

**Récit de pratique sur l'adaptation et sur l'utilisation  
d'un dispositif de formation au jugement évaluatif  
basé sur la concordance**

**Practice narrative on the adaptation and use  
of a training device for evaluative judgment based  
on concordance**

**Relato prático sobre a adaptação e utilização  
de um dispositivo de formação do juízo avaliativo  
baseado na concordância**

**Marie-France Deschênes**  
ID ORCID : 0000-0002-7949-4466  
*Université de Montréal*

**Eric Dionne**  
ID ORCID : 0000-0003-3046-4386  
*Université d'Ottawa*

**Michelle Dorion**  
ID ORCID : 0000-0001-5173-5435  
*Université d'Ottawa*



**MOTS CLÉS :** dispositif éducatif, évaluation des apprentissages, formation par concordance, innovation pédagogique, jugement évaluatif

*Les futurs enseignants ont rarement l'occasion de tester leurs interprétations de questions complexes, incertaines ou ambiguës en évaluation des apprentissages et d'obtenir une rétroaction formative sur ces questions. Dans cet article, nous présentons un récit de pratique sur l'adaptation et sur l'utilisation d'un dispositif de formation au jugement évaluatif de futurs enseignants basé sur la concordance de jugement. Les assises théoriques du dispositif, les principes de sa conception ainsi que les facteurs contextuels ayant influencé son adaptation sont abordés. Un tel exercice nous permet de cerner quelques avantages de l'utilisation du dispositif pour favoriser le développement du jugement évaluatif en éducation.*

**KEY WORDS:** educational device, evaluative judgment, learning assessment, learning by concordance, pedagogical innovation

*There remains a scarcity of opportunities for future educators to test their interpretations of complex, uncertain, or ambiguous questions in learning assessment and to obtain formative feedback on these questions. This article proposes a pedagogical practice based on the adaptation and the use of a Concordance-of-Judgment educational device to exercise future educators' evaluative judgment. The theoretical foundations of the educational device, the principles of its design as well as the contextual factors that have influenced its adaptation in an educational context are discussed. Such exercise allows us to identify some of the benefits of using the device to promote the development of educators' evaluative judgment.*

**PALAVRAS CHAVE:** avaliação da aprendizagem, dispositivo educativo, formação por concordância, inovação pedagógica, juízo avaliativo

*Os futuros professores raramente têm a oportunidade de testar as suas interpretações de questões complexas, incertas ou ambíguas em avaliação das aprendizagens e de receber feedback formativo sobre estas questões. Neste artigo, apresentamos uma narrativa de prática sobre a adaptação e o uso de um dispositivo de formação no juízo avaliativo de futuros professores baseado na concordância de juízo. Discutimos as bases teóricas do dispositivo, os princípios da sua conceção, bem como os fatores contextuais que influenciaram sua adaptação. Tal exercício permite identificar algumas vantagens do uso do dispositivo para promover o desenvolvimento do juízo avaliativo em educação.*

---

Note des auteurs : La correspondance liée à cet article peut être adressée à [marie-france.deschenes@umontreal.ca](mailto:marie-france.deschenes@umontreal.ca)

## Introduction

Dans sa pratique, tout enseignant est, à un moment ou à un autre, confronté à des situations complexes ou ambiguës en évaluation des apprentissages dans lesquelles il est amené à peser ses décisions pédagogiques et à exercer un jugement évaluatif (Chaumont & Leroux, 2018; Dionne & Simões Forte, 2013; Jabůrek et al., 2022; Leroux & Bélair, 2015). Le jugement évaluatif, soit le jugement professionnel exercé dans un contexte d'évaluation des apprentissages (Maes et al., 2019; Maes et al., 2020; Mottier Lopez & Allal, 2008), représente une compétence fondamentale à développer chez les enseignants. En effet, la plupart des politiques en évaluation des apprentissages, du moins au Canada, mettent en lumière l'importance du jugement évaluatif (Gouvernement de l'Ontario, 2010; Ministère de l'Éducation du Québec, 2003). Or, rares sont les occasions où les futurs enseignants peuvent tester leurs interprétations de situations d'évaluation dans lesquelles des éléments de complexité, d'ambiguïté ou d'incertitude font partie des décisions pédagogiques (Dionne & Simões Forte, 2013; Fives & Barnes, 2020; Smith, 2017).

Dans cet article, nous présentons le récit d'une pratique pédagogique abordant l'adaptation et l'utilisation d'un dispositif de formation au jugement évaluatif basé sur la concordance de jugement. Ce dispositif, le test de jugement professionnel en évaluation des apprentissages (TJPEA), a été conçu pour développer la compétence à évaluer les apprentissages des élèves au secondaire dans un contexte de la formation à l'enseignement. Le TJPEA inclut 25 situations de la pratique évaluative en enseignement (Dionne & Simões Forte, 2013). L'évaluation des apprentissages est le domaine exclusif utilisé dans le test, étant donné qu'il incarne bien la complexité et l'omniprésence de l'exercice du jugement professionnel en enseignement (Laveault, 2005, 2008; Leroux & Bélair, 2015; Maes et al., 2019). Dans les paragraphes suivants, nous abordons le concept du jugement évaluatif en enseignement et quelques résultats d'études rapportant le malaise persistant des enseignants face à son exercice. Ensuite, nous décrivons le récit de pratique pédagogique. Afin de structurer cette

description, nous nous appuyons sur le cadre de Cianciolo et Regehr (2019) proposant l'analyse stratifiée d'une pratique ou d'une intervention pédagogique innovante. Ce cadre nous permet d'examiner les interactions entre les assises théoriques qui sous-tendent l'utilisation du dispositif de formation, les principes de sa conception et de son adaptation ainsi que les facteurs contextuels influençant son utilisation pour répondre aux intentions pédagogiques ciblées. Grâce à cet exercice, nous avons pu cerner quelques avantages liés à l'utilisation du dispositif de formation pour développer le jugement évaluatif de futurs enseignants. Enfin, nous abordons les limites de l'utilisation du dispositif de formation ainsi que des perspectives pour accroître son usage en éducation.

### ***Le jugement évaluatif en enseignement : une compétence à développer chez les futurs enseignants***

Évaluer les apprentissages des élèves en approche par compétences représente un défi de taille pour les enseignants (Chaumont & Leroux, 2018; Leroux & Bélair, 2015; Tourmen, 2015). Dans cette approche, l'évaluation des apprentissages fait appel à des conditions authentiques de réalisation d'une tâche complexe ou d'une production concrète par les élèves (Tardif, 2017). Pour évaluer les apprentissages à partir de cette tâche, les enseignants doivent exercer leur jugement évaluatif au regard de la mobilisation et de l'utilisation efficaces des ressources par les élèves, et du niveau de développement des compétences visées (Bélair, 2014; Scallon, 2015; Tardif, 2006, 2017).

Le jugement évaluatif, selon Mottier Lopez et Allal (2008), se traduit par des démarches heuristiques dans lesquelles les enseignants réunissent et confrontent les preuves fournies par l'élève, combinent ces preuves et les interprètent afin de prendre une décision éclairée quant au niveau de développement des compétences. Aux fins de ces décisions, les enseignants s'appuient sur des procédures d'évaluation rigoureuses, transparentes, justes et équitables, tout en s'assurant de leur adaptation aux situations singulières qui l'exigent. Le jugement évaluatif se conçoit aussi comme un acte cognitif individuel chez l'enseignant. Cet acte s'intègre au sein d'un réseau de communication, dans un contexte social et institutionnel, et respecte les exigences du système scolaire, la culture de l'établissement, ainsi que les valeurs et les stratégies pédagogiques des enseignants (Chaumont & Leroux, 2018; Mottier Lopez & Allal, 2008; Tourmen, 2014).

Or, les décisions pédagogiques en matière d'évaluation des apprentissages ne sont pas exclusivement guidées par un processus rationnel de traitement et d'analyse des données. L'enseignant exerce aussi son jugement évaluatif à travers ses conceptions de l'enseignement, ses valeurs et ses expériences antérieures (Laveault, 2008 ; Leroux & Bélair, 2015 ; Piot, 2008 ; Tourmen, 2015). Il prend certaines décisions en se fiant à son intuition (Vanlommel et al., 2018) ou à ses représentations individuelles quant à la nature même de la compétence (Tourmen, 2015), ce qui rend difficile l'explication des motifs, des intentions ou des circonstances qui soutiennent ses décisions. Ceci est particulièrement vrai dans des situations ambiguës et dans lesquelles il n'existe pas de solution univoque ou endossée par l'ensemble de la communauté scientifique (Merle, 2012). Complexe, l'exercice du jugement évaluatif implique inéluctablement une part de subjectivité chez les enseignants, où certains biais cognitifs peuvent se manifester dans le processus d'évaluation (Campbell, 2015 ; Merle, 2012, 2018). Plusieurs chercheurs en éducation ont d'ailleurs proposé des pratiques évaluatives pour encadrer ou pour minimiser cette subjectivité, comme le recours à des preuves suffisantes et pertinentes pour attester le développement des compétences, l'utilisation de grilles critériées et de pratiques évaluatives incorporant les principes de rigueur, de transparence et de cohérence, et ce, surtout dans le contexte de l'attribution de notes (Chaumont & Leroux, 2018 ; Leroux & Bélair, 2015 ; Seden & Svaricek, 2018).

Processus dynamique et itératif, le jugement évaluatif suppose ainsi une interprétation fine des preuves recueillies ou des indices observés chez les élèves, en cohérence avec les intentions pédagogiques préalablement ciblées. En outre, il implique fréquemment l'emploi d'outils de mesure rigoureux pour évaluer les apprentissages et pour documenter fidèlement l'atteinte d'un niveau de développement des compétences (Leroux & Bélair, 2015 ; Loye & Fontaine, 2018). Or, même si l'évaluation des apprentissages ou des compétences s'engage dans une activité outillée, par l'usage de grilles critériées ou de référentiels de compétences, ces outils ne résolvent pas les difficultés éprouvées par les enseignants (Tourmen, 2014 ; Tourmen, 2015).

Complexe, le jugement évaluatif est fréquemment une source de questionnements, d'inconfort ou d'incertitude chez les enseignants (Dionne & Simões Forte, 2013 ; Leroux & Bélair, 2015 ; Maes et al., 2019 ; Maes et al., 2020 ; Smith, 2017). Malgré une documentation scientifique de plus en plus abondante sur le thème, des résultats d'études réitèrent le malaise persistant qu'éprouvent les enseignants face à l'exercice du jugement évaluatif

(Baribeau, 2015, 2020; Chaumont & Leroux, 2018; Leroux & Bélair, 2015; Maes et al., 2019; Maes et al., 2020). Dans une récente étude qualitative exploratoire, Maes et al. (2019) se sont intéressés à la construction du jugement évaluatif de superviseurs de stage en formation initiale des enseignants et ont mené des entretiens individuels auprès de huit superviseurs. Les résultats de l'étude mettent en évidence le caractère situé du jugement évaluatif, exemplifié par la prise en compte par les superviseurs d'aspects singuliers relatifs au contexte de stage. Les résultats montrent aussi que le jugement évaluatif final se construit sur une série de jugements d'appréciation partiels et provisoires. Malgré cette pratique, la majorité des superviseurs interviewés ont évoqué des zones d'inconfort quant à l'exercice de leur jugement évaluatif, notamment au regard de la façon de transmettre le message ou de la remise en question de leur jugement (Maes et al., 2019; Maes et al., 2020).

Dans son étude doctorale, Baribeau (2015) a réalisé des entretiens semi-dirigés sous forme d'analyse de récits de pratique professionnelle auprès d'enseignants ( $n = 12$ ). Le but de cette étude qualitative interprétative consistait à analyser les pratiques déclarées d'évaluation dans des décisions sommatives de certification des apprentissages d'élèves du secondaire. La chercheuse voulait comprendre de quelle manière se construit l'agir évaluatif de ces enseignants. Les résultats des entretiens ont été présentés ultérieurement à d'autres enseignants ( $n = 10$ ) afin d'obtenir une perspective plus globale du phénomène d'intérêt. Les résultats de l'étude font ressortir des obstacles et des tensions quant aux pratiques d'évaluation des apprentissages adoptées par les enseignants. Les résultats montrent que ces pratiques contiennent quelquefois une part d'arbitraire, où les enseignants s'appuient sur leur connaissance de l'élève et sur leurs impressions quant à l'engagement de l'élève dans ses apprentissages. La chercheuse affirme que cette situation pourrait s'expliquer, entre autres, par le manque de dispositifs de formation professionnalisante en évaluation des apprentissages (Baribeau, 2015, 2020).

Ainsi, la formation au jugement évaluatif des futurs enseignants s'avère essentielle dans les programmes de formation en enseignement. Les cours théoriques et les stages pratiques sont des stratégies éducatives utilisées pour favoriser le développement du jugement évaluatif des futurs enseignants (Leroux, 2019; Maes et al., 2019; Maes et al., 2020). Toutefois, malgré leur pertinence, ces stratégies semblent insuffisantes, considérant le malaise persistant documenté dans la recherche. Qui plus est, pendant leur

formation, les étudiants ont peu d'occasions qui favorisent le développement du jugement professionnel dans lesquelles des éléments d'incertitude et d'ambiguïté font partie des décisions pédagogiques (Deschênes et al., 2022; Dionne & Simões Forte, 2013; Smith, 2017). Rares sont les mécanismes permettant aux futurs enseignants de tester leurs interprétations de situations pédagogiques complexes, de réfléchir aux enjeux soulevés dans ces situations et d'obtenir une rétroaction formative. Bref, l'instrumentation comme appui à la formation en évaluation des futurs enseignants demeure insuffisante pour développer la compétence en évaluation des apprentissages (Baribeau, 2015, 2020).

### ***Un dispositif de formation pour exercer le jugement professionnel en évaluation des apprentissages***

Afin de favoriser le développement de la compétence à évaluer les apprentissages, un dispositif de formation au jugement évaluatif basé sur la concordance de jugement a été conçu, le TJPEA. Nous avons opté pour l'utilisation du terme dispositif de formation, car l'objet vise un soutien à l'apprentissage à l'aide d'une formation structurée et caractérisée par la combinaison délibérée d'interventions en ligne et en classe (Boelens et al., 2017; Sacré et al., 2019). Cette conception se distingue de l'usage d'une pratique pédagogique qui représente les différentes actions réalisées par l'enseignant, de manière plus ou moins consciente, en vue de favoriser les apprentissages des élèves (Duget, 2014). L'intention pédagogique liée à l'utilisation du dispositif visait le développement de la compétence à évaluer les apprentissages des élèves au secondaire dans un contexte de formation à l'enseignement. En cohérence avec le cadre de Cianciolo et Regehr (2019) proposant l'analyse stratifiée d'une pratique ou d'une intervention pédagogique innovante, nous décrivons, dans les prochains paragraphes, 1) les assises théoriques du dispositif de formation, 2) les aspects structurels de la pratique ou les principes de sa conception, ainsi que 3) les facteurs contextuels ayant influencé son adaptation en contexte éducatif. Une telle analyse aide à voir si la pratique pédagogique a répondu à l'intention pédagogique souhaitée et à en tirer des conclusions plausibles sur les composantes auxquelles ce résultat peut être attribué.

#### ***1) Les assises théoriques du dispositif de formation***

Le TJPEA est une adaptation du test de concordance de script (TCS), outil d'évaluation du raisonnement clinique, développé à la fin des années 1990 et utilisé en éducation médicale et, plus largement, en sciences de la

santé (Dory et al., 2012; Lubarsky et al., 2011; Lubarsky et al., 2013). La notion de concordance du TCS est liée au fait qu'en faisant le test, les répondants peuvent vérifier à quel point leurs réponses se rapprochent ou s'éloignent de celles de personnes expertes ayant préalablement répondu aux mêmes questions ou encore, si elles concordent avec leurs réponses. Le TCS est basé sur la théorie des scripts qui postule que toute personne n'appréhende pas les situations de la vie dans un vide cognitif ou sans s'appuyer sur ses représentations mentales (Abelson, 1975; Schank & Abelson, 1977). Elle réagit selon ce qui lui semble le plus pertinent à partir de son bagage de connaissances mémorisées sous forme de modèles ou de schémas. Un script est défini comme un schéma de représentation mentale qui stéréotype en unités de sens différentes informations liées à une situation (Schank & Abelson, 1977). Il se conçoit par des réseaux de connaissances finement organisés et élaborés dans la mémoire à long terme (Abelson, 1975; Schank & Abelson, 1977).

Au cours des années 1980, la théorie des scripts a été transposée en formation médicale pour décrire le raisonnement clinique, notamment à l'aide du concept de script-maladie. En cohérence avec les assises cognitivistes, le concept de script-maladie contient des liens associatifs entre les différentes connaissances (théoriques et cliniques) mobilisées dans le processus de raisonnement clinique du médecin. Cette mobilisation des connaissances permet de poser un diagnostic juste, d'investiguer et de traiter la personne (Charlin et al., 2000; Schmidt et al., 1990). Par exemple, le concept de script-infarctus du myocarde contient des liens associatifs entre les connaissances et les données suivantes : 1) des signes cliniques typiques (p. ex., une douleur à la poitrine) et d'autres signes atypiques, 2) des facteurs précipitants et d'autres prédisposants, 3) des antécédents de maladies cardiaques, 4) des examens et des traitements immédiats nécessaires, 5) des conséquences potentielles de la maladie, etc. Sur le plan cognitif, les scripts permettent au médecin d'enclencher un raisonnement clinique de type hypothético-déductif. Concrètement, il confronte les données d'une situation clinique avec ses propres scripts, ce qui l'aide à reconnaître des *patterns*, des similitudes ou des éléments saillants dans la situation pour orienter son raisonnement clinique. Ensuite, il recherche des données supplémentaires pour minimiser, renforcer, voire prioriser des hypothèses cliniques. Dans cette posture cognitiviste, l'expertise en matière de raisonnement clinique est liée



à la richesse des connaissances du médecin, mais aussi à la façon dont les connaissances sont organisées de façon intelligible dans sa mémoire à long terme (Charlin et al., 2000; Custers, 2015).

Dans le domaine de l'éducation, la question est de savoir s'il y a de tels scripts et de quelle manière ils se conçoivent dans l'exercice du rôle en enseignement. Pour ce faire, nous abordons la question à partir de la conception du jugement évaluatif de Mottier Lopez et Allal (2008) présentée précédemment. Cette définition conçoit bien la présence des processus décisionnels de l'enseignant et la mobilisation des connaissances issues de son expertise (expérience et formation). Comme dans tout processus de raisonnement, l'enseignant émet, plus ou moins consciemment, des hypothèses d'interventions pédagogiques à la lumière des informations qu'il détient dans les situations de sa pratique professionnelle. Grâce à son expertise, il peut aussi identifier des éléments déterminants (valeurs ou principes pédagogiques interpellés) pour guider son jugement selon le contexte dans lequel il œuvre. Par exemple, le script d'évaluation des apprentissages de l'enseignant pourrait comprendre les attributs suivants : 1) les principes d'équité, de justice et de transparence dans l'évaluation des apprentissages, 2) les exigences institutionnelles formalisées (règles ou politiques en matière d'évaluation des apprentissages), 3) les ressources dont il dispose et 4) les conséquences possibles de ses décisions sur l'élève (motivation, parcours scolaire, etc.) (Chaumont & Leroux, 2018; Dionne & Simões Forte, 2013; Leroux & Bélair, 2015; Smith, 2017).

Que cela soit en médecine ou en éducation, le construit ou le trait sollicité est complexe, même si l'un se décline en raisonnement clinique et l'autre en jugement évaluatif. Les situations rencontrées dans l'un ou l'autre des contextes professionnels amènent à interpréter des données, à prendre des décisions en situation d'incertitude et à composer avec des informations quelques fois manquantes, incomplètes ou ambiguës. En définitive, bien qu'il y ait des différences, les ressemblances sont assez nombreuses pour entrevoir l'adaptation et l'utilisation du dispositif de formation par concordance en contexte éducatif (Deschênes et al., 2022; Dionne & Simões Forte, 2013; Smith, 2017).

## ***2) Les principes pour concevoir le dispositif de formation***

Pour concevoir le dispositif de développement du jugement évaluatif par concordance de jugement, nous avons suivi les démarches proposées dans la documentation scientifique en formation des sciences de la santé abordant

les TCS (Dory et al., 2012 ; Lubarsky et al., 2013). Nous nous sommes aussi appuyés sur notre expérience de conception de tels dispositifs. Dans les prochaines lignes, nous présentons en quoi consiste le dispositif de formation, ses principes de conception et son adaptation en contexte éducatif.

### *Le TJPEA et ses principes de conception*

C'est à partir des similitudes observées entre la complexité du jugement professionnel en enseignement et le processus de raisonnement clinique qu'un TJPEA a été développé en 2008. Le dispositif de formation contient 25 vignettes pour sonder le jugement évaluatif dans des situations fréquentes de la pratique comme la qualité de la rédaction des items d'un outil de mesure, la planification d'une évaluation, les concepts de validité et de fidélité d'une évaluation, l'interprétation des scores et la communication des résultats. Une table de spécification élaborée en amont du dispositif de formation met en parallèle les situations authentiques de la pratique en enseignement et des composantes du processus du jugement évaluatif dans les items du TJPEA (p. ex., la planification, la prise d'informations et l'interprétation, le jugement, la décision et l'action).

Dans le TJPEA, chacune des vignettes contient des situations qui sont suivies d'items ( $n = 3$  à  $5$ ) afin de solliciter des microdécisions<sup>1</sup> chez les étudiants. Chaque vignette est composée de quatre sections, tel qu'illustré par la figure 1. La section 1 présente une courte situation problématique. La situation doit évoquer un problème complexe, mal défini (Mayer & Wittrock, 2006) ou mal structuré (Jonassen, 2011 ; Xun & Land, 2004). Un problème mal défini ou mal structuré représente une situation authentique de la pratique dans laquelle des éléments d'incertitude ou d'incomplétude subsistent à la prise de décision (Jonassen, 2011). En éducation ou, plus généralement, en sciences sociales, de telles situations ne possèdent habituellement pas de solution univoque ou endossée par l'ensemble de la communauté scientifique (Voss, 1988). On trouve ainsi des informations parcellaires, incomplètes ou encore volontairement ambiguës dans les situations des vignettes du dispositif de formation. Le choix de situations problématiques courantes, représentatives ou à forts enjeux, est encouragé lors de la rédaction des situations. Ceci se rapporte au principe de signifiante ou d'authenticité de la tâche cognitive dévolue à l'étudiant (Ashford-Rowe et al., 2014 ; Wiggins, 1993, 2011).

---

1. Une microdécision est une prise de décision malgré l'absence de l'ensemble des informations propres à une situation pour exercer un jugement définitif ou arrêté.

Les sections 2 et 3 portent sur les items menant les étudiants à prendre des microdécisions afin de résoudre, en partie, la situation problématique. Des hypothèses sont proposées (si vous pensiez à ...), lesquelles sont suivies d'une nouvelle information (suivant votre intervention, vous observez que ...). Indépendantes les unes des autres, les hypothèses présentées dans la section 2 sont plausibles; elles représentent des situations que les étudiants pourraient rencontrer en milieu écologique. En outre, les nouvelles informations dévoilées dans la section 3 visent à aiguiller la microdécision.

La section 4 permet aux étudiants de se prononcer sur chacune des hypothèses (votre intervention était ...). Lors de la conception du dispositif de formation, le libellé des catégories de choix de réponses peut être adapté selon la cible du jugement professionnel sollicité. La pertinence, l'acceptabilité ou l'utilité des hypothèses d'intervention représentent des caractéristiques fréquemment utilisées dans les catégories de choix de réponses du dispositif. Le libellé doit être clairement rédigé pour éviter que les étudiants fassent des choix hasardeux. En outre, la possibilité de donner une réponse neutre (p. ex., ni plus ni moins utile) est fréquemment retirée afin d'éviter des valeurs de refuge sur des questions éthiques, déontologiques ou qui touchent le professionnalisme. Des personnes expertes détenant quelques années d'expérience dans le domaine de l'enseignement répondent préalablement à tous les items du dispositif de formation. Ce sont des personnes crédibles et susceptibles de juger pertinemment les situations et les hypothèses suggérées dans les vignettes. Partenaires dans le processus de conception du dispositif de formation, elles répondent individuellement aux items du dispositif, sans consulter des collègues ou des ouvrages de référence (Dory et al., 2012; Lubarsky et al., 2013)<sup>2</sup>. Cette démarche permet d'obtenir une variabilité des choix de réponses aux items par les personnes expertes. Cela illustre que plus d'une option peut être envisagée dans les décisions complexes en contexte d'incertitude et d'ambiguïté.

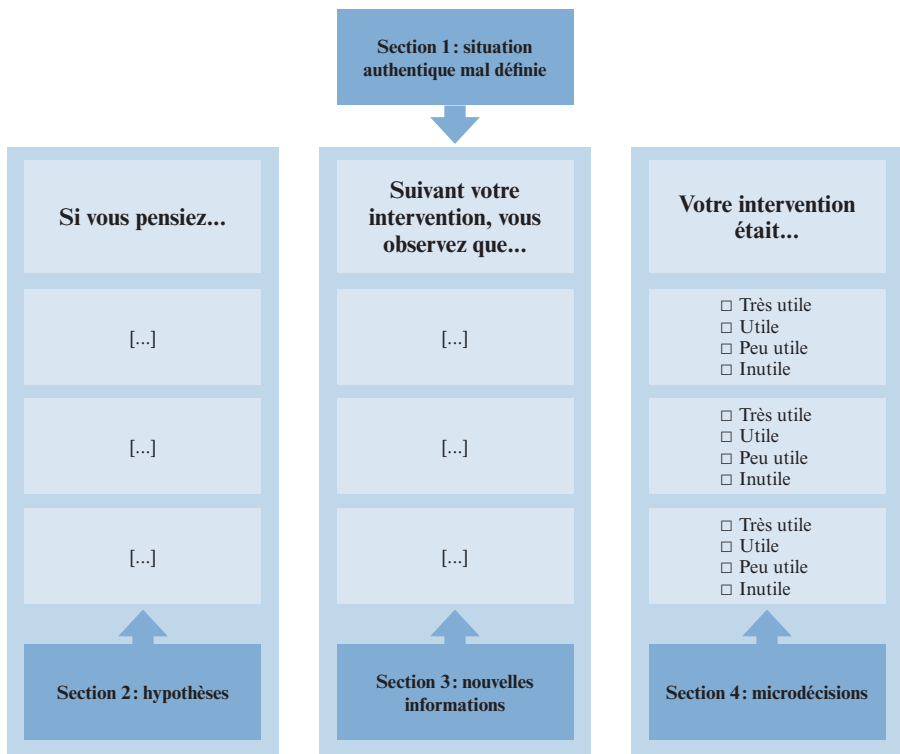
En répondant aux questions du dispositif de formation, les étudiants peuvent calibrer jusqu'à quel point leurs choix de réponses concordent ou non avec ceux des personnes expertes ayant préalablement répondu aux questions. L'originalité du TCS est qu'il tient compte, dans le calcul du score de l'étudiant, des réponses préalablement données par les personnes

---

2. Des environnements numériques à l'apprentissage (ENA) permettent aux personnes expertes de répondre en ligne aux items afin de générer automatiquement les rétroactions aux étudiants.

expertes. Plus l'étudiant donne une réponse qui concorde avec la réponse modale des personnes expertes, plus son score sera élevé. À l'inverse, plus la catégorie sélectionnée s'éloigne de la réponse modale des personnes expertes, plus son score sera faible (Dory et al., 2012 ; Lubarsky et al., 2011). Toutefois, cette méthode d'attribution des scores génère des scores différents, selon le nombre et les caractéristiques des personnes expertes consultées. Ceci pose un problème si le but est de donner un score absolu aux étudiants (Loye & Fontaine, 2018).

Figure 1  
*Sections d'une vignette dans le dispositif de formation*



### *Les adaptations du dispositif de formation*

La conception du dispositif de formation pour exercer le jugement professionnel en évaluation des apprentissages permet de constater de la transférabilité du TCS d'autres domaines professionnels, notamment pour

exercer le jugement professionnel en évaluation des apprentissages. Or, ceci soulève des questions quant au construit (ou au trait) à solliciter dans le contenu des situations et des items du dispositif de formation.

En formation médicale, la conception du script-maladie représente souvent une approche assez algorithmique des décisions à partir d'éléments clés pour catégoriser les manifestations ou l'histoire d'une personne en lien avec une maladie. La situation d'une personne subissant un infarctus du myocarde en est un exemple. Pour des situations problématiques en éducation, les réseaux de connaissances des enseignants ne semblent pas exclusivement reposer sur des connaissances menant à catégoriser une situation éducative. Comme l'ont mentionné plusieurs auteurs (Chaumont & Leroux, 2018; Laveault, 2008; Maes et al., 2019; Maes et al., 2020; Mottier Lopez & Allal, 2008), enseigner est une pratique professionnelle socialement située. La plupart des décisions pédagogiques se prennent en considérant le contexte scolaire, les règles institutionnelles, les savoirs issus de la théorie ainsi que les avis des collègues (Chaumont & Leroux, 2018; Laveault, 2008; Leroux & Bélair, 2015; Mottier Lopez & Allal, 2008).

Les principaux défis rencontrés lors de la rédaction du contenu du dispositif de formation touchent deux aspects: 1) le niveau d'incertitude requis pour répondre à la condition souhaitée d'authenticité des situations des vignettes et 2) l'identification du construit sollicité dans les items. Afin d'illustrer notre propos, nous proposons l'exemple de l'encadré 1.

La situation problématique sollicite des microdécisions en évaluation des apprentissages à partir de la mobilisation des connaissances issues de l'expérience et de la formation des répondants (personnes expertes et étudiants). Dans de telles situations authentiques, l'enseignant est appelé à faire preuve de professionnalisme, à connaître les règles et les principes d'équité en évaluation des apprentissages de son institution, à travailler en collégialité avec son groupe de pairs. Or, le questionnement et les plaintes des élèves, tout comme le contexte (p. ex., les politiques institutionnelles en évaluation des apprentissages, les pratiques du groupe professionnel) jouent un rôle dans les microdécisions sollicitées. Dans cette situation et dans d'autres situations similaires, le jugement de l'enseignant n'est pas exclusivement égocentré, il est plutôt « écocentré ». Il s'incarne dans un environnement singulier en tenant compte des aspects contextuels, sociaux-matériels, institutionnels, etc. d'une situation pédagogique. Bref, « le jugement professionnel n'est pas isolé du contexte dans lequel il s'exerce » (Leroux & Bélair, 2015, p. 99).

**Encadré 1. Exemple d'une vignette en contexte d'évaluation des apprentissages**

Après avoir assigné une évaluation sommative à vos élèves, l'un d'entre eux vous demande s'il peut vous soumettre un travail qu'il avait déjà effectué pour un autre cours, mais qui répond aux exigences du vôtre.

<b>Si vous pensez à...</b>	<b>À la suite de votre intervention...</b>
1. ... permettre à l'élève de vous remettre ce travail.	... vous corrigez le travail et remarquez que l'élève maîtrise effectivement la matière et qu'il répond aux attentes.
<b>Vous considérez que votre intervention était :</b>	<input type="checkbox"/> <b>Très utile</b> <input type="checkbox"/> <b>Utile</b> <input type="checkbox"/> <b>Peu utile</b> <input type="checkbox"/> <b>Inutile</b>
2. ... permettre à l'élève de remettre ce travail, mais vous lui demandez d'y ajouter une courte section expliquant les différences entre la façon dont la matière a été abordée au sein de l'autre cours et dans le vôtre.	... l'élève obtient une bonne note sur le travail, mais certains élèves se plaignent que cette permission spéciale est injuste.
<b>Vous considérez que votre intervention était :</b>	<input type="checkbox"/> <b>Très utile</b> <input type="checkbox"/> <b>Utile</b> <input type="checkbox"/> <b>Peu utile</b> <input type="checkbox"/> <b>Inutile</b>
3. ... ne pas permettre à l'élève de soumettre le travail déjà fait.	... l'élève conteste votre décision en expliquant que le but de votre évaluation est de vérifier l'apprentissage, et non pas le contexte dans lequel cet apprentissage a eu lieu.
<b>Vous considérez que votre intervention était :</b>	<input type="checkbox"/> <b>Très utile</b> <input type="checkbox"/> <b>Utile</b> <input type="checkbox"/> <b>Peu utile</b> <input type="checkbox"/> <b>Inutile</b>

**3) Les facteurs contextuels d'utilisation du dispositif de formation**

Le TJPEA a été utilisé à des fins de formation d'étudiants dans un programme menant à la profession enseignante (Dionne & Simões Forte, 2013). Dans les prochains paragraphes, nous abordons les éléments contextuels influençant les retombées ou les avantages observés au regard de l'utilisation du dispositif de formation.

Dionne et Simões Forte (2013) ont utilisé le TJPEA avec 72 étudiants répartis dans deux groupes inscrits dans un programme de formation d'une université canadienne menant à la profession enseignante au secondaire. Chaque groupe d'étudiants s'est vu proposer quatre items au cours de la session. Lors de la pratique pédagogique, les vignettes du dispositif étaient

projetées en classe, sans la rétroaction de personnes expertes. À l'aide de télévotants<sup>3</sup>, les étudiants indiquaient leurs réponses aux items. Ils étaient invités à juger l'utilité des interventions pédagogiques proposées et la probabilité qu'ils réalisent de telles interventions en contexte écologique (voir Encadré 2). La compilation des réponses des étudiants était ensuite présentée sous la forme de graphiques projetés à l'écran. Une discussion était dès lors lancée entre les étudiants et le professeur. À chaque question, les étudiants étaient invités à préciser à voix haute les données manquantes ou les données ambiguës décelées dans les vignettes, ainsi que des aspects qu'ils changeraient dans la situation présentée. Un devis qualitatif interprétatif mettant l'accent sur l'analyse des commentaires des étudiants a été utilisé.

### *Encadré 2. Vignette en évaluation des apprentissages*

En créant une tâche complexe menant à l'évaluation certificative des apprentissages, votre collègue vous suggère de demander aux étudiants de remplir une fiche d'auto-évaluation et d'évaluation des pairs au sein de leur groupe de travail.

<b>Si vous pensez à...</b>	<b>À la suite de votre intervention, vous observez que...</b>			
... accepter de faire remplir la fiche par les étudiants. Vous les avisez aussi que les commentaires de leurs pairs peuvent influencer leur résultat.	... vos étudiants démontrent de la confusion quant à la pondération de cette auto-évaluation par les pairs.			
<b>A. Vous considérez que votre intervention était :</b>	<input type="checkbox"/> Très utile	<input type="checkbox"/> Utile	<input type="checkbox"/> Peu utile	<input type="checkbox"/> Inutile
<b>B. La probabilité que vous ayez réalisé cette intervention est :</b>	<input type="checkbox"/> Très probable	<input type="checkbox"/> Probable	<input type="checkbox"/> Peu probable	<input type="checkbox"/> Très improbable

Suivant l'utilisation du dispositif, Dionne et Simões Forte (2013) ont sondé le niveau d'appréciation des étudiants à l'aide d'un questionnaire anonyme pour juger, sur une échelle de Libert à quatre niveaux (de parfaitement en désaccord à parfaitement d'accord), l'authenticité des situations,

3. Dispositifs permettant aux étudiant de voter de manière synchrone et directe pendant un cours. Il peut s'agir d'appareils spécialisés ou simplement d'appareils génériques (ex. : tablettes, portables, téléphones) capables de se connecter sur le web.

la pertinence des items et l'intérêt envers le TJPEA pour favoriser le développement de la compétence en évaluation des apprentissages. Les étudiants avaient aussi la possibilité d'ajouter des commentaires explicatifs pour documenter leur appréciation du TJPEA. Les résultats montrent que les étudiants ont jugé réalistes les items ( $n = 62$ ; 86%) dans un contexte réel d'enseignement. En ce qui concerne la pertinence des items, les résultats sont plus mitigés. Ainsi, 48 étudiants (66%) trouvaient les items pertinents, alors que 24 (33%) ont indiqué que les items ne leur ont pas permis de mieux comprendre les concepts associés à l'évaluation des apprentissages. Finalement, la majorité des étudiants ont trouvé l'utilisation du TJPEA motivante pour approfondir les connaissances sur le thème d'intérêt ( $n = 57$ ; 79%). Bref, les résultats montrent que les étudiants jugent favorablement l'utilisation du dispositif de formation. Ils soulignent la pertinence du contenu et la variété des situations dans les vignettes pour susciter la réflexion sur des décisions pédagogiques courantes. Ils apprécient l'effort cognitif menant à la résolution de situations problématiques, notamment grâce à la période de discussion en classe à propos des enjeux décelés dans les situations du dispositif. Les commentaires recueillis témoignent aussi du souci des étudiants de mieux comprendre la situation présentée, de tenter de mieux la définir en plus d'essayer d'y apporter des solutions. À ce sujet, il est intéressant de constater que les échanges avec les étudiants ont mis en lumière des aspects du contrat d'évaluation proposé par Dionne et Chanelière (2022). En effet, les discussions ont permis, entre autres, d'identifier les informations volontairement manquantes, ambiguës ou incomplètes, caractéristiques du test ou de la formation par concordance. Autrement dit, les étudiants ont pu exprimer leurs incompréhensions ou, au contraire, exprimer des éléments contextuels absents de la vignette qui ont justifié leur réponse. Nous avons ainsi été en mesure de constater que l'instrument lui-même permettait de capter une partie importante de l'information, mais la phase qualitative s'est avérée névralgique pour colliger des éléments clés du jugement évaluatif des enseignants novices.

En cohérence avec la pratique socialement située du jugement professionnel en évaluation des apprentissages, les résultats de cette pratique pédagogique nous amènent à suggérer que la situation et les items des vignettes du TJPEA pourraient être un peu plus denses en contenu comparativement aux modèles que l'on trouve en éducation médicale. Le but est d'éviter une polarisation des avis des répondants (personnes expertes ou étudiants) dans des situations qui ne sont pas suffisamment informatives ou documentées



pour juger avec finesse de la pertinence ou de l'acceptabilité d'une intervention pédagogique, par exemple. Le processus de réponse aux items peut alors en être affecté. Autrement dit, les étudiants peuvent faire des choix de réponses hasardeux, ce qui n'illustre pas une incertitude décisionnelle, mais plutôt une incompréhension des concepts ou de l'orientation souhaitée du jugement évaluatif dans les situations du dispositif de formation.

## Conclusion

Cet article reposait sur la présentation d'un récit d'une pratique pédagogique basée sur l'utilisation d'un dispositif de formation ayant pour but de favoriser le développement du jugement évaluatif en enseignement. Le cadre de Cianciolo et Regehr (2019) a permis d'examiner les interactions entre les assises théoriques qui sous-tendent l'utilisation du dispositif de formation, les principes de sa conception ainsi que les facteurs contextuels influençant son adaptation pour répondre aux intentions pédagogiques ciblées. Grâce à un tel exercice, nous avons été en mesure d'étayer nos réflexions sur l'attrait du dispositif pour exercer le jugement professionnel en évaluation des apprentissages. À la conception du dispositif, nous avons observé la présence incontournable d'éléments contextuels qui influencent le jugement évaluatif, ce qui teinte la rédaction et la densité du contenu des situations et des items des vignettes. En outre, l'utilisation du dispositif nous a permis de constater sa pertinence pour la formation au jugement évaluatif des futurs enseignants, notamment par des moments de réflexion individuelle ou de groupe (Dionne & Simões Forte, 2013).

### *Limites et perspectives*

Une limite de ce récit de pratique pédagogique est liée à sa dimension exploratoire. Cette limite fait naître l'intérêt de dépasser les contextes présentés au regard de l'utilisation du dispositif de formation et d'explorer davantage son utilisation en recherche et en enseignement. À ce titre, Smith (2017) a mené une étude exploratoire afin de comprendre le processus de développement du jugement d'enseignants et de futurs enseignants en utilisant le TJPEA. Des entrevues individuelles semi-dirigées ont été menées en utilisant la technique de la pensée à voix haute (*think aloud*). Il s'agit d'une méthode de collecte des données verbales auprès d'une personne au moment où elle réalise une démarche cognitive complexe (Ericsson & Simon, 1980; Newell & Simon, 1972; Van Someren et al., 1994). Dans l'étude de Smith (2017), les participants (n = 12) énonçaient à voix haute les concepts clés sollicités dans les situations du TJPEA. Ils étaient aussi invités à juger la pertinence des hypothèses d'interventions pédagogiques

proposées et la probabilité qu'ils réalisent de telles interventions en contexte écologique. La question relative à la probabilité visait à déceler les signes de désirabilité sociale chez les participants, résultant de leur volonté à se montrer sous un jour favorable (Crowne & Marlowe, 1960). Les 12 participants étaient catégorisés selon trois profils : 1) des étudiants novices, 2) des enseignants praticiens d'expérience et 3) des enseignants experts en évaluation des apprentissages. Les résultats de l'étude de Smith (2017) montrent que les praticiens et les experts sont davantage en mesure de nommer les concepts clés en évaluation des apprentissages que les novices. Ils repèrent aussi plus fréquemment les données manquantes dans les vignettes qui seraient, par ailleurs, utiles pour aiguiller leur jugement professionnel. Smith (2017) recommande d'explorer les concepts théoriques et éthiques sollicités à l'aide du dispositif afin de modéliser les processus cognitifs des étudiants. L'utilisation du dispositif en cours de formation pourrait aussi aider à guider les pratiques évaluatives des étudiants et, ainsi, soutenir le développement de leur jugement professionnel. En outre, d'autres expérimentations doivent être menées pour sonder, à plus ou moins long terme, les avantages de l'utilisation du dispositif de formation pour le développement réel du jugement professionnel en évaluation des apprentissages.

D'autre part, l'une des limites dans l'utilisation du dispositif est liée au choix des personnes expertes, partenaires dans la conception du dispositif de formation. Pour les recruter, il est recommandé de faire appel à des personnes crédibles et susceptibles de juger pertinemment les situations et les hypothèses suggérées dans les vignettes du dispositif de formation (Dory et al., 2012 ; Lubarsky et al., 2011 ; Lubarsky et al., 2013). Or, les critères pour déterminer les attributs ou les caractéristiques des personnes expertes demeurent ambigus ou peu documentés. Quels devraient être les critères pour préciser l'expertise d'un enseignant ? La précise-t-on à partir de son expérience ? De ses titres universitaires ? De la reconnaissance de sa crédibilité vis-à-vis ses pairs ? En effet, en éducation, il est difficile d'identifier aisément les traits distinctifs des personnes expertes, notamment parce que les situations pédagogiques sont caractérisées par leur nature écologique.

L'instrumentation liée au développement du jugement évaluatif est plutôt rare en formation à l'enseignement, ce qui montre le caractère novateur du dispositif de formation créé. Compte tenu des perspectives d'utilisation du dispositif, force est de constater que les possibilités sont vastes et prometteuses. Dans un premier temps, il serait judicieux de créer une formation en ligne à l'aide du TJPEA en intégrant les avis de personnes expertes en matière

d'évaluation des apprentissages. Jusqu'à maintenant, le TJPEA a servi pour sonder le niveau d'appréciation des étudiants ou leurs processus cognitifs sollicités. Concrètement, des personnes expertes pourraient répondre aux items du TJPEA et les commenter pour créer les rétroactions formatives aux étudiants et ainsi créer un système de formation en ligne. À la manière des pratiques en éducation médicale, les étudiants pourraient bénéficier de rétroactions automatisées, qui seraient composées des réponses des personnes expertes et de leurs commentaires explicatifs. Ces commentaires pourraient s'avérer très riches pour les apprentissages des étudiants, notamment par les nuances et par les subtilités documentées dans le processus de jugement (Charlin et al., 2018; Charlin et al., 2021; Fernandez et al., 2023). Enfin, à ces rétroactions s'ajoutent d'autres ressources clés à consulter en matière d'évaluation des apprentissages (exigences institutionnelles formalisées, ouvrages de références clés, etc.). Cette démarche a pour but de guider les apprentissages des étudiants et de faire une synthèse des intentions pédagogiques sous-jacentes à la résolution des situations. Le système de formation ainsi créé pourrait permettre aux étudiants de réfléchir sur diverses problématiques, d'obtenir une rétroaction formative sur des questions complexes ou incertaines auxquelles ils seront confrontés dans leur pratique professionnelle, de comprendre que toutes les solutions ne sont pas univoques dans la pratique et, ainsi, de mieux appréhender l'incertitude de certaines décisions pédagogiques.

En dépit de ces limites, l'analyse stratifiée de la pratique pédagogique réalisée dans ce texte nous a permis d'identifier les repères théoriques, les principes de conception du dispositif et ses adaptations en contexte éducatif. En outre, cette analyse nous a aidés à affiner la conception du dispositif de formation pour faire en sorte qu'il s'aligne avec les principes du jugement évaluatif. L'identification des difficultés rencontrées lors de la rédaction des situations et des items du dispositif en contexte éducatif renforce la nature complexe du jugement professionnel en enseignement.

Révision linguistique : Marie-Claire Legaré

Mise en page : Emmanuel Gagnon

Résumé en portugais : Eusébio André Machado

Réception : 23 mars 2022

Version finale : 17 janvier 2024

Acceptation : 24 janvier 2024

## LISTE DES RÉFÉRENCES

- Abelson, R. P. (1975). Concepts for representing mundane reality in plans. Dans D. G. Bobrow et A. Collins (dir.), *Representation and understanding: Studies in cognitive science* (p. 273-309). Academic Press.
- Ashford-Rowe, K., Herrington, J. & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205-222. <https://doi.org/10.1080/02602938.2013.819566>
- Baribeau, A. (2015). Analyse des pratiques d'évaluation d'enseignants du secondaire IV et V dans des décisions sommatives de certification des apprentissages des élèves [Thèse de doctorat, Université du Québec à Trois-Rivières et Université du Québec à Montréal].
- Baribeau, A. (2020). La professionnalité de l'agir évaluatif de l'enseignant du secondaire dans le contexte québécois. *Administration et Éducation*, 165, 233-239. <https://doi.org/10.3917/admed.165.0233>
- Bélair, L. (2014). Évaluer ce qu'ils ont appris. Dans L. Ménard et L. St-Pierre (dir.), *Se former à la pédagogie de l'enseignement supérieur* (p. 355-380). Association québécoise de pédagogie collégiale (AQPC).
- Boelens, R., De Wever, B. & Voet, M. (2017). Four key challenges to the design of blended learning: A systematic literature review. *Educational Research Review*, 22, 1-18. <https://doi.org/10.1016/j.edurev.2017.06.001>
- Campbell, T. (2015). Stereotyped at Seven? Biases in Teacher Judgement of Pupils' Ability and Attainment. *Journal of Social Policy*, 44(3), 517-547. <https://doi.org/10.1017/S0047279415000227>
- Charlin, B., Deschênes, M.-F., Dumas, J.-P., Lecours, J., Vincent, A.-M., Kassis, J., Guertin, L., Gagnon, R., Robert, D., Foucault, A., Lubarsky, S. & Fernandez, N. (2018). Concevoir une formation par concordance pour développer le raisonnement professionnel: quelles étapes faut-il parcourir? *Pédagogie Médicale*, 19, 143-149. <https://doi.org/10.1051/pmed/2019019>
- Charlin, B., Deschênes, M.-F. & Fernandez, N. (2021). Learning by concordance (LbC) to develop professional reasoning skills: AMEE Guide No. 141. *Medical Teacher*, 614-621. <https://doi.org/10.1080/0142159X.2021.1900554>
- Charlin, B., Tardif, J. & Boshuizen, H. P. A. (2000). Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. *Academic Medicine*, 75(2), 182-190. <https://doi.org/10.1097/00001888-200002000-00020>
- Chaumont, M. & Leroux, J. L. (2018). Le jugement évaluatif: subjectivité, biais cognitifs et postures du professeur. *Pédagogie collégiale*, 31(3), 27-33.
- Cianciolo, A. T. & Regehr, G. (2019). Learning Theory and Educational Intervention: Producing Meaningful Evidence of Impact Through Layered Analysis. *Academic Medicine*, 94(6). [https://journals.lww.com/academicmedicine/Fulltext/2019/06000/Learning\\_Theory\\_and\\_Educational\\_Intervention\\_.31.aspx](https://journals.lww.com/academicmedicine/Fulltext/2019/06000/Learning_Theory_and_Educational_Intervention_.31.aspx)
- Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of consulting psychology*, 24(4), 349-354.

- Custers, E. J. (2015). Thirty years of illness scripts: Theoretical origins and practical applications. *Medical Teacher*, 37(5), 457-462. <https://doi.org/10.3109/0142159X.2014.956052>
- Deschênes, M.-F., Tremblay, K., Pelletier, I., Charlin, B. & Fernandez, N. (2022). Expérience d'enseignants à la réalisation d'un dispositif numérique de formation basée sur la concordance de jugement. *Formation et Profession*, 30(2), 1-13. <https://doi.org/10.18162/fp.2022.704>
- Dionne, E. & Chanelière, M. (2022). Du contrat didactique au contrat évaluatif : une grille de lecture pour mieux appréhender l'évaluation des apprentissages en pédagogie médicale. *Pédagogie médicale*, 23(1), 85-92. <https://doi.org/10.1051/pmed/2021030>
- Dionne, É. et Simões Forte, L. A. (2013). Le test de jugement professionnel en évaluation des apprentissages (TJPEA) : un outil visant à mesurer des compétences avec des télévotants. Dans *Actes du Colloque scientifique international sur les TIC en éducation: bilan, enjeux actuels et perspectives futures*, Montréal.
- Dory, V., Gagnon, R., Vanpee, D. & Charlin, B. (2012). How to construct and implement script concordance tests: insights from a systematic review. *Medical Education*, 46(6), 552-563. <https://doi.org/10.1111/j.1365-2923.2011.04211.x>
- Duget, A. (2014). Les pratiques pédagogiques en première année universitaire : description et analyse de leurs implications sur la scolarité des étudiants [Thèse de doctorat, Université de Bourgogne].
- Ericsson, K. A. & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87, 215-251. <https://doi.org/10.1037/0033-295x.87.3.215>
- Fernandez, N., Deschênes, M.-F., Akremi, H., Lecours, L., Jobin, V. & Charlin, B. (2023). What can Designing Learning-by-Concordance Clinical Reasoning Cases Teach Us about Instruction in the Health Sciences? *Perspectives on Medical Education*, 12(1), 160-168. <https://doi.org/10.5334/pme.898>
- Fives, H. & Barnes, N. (2020). Navigating the complex cognitive task of classroom assessment. *Teaching and Teacher Education*, 92, 103063. <https://doi.org/10.1016/j.tate.2020.103063>
- Gouvernement de l'Ontario. (2010). *Faire croître le succès : évaluation et communication du rendement des élèves fréquentant les écoles de l'Ontario*. Ministère de l'Éducation de l'Ontario (MEO). <http://www.edu.gov.on.ca/fre/policyfunding/growSuccessfr.pdf>
- Jabůrek, M., Cígler, H., Valešová, T. & Portešová, Š. (2022). What is the basis of teacher judgment of student cognitive abilities and academic achievement and what affects its accuracy? *Contemporary Educational Psychology*, 69, 102068. <https://doi.org/10.1016/j.cedpsych.2022.102068>
- Jonassen, D. H. (2011). *Learning to solve problems: A handbook for designing problem-solving learning environments*. Routledge.
- Laveault, D. (2005). Le jugement professionnel de l'enseignant : quel impact sur l'acte d'évaluer ? *Mesure et évaluation en éducation*, 28(2), 93-114. <https://doi.org/10.7202/1087226ar>
- Laveault, D. (2008). Le jugement professionnel : foyer de tensions et de synergies nouvelles en évaluation scolaire. *Revue suisse des sciences de l'éducation*, 30(3), 483-500. <https://doi.org/10.25656/01:4230>
- Leroux, J. (2019). Les multiples facettes du travail des formateurs de stagiaires : quels enjeux et quels défis? *Phronesis*, 8(1-2), 111-126. <https://doi.org/10.7202/1066588ar>

- Leroux, J. & Bélair, L. (2015). Exercer son jugement professionnel en enseignement supérieur. Dans J. Leroux (dir.), *Évaluer les compétences au collégial et à l'université : un guide pratique* (p. 67-104). Association québécoise de pédagogie collégiale (AQPC).
- Loye, N. & Fontaine, S. (2018). S'instrumenter pour évaluer. *Pédagogie médicale*, 19(2), 95-107. <https://doi.org/10.1051/pmed/2019018>
- Lubarsky, S., Charlin, B., Cook, D., Chalk, C. & Van der Vleuten, C. (2011). Script concordance testing: a review of published validity evidence. *Medical Education*, 45(4), 328-338. <https://doi.org/10.1111/j.1365-2923.2010.03863.x>
- Lubarsky, S., Dory, V., Duggan, P., Gagnon, R. & Charlin, B. (2013). Script concordance testing: from theory to practice: AMEE guide no. 75. *Medical Teacher*, 35(3), 184-193. <https://doi.org/10.3109/0142159X.2013.760036>
- Maes, O., Colognesi, S. & Van Nieuwenhoven, C. (2019). Le processus de construction du jugement évaluatif par les superviseurs de stage en enseignement. *Mesure et évaluation en éducation*, 42(1), 35-61. <https://doi.org/10.7202/1066597ar>
- Maes, O., Van Nieuwenhoven, C. & Colognesi, S. (2020). La dynamique de construction du jugement évaluatif du superviseur lors de l'évaluation de stages en enseignement. *Canadian Journal of Education*, 43(2), 522-547. <https://www.jstor.org/stable/26954697>
- Mayer, R. E. & Wittrock, M. C. (2006). Problem solving. Dans P. A. Alexander & P. H. Winne (dir.), *Handbook of educational psychology* (2<sup>e</sup> éd., p. 287-303). Lawrence Erlbaum.
- Merle, P. (2012). L'évaluation des élèves. Une modélisation interactionniste des pratiques professorales. Dans L. M. Lopez & G. Figari (dir.), *Modélisations de l'évaluation en éducation : questionnements épistémologiques* (p. 81-96). De Boeck Supérieur.
- Merle, P. (2018). Normes et biais généraux d'évaluation. Dans P. Merle (dir.), *Les pratiques d'évaluation scolaire : historique, difficultés, perspective*. Presses Universitaires de France.
- Ministère de l'Éducation du Québec. (2003). Politique d'évaluation des apprentissages : Formation générale des jeunes, formation générale des adultes, formation professionnelle. Ministère de l'Éducation du Québec. [https://www.education.gouv.qc.ca/fileadmin/site\\_web/documents/dpse/evaluation/13-4602.pdf](https://www.education.gouv.qc.ca/fileadmin/site_web/documents/dpse/evaluation/13-4602.pdf)
- Mottier Lopez, L. & Allal, L. (2008). Le jugement professionnel en évaluation : un acte cognitif et une pratique sociale située. *Revue suisse des sciences de l'éducation*, 30(3), 465-482. <https://doi.org/10.24452/sjer.30.3.4798>
- Newell, A. & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall
- Piot, T. (2008). La construction des compétences pour enseigner. *McGill Journal of Education/Revue des sciences de l'éducation de McGill*, 43(2), 95-110. <https://doi.org/10.7202/019577ar>
- Sacré, M., Lafontaine, D. & Toczek, M.-C. (2019). Liens entre les composantes des dispositifs d'enseignement hybride et les performances des étudiants de l'enseignement supérieur : une revue systématique. *Mesure et évaluation en éducation*, 42(3), 109-152. <https://doi.org/10.7202/1074105ar>
- Scallon, G. (2015). Des savoirs aux compétences : exploration en évaluation des apprentissages. De Boeck Supérieur.
- Schank, R. & Abelson, R. (1977). *Script, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Lawrence Erlbaum Associates.

- Schmidt, H. G., Norman, G. R. & Boshuizen, H. P. (1990). A cognitive perspective on medical expertise: theory and implication. *Academic Medicine*, 65(10), 611-621.
- Seden, K. & Svaricek, R. (2018). Teacher subjectivity regarding assessment: Exploring English as a foreign language teachers' conceptions of assessment theories that influence student learning. *CEPS Journal*, 8(3), 119-139. <https://doi.org/10.25656/01:16008>
- Smith, J. (2017). Application d'un test de concordance : étude exploratoire du développement de la compétence à évaluer les élèves chez les enseignants [Mémoire de maîtrise, Université d'Ottawa]. <http://hdl.handle.net/10393/36510>
- Tardif, J. (2006). L'évaluation des compétences : documenter le parcours de développement. Chenelière éducation.
- Tardif, J. (2017). Des repères conceptuels à propos de la notion de compétence, de son développement et de son évaluation. Dans P. J. Marianne, Tardif & G. François (dir.), *Organiser la formation à partir des compétences. Un pari gagnant pour l'apprentissage dans le supérieur* (p. 15-37). De Boeck Supérieur.
- Tourmen, C. (2014). Contributions des sciences de l'éducation à la compréhension de la pratique évaluative. *Politiques et management public*, 31(1), 69-85. <http://journals.openedition.org/pmp/6970>
- Tourmen, C. (2015). L'évaluation des compétences professionnelles : apports croisés de la littérature en évaluation, en éducation et en psychologie du travail. *Mesure et évaluation en éducation*, 38(2), 111-144. <https://doi.org/10.7202/1036765ar>
- Van Someren, M., Barnard, Y. & Sandberg, J. (1994). Studying the content of cognitive processes. Dans M. Van Someren, Y. Barnard & J. Sandberg (dir.), *The Think Aloud Method: A practical guide to modelling cognitive processes* (p. 13-26). Academic Press.
- Vanlommel, K., Van Gasse, R., Vanhoof, J. & Petegem, P. V. (2018). Teachers' high-stakes decision making. How teaching approaches affect rational and intuitive data collection. *Teaching and Teacher Education*, 71, 108-119. <https://doi.org/10.1016/j.tate.2017.12.011>
- Voss, J. F. (1988). Problem solving and reasoning in ill-structured domains. Dans *Analysing everyday explanation: A casebook of methods*. (p. 74-93). SAGE Publications.
- Wiggins, G. (1993). Assessment: Authenticity, Context, and Validity. *The Phi Delta Kappan*, 75(3), 200-214. <http://www.jstor.org/stable/20405066>
- Wiggins, G. (2011). A True Test: Toward More Authentic and Equitable Assessment. *The Phi Delta Kappan*, 92(7), 81-93. <https://doi.org/10.1177/003172171109200721>
- Xun, G. E. & Land, S. M. (2004). A conceptual framework for scaffolding ill-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research and Development*, 52(2), 5-22. <https://doi.org/10.1007/BF02504836>