

Utilisation des Tests de Concordance de Scripts pour l'évaluation en informatique

Morgan MAGNIN (Ecole Centrale de Nantes – IRCCyN, Nantes), Guillaume MOREAU (Ecole Centrale de Nantes – CERMA, Nantes)

■ **RÉSUMÉ** : L'évaluation des étudiants doit porter non seulement sur les connaissances qu'ils ont acquises, mais également sur les compétences qu'ils ont développées. C'est pour améliorer l'évaluation du savoir-faire de nos élèves en informatique que nous avons choisi d'adapter le principe des tests de concordance de scripts (TCS) à ce domaine. Les TCS ont pour objectif d'évaluer le raisonnement d'étudiants placés dans une situation professionnelle. Nous avons défini un nouveau champ d'application pour les TCS, nous avons conçu des TCS et nous les avons expérimentés. Cette expérience nous a permis d'établir les forces et les limites des TCS en informatique, une analyse qui ouvre la porte à un usage plus large de ce dispositif dans d'autres disciplines.

■ **MOTS CLÉS** : évaluation de compétences, auto-évaluation, savoir-faire, pédagogie par problème.

■ **ABSTRACT** : Student assessment must not only take into account their acquired knowledge but also their skills. To improve the assessment of their professional skills in computer science, we have chosen to adapt the concepts of Script Concordance Tests (SCT) to the scientific domain. In this paper, we describe how we adapted SCT to computer science. We propose some SCT and an experiment. The experiment has shown advantages and drawbacks of SCTs, that allows to think of a wider use of SCT in other disciplines.

■ **KEYWORDS** : skill assessment, self-assessment, know-how, problem-based learning

- 1. Contexte
- 2. Les TCS
- 3. Objectifs et compétences ciblées
- 4. Processus de construction et d'adaptation des TCS à l'informatique
- 5. Retour d'expérience
- 6. Conclusions et perspectives
- 7. Remerciements
- BIBLIOGRAPHIE

1. Contexte

L'apprentissage par projet suscite un intérêt tout particulier dans le domaine des formations professionnalisantes telles qu'elles sont dispensées dans les écoles d'ingénieur comme Centrale Nantes. L'acquisition de compétences et d'aptitudes telle la capacité à chercher et utiliser des informations y est fondamentale. Comme l'expliquent (Ciocoiu et al., 2001) il s'agit plus d'apprendre à apprendre que d'apprendre. Face à l'essor des pédagogies constructivistes et à l'importance de certifier les compétences des apprenants, il importe de se doter d'outils pour l'évaluation (et l'auto-évaluation) du savoir-faire. Les stages ou projets en partenariat avec des industriels permettent de jauger un savoir-être et un savoir-faire global en entreprise. Mais il est difficile de vérifier l'acquisition, par l'étudiant, d'une compétence précise.

La littérature dans le domaine se focalise plutôt sur l'environnement technique permettant de superviser le travail des étudiants (Michel, 2009). Certains dispositifs formatifs (telle la revue de pairs mise en œuvre dans le cadre du Master de l'Université de Bretagne Occidentale (Saliou et Ribaud, 2005)) intègrent nativement des mécanismes d'auto-évaluation. Mais rares sont, à notre connaissance, les moyens de vérifier rapidement la bonne acquisition d'une compétence par un apprenant. Or, dans (Perrenoud, 2004),

l'auteur évoque la nécessité d'inventer de "vraies situations de raisonnement et d'action". Les spécificités de l'enseignement médical (contexte ouvert, importance des capacités de raisonnement permettant l'élaboration d'un diagnostic, etc.) ont concouru à l'émergence de différents outils utiles à l'évaluation du raisonnement. Dans ([Charlin et al., 2003](#)), les auteurs présentent un certain nombre de ces dispositifs – les examens oraux, les dissertations, les questions à appariement étendu, l'examen par éléments clefs, etc. – et leurs limites respectives (difficultés de mise en œuvre d'une grille d'évaluation, exigences en termes de temps de correction, travail de préparation préalable, etc.). Les tests de concordance de scripts (TCS), une forme de généralisation des questionnaires à choix multiples (QCM), y apparaissent comme une méthode à la fois novatrice, efficace, et médiane entre ces différents critères. Les TCS s'appliquent à des situations incertaines du milieu médical dont la nature nous est apparue très similaire à des cas pratiques de diagnostic en ingénierie.

C'est pourquoi, l'Ecole Centrale de Nantes a participé à un projet UNIT sur l'extension des tests de concordance de scripts (jusque là réservés à l'évaluation des compétences des étudiants en médecine) à l'évaluation en informatique. Comme leur nom le laisse penser, les TCS sont basés sur la théorie cognitive dite "des scripts", qui se focalise sur les réseaux de connaissances qu'utilisent les individus en situation de résolution de problème. Elle considère les réseaux de connaissances préalables (les scripts) que possède tout individu pour traiter les différentes données d'une situation problématique ([Charlin et al., 2006](#)). L'objectif de cette rubrique est de rendre compte des expérimentations menées à l'Ecole Centrale de Nantes autour d'une nouvelle déclinaison des TCS pour l'informatique au niveau L comme M.

La suite de cet article est organisée de la façon suivante : dans un premier temps, nous présentons la notion de TCS. Nous présentons ensuite notre analyse des TCS dans le contexte de l'évaluation des étudiants en informatique et les adaptations effectuées. Des exemples de TCS dans divers enseignements d'informatique sont ensuite décrits. Les retours d'expérience menés dans le cadre d'un examen donné à une centaine d'étudiants sur deux années sont ensuite présentés.

2. Les TCS

L'objectif des TCS est d'évaluer le raisonnement des étudiants placés dans une situation professionnelle, et non pas leurs connaissances (ce qui peut déjà être fait par des méthodes "classiques", allant du QCM au devoir sur table). Les TCS constituent donc un outil d'évaluation complémentaire des procédures déjà existantes. Des études ([Charlin et al., 2006](#)) ont montré la fidélité et la validité des TCS pour différencier les étudiants en fonction de leur niveau d'expertise. Ils permettent de tester un grand nombre d'étudiants de façon standardisée, discriminante et objective.

Ils sont déjà largement utilisés dans de nombreuses formations médicales. Ils permettent d'évaluer le raisonnement en situation incertaine correspondant à la vie professionnelle. Ils mettent en jeu des problèmes mal définis, avec des données incertaines et pour lesquels plusieurs solutions sont possibles. Les raisonnements des étudiants sont alors confrontés à ceux d'experts.

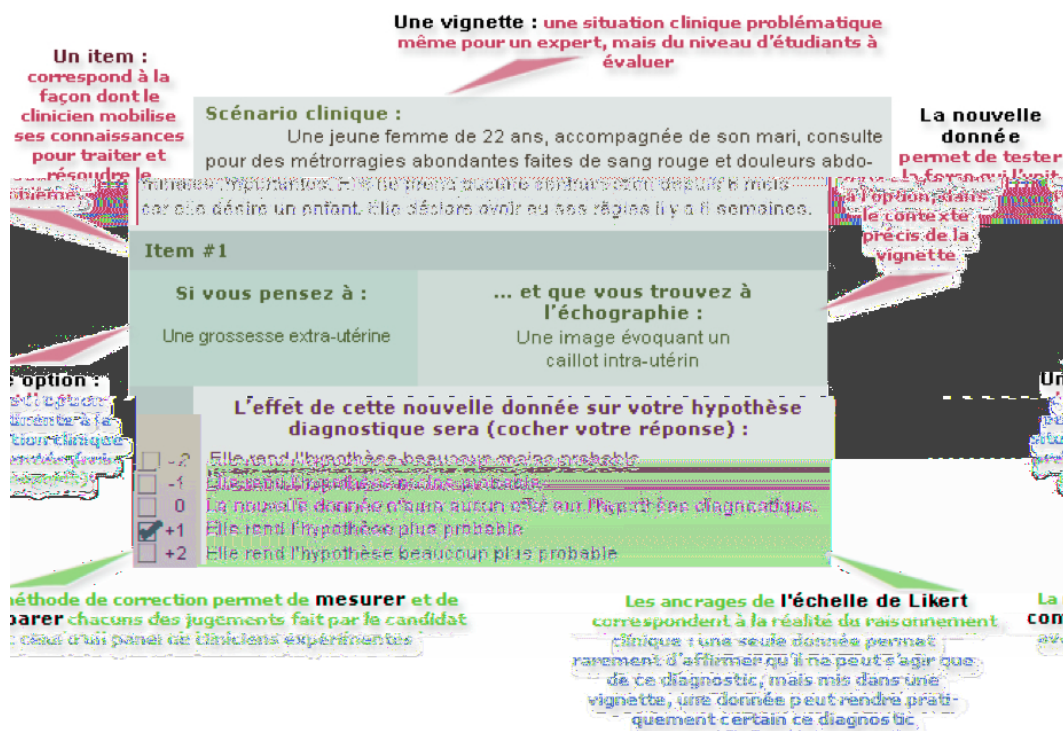


Figure 1 • TCS type du domaine médical, tiré de (TCS, 2011)

Le principe d'un TCS est le suivant (voir la figure 1) :

- On présente un scénario réel - une situation de la vie professionnelle problématique, même pour un expert du domaine. Il n'y a pas de bonne réponse unique : en présence de plusieurs experts, il y aura très probablement discussion.
- On propose une option, c'est-à-dire une hypothèse pertinente et une nouvelle donnée.
- On demande à l'étudiant de qualifier l'effet de cette nouvelle donnée sur l'hypothèse (-2, -1, 0, +1, +2)
- Le score attribué à l'étudiant dépend des réponses d'un groupe d'experts qui a répondu préalablement au TCS. L'étudiant gagne des points proportionnellement au nombre d'experts qui ont choisi la même réponse.

Le TCS permet d'évaluer le raisonnement en situation incertaine de manière isolée. Le secteur de la santé constitue, depuis le début des années 2000, un terrain d'exploitation privilégié pour les TCS. C'est ainsi que leur pertinence et leur efficacité a d'ores et déjà été prouvée en chirurgie, urologie, chirurgie, radiologie et, plus récemment, en neurologie (Lubarsky et al., 2009) et pédiatrie (Lemay et al., 2010).

Dans (Charlin et al., 2006), les auteurs reviennent sur l'origine des TCS pour jauger d'une capacité à raisonner en contexte d'incertitude. Partant des expériences menées dans le secteur de la santé, ils recommandent plus largement l'utilisation des TCS dans tout domaine qui implique de raisonner en contexte d'incertitude. Notre démarche - étendre les TCS à la discipline informatique - se place ainsi dans la droite lignée de ce conseil.

3. Objectifs et compétences ciblées

3.1. Enjeux

La principale motivation du projet est d'identifier un moyen efficace et pertinent d'évaluer les compétences des élèves en informatique, tout en promouvant un dispositif utilisable dans une démarche d'auto-évaluation comme la préparation au C2i (C2i, 2011). Les objectifs sont ainsi de :

- mesurer le raisonnement des étudiants en informatique pour mieux évaluer leurs compétences ;

- fournir aux étudiants un outil d'auto-évaluation sur leurs capacités afin de les aider à se situer et progresser ;
- à terme, mieux comprendre pourquoi un étudiant atteint les compétences demandées alors qu'un autre n'y arrive pas.

L'équipe pédagogique en informatique de l'École Centrale de Nantes avait déjà une expérience des QCM pour l'évaluation des connaissances. Les TCS présentent de nombreuses différences par rapport au QCM : le scénario correspond à une situation problématique (la situation décrite dans le QCM est en général simple), il n'est pas complètement décrit (dans le QCM l'énoncé contient tous les éléments nécessaires à la réponse). Il n'y a pas en soi de bonne réponse, puisque les experts ne donnent pas tous les mêmes réponses (dans le QCM, on force les experts à adopter un consensus sur la bonne réponse). Toute réponse donnée par un expert a une valeur même si les experts ne s'entendent pas.

Les TCS sont donc apparus comme prometteurs pour évaluer le savoir-faire des élèves. C'est dans ce contexte qu'un projet UNIT a été mené, rassemblant l'École des Mines de Nantes, l'École des Mines de Douai, l'École des Mines d'Alès, l'INSA de Rouen, l'Université de Bordeaux 2 et l'École Centrale de Nantes.

L'équipe pédagogique du département Informatique et Mathématiques de l'École Centrale de Nantes a rédigé plus de 60 TCS, qui se répartissent comme suit :

- Cours de Systèmes d'Information et Bases de Données - Élèves-Ingénieurs (E.I.) de 1ère année (i.e. Bac+3) : 29 TCS
- Cours d'algorithmique et programmation - E.I. de 1ère année : 25 TCS
- Cours de méthodes logicielles – E.I. de 2e année (i.e. Bac+4) : 10 TCS

3.2. Construction du référentiel de compétences

Chaque TCS vise à vérifier la bonne acquisition d'une compétence par l'apprenant. Si le référentiel de compétences n'existe pas, il convient donc de commencer par le construire avant d'écrire, dans un second temps, les TCS associés à chaque compétence. Nous disposions déjà d'un référentiel détaillé, via le programme des enseignements de notre établissement, établi conjointement par la Direction des Études, l'équipe pédagogique en informatique et des professionnels du secteur. Afin de ne pas balayer un spectre de compétences trop large au cours de cette première expérimentation des TCS, nous avons restreint le référentiel à certaines compétences identifiées comme "majeures" dans chaque domaine ciblé.

La figure 2 en donne un aperçu. Jusque là, certaines compétences n'étaient validées qu'à travers des travaux pratiques (TP). Malheureusement, ces TP sont réalisés en binôme ou en trinôme, rendant impossible le suivi individualisé de l'acquisition de ces compétences. Grâce aux TCS, nous souhaitons pouvoir dresser une cartographie beaucoup plus précise des compétences développée par chacun des étudiants.

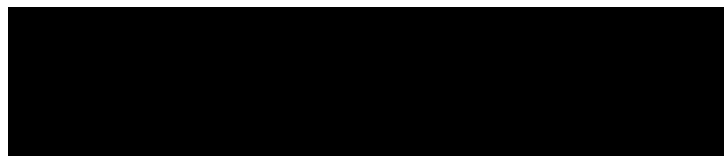


Figure 2 • Extrait du référentiel de compétences (et du nombre de TCS finalement élaborés pour chacune) dans le cadre de l'enseignement de méthode logicielles

3.3. Déroulé du projet

Partant du référentiel de compétences et de la connaissance préalable du principe des tests de concordance de scripts, le travail s'est ensuite porté sur l'élaboration de TCS. Le projet s'est donc décomposé suivant les différentes étapes ci-dessous :

- Acquisition du savoir-faire en TCS : il s'agissait là d'apprendre ce qu'est un TCS et d'identifier les

enjeux associés (fin 2008) ;

- Établissement du référentiel de compétences : définition de la liste de compétences prioritaires à évaluer dans le cadre de cette première expérience autour des TCS (début 2009).

-

4.1. Choix du cours : motivations

Nos premières expérimentations à base de TCS se sont déroulées dans le cadre d'un cours de "Système d'information et base de données" choisi en particulier pour sa mixité entre composantes techniques et non techniques. En effet, les problématiques des systèmes d'information (SI) sont au moins autant organisationnelles que techniques. Elles nous ont semblé de prime abord un terrain d'expérimentation privilégié parce que :

- les connaissances théoriques à elles seules ne peuvent en aucun cas suffire à résoudre un problème pratique comme ceux posés dans le cycle de vie d'un SI ; en effet, le SI est un modèle du système organisationnel pour lequel il est conçu, mais il est surtout un artefact de ce système avec une existence propre, parallèle à celle du système. Il est tout à fait concevable que les informations détenues par le système organisationnel et par le SI soient différentes, voire divergent. Il est important de faire comprendre ceci aux étudiants, dans le cours comme dans l'évaluation.
- le cours de Système d'information et base de données est conçu pour sensibiliser les étudiants, futurs ingénieurs, au paradigme DIKW (Data, Information, Knowledge, Wisdom) tel que défini par Ackoff dans (Ackoff, 1989) : si les données sont faciles à évaluer par des tests binaires comme les QCM, les niveaux supérieurs sont plus complexes à évaluer.
- enfin, peu de problèmes de la vie courante d'un système d'information peuvent se résoudre dans un cadre disciplinaire unique.

| | |
|-----------------|--|
| | de l'Ecole, le secrétariat Général édite la liste à partir du système d'information. Lors de cette édition, il apparaît un L. Baudoin et un Luc Baudoin. |
| Option | Si vous pensez à : Un champ prénom mal renseigné dans la base de données. |
| Nouvelle donnée | Ils ont tous les deux enseigné dans le département Info-Maths. |

Tableau 2 • Exemple 2 – SI (dérivé du cas précédent) – Compétence ciblée : Analyser un problème d'incohérence de données dans un système d'information

Cette fois, il est difficile de conclure : il existe évidemment un risque que le vacataire ait été entré deux fois (avec un prénom la première fois, avec une simple initiale la seconde), mais il est aussi possible d'avoir deux vacataires avec le même nom et la même initiale de prénom. Néanmoins, ce type de test est par trop contextualisé, nous en reparlerons par la suite.

Pour sortir un peu des limites du contexte, nous proposons maintenant des exemples qui ne relèvent pas d'une problématique SI de l'enseignement supérieur.

| | |
|-----------------|---|
| Scénario | Une requête sur la population d'une région ne renvoie pas le même résultat qu'une requête faisant la somme des populations des départements qui la composent. |
| Option | Si vous pensez à : Les données sont fausses quelque part. |
| Nouvelle donnée | Les gestionnaires concernés dans les départements et les régions vous assurent des données entrées. |

Tableau 3 • Exemple 3 – SI – Compétence ciblée : Analyser un problème d'incohérence de données dans un système d'information

Ce cas est encore un cas classique des systèmes d'information, facile à reproduire. Bien entendu, aucun département ne diffuse de données fausses aux populations concernées : les données démographiques sont prises à des instants différents et ne représentent donc pas la même information. Pour autant difficile de conclure sur un TCS, sans cette information, qui n'est d'ailleurs peut-être pas la seule explication.

| | |
|----------|--|
| Scénario | Vous avez mis en place un logiciel de gestion des frais de déplacements il y a un an pour accélérer les remboursement de frais de votre personnel et pour contrôler l'opportunité de certaines dépenses. Lors des contrôles, vous détectez une différence entre les montants définis dans l'application et les remboursements dans le logiciel de compatibilité. |
|----------|--|

| | |
|-----------------|--|
| Option | Si vous pensez que : Une fraude a eu lieu |
| Nouvelle donnée | Les logiciels de gestion des frais de déplacement comporte 165 déplacements, la comptabilité fait état de 2315 remboursements. |

Tableau 4 • Exemple 4 – SI – Compétence ciblée : Distinguer les problèmes logistiques des erreurs informatiques dans un système d'information

L'objectif ici était de montrer que les problématiques SI peuvent dépasser le cadre informatique : en effet, rien ne prouve qu'un logiciel, quelle que soit son utilité, soit réellement utilisé. Une des questions à se poser ici, compte tenu de la différence entre le nombre de déplacements et le nombre de remboursements est de savoir si tous les déplacements ont été effectivement saisis dans le logiciel, avant de penser à la moindre fraude. Si le logiciel a posé un problème d'appropriation (manque de formation, complexité, problème de conduite du changement), ceci est une hypothèse réaliste.

Nous avons ensuite abordé les problèmes d'accès concurrents au SI.

| | |
|-----------------|---|
| Scénario | M et R réservent tous les deux une place dans le train Nantes-Paris de 7h00 le 26 juin 2009. Lors de leur arrivée à la gare, ils se retrouvent avec la même place dans le même wagon. |
| Option | Si vous pensez à : Le logiciel de réservation comporte un bug au niveau de la vérification d'attribution des places. |
| Nouvelle donnée | Le problème ne se produit qu'une fois par an. |

Tableau 5 • Exemple 5 – SI – Compétence ciblée : Identifier les situations d'accès concurrents à une même ressource

On recherche les problèmes plutôt du côté du module d'attribution des places. Typiquement, ce TCS est très (trop) contextualisé par le cours où la question de l'accès en exclusion mutuelle et des sémaphores a été abordée à travers un exemple similaire. La réponse n'a finalement rien d'évident : s'agit-il s'un bug dans l'attribution ou justement ce fameux problème d'exclusion mutuelle typique des systèmes web mal conçus ? Aucun élément ne permet de répondre.

4.3. Exemples de TCS "Droit de l'informatique»

Rappelons que le référentiel CTI (Commission des Titres d'Ingénieur) qui définit les compétences de l'ingénieur inclut l'insertion des activités de celui-ci dans le cadre légal qui les régit. À l'instar des systèmes d'information, le droit informatique constitue un domaine pertinent d'exploitation des TCS. En effet, en situation professionnelle, il arrive souvent que nous ne possédions pas l'intégralité des informations nécessaires à la prise d'une décision parfaitement déterministe dans le domaine juridique. Il importe donc de raisonner sur les éléments à notre disposition, en étant le plus proche possible de l'esprit du législateur. Les TCS sont donc adaptés en ce qu'ils permettent d'évaluer des compétences dans des situations ouvertes.

| | |
|----------|--|
| Scénario | Vous venez d'être nommé Correspondant Informatique et Liberté dans votre |
|----------|--|

| | |
|-----------------|---|
| | entreprise. Parmi les adresses e-mail figurant dans les fichiers de prospection publicitaire de l'entreprise, vous soupçonnez que certaines ont été collectées sans l'accord préalable des personnes à recevoir des messages commerciaux. Mais les fichiers ne contiennent aucune information sur le fait que les personnes ont - ou non - donné leur consentement. |
| Option | Si vous pensez que : Il est nécessaire de supprimer l'intégralité des fichiers d'adresses e-mail. |
| Nouvelle donnée | Vous apprenez que ces adresses e-mail appartiennent majoritairement à des personnes morales. Le reste des adresses e-mail appartiennent à d'anciens clients de l'entreprise. |

Tableau 6 • Exemple 1 – Droit – Compétence ciblée : Reprendre un système d'information et le mettre en conformité avec la loi

Profitons de cet exemple pour illustrer la discussion que nous avons tenue en section 3.4, sur le choix d'une échelle adaptée entre les différentes modalités.

Dans le TCS présenté ici, nous devrions aboutir à 100% des experts sur la formulation "Elle rend l'hypothèse moins probable" si c'est la première échelle qui est utilisée ("exclut/confirme l'hypothèse"). En effet la nouvelle donnée a un effet sur l'hypothèse : on y apprend que, potentiellement, 100% du fichier de prospection publicitaire est légal (le droit français accorde l'autorisation de prospecter par courriel - sans accord préalable - des personnes morales et d'anciens clients n'ayant pas manifesté leur refus par une demande de désinscription). Les cas +2, +1 et 0 sont donc exclus. Reste le débat entre -1 et -2.

Dans le cas où le -2 est exprimé sous la forme "exclut l'hypothèse" : un expert devrait normalement la réfuter car il manque des informations pour décider définitivement (certains clients peuvent avoir manifesté leur souhait de ne plus recevoir de courriels publicitaires). Ce qui doit conduire un expert à choisir automatiquement -1.

Dans le cas où le -2 est exprimé sous la forme "beaucoup plus probable" : on devrait obtenir une répartition plus uniforme sur les -1 et -2 car le -2 n'est, ici, pas totalement exclu. Cela dépend de la sensibilité de l'expert. Autrement dit : la répartition des réponses sera un peu meilleure qu'avec l'autre échelle.

La seconde échelle laisse plus de possibilité d'expression aux experts, ce qui est a priori plus semblable à ce qui peut se passer dans le "monde réel".

Envisageons maintenant un second TCS, présenté dans le tableau 7, autour de la législation sur les données à caractère personnel. Les transferts de flux de données vers un pays n'appartenant pas à l'Union Européenne exigent des dispositions particulières. De plus, même si la Suisse est connue pour son secret bancaire, la législation suisse est différente de la législation française. Un expert verra ainsi immédiatement qu'un tel scénario nécessite de la prudence, et qu'il vaut mieux être mesuré quant au choix d'un prestataire helvète. Toutefois, comme nous le verrons dans la partie 5, un certain nombre d'étudiants oublient de traiter ce scénario sous l'angle légal, et se laissent emporter par leurs préjugés.

| | |
|-----------------|---|
| Scénario | Vous avez été mandaté pour choisir un prestataire externe pour la gestion des paies de votre entreprise (située en France). L'offre la plus intéressante vous a été adressée par l'entreprise La Bonne Paie, acteur reconnu dans le domaine, qui est notamment réputé pour la fiabilité et la sécurité des procédures qu'elle met en place. |
| Option | Si vous pensez que vous allez choisir ce prestataire. |
| Nouvelle donnée | Les serveurs de stockage de données utilisées par La Bonne Paie sont localisés en Suisse. |

Tableau 7 • Exemple 2 – Droit – Compétence ciblée : Déterminer la conformité d'un transfert de données au regard de la loi

4.4. Exemples de TCS "programmation"

Si les systèmes d'information constituent de prime abord un exemple très approprié d'utilisation des TCS puisqu'ils incorporent une notion de "diagnostic" en cas d'incident, il existe des cas où les TCS sont également utiles dans des démarches a priori plus *exactes* comme la programmation.

| | |
|-----------------|---|
| Scénario | A et B ont écrit chacun la moitié du programme dans deux fichiers x.c et y.c. Une fois compilé correctement et les liens édités sans problème, le programme plante. |
| Option | Si vous pensez que : A a tort car il est habituellement moins fort en programmation. |
| Nouvelle donnée | B a vérifié le fonctionnement de toutes ses fonctions. |

Tableau 8 • Exemple 1 – Programmation – Compétence ciblée : Distinguer les tests unitaires des tests d'intégration

L'exemple propose également une démarche de type diagnostic : il s'agit de comprendre les causes du non-fonctionnement d'un programme. Il y a deux grandes hypothèses restantes ici : soit c'est A qui n'a pas réalisé de tests unitaires suffisants sur les fonctions qu'il a écrites, soit c'est l'intégration qui a posé problème.

Si les exemples autour du diagnostic sont assez faciles à trouver, on peut trouver également des cas d'application en conception des structures de données.

| | |
|----------|---|
| Scénario | Vous cherchez une structure de données appropriée pour stocker des éléments munis d'une relation d'ordre. Ces éléments correspondent à des objets que vous avez |
|----------|---|

| | |
|-----------------|--|
| | définis vous-mêmes. |
| Option | Si vous pensez que : Vous pensez choisir la structure "arbre binaire de recherche". |
| Nouvelle donnée | Le coût de comparaison entre deux éléments est important. |

Tableau 9 • Exemple 2 – Programmation – Compétence ciblée : Choisir une structure de donnée appropriée

L'exemple précédent porte sur le choix d'une structure de données. La structure arbre binaire de recherche (ABR) est appropriée pour stocker les objets munis d'une relation d'ordre, mais elle est coûteuse en insertion/équilibre, phénomène aggravé par le coût de la comparaison entre les objets. Sans hypothèses supplémentaires sur l'utilisation de la structure de données, il est difficile de conclure.

5. Retour d'expérience

Comme nous l'avons évoqué dans l'introduction de cet article, le recours aux TCS en informatique vise à rendre possible l'évaluation du savoir-faire des étudiants à travers un outil complémentaire des approches traditionnelles (travaux pratiques, évaluation en stage, etc.). Nous avons donc souhaité, enfin, les mettre en œuvre dans le cadre de procédures d'évaluation des compétences en informatique. Dans cette partie, nous présenterons les expériences que nous avons menées, puis les leçons que nous en avons retirées.

5.1. Description des expérimentations

Nos deux premières expérimentations ont porté sur l'enseignement de "Systèmes d'Informations et Bases de Données", un cours de première année (équivalent Bac + 3) à Centrale Nantes. Le choix de cette matière tient à deux raisons principales :

- Contrairement à certains domaines plus théoriques (algorithmique notamment), le cours vise des compétences pratiques qu'il est facile d'éprouver par des études de cas. La conception de TCS en a été facilitée.
- Les responsables de cet enseignement étaient par ailleurs porteurs du projet "TCS" au sein de l'établissement, rendant plus aisée l'expérimentation des TCS en évaluation.

Nous avons mené une première expérimentation en juin 2009, sur une promotion de 110 étudiants, dans un examen qui comprenait, entre autres, 4 TCS. Nous avons reconduit cette expérience en juin 2010 sur la même matière, mais en augmentant le nombre (7) de TCS posés à une promotion de 108 élèves. Cette augmentation progressive du nombre de TCS dans un même devoir est issue d'une volonté d'éprouver le dispositif. La proportion du nombre de TCS dans nos sujets d'examens est désormais constante, et correspond à environ un quart du devoir surveillé (tout aussi bien en termes de points distribués que de temps pendant lequel les étudiants sont sensés plancher sur ces questions).

Le recours aux TCS pour l'évaluation du raisonnement en contexte d'incertitude soulève généralement deux limites : d'une part l'expertise présumée des experts, d'autre part le peu d'informations retirées sur le processus de raisonnement de l'étudiant.

La première limite peut être surmontée en composant avec soin le panel d'experts. Pour ce faire, nous avons sollicité des personnes avec un diplôme de niveau Master et au moins trois ans d'expérience pratique dans le domaine. Dans le secteur médical, (Charlin et al., 2007) ont constaté une légère différence dans les résultats des expertises des enseignants et celles des professionnels n'appartenant pas au secteur académique. Forts de leurs recommandations, nous avons constitué un panel mixte, composé de 80% d'enseignants(-chercheurs) et de 20% d'ingénieurs – ou de docteurs – appartenant au tissu industriel. De plus, nous avons constaté que les réponses des experts se concentraient généralement sur

deux ou trois éventualités, corroborant l'idée de processus de réflexion communs aux personnes constituant notre panel de référence.

La seconde limite est plus délicate à prendre en compte. Dans le secteur médical, il a été démontré ([Grant et Marsden, 1988](#)) que des médecins, confrontés à une même situation, pourront avoir des processus de raisonnement (i.e. des scripts) différents, mais aboutir à des conclusions similaires. Cela a été corroboré par les résultats obtenus sur des TCS, par exemple en neurologie ([Lubarsky et al., 2009](#)).

Nous prévoyons maintenant de creuser cette validation dans le domaine de l'ingénierie, en faisant passer nos TCS à des groupes d'experts avec des profils variés (ingénieurs et enseignants dans différentes formations d'ingénieur dans le monde francophone) et en comparant les résultats alors obtenus. Mais, dans une première approche, nous avons décidé d'enrichir le principe de base des TCS d'un champ "justification". Notre objectif était de comprendre ainsi le raisonnement qui avait conduit les étudiants à opter pour telle ou telle réponse. Ce champ supplémentaire nous a donné de précieuses informations sur la manière dont les étudiants comprenaient et analysaient les cas d'étude qui leur étaient ainsi posés.

La principale limite de l'analyse qui suit réside dans notre incapacité de comparer les résultats obtenus d'une part par une population à qui aurait été soumis un sujet sans TCS et, d'autre part, par une population devant traiter des TCS : pour des raisons d'équité, nous devons proposer le même sujet à l'ensemble d'une promotion. Pour contourner cet obstacle, nous envisageons, dans un avenir proche, d'utiliser les TCS non seulement dans des procédures d'évaluation, mais également dans des dispositifs d'auto-évaluation permettant aux étudiants d'améliorer leurs compétences avant un examen.

5.2. Réponses des experts

Pour cette première expérimentation, nous avons préalablement recueilli le retour de 10 à 12 experts par question, ce qui constitue la borne inférieure du nombre d'experts souhaitables pour qu'un TCS puisse être valide : dans ([Gagnon et al., 2005](#)), les auteurs ont démontré que, pour obtenir une consolidation satisfaisante des notes dans le cas d'examens cruciaux (tels que ceux visant à délivrer une certification aux étudiants), le groupe d'experts doit être constitué d'au moins 15 membres. Dans cette première phase d'expérimentation des TCS, nous n'avons pas pu mener un test d'une telle ampleur, faute d'experts en nombre suffisant. Mais les auteurs de ([Gagnon et al., 2005](#)) précisent également que, pour des enjeux moindres (tels que cette expérimentation), un panel d'expert de 8 à 10 membres peut suffire. Depuis, nous sommes parvenus à élargir notre panel d'experts, ce qui nous permettra des évaluations plus fines à l'avenir. Notons que la cotation des TCS n'est pas anodine et qu'elle requiert un investissement en temps de la part des experts. Ceux-ci ont généralement passé entre 1h35 et 1h50 à évaluer les 64 vignettes qui leur étaient proposées, soit généralement plus de 90 secondes par vignette.

Nous présentons dans les graphiques suivants, les réponses comparées entre experts et étudiants pour quelques-uns des exemples précédents. Les résultats, donnés en pourcentage, correspondent aux réponses recueillies auprès des experts et de plus d'une centaine d'étudiants sur deux années universitaires. Nous n'avons pas encore pu consolider les réponses des étudiants sur un jeu de questions qui seraient posées à l'identique à des promotions différentes. En effet, les étudiants de l'année (n+1) ayant accès aux sujets d'examen donnés à l'année (n), nous ne souhaitons pas introduire de biais méthodologique dans notre approche.

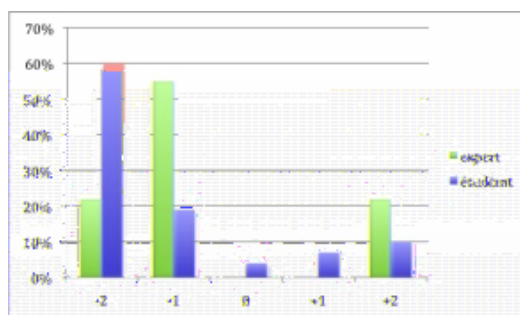


Figure 3 • Répartition des réponses exemple 1 – SI (Tableau 1)

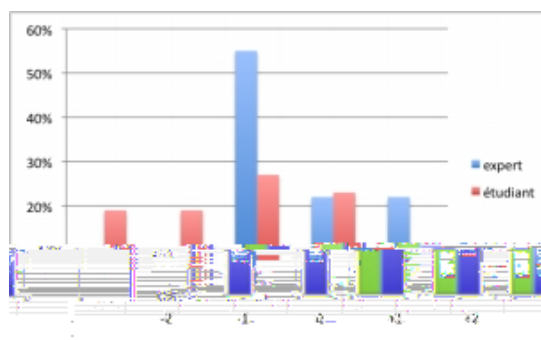


Figure 4 • Répartition des réponses exemple 5 – SI (Tableau 5)

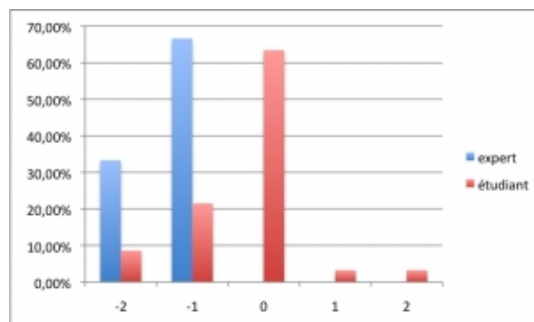


Figure 5 • Répartition des réponses exemple 2 – Droit (Tableau 7)

Nous avons constaté que, sur certaines questions, les experts étaient moins catégoriques que les étudiants. Il en va ainsi de notre question "exemple 1 – SI". Par rapport à l'effet de la nouvelle donnée sur l'hypothèse formulée (voir Figure 3), plus de 20% des experts la jugent "beaucoup plus probable" tandis que 60% des experts préfèrent un avis plus nuancé en répondant seulement "moins probable". Les étudiants, eux, choisissent, pour 60% d'entre eux, la formulation "beaucoup plus probable" et 19% l'hypothèse "moins probable". On assiste là à une inversion presque parfaite des réponses entre experts et étudiants. Si les deux réponses vont dans le même sens (l'hypothèse paraît confirmée par la nouvelle donnée), leur gradation n'est pas envisagée de la même manière par les experts d'une part et les étudiants d'autre part. Cette tendance des étudiants à choisir des éventualités plus fortes que les experts est confirmée sur les autres questions. Échaudés par des cas pratiques souvent difficiles à dénouer, les experts ont certainement plus tendance à une certaine réserve que des étudiants habitués, de par leur formation passée (une très large majorité de nos élèves sont issus de classes préparatoires scientifiques), à des devoirs "parfaitement posés" et à répondre à des problèmes. La question 1 montre un regroupement massif des étudiants sur l'éventualité "moins probable", là où les experts sont moins tranchés. Ce phénomène nous conduit à émettre deux remarques :

- nous pensons qu'il faut un panel d'experts d'une vingtaine de personnes pour être vraiment significatif et qu'il faudra respecter quelques règles sur la composition de l'échantillon. Il semble notamment important d'interroger des personnes extérieures à l'équipe pédagogique pour vérifier que les situations décrites dans le scénario du TCS correspondent bien à des configurations objectivement compréhensibles ;
- ce déséquilibre pose question quant à la notation à pondérer en fonction des avis des experts.

Sur un certain nombre de TCS (voir, par exemple, la figure 4), les experts ne se rassemblent pas tous sur la même éventualité, mais leur réponse indique un même spectre de tendances. Ce qui signifie qu'au moment de sa réponse, l'étudiant aura la possibilité de nuancer sa réponse sans que cela ne lui porte préjudice au moment de l'évaluation.

Sur quelques TCS (voir, par exemple, la figure 5), les experts sont quasiment unanimes. Les situations associées à ces TCS sont bien ouvertes, mais correspondent à des cas suffisamment explicites pour que les experts se retrouvent tous avec une analyse similaire. Comme nous l'avons signalé dans le paragraphe

précédent, il est notable de voir que les experts ne se rassemblent pas forcément sur une des hypothèses extrêmes : ainsi, c'est parfois une hypothèse nuancée qui fait l'unanimité.

Lors des cotations, il est apparu que certaines questions pouvaient aboutir à des résultats différents en fonction du contexte pédagogique. Ainsi, si un enseignement se restreint à un champ spécifique de connaissances (plutôt que d'embrasser toute la généralité du domaine), l'adéquation entre la bonne acquisition d'une compétence et les réponses des experts risque d'être remise en cause. Nous avons rencontré ce cas sur des TCS dédiés à la programmation, et plus spécifiquement à la manière dont une spécification algorithmique doit être traduite en C. En première année, à l'école Centrale, nous n'enseignons que le passage des paramètres par adresse ou par valeur, et non pas par référence. Il en résulte de possibles différences dans les réponses aux TCS afférents, en fonction d'une part de la familiarité des experts avec le contexte pédagogique dans notre établissement, d'autre part des expériences précédentes des étudiants dans le domaine visé. Certains élèves (issus notamment de cursus suivis à l'étranger) ont déjà acquis des compétences dans les matières enseignées en France, et vont, de fait, connaître ce qu'est le passage par référence. Ils vont être amenés à des réponses différentes de leur camarade et, potentiellement, des experts (en fonction du panel d'experts qui avait préalablement été constitué). Or cette réponse peut être tout à fait recevable.

Il en résulte la nécessité d'écrire des TCS suffisamment objectifs, qui s'abstraient de certaines spécificités de l'enseignement dans nos établissements respectifs. Et pour vérifier l'objectivité des TCS en question, il est indispensable de recueillir les réponses d'un large panel d'experts, mêlant des enseignants internes à l'établissement avec des personnes externes. Surtout, il apparaît nécessaire de créer des moments de discussions autour de certains énoncés équivoques et, de fait, de favoriser les échanges entre les experts.

5.3. Evaluation des étudiants

Dans le cas de TCS dans des domaines d'application de l'informatique (les systèmes d'information, le droit dans la société de l'information, etc.), la réponse peut être influencée par la connaissance que le candidat peut avoir de certaines informations complémentaires.

Prenons l'exemple de la question "exemple 3 – SI" et considérons le cas d'un étudiant qui connaîtrait bien les enseignants de l'école. Si nous parlons de la matière Systèmes d'Information et Bases de Données (effectivement enseignée dans l'établissement où le TCS est donné) et si l'étudiant émet une hypothèse basée sur ce qu'il a constaté du responsable, la réponse pourrait être différente de celle d'un étudiant qui n'a aucune connaissance sur ce responsable. Le même problème se pose dans le secteur médical, si le scénario proposé aux étudiants fait état d'un patient qu'ils auraient déjà eu l'occasion de rencontrer lors de leurs stages cliniques. C'est pourquoi les pathologies sont généralement décrites de la manière la plus objective possible. Autrement dit, pour renforcer la validité des TCS, il apparaît important de les décontextualiser le plus possible. Pour illustrer cet effort, citons le TCS du tableau 10 qui généralise l'exemple présenté sur le tableau 1.

| | |
|-----------------|---|
| Scénario | Pour payer ses personnels dans une société de travail par intérim, le responsable de la paie édite la liste des heures effectuées par personne à partir des éléments contenus dans le système d'information. Lors de cette édition, Luc Baudoin apparaît deux fois. |
| Option | Si vous pensez à : un doublon dans la base de données |
| Nouvelle donnée | La date de naissance est différente |

Tableau 10 • Reprise de l'exemple 1 - SI

Il est ainsi primordial de présenter les cas d'étude de la manière la plus universelle possible et de bannir

toute connaissance subjective du cas posé : interroger les étudiants de l'Ecole Centrale à propos de la gestion même de l'Ecole Centrale comporte trop de risques en termes d'interprétations "latentes".

La formulation d'une question revêt ainsi une importance capitale dans la pertinence et l'efficacité d'un TCS pour tester la compétence visée. Il en va également de la répétition (ou la non-répétition) d'un même mot. La question "exemple 5 – SI" est ambiguë car chaque individu pourra répondre en considérant qu'un déplacement équivaut à un remboursement, tout comme un autre considérera qu'un déplacement comprend plusieurs remboursements. Bref les étudiants pourront répondre correctement mais en ayant considéré un contexte différent, rendant ainsi leur réponse fautive (voir répartition des réponses sur la figure 3). Ceci met en avant, à notre sens, la nécessité d'une justification de la réponse donnée par l'étudiant.

Utilisés à la fin d'un enseignement, les TCS sont un outil précieux pour identifier des compétences non acquises et, ainsi, introduire des actions correctives. Illustrons notre propos sur le TCS du tableau 7 (droit) : il y a une forte différence entre les réponses des experts et celles des étudiants (cf. figure 5) Mais en examinant les justifications des élèves, on constate que cela ne vient pas d'une inadéquation du TCS avec la compétence ciblée, mais d'une compétence non-acquise par une majorité des étudiants. Dans leurs explications, les étudiants passent complètement à côté du problème et ne se posent pas de question. Florilège des justifications les plus vues : *"la sécurité des données bancaires en Suisse ; le réseau électrique est plus fiable en Suisse ; les Suisses sont des gens compétents ; la Suisse est en Europe ; il y a des différences de vitesse sur des transferts de données France-France et France-Suisse ; l'éloignement des serveurs peut induire des pertes de données"*. Grâce à un TCS mettant si bien en évidence une compétence non acquise par un grand nombre d'élèves, il est possible d'ajuster le tir et d'aider les étudiants dans leur apprentissage. En l'occurrence, nous avons fait parvenir aux étudiants un mémo insistant sur ces problèmes de transferts de données à l'étranger.

À la fin de chacun des examens dans lequel nous avons expérimenté les TCS, nous avons interrogé une douzaine d'étudiants sur leur ressenti face à ce nouvel exercice. Un questionnaire plus complet pour recueillir l'avis des élèves est d'ores et déjà prévu pour les prochaines utilisations de TCS au sein de nos examens.

Certains étudiants ont jugé les TCS "plus difficiles" que des QCM car cela force à se poser des questions, là où les QCM fonctionnent sur un mode généralement binaire "je connais la réponse/je ne la connais pas".

Ce sont les étudiants étrangers (11 sur 110 en 2009, 14 sur 108 en 2010) qui ont été le plus déstabilisés par l'exercice. En cause, la rédaction de certaines questions ; chaque mot décrivant les cas d'étude a son importance. Ainsi, en 2009, un quart des élèves étrangers ont trébuché sur le sens du mot "doublon" qui apparaissait comme hypothèse d'un des TCS posés. Il s'agit toutefois d'un terme technique que les étudiants sont supposés avoir rencontré au cours de l'enseignement, et donc maîtriser. Cela ne nécessite donc pas une réécriture du TCS. À l'inverse, citons le terme "convention collective", qui apparaissait dans un énoncé et qui a perturbé une même proportion d'étudiants étrangers. Il s'agit là d'un élément de vocabulaire sans lien avec le cœur de la matière, et qu'il peut donc être utile de clarifier dans l'énoncé. À noter, enfin, que les différences d'interprétation entre étudiants français et étrangers ne réside pas seulement dans le vocabulaire utilisé pour décrire un scénario, mais également dans l'utilisation de certains éléments "culturels" aussi basiques qu'un prénom ou un nom de famille. Reprenons le TCS présenté sur le tableau 1 : la simple modification du nom utilisé dans le TCS – par exemple "Rémi Rutaberge de Fontvielle" au lieu de "Luc Baudoin" – peut conduire à une modification de son interprétation : la fréquence d'apparition d'un nom tel que "Luc Baudoin" dans la population est (bien) plus élevée que celle de "Rémi Rutaberge de Fontvielle", laissant ainsi penser qu'il peut y avoir homonymie entre deux personnes dans le premier cas, erreur de saisie de la date de naissance dans le second cas. Nous avons constaté que les étudiants étrangers (chinois notamment) avaient ainsi du mal à prendre en compte ce paramètre dans leur raisonnement.

Depuis cette première expérimentation, nous avons effectué un effort particulier sur la formulation de nos TCS, afin de limiter les difficultés de compréhension liées à la langue. La population d'étudiants étrangers étant encore relativement réduite comparativement au nombre d'élèves français, il est difficile de tirer des conclusions définitives sur la pertinence du dispositif. Cela dit, nous avons constaté, dans nos

expérimentations ultérieures, que les réponses des étudiants étrangers se répartissent, globalement, comme celles de leurs collègues français.

Lors de l'examen, les étudiants étrangers ont la possibilité d'interroger les surveillants pour connaître la signification exacte d'un mot, mais ce droit n'est pas systématiquement utilisé. Ainsi, pour que les TCS évaluent bien les compétences des étudiants étrangers en informatique – et non seulement leur bonne compréhension des termes de la langue française – le concepteur de TCS doit prendre garde au jargon et aux références culturelles qu'il utilise. Encore plus que dans un sujet d'examen classique, il convient de porter attention à chacun des termes utilisés.

6. Conclusions et perspectives

Pour adapter les TCS à l'informatique, nous avons identifié des forces et des limites à cette approche.

Les TCS (dont l'efficacité a déjà été prouvée par des études scientifiques dans le domaine médical) constituent un outil efficace et pertinent pour évaluer le raisonnement/savoir-faire d'étudiants placés dans une situation professionnelle. L'évaluation d'une proposition ne se fait pas sur la base d'une unique réponse correcte, mais en considérant une batterie de réponses acquises auprès d'un panel d'experts. Autrement dit, il n'y a pas de réponse unique, mais il y a des réponses plus pertinentes que d'autres.

Les TCS ne sont pas adaptés à toutes les situations : ils visent plutôt des configurations incertaines et/ou ouvertes, dans lesquelles on n'a pas forcément accès à toutes les données. Ils ne sont pas pertinents dans des configurations parfaitement définies. Ils vont dénoter d'un savoir-faire par rapport à une situation incertaine, là où les QCM "classiques" vont permettre d'évaluer un savoir par rapport à une situation très bien définie. C'est un outil complémentaire d'approches plus traditionnelles.

Les TCS sont particulièrement adaptés dans des domaines tels que l'analyse des systèmes d'information, la programmation (débugage) ou le droit informatique. Ils s'appliquent efficacement pour la conception et le diagnostic. Nous comptons évaluer l'intérêt des TCS dans une démarche d'auto-évaluation (par exemple dans la perspective de valider des compétences liées au référentiel C2i). Nous prévoyons des les coupler au déploiement d'une plate-forme informatique permettant aux étudiants de passer ces tests en autonomie. Enfin, nous souhaiterions élargir le dispositif et étudier l'intérêt des TCS dans d'autres disciplines des sciences de l'ingénieur.

7. Remerciements

Les auteurs remercient D. Lime, O. Roux, M. Servières et V. Tourre pour avoir pris part aux réflexions sur l'usage des TCS en informatique à Centrale Nantes et S. Lorenzo pour la coordination du projet UNIT initial.

BIBLIOGRAPHIE

ACKOFF, R. L. (1989). From Data to Wisdom. *Journal of Applied Systems Analysis*, Vol. 16, p 3-9.

CHARLIN B, BORDAGE G. and VAN DER VLEUTEN C. (2003). L'évaluation du raisonnement Clinique. *Pédagogie Médicale*, Vol.4, p.42-51.

CHARLIN B., GAGNON R., KAZI-TANI D. and THIVIERGE R. (2006). Le test concordance comme outil d'évaluation en ligne du raisonnement des professionnels en situation d'incertitude. *Revue internationale des technologies en pédagogie universitaire*, Vol. 2(1).

CHARLIN B., GAGNON R., SAUVÉ É. and COLETTI M. (2007). Composition of the panel of reference for concordance tests: Do teaching functions have an impact on examinees' ranks and absolute scores? *Medical Teacher*, Vol. 29(1), p. 49-53

CIOCOIU C., PLOIX S., MICHAU F. (2001). Environnement d'apprentissage coopératif pour élèves ingénieurs. *Colloque Questions de Pédagogie dans l'enseignement supérieur*, Brest, France, 7 pages.

Certificat Informatique et Internet, <http://www2.c2i.education.fr/> (consulté le 12 février 2011)

GAGNON R., CHARLIN B., COLETTI M., SAUVÉ E., VAN DER VLEUTEN C. (2005). Assessment in context of uncertainty: How many members are needed on the panel of reference of a script concordance test? *Medical Education*, Vol. 39, p. 204-291.

GRANT J, MARSDEN P. (1988). Primary knowledge, medical education and consultant expertise. *Med Educ.* Vol. 22, p.173-179.

LEMAY JF., DONNON T., CHARLIN B. (2010). The reliability and validity of a paediatric script concordance test with medical students, paediatric residents and experienced paediatricians. *CMEJ 2010*, Vol. 1(2), p. e89-e95.

LUBARSKY S., CHARLK C., KAZITANI D., GAGNON R., CHARLIN B. (2009). The script concordance test: A new tool assessing clinical judgement in neurology. *Can. J. Neurol. Sci.* 2009, Vol. 36, p. 326-331

MICHEL C. (2009). Dispositif de supervision pour les tuteurs impliqués dans un apprentissage à la gestion de projets, *Conférence EIAH'2009 Environnements Informatiques pour l'Apprentissage Humain*, Le Mans, France, p.1-5.

PERRENOUD P. (2004). Évaluer des compétences. L'éducateur, p. 8-11. Disponible sur Internet : http://www.unige.ch/fapse/SSE/teachers/perrenoud/php_main/php_2004/2004_01.html (consulté le 12 février 2011).

SALIOU P. et RIBAUD V. (2005). La revue de pairs, un support au transfert de compétences inter-étudiants dans un paradigme d'apprentissage par projet. *Colloque Pédagogie dans l'enseignement supérieur: Nouveaux contextes, nouvelles compétences*. Lille. p. 160-165.

Test de concordance de script, <http://www.cpass.umontreal.ca/tcs.html> (consulté le 12 février 2011)

Courriel : A propos des auteurs

Morgan MAGNIN est maître de conférences en informatique à l'École Centrale de Nantes depuis 2008. Il effectue ses recherches au sein du laboratoire IRCCyN (Institut de Recherche en Communications et Cybernétique de Nantes) en bio-informatique, plus précisément dans le domaine de la biologie des systèmes. Il est également en charge de la mission EAT-TICE (pour "Enseigner et Apprendre avec les Technologies - Technologies de l'Information et de la Communication pour l'Enseignement") de l'Ecole Centrale de Nantes depuis 2010. Convaincu par le potentiel des technologies de l'information et de la communication dans le secteur de l'enseignement et de la recherche, il se concentre sur les initiatives et projets destinés à améliorer les démarches pédagogiques et les processus d'apprentissage.

Adresse : Ecole Centrale de Nantes. 1 rue de la Noë, BP 92101, 44321 Nantes Cedex 3

Toile : morgan.magnin@ec-nantes.fr<http://www.morganmagnin.net/>

Guillaume MOREAU a obtenu son doctorat en informatique en 1998 et son habilitation à diriger des recherches en 2009. Professeur des universités en informatique à l'Ecole Centrale de Nantes, il effectue ses recherches au sein du laboratoire CERMA (Centre de Recherche Méthodologique d'Architecture) sur le thème de l'étude du lien entre images numériques et systèmes d'information à travers les techniques de la réalité virtuelle et de la réalité augmentée. Il est aussi directeur du système d'information de l'Ecole Centrale de Nantes depuis 2006 après avoir été chargé de mission TICE. Dans ce cadre, il s'intéresse fortement à l'utilisation et à l'appropriation des nouvelles technologies dans le cadre de l'enseignement.

Adresse : Ecole Centrale de Nantes. 1 rue de la Noë, BP 92101, 44321 Nantes Cedex 3

Courriel :

Toile : guillaume.moreau@ec-nantes.fr<http://moreaug44.free.fr/blog/>

Référence de l'article :

Morgan MAGNIN, Guillaume MOREAU, Utilisation des Tests de Concordance de Scripts pour l'évaluation en informatique, Rubrique de la revue *STICEF*, Volume 18, 2011, ISSN : 1764-7223, mis en ligne le 7/12/2011, <http://sticf.org>

© Revue Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation, 2011