ReCALL **16** (2): 475–487. © 2004 Cambridge University Press
DOI: 10.1017/S0958344004001521 Printed in the United Kingdom

Évaluation et multimédia dans l'apprentissage d'une L2

MICHEL LAURIER
Faculty of Education, University of Montreal, PO Box 6132, Stn Centreville, Montréal (Quebec), Canada
(*e-mail:Michel.d.laurier@umontreal.ca*)

Abstract

In the first part of this paper different areas where technology may be used for second language assessment are described. First, item banking operations, which are generally based on Item Response Theory but not necessarily restricted to dichotomously scored items, facilitate assessment task organization and require technological support. Second, technology may help to design more authentic assessment tasks or may be needed in some direct testing situations. Third, the assessment environment may be more adapted and more stimulating when technology is used to give the student more control. The second part of the paper presents different functions of assessment. The monitoring function (often called formative assessment) aims at adapting the classroom activities to students and to provide continuous feedback. Technology may be used to train the teachers in monitoring techniques, to organize data or to produce diagnostic information; electronic portfolios or quizzes that are built in some educational software may also be used for monitoring. The placement function is probably the one in which the application of computer adaptive testing procedures (e.g. French CAPT) is the most appropriate. Automatic scoring devices may also be used for placement purposes. Finally the certification function requires more valid and more reliable tools. Technology may be used to enhance the testing situation (to make it more authentic) or to facilitate data processing during the construction of a test. Almond *et al.* (2002) propose a four component model (Selection, Presentation, Scoring and Response) for designing assessment systems. Each component must be planned taking into account the assessment function.

Il semble bien que la recherche et le développement en ce qui a trait à l'utilisation des technologies de l'information et de la communication (TIC) aient atteint une certaine maturité. Celle-ci se manifeste d'abord par l'élaboration d'environnements d'apprentissage diversifiés et riches. La quête de la nouveauté a été remplacée par un souci de mieux utiliser les ressources en tenant compte des limites des technologies existantes. Non seulement voit-on maintenant les possibilités qui s'offrent mais aussi les difficultés qui se posent. C'est aussi vrai des applications reliées à l'évaluation des apprentissages.

https://doi.org/10.1017/S0958344004001521 Published online by Cambridge University Press

De fait, on se rapproche de la suggestion de Salaberry (1996) qui prônait le retour vers des objectifs modestes en tablant sur l'ordinateur comme outil pour emmagasiner de grandes bases de données. Cela permet déjà beaucoup de choses à la condition qu'on utilise l'ordinateur pour ce qu'il est et pour ce qu'il peut faire. Dans ce contexte, Salaberry envisage quatre types d'innovations:

- *Le traitement de données* – L'ordinateur est un outil dont on reconnaît la valeur pour recueillir, emmagasiner et traiter des données. Ces qualités sont particulièrement intéressantes pour ce qui est de l'évaluation de la compétence langagière. Les spécialistes du testing savent depuis longtemps qu'ils peuvent largement miser sur cette capacité de l'ordinateur de traiter des quantités importantes de données en effectuant des analyses relativement complexes.
- *La modélisation du cheminement de l'apprenant* – Il est possible de garder la trace des activités des élèves: par exemple, le nombre d'exercices complétés et réussis ou même, à un niveau plus fin, le nombre d'accès aux écrans d'aide ou d'effacements à différents moments. Cette trace peut servir à construire un «modèle» de l'évolution des compétences langagières. De fait, cette application sert non seulement à renseigner sur le cheminement de l'apprenant, mais elle peut aider le chercheur à mieux comprendre comment les élèves utilisent un logiciel (Renié, 2000)
- *La gestion des ressources éducationnelles* – L'ordinateur fournit une aide précieuse à l'enseignant dans sa planification de classe bien que cette aide ne le remplace pas puisqu'il reste le responsable de son enseignement. L'ordinateur peut notamment servir de support à une activité avec les élèves dans la perspective de scénarios d'apprentissage (Laurier & Viens, 2000). Dans une autre optique, l'ordinateur peut servir à la réalisation de tâches administratives telles que la consignation d'observations de classe ou la circulation du matériel pédagogique.
- *L'interaction de diverses sources multimédias* – L'ordinateur sert à mettre en relation diverses sources d'information: vidéo, musique, texte... On peut même ajouter les communications interpersonnelles que l'ordinateur sert à médiatiser. L'apprenant est au centre de ces interactions entre différents médias. L'exploitation optimale de cette possibilité suppose évidemment un certain engagement de l'élève dans le processus d'apprentissage.

Malgré l'importance qu'on accorde généralement à l'évaluation, on trouve relativement peu de recherches dans le domaine de l'application des technologies pour l'évaluation de la compétence langagière en langue seconde. Pourtant, les axes d'innovation que trace Salaberry correspondent effectivement à des aspects où l'évaluation est directement concernée.

Cet article commencera par quelques considérations autour d'aspects pour lesquels la contribution des technologies à l'évaluation est ou pourrait être importante: la gestion des tâches évaluatives, l'évaluation authentique des apprentissages et le contrôle par l'apprenant. Dans la seconde partie, nous rappellerons les différentes fonctions de l'évaluation pour préciser ce qu'on peut attendre de l'utilisation des technologies à cet égard.

https://doi.org/10.1017/S0958344004001521 Published online by Cambridge University Press

1 La gestion des tâches évaluatives

Quand on pense à l'intérêt que peuvent présenter les TIC pour l'évaluation, on a généralement en tête le premier type d'innovation de Salaberry soit la capacité de traitement de données de l'ordinateur. C'est cette capacité qui a permis le développement des techniques de testing adaptatif et il ne fait pas de doute qu'il s'agit d'une exploitation judicieuse de l'ordinateur. De fait, c'est le développement technologique qui a permis les applications de la théorie de réponses aux items (TRI) puisque sans la capacité de traitement que permet l'ordinateur, les principes de cette théorie psychométrique n'auraient jamais trouvé d'application concrète. Il ne faut toutefois pas perdre de vue les limites qu'impose cette théorie (Wainer & Mislevy, 2000)

Dans le contexte de la TRI, un item se définit comme une question à laquelle l'apprenant répond et dont la correction est habituellement dichotomique (correct/incorrect). De fait, dans la majorité des applications, on trouve des items à choix multiple. Dans le domaine de l'évaluation de la compétence langagière, on peut imaginer une série d'items qui sont composés de phrases lacunaires pour lesquelles l'élève doit choisir le mot qui convient dans la liste des mots suggérés. Il peut aussi s'agir de courts paragraphes à lire suivis d'une question à choix multiple qui en vérifie la compréhension. Lorsqu'on réunit un certain nombre d'items de ce type, on peut mettre sur pied des banques d'items. À chacun des items d'une banque sont associées certaines caractéristiques métrologiques telles que sa difficulté et sa discrimination. S'il demeure possible de déterminer ces caractéristiques à l'aide des techniques classiques de la docimologie, l'établissement et l'expansion des banques peuvent se réaliser de façon plus efficace en appliquant les principes de la TRI car les caractéristiques obtenues sont moins tributaires des spécificités de l'échantillon utilisé pour leur mise à l'essai. Les algorithmes qui servent à déterminer les caractéristiques métrologiques sont complexes et le calcul requiert la capacité de traitement de l'ordinateur. Depuis quelques années, le concepteur de test peut choisir parmi plusieurs logiciels qui effectuent ces analyses psychométriques et qui permettent souvent, certes au prix d'une certaine lourdeur et d'une complexité parfois rebutante, d'analyser d'autres types de réponse que celles des items à correction dichotomique. On peut par exemple, à partir des jugements des correcteurs, déterminer les caractéristiques métrologiques de tâches de production orale ou écrite de façon à créer des banques de tâches complexes plutôt que des banques d'items à correction dichotomique.

Les banques d'items peuvent servir à la préparation de tests qui s'administrent autant sous une forme conventionnelle («papier-crayon») que sous une forme informatisée. Dans ce dernier cas, l'ordinateur permet de choisir dans la banque, l'item qui est le plus approprié. C'est ce principe qui sous-tend le fonctionnement des tests adaptatifs. Dans un test adaptatif, l'ordinateur choisit, dans la banque, sur la base des informations précédentes, l'item qui correspond le mieux au niveau estimé. Celui-ci est réestimé à chaque nouvelle réponse de l'élève de façon à choisir l'item suivant. Normalement, le processus se termine lorsque l'erreur de mesure a été réduite à une marge acceptable. Brown (1997) et Dunkel (1999) décrivent le testing adaptatif comme l'illustration la plus éloquente des possibilités qu'offrent les TIC pour ce qui est de l'évaluation de la compétence langagière. L'élaboration de tels instruments présente un intérêt indéniable

https://doi.org/10.1017/S0958344004001521 Published online by Cambridge University Press

sur le plan de la recherche car elle ouvre la voie à de nouvelles perpectives de modélisation de l'apprenant (Laurier, 1999; Chalhoub-Deville, 2001). Sur le plan pratique, les avantages sont clairs : précision de la mesure, passation individuelle, correction immédiate, frustration réduite chez l'élève (surtout s'il est débutant).

2 L'évaluation authentique des apprentissages

Il convient toutefois de tempérer cet optimisme en rappelant les limites sérieuses qui viennent de la TRI qui, malgré les nouvelles techniques d'analyse que nous avons évoquées ci-dessus, rendent difficile la pratique d'une évaluation authentique (Laurier, 2002). La mise en place d'une évaluation authentique requiert en effet des tâches complexes et contextualisées qui ne se conforment pas toujours aux postulats contraignants de la TRI. On peut déplorer que le testing adaptatif ait autant retenu l'attention des chercheurs qui se sont intéressés à la contribution des TIC en ce qui a trait à l'évaluation de la compétence langagière. Cette insistance a peut-être fait perdre de vue les avenues prometteuses qui s'ouvrent avec des environnements multimédias interactifs où il est possible de recréer des situations apparentées aux situations de vie dans lesquelles s'exerce la compétence. S'inspirant de courants actuels en évaluation des compétences, on peut maintenant créer des situations de test plus authentiques.

La démarche évaluative se définit comme un processus qui vise à inférer à partir d'une performance particulière que la compétence qui sous-tend cette performance permettra à l'apprenant de faire face à des situations réelles semblables. Un des problèmes auquel se trouve souvent confronté l'évaluateur est celui d'assurer une correspondance entre les caractéristiques de la situation de l'évaluation et celles des situations réelles visées. Grâce à des environnements multimédias interactifs, il devient possible de reconstituer une situation qu'on ne peut pas trouver dans une situation d'évaluation traditionnelle. Il s'agit donc en quelque sorte de repousser les contraintes de la situation évaluative en recréant les conditions d'utilisation authentique pour ainsi améliorer la validité des moyens mis en œuvre pour l'évaluation. On peut, par exemple, imaginer des tests de compréhension auditive où l'élève choisit la situation qui correspond le plus à celle qu'il est susceptible de rencontrer. On pourrait aussi concevoir des dispositifs d'évaluation de l'expression orale sur le modèle des entrevues semi-directes que propose Educational Testing Services où l'on simule un dialogue avec des questions pré-établies. Les TIC permettent alors de fournir des images afin de rendre la situation plus authentique et de prendre en compte certains paramètres pour l'adapter aux besoins particuliers de l'apprenant.

Par ailleurs, il faut souligner qu'il est de plus en plus fréquent qu'un sujet doive mobiliser les éléments d'une compétence langagière dans des situations qui intègrent les environnements informatisés ou multimédias. Par exemple, retrouver un document en langue seconde ou étrangère sur la Toile, résumer l'information ou l'intégrer dans un texte qu'elle doit organiser, mettre en page et réviser sont toutes des tâches qui sont normalement exécutées devant l'ordinateur. De telles situations sont courantes particulièrement dans l'exercice de tâches professionnelles ou académiques par des adultes. Nous sommes alors en présence d'une évaluation directe: l'évaluateur observe la performance dans une situation tout à fait authentique et porte un jugement sur la compétence qui sous-tend cette performance. Il ne s'agit plus de recréer une situation

https://doi.org/10.1017/S0958344004001521 Published online by Cambridge University Press

puisque la situation est en elle-même une situation de vie.

L'idée de placer l'apprenant dans une situation que se rapproche le plus possible d'une situation réelle est un aspect important du courant de l'évaluation authentique qui s'est développé au cours de la dernière décennie et dont Wiggins (1993) s'est fait le promoteur.

3 Le contrôle par l'apprenant

Parmi les propositions de Wiggins, on trouve aussi le recours au *portfolio* (le dossier d'apprentissage). L'usage du *portfolio* pour des fins d'évaluation s'accorde avec les principes socio-constructivistes dont se sont inspirées plusieurs révisions récentes de curriculum. Les *portfolios* ont connu un succès notable particulièrement dans l'enseignement des langues en milieu scolaire. Le *portfolio* peut se définir comme un ensemble de pièces qui manifestent la compétence de l'individu et qui est accompagné d'indications sur son cheminement. Ce dossier se construit en cours d'apprentissage et peut servir tant pour des fins de régulation des apprentissages que de certification des compétences. Par exemple, dans un cours d'écrit, l'élève peut choisir ses meilleures pièces (sur la base de son auto-évaluation, de l'évaluation de l'enseignant et de celle de ses pairs), justifier leur choix dans un paragraphe puis présenter ce *portfolio* comme démonstration de sa compétence. Le *portfolio* est un objet dynamique qui suit l'élève et qui change en cours d'apprentissage. C'est pourquoi il est intéressant que l'élève puisse utiliser l'ordinateur pour modifier les documents qui s'y trouvent, pour les organiser et pour les présenter. On voit donc maintenant se répandre l'idée d'un « *portfolio* électronique ». L'élève construit son dossier d'apprentissage en réunissant les fichiers de ses textes les plus représentatifs, les fichiers des projets dont il est le plus fier, les commentaires sur ces pièces; il peut choisir de présenter l'ensemble à l'aide d'un logiciel de présentation, d'établir des hyperliens entre les composantes et même de placer son *portfolio* sur la Toile.

Les enseignants qui ont utilisé le *portfolio* (électronique ou non) avec leurs élèves (jeunes ou adultes) savent que ceux-ci sont grandement motivés par sa construction. Cela tient en grande partie, au fait largement documenté dans la littérature sur la motivation (Schuman 1997) que l'individu qui apprend doit avoir un certain contrôle de la situation et peut ainsi se placer devant des tâches qu'il est susceptible de réussir. La technologie n'intervient donc pas seulement pour décider ce qui est le mieux pour l'élève (comme le fait le testing adaptatif) mais permet aussi certains choix pour modifier la situation. Ce concept apparaît dans un logiciel d'évaluation de la compétence à l'oral comme le *Computer Oral Proficiency Instrument* (COPI) développé par le *Centre for Applied Linguistics* (CAL) sur la base des lignes directrices de l'ACTFL (Kenyon & Malabonga, 2001). Le COPI commence par une auto-évaluation et l'élève peut choisir de faire des tâches plus ou moins difficiles en cours de route. Il peut aussi déterminer le temps dont il a besoin et sélectionner le thème de la tâche selon ses préférences.

4 Les fonctions de l'évaluation

Laurier et Dassa (2003) font une distinction entre diverses fonctions de l'évaluation des apprentissages. Chacune de ces fonctions correspond à une décision différente en vue de

https://doi.org/10.1017/S0958344004001521 Published online by Cambridge University Press

laquelle est mis en place un mécanisme d'évaluation. Ils distinguent ainsi six fonctions:

- *La régulation* – La régulation se réfère au type de décision que prend l'enseignant dans le cadre d'une démarche qui correspond à ce que plusieurs chercheurs francophones ont associé à une «évaluation formative». Scallon (2000:21) définit ainsi cette dernière: *processus d'évaluation continue ayant pour objectif d'assurer la progression des individus engagés dans une démarche d'apprentissage ou de formation, selon deux voies possibles: soit par des modifications de la situation ou du contexte pédagogique, soit en offrant à chaque individu l'aide dont il a besoin pour progresser, et ce, dans chacun des cas, pour apporter s'il y a lieu, des améliorations ou des correctifs appropriés. La « décision action », c'est-à-dire la régulation a pour objet, soit la situation d'apprentissage, soit l'individu lui-même*. La régulation est donc un ajustement qui doit être opéré de façon immédiate par le professeur en collaboration avec l'élève. En ce sens, elle est étroitement liée au processus d'enseignement et d'apprentissage. Elle peut tantôt porter sur des éléments très pointus de la compétence (le vocabulaire d'un secteur d'activité ou la conjugaison d'un verbe, par exemple) tantôt prendre une perspective plus globale. Cependant l'aspect diagnostique reste omniprésent. De ce point de vue, il est souvent utile de distinguer deux formes complémentaires d'évaluation formative: celle qui vise la «micro-régulation» et celle qui vise la «macro-régulation».
- *Le bilan périodique* – Dans la plupart des systèmes scolaires ou des programmes de formation, il est nécessaire, à différents moments du cursus (trois fois par année, par exemple), de renseigner un certain nombre d'acteurs sur le déroulement des apprentissages. Ce bilan, même s'il intéresse au premier chef l'élève, peut aussi s'adresser à des parents, à des responsables de programmes, à des équipes d'enseignants… Au cours des dernières années, on a vu de plus en plus s'imposer des balises (sous forme de *standards* ou de *benchmarks*) qui servent de référence à l'évaluation périodique des apprentissages. Il est certain que des observations recueillies dans le cadre d'une «macro-régulation» peuvent être réinvesties dans le bilan. Toutefois, l'information du bilan ne porte pas exclusivement sur les apprentissages réalisés mais peut prendre en compte des aspects comme le comportement, le progrès, l'attitude etc.
- *Le classement* – Lorsqu'un élève se présente dans un programme de formation et que les responsables de ce programme disposent de peu d'information à son sujet, il est important d'évaluer ses compétences afin de le placer dans le groupe qui correspond le plus possible à son niveau. Quand il s'agit de cours de langue, cette fonction est particulièrement importante. Sans viser la création de groupes parfaitement homogènes, on cherche alors à placer l'élève dans un environnement de classe susceptible de favoriser son apprentissage. De façon générale, la décision de classement consiste à situer un élève sur l'axe de développement d'une compétence. Il est important de souligner que cette décision s'exerce souvent dans un contexte où les considérations pratiques sont déterminantes. Ainsi, beaucoup d'établissements se trouvent souvent confrontés à devoir assigner un groupe à un grand nombre d'élèves, qui se présentent souvent tous au même moment, de façon à ce que la formation puisse démarrer

https://doi.org/10.1017/S0958344004001521 Published online by Cambridge University Press

sans délai. Le test de classement doit pouvoir rapidement attribuer un niveau à chacun, qu'il soit débutant, intermédiaire ou avancé.

- *La certification* – Les décisions de certification ont souvent des conséquences déterminantes pour l'avenir des individus. Ce sont en effet des épreuves de certification qui attestent de l'atteinte d'un niveau de compétence suffisant pour que, par exemple, un immigrant puisse exercer sa profession ou qu'un élève reçoive son diplôme. Étant donnée l'importance de la décision, la construction des instruments qui servent à la certification doit obéir à des exigences métrologiques élevées en termes de validité et de fidélité. Qui plus est, l'administration doit se dérouler dans des conditions uniformes qui garantissent l'équité. Contrairement aux instruments de classement, ces épreuves sont ciblées autour des tâches représentatives d'un niveau, celui dont on vise à attester de l'atteinte.
- *La sélection* – Bien qu'elle s'apparente à certains égard à la décision de certification, la décision de sélection s'en distingue du fait que l'interprétation se fait généralement de façon normative c'est-à-dire en comparant les résultats des candidats les uns par rapport aux autres. Les épreuves de sélection servent ainsi à déterminer les candidats qui méritent une bourse ou ceux qui peuvent être admis à un programme contingenté. Bien que relativement peu utilisés dans les programmes de formation linguistique, les instruments de sélection jouent un rôle majeur dans la plupart des systèmes éducatifs.
- *L'évaluation de programme* – Sans vouloir limiter l'évaluation de programme à l'évaluation des apprentissages, celle-ci n'en constitue pas moins une dimension importante. Il s'agit ici d'une fonction de pilotage de système qui permet de prendre des décisions non pas par rapport à des individus mais par rapport à une organisation, un type de matériel didactique, une approche pédagogique, un personnel enseignant etc. Dans un contexte de reddition de compte des organismes publics, cette fonction devient de plus en plus présente. Elle permet de gérer les ressources éducatives plus efficacement de manière à mieux prendre en compte les attentes des citoyens et des bénéficiaires.

5 Le rôle des technologies

De ces six fonctions, trois méritent ici une attention particulière dans la mesure où elles sont prépondérantes dans le domaine de l'enseignement des langues et dans la mesure où les technologies sont susceptibles de contribuer de façon significative à les bien réaliser: la régulation, le classement et la certification.

5.1 Régulation

Pour ce qui est de la régulation, il faut se rappeler d'abord que l'enseignant est en grande partie responsable de cette fonction. On a souvent évoqué la crainte que suscitent les TIC chez plusieurs enseignants que ni la formation ni l'expérience n'ont préparés à une utilisation confiante et intensive de ces nouvelles ressources pédagogiques. Pourtant, Lam (2000) conteste l'argument de la «technophobie» et suggère que beaucoup d'enseignants de langue utilisent relativement peu les TIC non pas parce qu'ils en ont peur, mais parce qu'ils ne voient pas en quoi elles favorisent

https://doi.org/10.1017/S0958344004001521 Published online by Cambridge University Press

l'apprentissage d'une langue. Dans le cadre d'une enquête menée dans l'État américain du Texas, Moore, Morales et Carel (1998) ont montré que les enseignants de langue utilisent peu les technologies. Cette étude confirmait les impressions de beaucoup d'intervenants et on pourrait sans doute généraliser cette conclusion ailleurs. Les auteurs en arrivaient à dire que cette sous-utilisation des TIC était en grande partie attribuable au fait que les enseignants ne savent pas comment les intégrer dans la classe. Au-delà des limites incontestables des technologies pour ce qui est d'apprendre une langue, il n'en reste pas moins que peu d'enseignants arrivent à les intégrer dans leur classe pour en tirer tous les avantages qu'ils pourraient en tirer. Les activités qui utilisent les TIC apparaissent souvent comme un peu plaquées, sans lien avec la planification de classe. Cela peut surprendre quand on connaît toutes les informations qui sont aujourd'hui disponibles sur Internet dans différentes langues et les possibilités d'interaction langagière qu'offrent les TIC. Du point de vue de l'évaluation des apprentissages, il est étonnant de voir aussi peu exploitées de telles ressources.

Pourtant, on peut imaginer de nombreuses applications des technologies à des fins de régulation des apprentissages. Certaines visent à créer des conditions propices pour induire des pratiques d'évaluation plus adéquates; d'autres sont directement reliées à la nature même des activités évaluatives.

- *Former les maîtres à la pratique de l'évaluation formative* – C'est une application qui découle du constat que nous venons d'établir mais qui se distingue des utilisations que nous évoquerons par la suite. On peut affirmer que les technologies sont utiles pour la formation des formateurs, notamment pour le développement de compétences professionnelles complexes comme celle de l'évaluation. Ainsi, dans un projet que nous avons mené à l'Université de Montréal dans le cadre des programmes de formation des maîtres (Bélanger, Blais, Giardina & Laurier, 2002) nous avons conçu un environnement (*RégulUM*) qui amène les étudiants à analyser les gestes professionnels associés à différentes situations d'évaluation formative. Cela se réalise à travers l'utilisation de récits informatisés que les étudiants peuvent modifier pour définir le geste professionnel le plus approprié.
- *Gérer les résultats de l'évaluation* – Au lieu de s'en tenir au cahier de bord traditionnel, beaucoup d'enseignants consignent les observations qu'ils font dans une base de données construite à l'aide de logiciels classiques de bureautique ou de logiciels spécialisés. Il ne faut pas sous-estimer l'importance de conserver les évaluations ponctuelles particulièrement celles qui sont recueillies à l'aide de grilles d'appréciation pour juger des performances des élèves (exposé oral, projet d'équipe, composition…). Dans certaines conditions, les données de l'évaluation formative peuvent alors servir au bilan périodique.
- *Amener les élèves à utiliser le portfolio électronique* – L'utilisation du *portfolio* pose des problèmes dans des situations d'évaluation à «enjeux critiques» comme c'est le cas en certification. Par contre, la démarche d'élaboration du *portfolio* s'avère un mécanisme de régulation très efficace particulièrement pour la production écrite. L'élève produit des textes, les modifie et intègre ceux qu'il juge les plus représentatifs dans son dossier d'apprentissage. Cette démarche motive l'élève et l'amène à améliorer sa capacité d'auto-évaluation et à développer des

https://doi.org/10.1017/S0958344004001521 Published online by Cambridge University Press

habiletés métacognitives. Des logiciels facilitent la construction du *portfolio*. Par exemple, la Société GRICS [http://www.grics.qc.ca] qui travaille en partenariat avec les commissions scolaires et le ministère de l'Éducation du Québec a mis au point un logiciel de *portfolio* électronique particulièrement intéressant.

- *Utiliser les mécanismes de mesure qui sont souvent intégrés dans les didacticiels* – Beaucoup de logiciels éducatifs couramment utilisés par les enseignants comportent des procédures qui comptent le nombre de bonnes réponses ou accumulent des points de sorte que l'aspect ludique se combine à la rétroaction sur la performance. Bien qu'il soit souvent difficile de faire des inférences justes sur la compétence sous-jacente à partir de cette information, il y a parfois dans ces scores une information utile.
- *Concevoir des applications pour l'évaluation diagnostique* – Il existe maintenant des logiciels qui permettent au professeur de concevoir des tests qui permettent de vérifier certaines acquisitions. On vise ici généralement la «micro-régulation» c'est-à-dire qu'on essaie d'évaluer des éléments assez pointus de la compétence. Dans cette perspective, deux produits intéressants ont été mis au point en ayant en tête des applications pour l'apprentissage des langues: *Hot Potatoes* qui a été développé par le Humanities Computing and Media Centre de l'Université de Victoria (commercialisés par Half-Baked Software) [http://www.halfbakedsoftware. com] et *NetQuiz* [http://www.ccdmd.qc.ca/netquiz] conçu par une équipe de professeurs du collège Maisonneuve avec le soutien du Centre collégial de développement de matériel pédagogique qui le distribue. Ces logiciels permettent de créer des exercices à trous, des appariements, des dictées…, de poser des questions à choix multiple sur un texte ou un passage audio ou vidéo etc. Ils permettent surtout l'établissement d'un diagnostic personnalisé qui informe l'élève et le professeur sur la maîtrise de certains aspects particuliers de la langue. Comme il est possible de les placer sur la Toile, on peut ainsi gérer un enseignement à distance.
- *Situer l'élève en fonction de différents aspects de la compétence langagière* – Il s'agit ici d'une épreuve standardisée en ligne qui permet de faire un diagnostic plus global pour situer l'élève dans le développement de ses apprentissages selon différentes dimensions. C'est ce que fait le système DIALANG, un test diagnostique qui a été mis au point dans le cadre d'un projet de la communauté (Alderson, 2000). Ce test est disponible dans une douzaine de langues différentes. Il fournit un diagnostic dans cinq dimensions différentes en fonction des niveaux du Conseil de l'Europe: lecture, rédaction, écoute, grammaire et vocabulaire. Comme il sert à la régulation, aucun problème de sécurité ne se pose et on peut donc le télécharger librement. De plus, le système repose sur la coopération de l'apprenant de sorte que l'évaluation fait appel à des jugements d'auto-évaluation de la part de l'élève. Ainsi, la partie d'écoute utilise les jugements de l'élève sur sa propre capacité à faire face à différentes situations de communication; l'évaluation de l'écrit se fait en demandant à l'élève de comparer sa production avec des productions typiques de différents niveaux.

5.2 Classement

La deuxième fonction qui ressort est celle du classement. Même si un système comme

https://doi.org/10.1017/S0958344004001521 Published online by Cambridge University Press

DIALANG peut sembler une solution au classement, il s'agit d'une utilisation problématique dans la mesure où les items ne sont pas protégés, qu'aucun mécanisme n'est prévu pour éviter la sous-évaluation ou la sur-évaluation et que la combinaison des éléments du diagnostic en vue d'inférer la maîtrise générale dans une langue pose problème.

Le classement apparaît comme la fonction par excellence pour appliquer les techniques de testing adaptatif. La construction de ces tests suppose la mise en place d'une banque d'items qui, bien qu'ils puissent constituer une mesure indirecte de la compétence, montrent un lien étroit avec le trait à l'évaluer, en l'occurrence une certaine maîtrise générale de la langue. Les items sont expérimentés auprès de nombreux sujets en vue de leur calibration. Cette opération permet d'assigner un niveau de difficulté à chaque item sur une échelle à laquelle on se référera également pour estimer, en cours d'administration, le niveau de l'apprenant après chaque réponse de chaque apprenant. La stratégie d'administration adaptative convient bien à des besoins de classement car elle évite de soumettre des items difficiles à des élèves débutants et des items faciles à des élèves avancés. Le test est une expérience plus agréable pour l'apprenant non seulement parce que les tâches sont d'un niveau approprié mais aussi parce l'épreuve est beaucoup plus courte qu'une épreuve conventionnelle. L'algorithme de correction permet en effet d'estimer le niveau bien que les items présentés varient d'un apprenant à l'autre. Le test BULATS [http://www.bulats.org] mis au point par les services d'évaluation de l'Université de Cambridge est disponible dans une version adaptative qui se prête particulièrement bien à des décisions de classement même si ses promoteurs prétendent qu'il peut aussi servir à la sélection et à la certification. Comme DIALANG, le BULATS se réfère aux six niveaux du Conseil de l'Europe. Il est disponible en anglais, en français, en allemand et en espagnol.

Un autre exemple de test adaptatif est le *French CAPT*, un test de classement que nous avons développé, il y a quelques années, pour le classement d'étudiants anglophones qui s'inscrivent à des cours de français dans des établissements post-secondaires (Laurier & Perron, 1996). Ce test comporte cinq sections: lecture d'un court paragraphe, choix d'un énoncé dans une situation de communication, phrases lacunaires, compréhension auditive et auto-évaluation de l'expression orale. Les estimations sont combinées afin de déterminer le niveau général de maîtrise sur une échelle à 14 niveaux.

Comme une épreuve de classement peut difficilement être considérée comme une évaluation à « enjeu critique », il est aussi possible de recourir à des innovations qui, bien qu'elles ne permettent pas de satisfaire des exigences métrologiques élevées, présentent des avantages appréciables sur le plan pratique. C'est dans cette optique qu'est apparue une solution comme *Phone Pass* [http://www.ordinate.com], un test qui peut être administré par téléphone. L'épreuve dure une quinzaine de minutes et comprend cinq parties: lecture à haute voix, répétition de phrases, énoncé de contraires, questions à réponses courtes et questions à réponses ouvertes. Le système de correction automatisée fait appel à un dispositif de reconnaissance de la parole et tient compte de la compréhension du vocabulaire, de l'exactitude de la répétition, de la prononciation, du débit en lecture et du débit en répétant. Il semble qu'on arrive à une validité concurrente satisfaisante, ce qui est souvent suffisant pour des décisions de classement.

5.3 Certification

En ce qui a trait à la fonction de certification, l'utilisation des tests adaptatifs n'est

https://doi.org/10.1017/S0958344004001521 Published online by Cambridge University Press

souvent pas appropriée puisqu'il s'agit souvent d'évaluations ciblées vers un niveau de maîtrise attendu; l'épreuve s'adresse donc à des candidats dont le niveau de compétence ne couvre qu'une bande relativement étroite du spectre. Qui plus est, une épreuve de certification, qui est toujours une évaluation à «enjeux critiques», doit se faire dans des conditions où il est possible de garantir que la personne qui y répond est bien celle qui doit répondre. De plus, il est important que les questions ne circulent pas. Néanmoins, on trouve une version informatisée du TOEFL qui recourt à certains procédés hérités du testing adaptatif et est administrée dans des conditions rigoureusement contrôlées (ETS, 1998). Un des avantages que font valoir les auteurs est la possibilité de concevoir des tâches innovatrices qui s'apparentent davantage à des conditions d'utilisation effective de la langue. La technologie est ainsi mise à contribution afin de créer des environnements virtuels qui reproduisent les environnements réels. Dans la même veine, on pourrait imaginer des épreuves de certification qui recréent des situations d'interaction orale comme le font certains logiciels d'apprentissage où l'on reconstitue, par exemple, le décor d'un hôtel, et où le candidat répond à des employés qu apparaissent à l'écran. Cette authenticité est un élément de la validité. Le gain en validité est encore plus net quand on dépasse des situations simulées pour s'approcher de situation de testing direct. C'est le cas quand on observe, à des fins d'évaluation, l'exécution d'une tâche langagière qui doit normalement s'exécuter à l'ordinateur – une tâche de secrétariat, par exemple. À cet égard, on pourrait penser que la plupart des épreuves d'écriture devraient se faire devant un ordinateur puisque c'est maintenant la forme d'écriture la plus courante.

Par ailleurs la fonction de certification est certainement celle où l'apport des technologies sur le plan du traitement des données est le plus significatif. Il s'agit d'applications que les spécialistes de la mesure connaissent depuis longtemps. Ainsi les techniques d'analyse psychométrique issues de la TRI n'auraient jamais pu être élaborées sans la capacité de traitement numérique des ordinateurs. Les technologies actuelles ont rendu ces techniques d'analyse plus accessibles non seulement en rendant disponibles des machines de grande puissance mais en permettant la création d'environnements plus conviviaux même s'il reste beaucoup à faire à cet égard. Ces techniques vont bien au-delà de l'analyse classique des items et des statistiques descriptives des résultats de groupes expérimentaux. Elle permettent la génération de versions parallèles et l'analyse des jugements des correcteurs face à des performances complexes en s'appuyant sur différents critères. Grâce à ces analyses, on peut défendre, tant du point de vue de la fidélité que de la validité, des décisions de certification dont les conséquences sont importantes pour les individus.

6 Conclusion

Au terme de ce tour d'horizon, il apparaît que les technologies multimédias ont un rôle à jouer pour ce qui est de l'évaluation des apprentissages. Cette contribution se situe à différents niveaux et varie selon la fonction visée. Une équipe de chercheurs américains associé au *Educational Testing Service* (Almond, Steinberg & Mislevy, 2002) propose une architecture de système d'évaluation, construite autour de quatre processus (Figure 1).

- Le processus de sélection de l'activité – C'est la composante qui sert à ordonner

https://doi.org/10.1017/S0958344004001521 Published online by Cambridge University Press

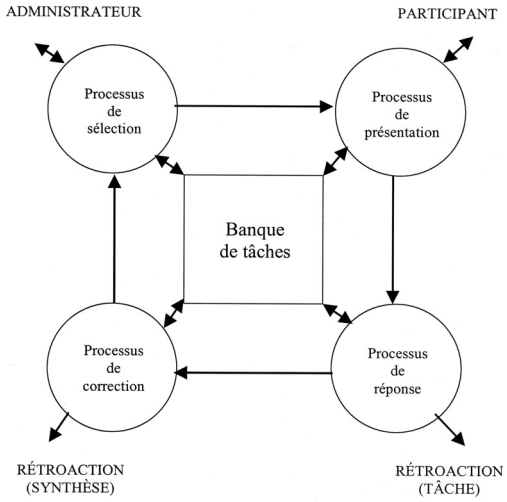


Fig. 1. Les quatre processus du cycle de l'evalaution. Traduit et adapté de Almond (Steinberg & Mislevy, 2000).

et choisir les tâches parmi celles qui sont disponibles dans la banque.

- *Le processus de présentation* – C'est cette composante qui sert à déterminer la forme que prend la tâche: type d'affichage, source audio ou vidéo, saisie de la réponse etc.
- *Le processus de traitement de la réponse* – Cette composante traite chaque réponse spécifique pour identifier les éléments qui sont pertinents dans le jugement évaluatif.
- *Le processus de synthèse de la correction* – Traite l'ensemble des observations pour constituer un jugement sur le développement de la compétence ou l'acquisition de ses éléments.

Selon le type de décision à prendre, ces processus se dérouleront de façon différente. Ainsi, pour la fonction de régulation, le processus de sélection se modèle sur l'approche pédagogique, le processus de présentation peut utiliser la Toile sans problème, le processus de traitement de la réponse peut prendre en compte l'auto-évaluation et le processus de synthèse peut produire une fiche diagnostique. Par contre, pour la fonction de certification, la sélection doit assurer la comparabilité des tâches entre les candidats, la présentation doit être mieux contrôlée, le traitement de la réponse se fera en fonction d'une réponse attendue et la synthèse devra fournir un score interprétable. Ce qui ressort de l'approche proposée c'est l'importance de concevoir l'évaluation en tenant compte des buts poursuivis. Ainsi, ce qui est tout à fait souhaitable pour la régulation ne l'est peut être pas pour le classement ou la certification.

Références

Alderson, J. C. (2000) Technology in testing: The present and the future. *System*, **28** (4): 593–603.
Almond, R. G., Steinberg, L. S. and Mislevy, R. J. (2002) Enhancing the design and delivery of assessment systems: A four process architecture. *Journal of Technology, Learning &*

https://doi.org/10.1017/S0958344004001521 Published online by Cambridge University Press

Assessment, **1** (5): 2-63 [http://www.jtla.org].

Bélanger, K., Blais, J.-G. and Laurier, M. (2002) L'évaluation formative et les récits d'expérience professionnelle. *Communication à la session d'études de l'ADMÉE*, Longueuil, QC.

Brown, J. D. (1997) Computers in language testing: Present research and some future directions. *Language Learning & Technology*, **1** (1): 4449 [http://llt.msu.edu/vol1num1].

Chalhoub-Deville, M. (2001) Language testing and technology: Past and future. *Language Learning & Technology*, **5** (2): 4449 [http://llt.msu.edu/vol5num2].

Dunkel, P. A. (1999) Considerations in developing or using second/foreign language proficiency computer-adaptive tests. *Language Learning & Technology*, **2** (2): 7793 [http://llt.msu.edu/vol2num2].

Kenyon, D. et Malabonga, V. (2001) Comparing examinee attitudes toward computer-assisted and other proficiency assessments. *Language Learning & Technology*, **5** (2): 60–83 [http://llt.msu.edu/vol5num2].

Laurier, M. (1999) Testing adaptatif et évaluation des processus cognitifs. Dans: Depover, C. & Noël, B. (dir.), *L'évaluation des compétences et des processus cognitifs: modèles, pratiques et contextes*. Paris et Bruxelles: De Boeck – Université, 255266.

Laurier, M. (2002) Can computerised testing be authentic? *ReCALL*, **12** (1): 93–104.

Laurier, M. et Dassa, C. (2003) Le diagnostic pédagogique comme moyen d'informer. Dans: Laurier, M. (dir.), *Évaluation et communication*. Montréal: Éditions Québécor, pp. 103–130.

Laurier, M. et Perron, M. (1996) Un test adaptatif pour le classement des élèves dans un cours de langue. Dans: L. Sauvé, M.-T. Bourbonnais & M. Laurier (dir.), *La technologie éducative en réseau*. Québec: CIPTE / Télé-Université, pp. 279–286.

Laurier M. and Viens, J. (2000) La familiarisation des futurs maîtres aux technologies de l'information et de la communication. Dans: Duquette, L. & Laurier, M. (dir.), *Apprendre une langue dans un environnement multimédia*. Montréal: Éditions Logiques, pp. 281–301.

Lam, Y. (2000) Technophilia vs Technophobia: A preliminary look at why second-language teachers do or do not use technology in their classrooms. *Canadian Modern Language Review*, **56** (3): 389–420.

Moore, Z., Morales, B. and Carel, S. (1998). Technology and Teaching Culture: Results of a State Survey of Foreign Language Teachers. *CALICO Journal*, **15** (1-3): 109–128.

Renié, D. (2000) Apport d'une trace informatique dans l'analyse du processus d'apprentissage d'une langue seconde ou étrangère. In: L. Duquette et M. Laurier (dir.), *Apprendre une langue dans un environnement multimédia*. Montréal: Éditions Logiques, pp. 281–301.

Salaberry, R. (1996) A theoretical foundation for the development of pedagogical tasks in computer mediated communication. *CALICO Journal*, **14** (1): 5–34.

Schuman, J. H. (1997) *The Neurobiology of Affect in Language*. Malden, MA: Blackwell Publishers.

Scallon, G. (2000) *L'évaluation formative*. Montréal: Éditions du Renouveau pédagogique.

Wainer, H. and Mislevy, R.J. (2000) Item response theory, Item calibration and Proficiency Estimation. In: Wainer, H. (ed.), *Computerized Adaptive Testing: A Primer* (2nd edition), Hillsdale, NJ: Lawrence Earlbaum, pp. 65–101.

Wiggins, G. (1993) *Assessing Student Performance*. San Francisco, CA : Jossey-Bass.

https://doi.org/10.1017/S0958344004001521 Published online by Cambridge University Press