

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

Magnin Morgan, Moreau Guillaume
{morgan.magnin|guillaume.moreau}@ec-nantes.fr
École Centrale de Nantes
1 rue de la Noë - BP 92101
44321 Nantes Cedex 3

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

L'École Centrale de Nantes est une école d'ingénieurs généraliste du Ministère de l'Enseignement Supérieur et de la Recherche accueillant environ 1700 étudiants dont un millier d'élèves ingénieurs, mais aussi des masters, des doctorants et des formations en alternance. Elle est membre du Groupe des Ecoles Centrale, regroupant 5 écoles en France auxquelles il faut ajouter l'École Centrale de Pékin. La vocation des Ecoles Centrales est de former des ingénieurs généralistes de haut niveau scientifique et technique, entrepreneurs et innovateurs. Cette formation passe par une pleine appropriation des technologies de l'information et de la communication (TIC). Ce choix est désigné sous le titre EAT-TICE : Enseigner-Apprendre avec les Technologies, Technologies de l'Information et la Communication au service de l'Enseignement. Depuis 2004, l'École Centrale de Nantes a développé une politique d'appropriation par les enseignants et les étudiants de ces technologies, mais aussi du développement de l'autonomie (25% du total de la formation) et du raisonnement à partir de cas (Problem Based-Learning) faisant le plus souvent appel aux TIC. Forte de plus de 450 cours en ligne pleinement utilisés par enseignants et étudiants, l'École Centrale de Nantes a participé à plusieurs projets dans le domaine des TICE ces dernières années : elle a notamment été lauréate en 2008 du prix HP « Technology for Teaching » sur un projet visant à développer l'usage des Tablets PC pour les enseignants comme les étudiants. Ce projet « Most complex systems with simplest Tools » vise la convergence entre le monde numérique (informatisé) et le

monde analogique (le manuscrit) grâce aux Tablets-PC. C'est la question de l'évaluation sur machine qui nous a conduit à réfléchir à nouvelles formes d'évaluation, notamment dans des matières comme l'informatique.

Dans le cadre des formations professionnalisantes que nous dispensons, il importe de disposer d'outils pour l'évaluation des compétences des élèves en situation professionnelle. 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. Afin de cibler certaines compétences primordiales, l'École Centrale de Nantes a participé à un projet UNIT sur les tests de concordance de scripts (TCS), une forme de généralisation des questionnaires à choix multiples (QCM), jusque là réservés au domaine médical pour l'évaluation des capacités de diagnostic des étudiants en médecine. Comme leur nom le laisse penser, les TCS sont basés sur la théorie cognitive dite « des scripts », qui se focalise sur les réseaux de connaissance 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 [2]. L'objectif de cet article est de rendre compte des expérimentations menées à l'École Centrale de Nantes autour d'une nouvelle déclinaison des TCS pour l'informatique, à la fois dans le cadre du projet UNIT et au delà, celui-ci ne visant que le niveau L. Nous avons étudié et conçu nos TCS également au niveau M.

La suite de cet article est organisée de la façon suivante : dans un premier temps, nous présentons la notion de TCS tels qu'ils sont utilisés dans le domaine médi-

cal. 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 2 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 [2] 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 universités relevant du secteur médical. 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.

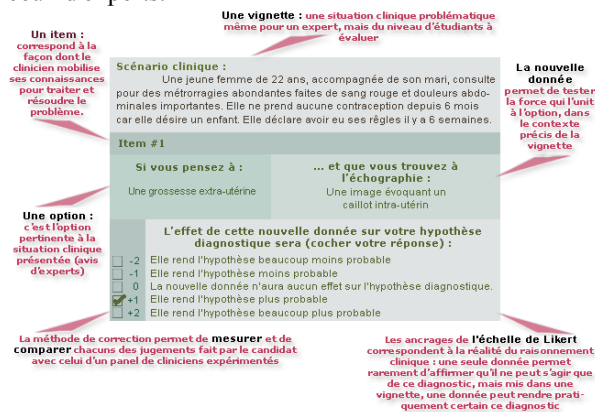


Figure 1: TCS type du domaine médical, tiré de [8]

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 2 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 peut s'effectuer manuellement, mais est plus souvent réalisé à l'aide de systèmes informatiques. En particulier, l'Université de Bordeaux 2 a développé un outil pour gérer entièrement les TCS (saisie des questions, passation du test et calcul des scores) : <http://www.evalutoile.u-bordeaux2.fr/>

Le TCS permet d'évaluer le raisonnement en situation incertaine de manière isolée. Une note de stage (ou d'autres mises en situation) est globale, elle contient beaucoup de choses : de la politesse et ponctualité à la pertinence scientifique en passant par le raisonnement. Le TCS permet, quant à lui, de cibler le raisonnement.

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 [6] et pédiatrie [5].

Dans [2], 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 TCS POUR L'INFORMATIQUE

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 [7]).

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 des étudiants sur la plate-forme Linux (manipulation de la ligne de commande, arborescence des fichiers, etc.). Mais nous ne disposons d'aucun outil pour évaluer des compétences en milieu professionnel.

Or 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 (i.e. Bac+3) : 25 TCS
- Cours de méthodes logicielles – E.I. de 2e année (i.e. Bac+4) : 10 TCS

3.2 Adaptation de la notion de TCS

Les discussions préliminaires sur le projet ont mis à jour la nécessité de bien situer une approche TCS par rapport à une approche QCM classique. L'un des enjeux principaux est de proposer des situations suffisamment ouvertes pour qu'un panel d'experts ne réponde pas de manière uniforme à un même test. En effet, dans un tel cas, on perdrait l'une des spécificités des TCS : l'évaluation du savoir-faire face à un problème ouvert dont toutes les données ne sont pas connues/définies (à l'instar du diagnostic médical qui ne se base que sur les éléments – incomplets – dont le médecin a connaissance). A cet égard, on peut voir un TCS très ciblé et à réponse unique comme un QCM !

Un autre enjeu du projet était de déterminer l'échelle la plus adaptée pour l'usage de TCS en informatique. Pour ce faire, nous avons du discuter les avantages et inconvénients des deux principales échelles suivantes :

- confirme l'hypothèse (+2) / rend l'hypothèse plus probable (+1) / n'a aucun effet sur l'hypothèse (0) / rend l'hypothèse moins probable (-1) / exclut l'hypothèse (-2) ;
- rend l'hypothèse beaucoup plus probable (+2) / rend l'hypothèse plus probable (+1) / n'a aucun effet sur l'hypothèse (0) / rend l'hypothèse moins probable (-1) / rend l'hypothèse beaucoup moins probable (-2).

Au final, la seconde échelle nous est apparue la plus pertinente. C'est d'ailleurs celle qui est privilégiée à l'Université de Montréal. Il est vrai que l'appréciation entre « beaucoup plus probable » et « plus probable » peut varier selon les personnes, comme nous le verrons d'ailleurs dans la section 5.2.

Toutefois, la première échelle (avec exclusion et confirmation) est plus « directive ». Cela a des avantages sur certains TCS, mais cela a été perçu par certains enseignants - dans le cas de problèmes véritablement ouverts - comme trop discriminant (car cela ne reflète

pas suffisamment la réponse). Et donc comme une échelle pouvant mener trop souvent à une répartition des réponses des experts allant à 100% sur une modalité et à 0% sur toutes les autres.

Or, nous l'avons vu, il est important d'éviter, dans le cas des TCS, des répartitions du type 0% / 100% / 0% / 0% / 0%. De plus, la formulation « confirme/exclut l'hypothèse » peut donner l'impression que nous avons affaire à une configuration où tout est, au final, pleinement formulé : le genre de cas qui pourrait être, de fait, décrit à l'aide de QCM plutôt que de TCS.

Nous illustrerons cette discussion sur un TCS écrit pour évaluer une compétence en droit informatique dans la section 4.3.

4 EXEMPLES

A ce jour, 64 TCS ont été rédigés dans le domaine de l'informatique. Après avoir exposé nos motivations, nous présentons dans les paragraphes qui suivent des exemples qui nous paraissent représentatifs des matières choisies.

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 » et en particulier pour sa composante non technique. 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 [1] : 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 : les TCS qui suivent font pour certains partie de l'expérience quotidienne du directeur du système d'informations de l'École Centrale de Nantes, impliqué dans le projet en tant qu'enseignant.

À ce jour, seul le cours de « Systèmes d'information et base de données » a fait l'objet d'expérimentations

grandeur nature. Néanmoins, d'autres TCS ont été rédigés pour les cours « d'Algorithmique et programmation » et « Méthodes logicielles ».

4.2 Exemples de TCS « SI »

Dans ces exemples, nous introduirons successivement le scénario, l'hypothèse de départ, l'information complémentaire puis une explication des différentes réponses possibles.

<i>Scénario</i>	Pour payer les enseignants non-permanents de l'Ecole, le secrétariat Général édite la liste à partir du système d'information. Lors de cette édition, Luc Baudoin apparaît deux fois
<i>Option</i>	Si vous pensez à : un doublon dans la base de données
<i>Nouvelle donnée</i>	la date de naissance est différente

Tableau 1 : exemple 1 – SI

A priori, ce premier test (Tableau 1) peut laisser penser à un doublon dans la base de données. Une date de naissance différente conduit naturellement à deux hypothèses (d'ailleurs non exclusives) : une simple homonymie ou une erreur sur une ou plusieurs dates de naissance. Ce dernier cas n'exclut pas le doublon, mais ne le valide pas totalement non plus : il n'est pas impossible, même si c'est très peu probable, d'avoir affaire à deux personnes de même nom et de dates de naissance identiques. Enfin, on peut aussi penser à une erreur de saisie des noms !

Il est assez facile de tirer des cas dérivés de ce premier test. Ils correspondent là encore à des cas vécus.

<i>Scénario</i>	Pour payer les enseignants non-permanents 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
<i>Option</i>	Si vous pensez à : Un champ prénom mal renseigné dans la base de données
<i>Nouvelle donnée</i>	Ils ont tous les deux enseigné dans le département Info-Maths

Tableau 2 : exemple 2 - SI (dérivé du cas précédent)

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. Avec près de 450 vacataires employés par an en plus des 120 enseignants-chercheurs permanents, l'Ecole Centrale de Nantes n'est pas à l'abri de ce type de risque. Néanmoins, ce type de test est par trop contextualisé, nous en reparlerons par la suite.

<i>Scénario</i>	Lors de l'édition du programme pédagogique de l'année suivante, les informa-
-----------------	--

	tions concernant la matière X sont visiblement fausses
<i>Option</i>	Si vous pensez à : Que l'enseignant responsable de la matière n'a pas entré les bonnes informations
<i>Nouvelle donnée</i>	Les logs de transaction vous disent qu'il a modifié la description de la matière la veille

Tableau 3 : exemple 3 – SI

Ici également, il est difficile de conclure : le travail effectué par l'enseignant la veille ne signifie pas qu'il ne s'est pas trompé et cela ne prouve pas non plus que l'enregistrement ait été effectué correctement. De même, la génération du programme pédagogique s'est-elle faite correctement ? Cela reste néanmoins assez lié au contexte Centrale Nantes.

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.

<i>Scénario</i>	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
<i>Option</i>	Si vous pensez à : Les données sont fausses quelque part
<i>Nouvelle donnée</i>	Les gestionnaires concernés dans les départements et les régions vous assurent des données entrées

Tableau 4 : exemple 4 - SI (intégrité des données)

Ce cas est encore un cas classique des systèmes d'information, facile à reproduire, en s'appuyant sur les sites web des conseils régionaux et départementaux. Bien entendu, aucun ne souhaite mettre à disposition des données fausses aux populations concernées. On peut bien sûr arguer de la non cohérence des sites. Référons-nous alors à l'encyclopédie en ligne Wikipedia pour consulter les démographies respectives de 5 départements de la région Pays de la Loire et celle totale de la région (date de consultation : 17 avril 2010). On aboutit bien à des valeurs différentes qui ne sont pas de simples erreurs d'arrondi. Dans le cas présent, un élément d'explication est donné directement sur le site : les données démographiques sont données à des instants différents (variant de 2007 à 2009). Pour autant difficile de conclure sur un TCS, sans cette information, qui n'est d'ailleurs peut-être pas la seule explication. On peut néanmoins s'étonner que la même différence existe entre superficie de la région et somme des superficies des départements qui la composent !

<i>Scénario</i>	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 dé-
-----------------	---

	tectez une différence entre les montants définis dans l'application et les remboursements dans le logiciel de compatibilité
<i>Option</i>	Si vous pensez à : Une fraude
<i>Nouvelle donnée</i>	Les logiciels de gestion des frais de déplacement comporte 165 déplacements, la comptabilité fait état de 2315 remboursements

Tableau 5 : exemple 5 – SI

L'objectif ici était de montrer que les problématiques SI peuvent dépasser le cadre informatique : en effet, rien ne prouve qu'un logiciel, quelque 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 raisonnable. Nous avons ensuite abordé les problèmes d'accès concurrents au SI.

<i>Scénario</i>	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.
<i>Option</i>	Si vous pensez à : Le logiciel de réservation comporte un bug au niveau de la vérification d'attribution des places
<i>Nouvelle donnée</i>	Le problème ne se produit qu'une fois par an

Tableau 6 : exemple 6 - SI (accès concurrents)

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 »

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.

<i>Scénario</i>	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.
<i>Option</i>	Si vous pensez à : Il est nécessaire de supprimer l'intégralité des fichiers d'adresses e-mail.
<i>Nouvelle donnée</i>	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 7 : exemple 1 - droit (commerce électronique)

Profitions 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. Rappelons que la discussion portait sur les deux formulations suivantes :

- confirme l'hypothèse (+2) / rend l'hypothèse plus probable (+1) / n'a aucun effet sur l'hypothèse (0) / rend l'hypothèse moins probable (-1) / exclut l'hypothèse (-2) ;
- rend l'hypothèse beaucoup plus probable (+2) / rend l'hypothèse plus probable (+1) / n'a aucun effet sur l'hypothèse (0) / rend l'hypothèse moins probable (-1) / rend l'hypothèse beaucoup moins probable (-2).

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 ».

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.

<i>Scénario</i>	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
<i>Option</i>	Si vous pensez à : A a tort car il est habituellement moins fort en programmation
<i>Nouvelle donnée</i>	B a vérifié le fonctionnement de toutes ses fonctions

Tableau 8 : exemple 1 – programmation

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.

<i>Scénario</i>	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
<i>Option</i>	Si vous pensez à : Vous pensez choisir la structure « arbre binaire de recherche ».
<i>Nouvelle donnée</i>	Le coût de comparaison entre deux éléments est important.

Tableau 9 : exemple 2 – programmation

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'EXPERIENCE

5.1 Généralités

Nous avons conduit ce projet en suivant les étapes suivantes :

1. 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) ;
2. Établissement du référentiel de compétences : définition de la liste de compétences à évaluer (début 2009).
3. Conception de questions : création d'une série de questions correspondant aux compétences visées (depuis le printemps 2009).
4. Recueil de l'expertise : exécution des TCS par les experts et ajustement de certaines questions en fonction du retour des experts (depuis le printemps 2009).
5. Premières expérimentations (juin 2009)

Nous avons mené une première expérimentation en juin 2009, sur une promotion de 110 étudiants, dans un examen qui comprenait, entre autres, 4 TCS. La matière concernée était « Systèmes d'Informations et Bases de Données », un enseignement de première année (équivalent Bac + 3) à Centrale Nantes. 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.

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. De plus, nous avons constaté que les réponses des experts se concentraient généralement sur deux ou trois éventualités, alors que les réponses des étudiants balaient plus largement le spectre des choix possibles. Cela corrobore l'idée d'une expertise commune des personnes nous ayant permis de constituer notre panel de référence.

La seconde limite est plus délicate à prendre en compte. Dans le secteur médical, il a été démontré [4] 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 [6].

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, pour le moment, 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 experts sur ces questions, ce qui constitue la borne inférieure du nombre d'experts souhaitables pour qu'un TCS puisse être valide : dans [3], 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 [3] précisent également que, pour des enjeux moindres (tels que cette expérimentation), un panel d'expert de 8 à 10 membres peut suffire. À terme, nous visons évidemment à élargir notre panel d'experts.

Nous présentons dans les graphiques suivants, les réponses comparées entre experts et étudiants pour quelques-uns des exemples précédents.

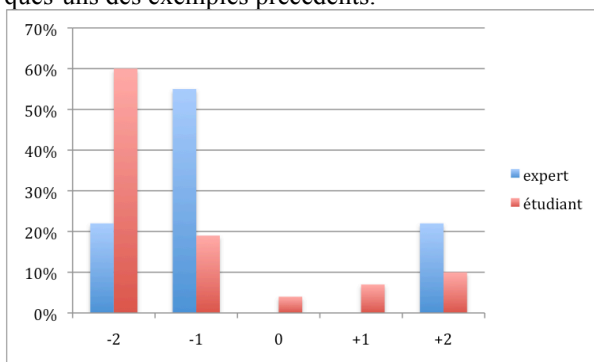


Figure 2 : Répartition des réponses exemple 1 - SI

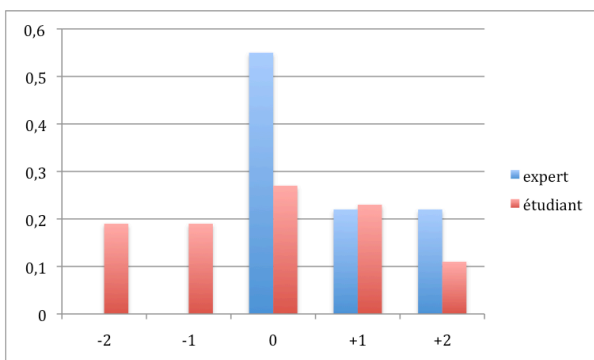


Figure 3 : répartition des réponses exemple 3 - SI

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 2), 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 enseignants 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 s'est retrouvée confirmée sur les autres questions. Échaudé 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 problèmes « 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.

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 dé-contextualiser le plus possible.

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 Figure 3). Ceci met en avant, à notre sens, la nécessité d'une justification de la réponse donnée par l'étudiant.

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, l'aspect « littéraire » de certaines questions ; chaque mot décrivant les cas d'étude ont leur 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. Une même proportion d'étudiants étrangers s'est retrouvée démunie devant le terme « convention collective » qui apparaissait dans un énoncé.

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 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, le droit informatique ou la programmation (débugage). 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). Enfin, nous souhaiterions élargir le dispositif et étudier l'intérêt des TCS dans d'autres disciplines des sciences de l'ingénieur.

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.

Références Bibliographiques et Webographiques

- [1] Ackoff, R. L., "From Data to Wisdom", *Journal of Applied Systems Analysis*, Volume 16, 1989 p 3-9.
- [2] 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*. 2(1).
- [3] 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*. 39: 204-291.
- [4] Grant J, Marsden P. "Primary knowledge, medical education and consultant expertise". *Med Educ*. 1988 ;22:173-9.
- [5] Lemay JF., Donnon T., Charlin B. "The reliability and validity of a paediatric script concordance test with medical students, paediatric residents and experienced paediatricians". *CMEJ* 2010;1(2):e89-e95.
- [6] Lubarsky S., Charlk C., Kazitani D., Gagnon R., Charlin B. "The script concordance test: A new tool assessing clinical judgement in neurology". *Can. J. Neurol. Sci*. 2009;36:326-331
- [7] <http://www2.c2i.education.fr/>
- [8] <http://www.cpass.umontreal.ca/tcs.html>