

CHARLIN B., BORDAGE G., VAN DER VLEUTEN C., 2003, L'évaluation du raisonnement clinique, *Pédagogie Médicale*, 4, 42-52.

CHARLIN B., VAN DER VLEUTEN C., 2004-a, Standardised assessment of reasoning in contexts of uncertainty : the script concordance approach. *Evaluation and the Health professions*, 27, 3, 304-319.

CHARLIN B., GAGNON R., 2004-b, The effect of variability of answers among criterion experts to detect expertise with a SCT, *Academic Medicine*, submitted.

CRONBACH L.J., 1951, Coefficient alpha and the internal structure of tests, *Psychometrika*, 16, 3, 297-334.

ELSTEIN A.S., SHULMAN L.S., SPRAFKA S.A. (1990), Medical problem solving, a ten years retrospective, *Evaluation and the Health Professions*, 13, 5-36.

FRENETTE-LECLERC C-A., 1992, Sur la route de l'expertise, *Nursing Québec*, 12, 1, 48-54.

GAGNON R., CHARLIN B., COLETTI M., SAUVE E., VAN DER VLEUTEN C., 2003, Assessment in context of uncertainty : how many experts are needed on the Script Concordance Test panel of reference ? *Medical Education*, submitted.

GUILBERT J-J., 1992, *Comment raisonnent les médecins, réflexions sur la formation médicale*, Genève, Editions médecine et hygiène.

JOUQUAN J., 2002, L'évaluation des apprentissages des étudiants en formation initiale médicale, *Pédagogie Médicale*, 3, 38-51.

IVERNOIS (d') J-F., GAGNAYRE R., MARCHAND C., 2002, *Guidance de recherche*, UFR SMBH Léonard de Vinci, Bobigny.

KASSIRER J.P., 1983, Teaching Clinical Medicine by Iterative Hypothesis Testing, *The new England Journal of Medicine*, 309, 921-923.

KASSIRER J.P., 1989, Diagnostic Reasoning, *Annals of Internal Medicine*, 110, 893-900.

KHONPUTSA P., BESINQUE K., FISHER D., GONG W.C., 2003, The Pharmacy Script Concordance Test : a pilot study assessing clinical pharmacy competence in Thailand, *Medical Teacher*, submitted.

LABELLE M., GAGNON R., THIVERGE R., LAPRISE R., SAINTE-MARIE L-G., CHARLIN B., 2003, Formation continue en petits groupes sur l'ostéoporose : comparaison d'un atelier basé sur le test de concordance de scripts (TCS) et d'un atelier classique, *Pédagogie Médicale*, 4, 145-153.

LANSAC J., BODY G., 1988, *Pratique de l'accouchement*, Simep, Paris.

LANSAC J., BERGER C., MAGNIN G., 1990, 2^{ème} édition, *Obstétrique pour le praticien*, Simep, paris.

LAUGIER J., ROZE J-C., 2001, *Soins aux nouveaux-nés. Avant, pendant et après la naissance*, Masson, Paris.

LE MAOUT G., 1999, *Le raisonnement médical en gynécologie obstétrique*, Paris, Maloine.

LLORCA G., ROY P., RICHE B., 2003, Evaluation de résolution de problèmes mal définis en éthique clinique : variation des scores selon les méthodes de correction et les caractéristiques des jurys. *Pédagogie Médicale*, 2, 80-87.

MESSICK S., 1995, Standards of validity and the validity of standards in performance assessment, *Educational Measurement*, 5-8.

PAOLAGGI J-B, COSTE, 2001, *Le raisonnement médical de la science à la pratique clinique*, Paris, Estem.

RAYNAL F., RIEUNIER A., 1998, 2^{ème} édition, *Pédagogie : dictionnaire des concepts clés, apprentissage, formation, psychologie cognitive*, Paris, ESF.

REGHER G., NORMAN G. R., 1996, Issues in Cognitive Psychology : Implications for Professional Education, *Academic Medicine*, 71, 998-1001

SCHMIDT H.G., NORMAN G.R., BOSHUIZEN H., 1990, A cognitive perspective on medical expertise : theory and implication, *Academic Medicine*, 65, 611-21.

SIBERT L., CHARLIN B., GAGNON R., CORCOS J., KHALAF A., GRISE P. 2001, Evaluation du raisonnement clinique en urologie, *Progrès en Urologie*, 11, 1213-1219.

SIBERT L., CHARLIN B., CORCOS J., LECHEVALLIER J., GRISE P., 2002, Assessment of clinical reasoning competence in urology with the Script Concordance Test : an exploratory study on two sites from different countries, *European Urology*, 41, 227-233.

ANNEXES

Annexe I

Exemple d'une vignette

Annexe II

Grille de recueil du matériel nécessaire à la construction de TCS

Annexe III

Table de spécification par domaines de compétences, répartition des vignettes

Annexe IV

Grille de contrôle qualité dans la construction du test

Annexe V

Fiche d'identification des participants

Annexe VI

Questionnaire d'opinion sur le TCS – feedback participants

Annexe VII

Tableau récapitulatif des membres du panel

Annexe VIII

Notes CCT (contrôle continu théorique), CL (clinique), RC (raisonnement clinique), TCS – ESF 3 et ESF 4.

Evaluation du raisonnement clinique des étudiants sages-femmes par le test de concordance de script

Introduction :

Comme les autres cliniciens, les sages-femmes rencontrent dans leur pratique des problèmes médicaux en contexte d'incertitude. Le concept de raisonnement clinique en contexte d'incertitude s'appuie à la fois sur la théorie hypothético-déductive du raisonnement clinique, et sur la théorie des scripts issue de la psychologie cognitive. Pour explorer les processus de raisonnement clinique il existe un instrument d'évaluation : le Test de Concordance de Script (TCS).

Contexte :

Les études de sages-femmes sont actuellement validées par différents types d'épreuves qui se complètent et visent l'évaluation sommative de connaissances et de compétences différentes. Toutes présentent des avantages et des inconvénients mais aucune ne permet pour l'instant une évaluation standardisée et objective du raisonnement clinique. L'ensemble des études publiées depuis 1998 sur le TCS le décrivent comme un instrument valide pour explorer les processus de raisonnement clinique en contexte d'incertitude.

Finalité :

Evaluer les processus de raisonnement clinique des étudiants sages-femmes en fin de cursus par l'introduction d'un Test de Concordance de Script et montrer l'intérêt d'intégrer cet instrument au dispositif d'évaluation des sages-femmes en formation initiale.

Matériels et méthode :

L'étude est une recherche-action évaluative et prospective qui s'est déroulée au deuxième trimestre 2004 à l'Ecole Régionale de Sages-femmes de Marseille. 50 sages-femmes et 54 étudiantes réparties en 3 groupes de niveaux d'expérience différents ont passé un TCS de 70 items couvrant les domaines de compétences habituels de sages-femmes hospitalières. 20 des 50 professionnelles ont été tirées au sort pour constituer le panel de référence dont les réponses ont permis d'établir les scores selon la méthode dite « des scores combinés ». Les données ont été saisies sur le site Internet du correcteur TCS et les analyses statistiques ont été faites à l'aide du logiciel SPSS.

L'étude présente des limites qui ont été identifiées et décrites.

Résultats :

Le TCS discrimine significativement le niveau d'expérience entre les étudiantes et les professionnelles en exercice depuis au moins 5 ans, ce qui vient confirmer la première hypothèse de recherche et établir la validité de construit du test.

Il n'existe aucune corrélation significative entre les résultats obtenus par les étudiants au TCS et aux autres épreuves habituelles (contrôle continu théorique et épreuves cliniques), ce qui vient infirmer la seconde hypothèse de recherche.

Le test construit ici s'avère plus pertinent et plus fiable pour les étudiants de troisième année (coefficient α de Cronbach de 0.69) que pour ceux de quatrième année.

Perspectives :

Plusieurs arguments sont développés en faveur de l'introduction du TCS à l'école de sages-femmes et deux axes d'actions sont définis : le premier s'oriente vers un réajustement de notre dispositif de formation consécutif aux résultats de cette étude, le second vers la poursuite des recherches sur le TCS pour les sages-femmes et autres professionnels de santé.

Conclusion :

Cette étude a permis de montrer que le TCS élaboré ici demande à être amélioré pour augmenter sa fiabilité mais peut déjà être considéré comme un instrument valide, pertinent et complémentaire dans l'évaluation sommative des compétences cliniques des sages-femmes en formation initiale. Son introduction augmenterait la cohérence globale du dispositif d'évaluation puisqu'il concerne une dimension jusqu'ici peu ou mal explorée : le raisonnement clinique des étudiants sages-femmes.

Mots clés : évaluation, sages-femmes, raisonnement clinique, contexte d'incertitude.