

# UN CANEVAS D'ITEM POUR ÉVALUER LA COMPÉTENCE D'INVESTIGATION SCIENTIFIQUE EN LABORATOIRE

ERIC DIONNE *University of Ottawa*

**RÉSUMÉ.** L'évaluation des compétences scientifiques chez les élèves du secondaire (12-16 ans) présente des défis importants, en particulier en ce qui concerne l'instrumentation. Cet article présente des propositions à caractère méthodologique permettant de développer et d'administrer des situations d'évaluation (SE) de la compétence d'investigation scientifique dans un contexte à enjeux critiques comme c'est le cas, par exemple, en évaluation sommative ou certificative. Ces propositions s'appuient sur les postulats de l'approche par compétence, l'approche préconisée dans plusieurs systèmes éducatifs nord-américains et européens.

## AN ITEM BLUEPRINT FOR ASSESSING SCIENTIFIC INQUIRY COMPETENCIES IN A LABORATORY SETTING

**ABSTRACT.** The assessment of secondary students' (12-16 years old) scientific competencies involves significant challenges, in particular instrumentation competencies. This article presents a methodological framework to develop and administer tools that can assess students' scientific investigation competencies in a high-stakes context, such as summative or certificative evaluations. The framework draws on the competency-based approach, which is the recommended approach in many North American and European educational systems.

Depuis une dizaine d'années, les systèmes scolaires sont les témoins de profonds changements, autant en Amérique du Nord qu'en Europe. Plus particulièrement, les décideurs apportent des changements importants à l'éducation fondamentale de masse autant pour les ordres du primaire que du secondaire. En effet, on assiste à la mise en œuvre de multiples décisions administratives et pédagogiques qui visent essentiellement à réguler les systèmes afin, semble-t-il, de les rendre plus efficaces. Le besoin de préparer les jeunes aux nombreux défis du 21<sup>e</sup> siècle n'est certes pas étranger à cette situation

comme le souligne par exemple le rapport à l'UNESCO de la Commission internationale sur l'éducation pour le vingt-et-unième siècle (Delors, 1999). À titre d'exemple, on peut penser à l'explosion des connaissances ayant cours depuis le début du siècle dernier suggérant ainsi de passer d'une pédagogie axée sur l'acquisition de connaissances à une pédagogie axée sur le traitement de ces dernières. Les programmes formulés par compétences en sont d'ailleurs de bons exemples puisqu'ils visent, pour la plupart, à mieux outiller les élèves à faire face aux multiples défis qu'ils auront à surmonter au cours du vingtième et unième siècle. À ce sujet, plusieurs juridictions ou états ont décidé d'aller de l'avant avec cette approche en proposant des programmes d'études qui visent le développement de compétences, et ce, autant en Amérique du Nord (par ex. Québec, Ontario, etc.) qu'en Europe (par ex. Belgique, Suisse, etc.). L'analyse de ces programmes révèle d'ailleurs de grandes disparités dans la définition même du concept de compétence et dans sa façon de l'opérationnaliser. Nous verrons plus loin que cet aspect ajoute à la complexité de l'évaluation des compétences. Plusieurs chercheurs<sup>1</sup> (Dionne, 2005 ; Rey, Carette, Defrance et Kahn, 2003 ; Scallon, 2004) soulèvent des questions concernant les démarches et méthodes à employer afin de développer une instrumentation pertinente pour juger de ces compétences dans des contextes scolaires qui sont souvent jalonnés d'obstacles au regard, entre autres, de la mise en œuvre de ces méthodes et particulièrement dans le contexte de l'évaluation sommative ou certificative.

Ce changement de paradigme, soit le passage d'une pédagogie axée sur la transmission de connaissance à une pédagogie axée sur le traitement de l'information, implique de nombreux changements et réajustements et en particulier au niveau de l'évaluation des apprentissages. Des instruments de collecte de données plus adaptés doivent être développés ou amendés puisque les outils traditionnellement employés par les enseignants (ex. épreuves à correction objective) possèdent des limites, d'autant plus quand les apprentissages mesurés sont variés et complexes (Laurier, Tousignant et Morrissette, 2005). Or, s'il existe de nombreux écrits sur les règles de construction des items (question à choix multiple, vrai ou faux, question à réponse courte, etc.) (voir par exemple Airasian, Engemann et Gallagher 2007 ; Conderman et Koroghlian, 2002 ; Durand et Chouinard, 2006 ; Haladyna, Downing et Rodriguez, 2002 ; McCowan, 1999 ; McMillan, 2001 ; Scallon, 2004), on en retrouve beaucoup moins concernant les règles de construction de situation d'évaluation qui visent à mesurer des performances représentatives d'un construit (ex. compétence) ou des apprentissages complexes. Ceci s'observe d'autant plus dans le cas où l'évaluation se déroule dans un contexte scolaire à enjeux critiques comme c'est le cas pour des épreuves administrées dans un contexte sommatif ou certificatif. Selon bien des auteurs (Baker, 1997 ; Hogan et Murphy, 2007 ; Scallon, 2004 ; Solano-Flores, Jovanovic et Shavelson, 1994 ; Solano-Flores et Shavelson, 1997 ; Solano-Flores, Shavelson et Schneider, 2001), il existe un besoin réel de mieux documenter le processus permettant de mettre en place

des situations d'évaluation qui mesurent ce genre de construit. Cet article a pour objectif de faire une proposition à caractère méthodologique visant à évaluer une compétence. Le contexte associé à cette proposition est l'éducation scientifique et nous nous sommes intéressés de façon plus particulière à l'investigation scientifique. Cette recherche a été menée au Québec et s'appuie sur les prescriptions ministérielles qui sont en vigueur pour l'ordre du secondaire dans cette juridiction depuis le milieu des années 2000. Nous postulons que les propositions que nous exposons dans cet article sont susceptibles d'intéresser autant les chercheurs que les praticiens qui s'intéressent aux méthodes permettant de mesurer des construits complexes. Bien que nous n'ayons aucune intention de généralisation à d'autres juridictions ou états, nous espérons que les réflexions condensées dans cet article puissent inspirer les intervenants qui oeuvrent dans d'autres juridictions ou états.

Dans cet article, nous tenterons de répondre à la question suivante : comment construire des situations d'évaluation qui permettent d'évaluer une compétence scientifique dans un contexte ordinaire d'enseignement et donc soumis à de nombreuses contraintes ?

Dans un premier temps, nous présenterons la problématique et les aspects contextuels pris en compte dans notre réflexion. Dans un second temps, nous discuterons des aspects théoriques abordés dans ce texte à savoir l'investigation scientifique, le concept de compétence et les contraintes inhérentes à la salle de classe. Enfin, nous présenterons le modèle que nous proposons en le définissant et en indiquant ses principales caractéristiques tout en montrant de quelle façon ce modèle pourrait tenir compte des différentes contraintes mentionnées dans cet article.

## PROBLÉMATIQUE ET ÉLÉMENTS DE CONTEXTE

### *L'évaluation sommative au secondaire : un contexte organisationnel contraignant*

Depuis quelques années, le ministère de l'Éducation, du Loisir et du Sport (MELS) demande aux enseignants de science et de technologie de développer et d'évaluer chez leurs élèves des compétences. Cependant, force est de constater qu'il existe peu de modèles qui présentent des exemples concrets sur lesquels les enseignants pourraient s'appuyer dans le cadre de la planification de leurs séquences d'enseignement ou de leurs séquences d'évaluation. Qui plus est, les propositions existantes ne tiennent pas toujours compte des nombreuses contraintes qui viennent complexifier la tâche des enseignants lorsqu'ils ont à réaliser ces opérations de planification. Dans le cadre de nos travaux, notre point de départ a été la prise en compte des multiples contraintes qui prévalent généralement dans les écoles dites ordinaires. En ce qui nous concerne, une école dite ordinaire possède trois caractéristiques : (1) il n'y a aucune opération de sélection des élèves, (2) le ratio étudiants / classe est semblable à ce que

l'on retrouve dans la moyenne des écoles de la province<sup>2</sup> et (3) le financement de l'école se situe dans les paramètres normaux de la moyenne des écoles comparables ailleurs dans la province. Les contraintes que nous allons évoquer peuvent être classées en deux grandes catégories : les contraintes pédagogiques et les contraintes organisationnelles. Le tableau 1 qui suit présente une liste non exhaustive des contraintes dont nous avons tenu compte pour l'élaboration des instruments. La prise en compte de ces dernières est une condition souvent jugée essentielle pour favoriser les changements de pratiques pédagogiques chez les enseignants de science (Solano-Flores, Shavelson et Bachman, 1999 ; Solano-Flores et Shavelson, 1997).

**TABLEAU 1. Liste de contraintes pédagogiques et organisationnelles**

Contraintes pédagogiques	Contraintes organisationnelles
Durée accordée à la situation d'évaluation (SE)	Nombre d'élèves par classe
Gestion des élèves en difficulté ou en grande difficulté	Présence d'un technicien de laboratoire
Menaces à la validité des inférences (ex. risque de tricherie)	Disponibilité du matériel de laboratoire
Menaces à la fidélité des scores	Disponibilité de la classe-laboratoire

Parmi les contraintes pédagogiques, on note : (1) la durée accordée à la SE, (2) la gestion des élèves en difficulté ou en grande difficulté, (3) les menaces à la validité et (4) les menaces à la fidélité. Examinons brièvement chacune de ces dernières. En contexte d'enseignement, la durée consacrée à l'évaluation est limitée. Les enseignants perçoivent parfois négativement la durée consacrée à l'évaluation. Pour plusieurs, ces périodes de temps pourraient être plutôt consacrées aux apprentissages, ce qui représente pour eux une forme de perte de durée d'enseignement. Aussi, nous savons que, plus les élèves sont jeunes, plus leur niveau de motivation sera difficile à soutenir sur des durées de temps importantes. Dans un contexte d'évaluation, la durée prend une très grande importance. À titre d'exemple, les évaluations qui ont lieu à la fin d'une année scolaire ou à la fin d'un cycle se réalisent généralement au mois de juin juste avant les vacances estivales. À ce moment, les enseignants doivent souvent accélérer le rythme d'enseignement afin de s'assurer que les élèves auront appris l'ensemble des éléments prévus au programme d'études. Traditionnellement, dans une approche par objectifs et dans une organisation scolaire annuelle, une période d'évaluation est prévue au mois de juin. On réserve souvent une dizaine de jours qui sont consacrés aux évaluations sommatives qui prennent

le plus souvent la forme d'examens visant à mesurer l'acquisition des apprentissages à l'égard des principaux objectifs du programme d'études.

Le développement de compétences impose des situations d'évaluation qui mettent l'accent sur la performance des élèves à résoudre une tâche ou une situation donnée. Cette performance doit respecter les principes de l'évaluation des apprentissages. Examinons trois menaces qui planent au-dessus de certains de ces principes. Premièrement, ce genre de situations d'évaluation prend généralement un temps considérable à réaliser. Tout en présentant des situations qui permettent de juger du développement d'une compétence, il faut également faire en sorte d'inclure volontairement des contraintes qui vont la limiter dans la durée. Deuxièmement, le contexte d'évaluation sommatif ou certificatif fait en sorte qu'il faut s'assurer que la production d'un élève lui appartient entièrement et n'est pas associée à d'autres élèves. Il s'agit de la même préoccupation que dans le cas d'un contexte de *testage* où il faut s'assurer d'éviter la tricherie. Autrement dit, pour des raisons pédagogiques, mais également juridiques, il faut être en mesure de prouver que le jugement de l'enseignant porte uniquement sur la performance de l'élève. Troisièmement, dans le cas de l'évaluation d'un construit complexe, les menaces à la fidélité prennent différentes formes. Comme nous l'avons déjà mentionné, la durée consacrée à l'évaluation est limitée. D'une part, il n'est pas possible d'administrer de façon réaliste plusieurs SE semblables qui permettraient aux enseignants d'exercer leur jugement en s'appuyant sur des instruments réputés stables. D'autre part, il est aussi très difficile voir impossible en contexte scolaire d'exercer un jugement qui s'appuie sur un processus de notation avec correcteurs multiples. Cela fait en sorte qu'il faut être encore plus prudent pour s'assurer de la validité et de la fidélité des interprétations en lien avec chacune de ces SE.

Les contraintes organisationnelles, quant à elles, sont surtout associées aux ressources nécessaires au travail pédagogique en classe-laboratoire. À titre d'exemple, on peut mentionner la disponibilité : (1) du technicien de laboratoire, (2) du matériel de laboratoire et (3) de la classe-laboratoire. En effet, les écoles sont habituellement équipées adéquatement pour offrir des cours de science et technologie. Nous nous sommes quand même assuré de ne pas recourir à des instruments ou des matériaux qui seraient difficilement accessibles pour certaines écoles. Aussi, les écoles ne sont généralement pas très bien pourvues en locaux permettant de consigner les travaux des élèves. Ce détail est de toute première importance puisque certaines activités évaluatives (ex. création d'une maquette) pourtant fort intéressantes sont difficilement réalisables dans la pratique, faute d'espace. En ce sens, les périodes d'expérimentation devraient être assez courtes afin d'être complétées à l'intérieur d'une plage horaire restreinte (ex. une période de 75 minutes). Enfin, l'organisation scolaire (ex. horaire, disponibilité des classes-laboratoires, etc.) amène fréquemment son lot de défis. Dans un monde idéal, il serait souhaitable que les élèves réalisent la situation

d'évaluation en continu (ex. l'équivalent de 4 périodes de 75 minutes) ce qui limiterait, entre autres, les risques de contamination. Cependant, les contraintes organisationnelles font souvent en sorte de rendre impossible un tel scénario. Un meilleur arrimage entre les aspects pédagogiques et organisationnels serait évidemment souhaitable. Nous avons choisi de considérer cette contrainte en espérant qu'elle se résorbe, éventuellement, par un meilleur arrimage. À ces contraintes viennent s'ajouter les effectifs étudiants qui représentent un enjeu réel pour les praticiens. En effet, le recours, par exemple, aux entrevues individuelles ne serait pas une solution acceptable pour des enseignants qui doivent porter leur attention sur des groupes souvent populeux.

Comme nous venons de le décrire, le contexte pédagogique associé à l'enseignement de la science et de la technologie impose de multiples contraintes. Nous pensons que des propositions méthodologiques qui ne prendraient pas en compte ces dernières ne peuvent qu'entraîner des effets démotivants et une perte d'intérêt chez les enseignants qui ne seraient alors pas tentés de mettre en œuvre ces propositions les jugeant inapplicables en situation réelle.

## LE CONSTRUIT MESURÉ : L'INVESTIGATION SCIENTIFIQUE

### *Quelques définitions*

Le *National Research Council* (1996) définit l'investigation scientifique comme :

une activité qui implique de réaliser des observations, poser des questions, prendre connaissance de livres et d'autres sources de renseignements afin de déterminer ce qui est déjà connu, confronter ce qui est déjà connu aux données expérimentales, utiliser des instruments afin de collecter des données, analyser et interpréter les données, proposer des solutions des explications et des prédictions et enfin communiquer les résultats. (p. 23; trad. auteur)

Cette définition met en lumière une démarche de résolution de problème dans un contexte scientifique en lien avec le courant du *hands-on science*. De leur côté, Lumpe et Oliver (1991) identifient trois dimensions importantes associées à l'investigation scientifique soit la dimension : (1) découverte, (2) structurelle et (3) expérimentale. La première dimension fait référence au fait que l'investigation doit représenter pour l'élève une réelle découverte. En ce sens, l'investigation va bien au-delà de la simple confirmation d'une théorie pouvant avoir été par exemple présentée en classe par l'enseignant. La dimension structurelle, quant à elle, touche à l'encadrement de l'élève dans son processus d'investigation. Dans une approche basée sur le développement « pur » de l'investigation scientifique, l'élève devrait avoir suffisamment de liberté et être en mesure de prendre des décisions significatives à différentes étapes de son processus d'investigation. Enfin, la dimension expérimentale est en lien avec le processus de preuve mis en place afin de donner de la crédibilité aux processus réalisés par l'élève. Les recommandations de ces auteurs ainsi que celles du *National Research Council* nous ont largement inspiré pour l'élaboration de

la structure des outils que nous présentons dans cet article. Il est important de bien définir ce que nous entendons par « investigation scientifique », car, bien que ce soit une activité pratiquement incontournable dans le paysage de l'enseignement des sciences, il n'en demeure pas moins qu'on remarque des acceptions très différentes. Pour certains, l'investigation scientifique se réduit à la maîtrise d'habiletés de manipulation dans un contexte de laboratoire (ex. manipuler adéquatement et de façon sécuritaire un bruleur au gaz) alors que, pour d'autres, il s'agit de résoudre de manière originale un problème inédit par le biais d'une démarche empirique. Confronter ces deux exemples aux recommandations que nous venons de présenter révèle le fossé qui sépare les conceptions souvent entretenues face à l'investigation scientifique.

L'investigation scientifique qui consiste à développer et à appliquer des stratégies de résolution de problème en laboratoire est certainement le type d'activité qui possède le plus de caractéristiques communes avec celles de la définition proposée par le *National Research Council*. Il s'agit essentiellement pour l'élève de mettre en œuvre une démarche de résolution de problème dans un contexte de classe laboratoire (ex. s'appuyer sur des données empiriques, établir des évidences basées sur des éléments observables, etc.). Une autre dimension que nous avons retenue consiste à offrir la possibilité aux élèves de faire des choix et de les justifier. Selon Duschl (2003) une façon de promouvoir le développement d'une compétence d'investigation scientifique consiste à sortir de la logique des activités d'investigation clé en main (*kit-based science investigation*) afin que les élèves se voient offrir des situations qui leur permettront de faire des choix et de les justifier ce qui est plus conforme à l'activité scientifique.

### ***La compétence : une tentative de définition***

D'aucuns reconnaissent que le concept de « compétence » est polysémique. Plusieurs propositions ont été formulées en éducation au cours des dernières années (Perrenoud, 1997 ; Tardif, 2003) afin de mieux déterminer le périmètre de ce concept. En ce qui nous concerne, nous avons retenu la proposition de Scallon (2004). Ce dernier propose que « la compétence est la possibilité, pour un individu, de mobiliser de manière intériorisée un ensemble intégré de ressources en vue de résoudre une famille de situations-problèmes » (Scallon, 2004, p. 105). À bien des égards, cette définition rejoint celle de Tardif (2006) qui propose, quant à lui, de définir la compétence comme « un savoir-agir complexe prenant appui sur la mobilisation et la combinaison efficace d'une variété de ressources internes et externes à l'intérieur d'une famille de situations » (p. 22). Parmi les points communs, on remarque l'aspect de mobilisation des ressources par le sujet et également le concept de famille de situation qui sera défini plus en profondeur dans la section qui suit.

La définition que propose Scallon (2004) met l'accent sur des éléments qui se retrouvent normalement dans une démarche d'investigation scientifique en laboratoire. L'idée de résoudre des problèmes apparentés (résoudre une famille

de situations-problèmes) en est à notre avis un bon exemple. Dans un contexte d'évaluation, le défi réside donc à définir des familles de situations-problèmes qui ont, de façon intrinsèque, des caractéristiques et des propriétés semblables. Par analogie, on peut comparer les situations d'une même famille à des tests parallèles dans la théorie classique des tests. L'autre idée importante qui se retrouve dans la définition de Scallon est que l'élève doit mobiliser des ressources dont certaines sont intériorisées. Il ne s'agit donc pas pour l'élève de suivre à la lettre les consignes de l'enseignant. L'élève doit donc apprendre et maîtriser des apprentissages, il doit également être en mesure de les transférer dans des situations significatives et authentiques. Les situations d'évaluation doivent donc permettre de mesurer et d'observer si les élèves sont en mesure de produire de telles actions. Les situations d'évaluation que nous avons construites, et en particulier pour le niveau 1 — que nous définirons subséquemment — prennent donc en considération les aspects que nous venons de présenter à l'égard du concept de compétence.

### *La notion de « famille de situations »*

Le concept de « famille de situations » est au coeur du modèle que nous proposons. Roegiers (2000) définit une famille de situations comme « un ensemble de situations proches l'une de l'autre » (p. 130). Le concept de « famille de situations » est important, car il permet de circonscrire les types de situations qui seront présentées aux élèves. Afin d'assurer la validité des situations d'évaluation, il importe de s'assurer que les élèves ont eu: (1) l'opportunité de développer leurs compétences et (2) l'occasion d'être confrontés à des situations semblables en contexte d'apprentissage. Roegiers propose deux paramètres afin de mieux définir les familles de situations : ce qu'on offre aux élèves et le type de tâche présenté. Ces deux paramètres ne nous apparaissent pas mutuellement exclusifs et ne nous semblent pas efficacement opérationnels afin de bien distinguer la nature d'une famille de situation. Quoi qu'il en soit, le premier paramètre fait surtout référence aux composantes et aux caractéristiques de la situation offerte aux élèves. Les contraintes (ex. le nombre de mots ou de pages accordés afin de répondre à la question posée) soigneusement choisies et enchâssées dans l'énoncé de la situation, en représentent de bons exemples. Le second paramètre, quant à lui, fait surtout référence aux conditions de passation de la situation. Il peut s'agir, par exemple, du matériel mis à la disposition des élèves, du degré de complexité de la tâche, etc.

### *Liens entre l'investigation scientifique, la compétence et les familles de situations*

Nos travaux ont été menés au Québec et, par conséquent, nous nous sommes appuyés sur les prescriptions ministérielles ayant force de loi dans cette juridiction. La figure 1 présente une modélisation de la compétence d'investigation scientifique que les élèves du Québec doivent développer tout au long de leur scolarité. La démarche qui nous intéresse est celle présentée dans la voie de



gauche et qui est composée des étapes suivantes : cerner un problème, choisir un scénario, concrétiser sa démarche, effectuer l'expérience, analyser ses résultats et faire un retour. Les nombreuses boucles de rétroaction montrent la non-linéarité de ce processus et mettent en évidence les allers-retours que les élèves peuvent réaliser. Sans reprendre tout ce qui a été dit dans cette section du texte, on peut voir que cette modélisation est en lien avec la nature de l'investigation scientifique telle que nous l'avons définie précédemment. Il s'agit pour les élèves de mettre en œuvre une démarche visant à trouver une ou des réponses à un problème authentique. Ils doivent donc mettre à profit une démarche cognitive complexe qui fera en sorte qu'ils devront réaliser de façon synergique chacune des étapes de la démarche illustrée à la figure 1. Cette dernière montre une démarche d'investigation scientifique, mais elle peut également illustrer d'autres démarches scientifiques telles que la démarche d'observation. En ce sens, il s'agit d'une démarche générique. Son caractère plus distinct se révèle au moyen des familles de situations. En effet, dans ce contexte, une des familles de situations en lien avec cette modélisation est la démarche d'investigation scientifique. Comme nous le verrons dans une section ultérieure, il est ainsi possible de construire des situations d'évaluation qui se ressemblent et qui prennent appui sur cette modélisation de la compétence d'investigation scientifique.

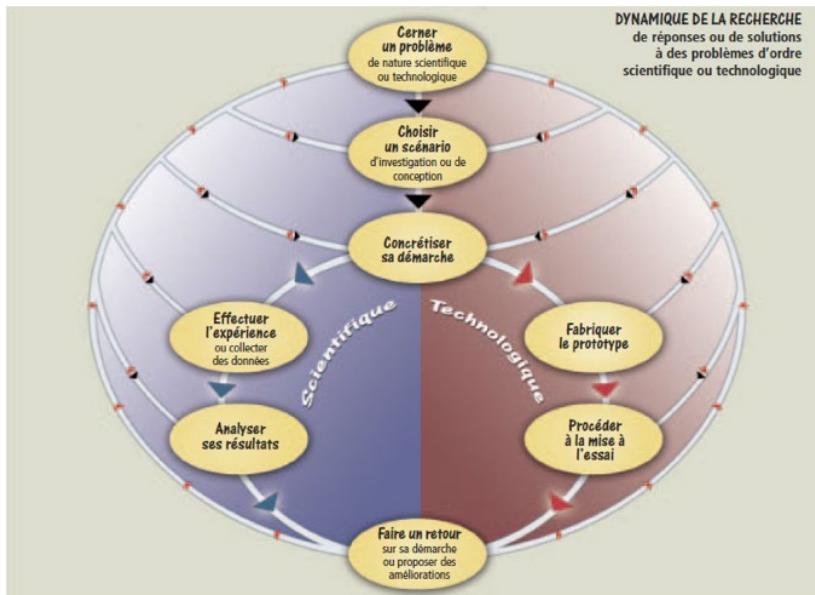


FIGURE 1. Modélisation de la compétence d'investigation scientifique telle que proposée par le ministère de l'Éducation du Loisir et du Sport (MELS, 2003).

## L'ÉVALUATION DANS UN CONTEXTE À ENJEUX CRITIQUES

Dans un monde idéal, les enseignants auraient à leur disposition une panoplie d'outils d'évaluation afin de mesurer toutes sortes d'apprentissages des plus simples aux plus complexes. Qui plus est, ils connaîtraient parfaitement les propriétés et les limites de ces différents instruments. L'enseignant aurait tout son temps afin de recueillir toutes les données dont il a besoin pour appuyer avec conviction son jugement. Il pourrait, par exemple, croiser les résultats de quelques situations d'évaluation avec des entrevues individuelles afin de bien saisir, et avec raffinement, le développement de la compétence qu'il souhaite évaluer chez l'élève. Cette situation idyllique est malheureusement loin de la réalité des enseignants de science et ce, peu importe le système scolaire dans lequel ils œuvrent. Pour ces derniers, simplement considérer de nouvelles façons d'évaluer passe obligatoirement par la prise en compte des nombreuses contraintes auxquelles ils doivent faire face jour après jour. De fait, nous sommes convaincu que la viabilité de nouveaux dispositifs d'évaluation ne peut se réaliser sans la considération de ces contraintes multiples et malheureusement incontournables. Dans la section qui suit, nous présenterons quelques-unes des contraintes que nous avons choisi de considérer pour la mise en place de situations d'évaluation de l'investigation scientifique en laboratoire.

Une activité d'évaluation qui se déroule dans un contexte à enjeux critiques signifie que les conséquences associées aux résultats obtenus par l'élève peuvent être lourdes pour ce dernier. Les évaluations se déroulant dans un contexte sommatif ou certificatif sont souvent associées à ce contexte. En contrepartie, on associe souvent les fonctions formatives ou diagnostiques à un contexte d'évaluation à faibles enjeux. Cette façon de catégoriser est discutable. Le but de cet article n'est pas d'aborder cette question. Dans un contexte pratique, on peut s'imaginer que la méthodologie présentée dans cet article peut s'appliquer à des situations d'évaluation utilisées, par exemple, dans un contexte sommatif pour la confection du bulletin ou pour un bilan formel. Nous avons délibérément choisi de situer nos travaux dans un contexte à enjeux critiques qui est certainement le contexte qui impose le plus de contraintes. En effet, il faut alors prendre en considération les possibilités de contamination (ex. tricherie), la gestion de la communication entre l'enseignant et les élèves (ex. offrir les mêmes consignes à tous) ou la comparabilité des situations (ex. situations ayant une difficulté comparable). Dans un contexte formatif où le but est d'aider chacun des élèves à améliorer ses apprentissages, il devient secondaire de s'assurer de la prise en compte de ces considérations.

Compte tenu du contexte que nous venons de définir, plusieurs principes ont dicté nos choix tout au long du processus d'élaboration de cette méthodologie. D'abord, nous devons nous assurer de la cohérence de l'alignement programme-apprentissage-évaluation. En effet, une des conditions permettant d'assurer la validité du processus d'évaluation est de faire en sorte que les

prescriptions tirées des programmes, les activités pédagogiques et didactiques et finalement les dispositifs d'évaluation soient les plus cohérents possible. Ensuite, nous devons nous assurer de la comparabilité des outils et des informations recueillies. Puisque nous souhaitons proposer aux élèves des SE différentes afin d'éviter le plus possible la contamination, nous devons en même temps nous assurer que les SE développées étaient relativement équivalentes autant en terme de difficulté ou encore en terme de durée de réalisation. Dans un contexte à enjeux critiques, la comparabilité des situations d'évaluation est, en effet, un facteur de validité incontournable.

## **UNE PROPOSITION POUR DÉVELOPPER UN INSTRUMENT MESURANT UNE COMPÉTENCE**

Nous rappelons au lecteur que le but de cet article est de faire état d'une méthodologie permettant de concevoir et d'administrer des situations d'évaluation permettant de juger du développement d'une compétence d'investigation scientifique. Le lecteur intéressé par les résultats de l'expérimentation, et en particulier les scores obtenus par les élèves et les propriétés métriques des SE de cette méthodologie peut se référer, entre autres, aux travaux de Dionne (2010). Dans la présente section, nous présenterons l'architecture des situations d'évaluation, nous ferons également une description de chacune d'elles. Par la suite, nous mettrons l'emphase sur les caractéristiques de ces situations et en particulier sur les niveaux que nous avons développés afin de nous assurer de recueillir des informations pour tous les élèves, peu importe leur niveau de maîtrise de la compétence. Enfin, nous indiquerons les modalités associées à la passation et à la correction de ces SE.

### *Architecture des situations d'évaluation*

Une situation d'évaluation de l'investigation scientifique devrait, selon Solano-Flores et Shavelson (1997) contenir trois composantes : « (1) un problème nécessitant un traitement en laboratoire, (2) une feuille réponse où l'élève peut exprimer sa démarche et sa solution et (3) un système d'assignation des scores » (p. 17). À ces composantes, nous jugeons qu'il est pertinent d'ajouter également les conditions qui permettent la mise en œuvre de la situation d'évaluation, par exemple un guide de passation destiné à l'enseignant ou au technicien de laboratoire. Cette condition nous apparaît importante afin d'identifier les obstacles à la mise en place de telles situations ainsi que les solutions alternatives permettant de contrer ces obstacles. Cette recommandation est d'autant plus importante quand le mode de passation déroge des formats habituels. Dans le cadre de cet article, nous désignons par dispositif d'évaluation l'ensemble de ces composantes.

Lors de l'élaboration de situations d'évaluation, les auteurs sont souvent confrontés rapidement à un problème important : comment s'assurer que différentes situations d'évaluation seront comparables les unes avec les autres ?

Autrement dit, comment s'assurer de l'équivalence des situations d'évaluation autant au niveau de leurs propriétés métriques (ex. difficulté, discrimination, etc.) qu'au niveau du construit que chacune de ces SE prétend mesurer ? Ce problème est étudié depuis les années 70 principalement dans le cas du développement d'items (Bormuth, 1970). Ce dernier a, en effet, proposé un algorithme complexe visant à produire des items relativement équivalents les uns aux autres. Il s'agissait alors d'un pas dans la bonne direction, mais la complexité de l'algorithme a tôt fait de le rendre plus ou moins applicable auprès des concepteurs d'item. Ces travaux ont inspiré dans les décennies qui ont suivi d'autres chercheurs, ce qui a amené le concept de *item shell* que nous traduisons par canevas d'item. Dans le cadre de nos travaux, nous avons retenu la proposition de Haladyna et Rodriguez (2013), que nous avons adaptée au contexte de l'évaluation de situations d'évaluation plutôt qu'au contexte d'items plus classiques (ex. items à choix multiples), le but étant de nous baser sur une méthode permettant de produire des situations d'évaluation différentes, mais néanmoins équivalentes. La méthode proposée par Haladyna, et que nous avons adaptée, se présente en six étapes :

1. Écrire un énoncé de problème ;
2. Identifier les mot-clés du problème ;
3. Énoncer différentes variations associées aux différents mot-clés ;
4. Indiquer une réponse attendue et acceptable ;
5. Écrire un nouvel énoncé de problème en prenant appui sur les variations associées aux différents mot-clés ;
6. Indiquer une réponse attendue et acceptable pour ce nouveau problème.

Afin d'illustrer la méthode que nous avons retenue, examinons le problème associé à la première situation d'évaluation. Dans les étapes énoncées précédemment, dans le cadre de cet article nous mettrons l'emphase sur le concept de canevas d'item (*item shell*) en discutant principalement des étapes correspondantes à savoir 1, 2, 3 et 5. La situation qui nous servira d'exemple est, quant à elle, inspirée de la proposition de Gott et Welford (1987). Il s'agit de choisir une variable qui peut influencer le rebond d'une balle de tennis et d'imaginer une expérimentation permettant de vérifier l'hypothèse émise au regard de la variable investiguée. L'intérêt d'une telle situation réside dans le fait que les élèves ne doivent pas obligatoirement maîtriser des concepts disciplinaires pour arriver quand même à résoudre le problème proposé. L'accent est davantage mis sur le processus de résolution de problème et sur la démarche empirique permettant d'apporter des données qui supportent les hypothèses formulées. Lors de la rédaction des problèmes, nous nous sommes basés sur les travaux d'Astolfi, Darot, Ginsburger-Vogel et Toussaint (1997) qui indiquent qu'un problème ouvert devrait avoir les caractéristiques suivantes : (1) un énoncé court qui n'induit ni méthode ni solution, (2) un domaine conceptuel avec lequel les élèves ont assez de familiarité. Ces caractéristiques nous ont guidé pour l'élaboration des SE. En voici un exemple :

Les balles jouent un rôle fort important au tennis. Elles doivent être bien conçues afin de résister aux chocs violents de la raquette en plus de bien rebondir. Déterminez une variable qui peut influencer le rebond de la balle et démontrez comment cette variable influence le rebond. (Dionne, 2008, p. 311)

La seconde étape consiste à identifier les mot-clés du problème qui permettront d'assurer une certaine variabilité d'un problème à l'autre tout en s'assurant de la consistance interne de ce dernier. Dans l'exemple qui nous occupe, les mot-clés sont ceux soulignés dans l'énoncé qui suit. Nous avons retenu trois blocs de mot-clés pour chacun des problèmes :

(1) Les balles jouent un rôle fort important au tennis. Elles doivent être bien conçues afin de résister aux chocs violents de la raquette en plus de bien rebondir. (2) Déterminez une variable qui peut influencer le rebond de la balle et (3) démontrez comment cette variable influence le rebond. (Dionne, 2008, p. 311)

Le premier bloc (1) est la mise en situation à caractère authentique qui permet à l'élève de situer le contexte général dans lequel le problème se situe. Le deuxième bloc (2) représente une consigne à caractère disciplinaire ici associée à la démarche d'investigation scientifique. Dans ce cas-ci, l'élève doit identifier une variable pertinente à étudier compte tenu du problème et de son contexte. On peut également présumer que l'élève va choisir cette variable en formulant une ou plusieurs hypothèse(s) pertinente(s). Enfin, le troisième bloc (3) présente la finalité de la tâche qui consiste à planifier et à construire un dispositif expérimental qui permettra de vérifier de façon empirique l'effet de la variable sur le rebond de la balle de tennis.

À la troisième étape, il s'agit de trouver des variantes aux mot-clés précédemment identifiés. Dans le cas du premier bloc, on peut imaginer n'importe quelle situation authentique où les élèves pourront mettre en place une démarche d'investigation scientifique. L'énoncé qui suit représente une variante du bloc 1 présenté précédemment :

(1) L'été approche et la saison de la pêche va reprendre de plus belle. Afin de te faire un peu d'argent de poche, tu décides de cultiver des lombrics pour les vendre aux pêcheurs. Tu voudrais bien leur offrir les meilleures conditions de vie possible afin d'obtenir de beaux gros lombrics bien en santé qui attireront les poissons !

Il s'agit donc du contexte ou de la mise en situation. Le deuxième bloc fait référence, quant à lui, aux consignes plus disciplinaires associées à la démarche d'investigation scientifique. Dans ce cas-ci, nous retenons :

(2) Parmi toutes les variables qui peuvent influencer l'abondance et la santé des lombrics, choisis-en une.

Et enfin le troisième bloc présente la démarche à réaliser à savoir :

(3) Démontre comment cette variable agit sur les lombrics.

Finalement, l'étape 5 consiste à formuler le nouvel item. La situation générée touche à la biologie. Il s'agit d'une situation originale développée par l'auteur et inspirée du modèle de canevas d'item.

(1) L'été approche et la saison de la pêche va reprendre de plus belle. Afin de te faire un peu d'argent de poche, tu décides de cultiver des lombrics pour les vendre aux pêcheurs. Tu voudrais bien leur offrir les meilleures conditions de vie possible afin d'obtenir de beaux gros lombrics bien en santé qui attireront les poissons ! (2) Parmi toutes les variables qui peuvent influencer l'abondance et la santé des lombrics, choisis-en une et (3) démontre comment cette variable agit sur les lombrics. (Dionne, 2008, p. 355)

Pour compléter la démarche, il faut proposer des solutions aux problèmes présentés, ce qui correspond aux étapes 4 et 6 de la démarche de Haladyna. Ces aspects dépassent le cadre de cet article et ne seront donc pas abordés.

### ***Démarches attendues de l'élève***

Comme nous l'avons observé dans une section précédente, la démarche d'investigation scientifique telle que développée chez les élèves québécois consiste à : (1) cerner un problème, (2) choisir un scénario, (3) concrétiser sa démarche, (4) effectuer l'expérience, (5) analyser ses résultats et (6) faire un retour. Pour les deux exemples précédemment présentés, les élèves devaient obligatoirement réaliser chacune de ces étapes de façon plutôt holistique ou plutôt séquentielle. Dans le cadre de la SE qui abordait la question du rebond d'une balle de tennis, l'élève doit d'abord cerner le problème. Il y a de nombreuses façons d'y arriver : un élève peut décider de reformuler le problème dans ses propres mots ou encore l'illustrer de façon schématique. La seconde étape consiste à choisir un scénario : l'élève doit alors proposer un plan d'investigation montrant, par exemple, ses hypothèses, les données à recueillir, etc. La troisième étape, qui peut être amalgamée à la quatrième, touche à la concrétisation de la démarche : il faut alors que l'élève exécute le scénario qu'il a proposé à l'étape 2. Il s'agit alors pour l'élève de réaliser l'expérience et de recueillir les données. Il peut aussi en profiter pour indiquer les sources d'erreur possible ou les amendements à son scénario initial si c'est le cas. La cinquième étape consiste à analyser les résultats. L'élève doit alors établir les liens pertinents en s'appuyant sur les données recueillies à l'étape 4. Enfin, l'étape 6 consiste à faire un retour : l'élève doit réaliser les inférences adéquates pour répondre à son questionnement initial (étape 1) et, si possible, proposer de nouvelles pistes de recherche en lien avec le problème qui lui était initialement proposé.

### ***La validité de contenu***

La confection de situations d'évaluation est une entreprise complexe qui est due, en partie, à la quantité impressionnante de décisions qu'il faut prendre

tout au long du processus de construction du dispositif d'évaluation. Ces décisions sont d'autant plus difficiles à prendre en l'absence de balises, ce qui est encore le cas aujourd'hui. Lorsque l'évaluation se réalise dans un contexte sommatif ou certificatif, il devient primordial de varier les SE afin d'éviter les risques de contamination. Afin de dupliquer le nombre de SE et de les rendre comparables les unes aux autres, il est utile de s'appuyer sur un canevas qui précise les dimensions importantes et les caractéristiques fondamentales de la famille de situations d'évaluation qui est conçue. Autrement dit, il faut un cadre faisant office de « squelette » et qui permet, selon le contexte disciplinaire, de construire des situations d'évaluation équivalentes (Haladyna et Shindoll, 1989 ; Solano-Flores et coll.,1999). En contexte d'évaluation sommative ou certificative, il importe de s'assurer que les SE proposées aux élèves soient comparables au regard, par exemple, de la difficulté ou du temps nécessaire à la réalisation. La validité de contenu est un processus qui doit permettre de discuter de ces aspects. Dans l'expérimentation qui nous sert ici d'exemple, nous avons documenté la validité de contenu en interrogeant des experts composés de didacticiens des sciences, de spécialistes de la mesure et de l'évaluation et de praticiens du milieu scolaire. Des expertises différentes, mais à la fois complémentaires, nous permettaient d'assurer une forme de régulation. Les SE ont été appréciées initialement au moyen d'une grille comportant 8 critères (Dionne, 2008). Au terme du processus, nous avons laissé tomber deux critères qui apportaient moins d'information pour en retenir seulement 6. Les critères retenus apparaissent au tableau 2 qui suit.

TABLEAU 2. *Critères retenus pour discuter de la validité de contenu*

Critères	Exemples de questions pour les experts
1. Authenticité	Est-ce que la SE ressemble à une situation de la vie courante ?
2. Complexité	Est-ce que la SE présente un défi et des contraintes réalistes pour des élèves de cet âge ?
3. Ouverture	Est-ce que les élèves peuvent faire des choix et prendre des décisions autant sur la production que sur le processus cognitif ?
4. Nature du construit	Est-ce que la SE permet de rendre compte de la compétence visée ?
5. Niveau cognitif	Est-ce que la SE impose aux élèves de recourir à toutes les dimensions d'une compétence (habiletés, techniques, savoirs, savoirs-agir, etc.)
6. Consignes	Est-ce que les consignes sont claires sans induire ni la démarche ni la réponse ?

### Les niveaux

Le modèle que nous avons développé repose sur des situations d'évaluation qui se déclinent en trois différentes « versions » que nous appellerons ici des niveaux (N). Au total, on comptait trois niveaux (N1, N2, N3) pour chacune des situations d'évaluation. Nous considérons que seul le niveau 1 présente une véritable situation permettant de rendre compte d'une compétence. En effet, c'est la seule version où l'élève doit mobiliser ses ressources afin d'offrir une performance qui sera alors jugée. Le niveau 2 présente une situation qui s'apparente à une tâche permettant de rendre compte d'une compétence sans en avoir toutes les caractéristiques. Il s'agit pour l'élève de suivre les étapes qui lui sont proposées tout en prenant certaines initiatives dictées par l'enseignant. Il s'agit en définitive d'une situation qui permet de rendre compte de la maîtrise d'une compétence, mais de façon dirigée. Enfin, le niveau 3 présente une situation qui vise à mesurer la maîtrise des habiletés inhérentes à la compétence déployée. Il s'agit alors pour l'élève de suivre pas à pas les étapes qui lui sont proposées. Le niveau 3 présente donc des situations d'évaluation qui s'éloignent largement d'une situation mesurant réellement une compétence. Cependant, puisque certains élèves sont souvent en difficulté face à des tâches complexes, il est nécessaire d'avoir des outils qui permettent de se prononcer sur leurs acquis. Cela est d'autant plus important dans le cadre d'une évaluation se réalisant dans un contexte à enjeux critiques. Le tableau 3 qui suit dresse une synthèse des neuf situations d'évaluation proposées.

TABLEAU 3. *Formats des situations d'évaluation selon les niveaux.*

Situations d'évaluation	Niveaux		
	N1	N2	N3
SE1	SE1-1	SE1-2	SE1-3
SE2	SE2-1	SE2-2	SE2-3
SE3	SE3-1	SE3-2	SE3-3

Afin de créer des situations d'évaluation comparables les unes aux autres, il a fallu déterminer avec le maximum de précision en quoi chacun des niveaux se distinguait des autres. Pour ce faire, nous avons identifié les indices pouvant être offerts aux élèves afin de les aider dans la résolution de la SE qui leur était offerte. À partir de l'ensemble de ces indices, nous les avons classés en cinq catégories : (1) les indices méthodologiques, (2) le matériel disponible, (3) la latitude quant aux choix pouvant être exercés par l'élève, (4) les indices conceptuels ou théoriques et (5) les indices métacognitifs. Ce sont ces cinq



catégories d'indices qui nous ont permis de distinguer les trois niveaux (N1, N2, N3). Le premier niveau (N1) présente uniquement l'énoncé de l'une des trois situations d'évaluation précédemment présentées. Il s'agit essentiellement de l'énoncé du problème avec un cahier de réponse vide ne suggérant aucune indication sauf les consignes usuelles (ex. nom de l'élève). Il s'agit du niveau qui respecte le mieux l'esprit de l'approche par compétences où l'élève doit mettre à profit les ressources internes (ex. stratégies, compréhension des concepts, etc.) et externes (ex. matériel de laboratoire). En définitive, aucune aide ne lui est accordée. Le deuxième niveau (N2) présente la même situation qu'au niveau 1, mais elle offre, en plus, des indices afin de guider l'élève dans la résolution de la situation qui lui est présentée. Le dosage des indices est important puisque le niveau 2 doit demeurer une situation mesurant une compétence. À titre d'exemple, il faut que l'élève puisse recourir à ses ressources internes et externes afin de résoudre la situation qui lui est proposée. Les indices ont été concoctés afin d'aiguiller l'élève sans lui suggérer une réponse en particulier. Dans le cas du niveau 3 (N3), on propose alors des indices suffisamment précis qui permettent, en théorie, aux élèves de s'engager dans une démarche qui permettra d'arriver à une solution unique. Dans le cas du niveau 3, la situation proposée ressemblait à une expérimentation de type recette qui est traditionnellement utilisée dans les cours de science (Dionne, 2000). Voyons un exemple illustrant la gradation des indices offerts en ce qui concerne les « indices méthodologiques ». Les indices méthodologiques renvoyaient aux étapes de la résolution de problème qui était proposée aux élèves. Dans le cas du niveau 1, il n'y avait aucun indice sur les étapes de la démarche scientifique (ex. identifier le problème, poser une hypothèse, établir un protocole de manipulation, etc.). Pour le niveau 2, on remettait un cahier à l'élève dans lequel il pouvait trouver les étapes génériques que nous avons présentées au point précédent. Enfin, le niveau 3 contenait un cahier de laboratoire dans lequel on pouvait trouver les étapes de réalisation de la situation d'évaluation ainsi que des questions permettant de guider la réflexion de l'élève. La même logique s'est appliquée pour la gradation des indices pour les autres catégories d'indice.

### *La passation des SE*

Chaque élève s'est vu offrir, au hasard, une des trois SE. Au départ, chaque élève recevait la situation de niveau 1 (ex. SE1-1, SE2-1<sup>3</sup>, SE3-1) c'est-à-dire celle qui est le plus susceptible d'évaluer une compétence. Le passage d'un niveau à l'autre (ex. SE1-1 à SE1-2 ou SE1-2 à SE1-3) est fondamentalement une opération basée sur le jugement professionnel des enseignants. Or, nous souhaitons baliser ce jugement afin de mieux encadrer l'administration des différents niveaux. En effet, nous souhaitons ne pas être confronté au fait qu'un enseignant plus expéditif administre le niveau 2 après les 10 premières minutes du premier cours alors qu'un collègue le réalise au terme du troisième cours, ce qui amènerait irrémédiablement des effets négatifs sur la comparabilité des scores. Dans tous les cas, la décision de passer d'un niveau à l'autre

était subordonnée à deux règles que nous avons fournies aux enseignants : (1) administrer le niveau subséquent (ex. passer du niveau 1 au niveau 2 ou du niveau 2 au niveau 3) si l'élève n'arrive pas à progresser adéquatement au terme d'une période de 75 minutes ou (2) administrer le niveau subséquent si l'élève manifeste de toute évidence des difficultés telles que les probabilités de réussite sont jugées faibles au regard du niveau proposé. Bien que ces règles ne garantissent pas la standardisation des modalités de passation, elles représentaient des balises qui nous apparaissaient suffisantes et réalistes compte tenu du contexte de passation. Ceci étant dit, nous sommes conscients qu'il faudrait éventuellement réfléchir à des règles de passation plus rigoureuses qui permettraient de mieux baliser ces opérations de changements de niveaux.

### **La correction des SE**

Afin de rendre compte du travail exécuté par les élèves, nous avons conçu une grille critériée descriptive permettant d'analyser les traces des élèves au regard de 12 critères de notation. Les critères étaient jugés avec une échelle de type Likert en 4 catégories : (0) nettement en deçà des attentes, (1) satisfait minimalement les attentes, (2) satisfait clairement les attentes et (3) dépasse les attentes. Les critères de notation que nous avons retenus sont présentés dans le tableau 4 qui suit.

Les dix premiers critères touchent spécifiquement à l'investigation scientifique. Les deux derniers critères, quant à eux, sont associés à la communication que nous avons considérée comme faisant partie intégrante du processus d'investigation scientifique. Cette grille de notation a été employée pour corriger les trois situations d'évaluation développées. En effet, nous souhaitions avoir un outil qui permettrait aux enseignants de corriger tous les élèves avec les mêmes critères de notation et la même échelle, et ce, peu importe la situation qui était donnée à chacun des élèves. Il faut aussi souligner que cette grille devait également pouvoir être utilisée avec chacun des niveaux (1, 2, 3) d'une même SE.

### **COMMENTAIRES DES ENSEIGNANTS QUI ONT EXPÉRIMENTÉ LES SITUATIONS D'ÉVALUATION**

Bien que l'objectif de cet article soit d'abord et avant tout de présenter une méthodologie pour la création de SE visant à documenter l'investigation scientifique en laboratoire, il nous est apparu pertinent de présenter quelques résultats qualitatifs relatifs aux commentaires des enseignants à la suite de l'expérimentation de ces SE. Lors de cette dernière, nous leur avons remis un journal de bord et nous les avons invités à consigner leurs réflexions et leurs commentaires à chaque cours (généralement une période de 75 minutes). Au total, 22 enseignants participaient à cette recherche et 14 d'entre eux ont remis leur journal de bord. Après avoir pris connaissance de ces commentaires, nous les avons codés de façon descriptive ; les analyses nous ont révélé que tous les commentaires pouvaient être regroupés en deux grandes catégories : (1)

gestion de classe et aspects techniques et (2) nature des situations d'évaluation. Examinons les résultats au regard de chacune de ces catégories.

**TABLEAU 4. Critères de notation des situations d'évaluation**

Critères
1. Reformulation du problème
2. Identification de la variable à examiner
3. Élaboration d'une liste de matériel
4. Formulation d'une hypothèse vérifiable et plausible
5. Formulation et présentation des manipulations
6. Consignation des données expérimentales et des observations
7. Identification des causes d'erreurs
8. Traitement des données expérimentales
9. Maîtrise des concepts, des règles et des techniques
10. Application des résultats expérimentaux dans la solution proposée
11. Présentation générale
12. Utilisation des symboles et des termes

**(1) Gestion de classe et aspects techniques**

La gestion de classe (4/14) et le suivi auprès des élèves (6/14) semblent avoir été difficiles pour plusieurs enseignants. L'un d'entre eux indique :

Beaucoup de déplacements, difficiles à gérer. Voir à ce que tout se passe bien au niveau de la sécurité. Superviser si les élèves font ce qu'ils ont à faire. S'arranger pour faire taire ceux qui parlent... En général assez bien sauf pour un groupe où ce fut très difficile. Bavardage, ils se lancent des objets, de l'eau, etc. Même en étant trois ce fut presque ingérable!!

Ce résultat est cohérent avec nos idées de départ à savoir que les contraintes à caractère pragmatique ne doivent pas être minimisées puisqu'elles peuvent avoir

un impact certain sur la volonté des enseignants à évaluer des compétences plutôt que des connaissances.

## *(2) Nature des situations d'évaluation*

La majorité des enseignants (8/14) ont trouvé difficile d'assigner une SE de niveau 2 ou une SE de niveau 3 à certains élèves. Pour plusieurs d'entre eux, le simple changement de niveau représentait en soi une forme d'échec et était perçu de façon négative.

Malgré les explications des trois niveaux, les élèves voient beaucoup le passage d'un niveau à l'autre comme un échec???

Ce résultat n'est pas très étonnant puisque la majorité des enseignants (9/14) ont mentionné que les SE étaient très différentes de celles présentées de façon ordinaire en classe cette année-là.

Les élèves ont réalisé plusieurs situations d'apprentissage, mais aucune d'elles ne ressemblait à celles proposées dans cette recherche. Les situations que j'ai fait faire aux élèves sont en quelque sorte « les classiques ». On donne un maximum d'informations et l'élève a juste à aller chercher les réponses et répondre aux questions.

Sept enseignants ont aussi indiqué que l'intégrité intellectuelle (effet de contamination) était problématique. Ils mentionnaient que cela n'était pas nécessairement une difficulté compte tenu du caractère formatif associé à la passation, mais qu'ils étaient perplexes dans un contexte sommatif.

Mes trois élèves qui avaient de la difficulté ont miraculeusement tout compris (je crois qu'il y a eu des fuites entre les périodes). D'autres étudiants ont changé leur façon de faire surtout pour les balles de tennis. J'en ai un qui a eu l'idée de mouiller sa balle et tous les autres ont fait la même chose.

Environ le tiers (5/14) des enseignants ont mentionné que les SE avaient été considérées « intéressantes » par les élèves.

J'ai fait un sondage dans mes trois groupes afin de déterminer leurs projets préférés et les moins bien aimés. Sur neuf projets (de janvier à juin), votre SE s'est classée en cinquième position! Pas si mal, hein?

## CONCLUSION

Cet article visait à répondre aux deux questions qui suivent : (1) comment construire des situations d'évaluation qui permettent de juger du développement d'une compétence scientifique dans un contexte ordinaire d'enseignement et donc soumis à de nombreuses contraintes ? et (2) quelles sont les difficultés à prévoir lorsqu'on élabore une méthodologie visant à évaluer une compétence d'investigation scientifique ? Le modèle que nous avons présenté nous apparaît une proposition intéressante et pertinente pour l'évaluation de la compétence d'investigation scientifique. Il est à souhaiter que d'autres propositions de ce genre soient présentées afin d'offrir différents choix aux praticiens.

L'élaboration de situations d'évaluation est une entreprise complexe et d'autant plus quand le contexte de passation est à enjeux critiques. Il faut alors non seulement proposer des situations d'évaluation valides, mais il faut également offrir un contexte de réalisation qui prenne en compte les nombreuses contraintes auxquelles les enseignants sont confrontés. La prise en compte de ces contraintes est également une dimension importante associée à la validité du processus d'évaluation.

Le modèle que nous avons présenté a été expérimenté uniquement au regard de la compétence d'investigation scientifique en laboratoire. Il serait intéressant de mettre à l'épreuve le modèle avec d'autres types de compétences qui s'opérationnalisent dans d'autres contextes d'enseignement comme l'analyse d'objets techniques ou la résolution de problèmes dans un contexte scientifique. Il serait aussi pertinent de voir jusque dans quelle mesure le modèle est généralisable à d'autres disciplines comme les langues ou les mathématiques. Aussi, nous n'avons pas abordé de façon spécifique la question des propriétés métriques des échelles utilisées pour la notation. Il serait pertinent de vérifier sur quel genre d'échelle se situent les données. Une étude qui s'intéresserait à cet aspect pourrait grandement nous éclairer sur la validité théorique de situations d'évaluation de ce genre. Enfin, une perspective de recherche intéressante serait d'informatiser le processus d'administration du prototype de SE que nous avons développé et présenté dans cet article. Il y a tout lieu de croire que l'automatisation de certaines actions, par exemple le processus de passation d'un niveau à un autre, pourrait contribuer à rendre la gestion de ce genre d'épreuve plus aisée pour les enseignants. Aussi, cela permettrait probablement de suivre plus facilement le cheminement cognitif de l'élève au travers des différents niveaux qui lui sont proposés.

## NOTES

1. Le masculin est utilisé dans ce texte dans le seul but d'en alléger la lecture.
2. Au Québec, le ratio élèves / classe au secondaire se situe autour de 30.
3. SE2-1 signifie situation d'évaluation 2 (la balle de tennis) et niveau 1 (situation de compétence).

## RÉFÉRENCES

- Airasian, P. W., Engemann, J. et Gallagher, T. L. (2007). *Classroom assessment: Concepts and applications* (Édition canadienne). Toronto, ON: McGraw Hill.
- Astolfi, J.-P., Darot, E., Ginsburger-Vogel, Y. et Toussaint, J. (1997). *Mots-clés de la didactique des sciences : Repères, définitions, bibliographies*. Bruxelles, BE : De Boeck Université.
- Baker, E. L. (1997). Model-based performance assessment. *Theory into Practice*, 36(4), 247-54.
- Bormuth, J. R. (1970). *On a theory of achievement test items*. Chicago, IL: University of Chicago Press.
- Conderman, G. et Koroghlanian, C. (2002). Writing test questions like a pro. *Intervention in School and Clinic*, 38(2), 83-87.
- Delors, J. (1999). *Rapport à l'UNESCO de la Commission internationale sur l'éducation pour le vingt-et-unième siècle, l'éducation : un trésor est caché dedans*. Paris, FR: Odile Jacob.

- Dionne, E. (2000). *État de l'application du constructivisme en sciences physiques dans une commission scolaire québécoise* (Thèse de maîtrise inédit). Université de Montréal, Montréal, Québec, Canada.
- Dionne, E. (2005). Enjeux en évaluation des apprentissages dans le cadre de programmes d'études formulés par compétences : l'exemple de science et technologie. *Mesure et évaluation en éducation*, 28(2), 49-66.
- Dionne, E. (2008). *Expérimentation d'un modèle d'évaluation permettant de juger du développement d'une compétence d'investigation scientifique en laboratoire* (Thèse de doctorat inédit). Université de Montréal, Montréal, Québec, Canada.
- Dionne, E., et Laurier, M. D. (2010). Expérimentation d'un modèle d'évaluation certificative dans un contexte d'enseignement scientifique. *Revue canadienne de l'éducation*, 33(1), 83-107.
- Durand, M. J. et Chouinard, R. (2006). *L'évaluation des apprentissages : de la planification de la démarche à la communication des résultats*. Montréal, QC: Hurtubise HMH.
- Duschl, R. A. (2003). Assessment of inquiry. Dans J. M. Atkin et J. E. Coffey (dir.), *Everyday assessment in the science classroom* (pp. 41-59). Arlington, VA : National Science Teachers Association Press.
- Gott, R. et Welford, G. (1987). The assessment of observation in science. *School science review*, 69(247), 217-227.
- Haladyna, T. M., Downing, S. M. et Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334.
- Haladyna, T. M. et Rodriguez, M. C. (2013). *Developing and validating multiple-choice test items*. New-York, NY: Routledge.
- Haladyna, T. M. et Shindoll, R. R. (1989). Shells: A method for writing effective multiple-choice test items. *Evaluation in the Health Professions*, 12, 97-104.
- Hogan, T. P. et Murphy, G. (2007). Recommendations for preparing and scoring constructed-response items: What the experts say. *Applied Measurement in Education*, 20(4), 427-441.
- Laurier, M. D., Tousignant, R. et Morissette, D. (2005). *Les principes de la mesure et de l'évaluation des apprentissages* (3<sup>e</sup> éd.). Montréal, QC : Gaëtan Morin.
- Lumpe, A. T., et Oliver, J. S. (1991). Dimensions of hands-on science. *The American Biology Teacher*, 53(6), 345-348.
- McCowan, R. J. (1999). *Developing multiple choice tests: Tips & techniques*. Consulté à partir : [http://www.vcu.edu/cte/workshops/workshop\\_list/references/mccowan.pdf](http://www.vcu.edu/cte/workshops/workshop_list/references/mccowan.pdf)
- McMillan, J. H. (2001). *Classroom assessment : Principles and practice for effective instruction* (2<sup>e</sup> éd.). Needham Heights, MA: Allyn and Bacon.
- Ministère de l'Éducation du Québec. (2003). *Programme de formation de l'école québécoise. Enseignement secondaire, premier cycle*. Québec, QC : Gouvernement du Québec.
- National Research Council (1996). *The National Science Education Standards*. Consulté à partir : [http://www.nap.edu/openbook.php?record\\_id=4962&page=19](http://www.nap.edu/openbook.php?record_id=4962&page=19)
- Perrenoud, P. (1997). *Construire des compétences dès l'école* (3<sup>e</sup> éd.), Paris, Éditions sociales françaises.
- Rey, B., Carette, V., Defrance, A. et Kahn, S. (2003). *Les compétences à l'école, apprentissage et évaluation*. Bruxelles, BE: DeBoeck.
- Roegiers, X. (2000). *Une pédagogie de l'intégration : compétences et intégration des acquis dans l'enseignement*. Bruxelles, BE: DeBoeck Université.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Montréal, QC: ERPI.
- Solano-Flores, G., Jovanovic, J. et Shavelson, R. J. (1994, avril). *Development of an item shell for the generation of performance assessments in physics*. Présentation d'un travail (de recherche) à le Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Solano-Flores, G. et Shavelson, R. J. (1997). Development of performance assessments in science: Conceptual, practical, and logistical issues. *Educational Measurement: Issues and Practice*, 16(3), 16-25.

Solano-Flores, G., Shavelson, R. J. et Bachman, M. (1999). On the development and evaluation of a shell for generating science performance assessments. *International Journal of Science Education*, 21(3), 293-315.

Solano-Flores, G., Shavelson, R. J., & Schneider, S. A. (2001). Expanding the notion of assessment shell: From task development tool to instrument for guiding the process of assessment development. *REDIE: Electronic Journal of Educational Research*, 3(1), 1 - 19. Consulté à partir: <http://redie.ens.uabc.mx>.

Tardif, J. (2003). Développer un programme de compétences : de l'intention à la mise en oeuvre. *Pédagogie collégiale*, 16(3), 36-45.

Tardif, J. (2006). L'évaluation des compétences : Documenter le parcours de développement. Montréal, QC : Chenelière.

ERIC DIONNE est professeur à la Faculté d'éducation de l'Université d'Ottawa. Les intérêts de recherche du professeur Dionne touchent à la modélisation des scores ainsi qu'au développement et à la validation d'outils de cueillette de données. [eric.dionne@uottawa.ca](mailto:eric.dionne@uottawa.ca)

ERIC DIONNE is a professor in the Faculty of Education at Ottawa University. His research interests are educational modelling and the development and validation of data collection tools. [eric.dionne@uottawa.ca](mailto:eric.dionne@uottawa.ca)

