

JENNIFER SMITH

Application du test de concordance : étude exploratoire du développement de la compétence à évaluer les élèves chez les enseignants

Thèse présentée à la
Faculté des études supérieures et postdoctorales
de l'Université d'Ottawa en vue de l'obtention
du grade de Maîtrise ès arts en éducation – M.A. (Éd.)

Faculté d'éducation
Université d'Ottawa

© Jennifer Smith, Ottawa, Canada, 2017

Résumé

Les approches contemporaines de l'éducation doivent répondre à un nombre de tensions suscitées par les changements technologiques, sociaux et économiques. Les questionnements sur la vision de l'apprentissage qui en découlent ont justifié l'implantation des programmes par compétences. Ce changement de paradigme a impliqué des réajustements sur le plan de l'évaluation des apprentissages. À la lumière de ces réalités, il est impératif de se pencher sur les dispositifs de formation au jugement évaluatif et de les bonifier. Inspiré par les points de convergence entre le développement du jugement clinique des médecins et celui du jugement en évaluation, Dionne (2005) a développé un instrument intitulé le test de jugement professionnel en évaluation des apprentissages (TJPEA). Composé de situations authentiques, il vise à mesurer le niveau de jugement des étudiants à la formation des maîtres en contexte formatif.

Cette étude exploratoire qualitative interprétative a été menée auprès de douze participants catégorisés en trois profils différents : quatre étudiants de première année à la formation des maîtres au secondaire (novices), quatre enseignants d'expérience oeuvrant dans une école secondaire québécoise (praticiens) et quatre professeurs universitaires spécialisés en évaluation des apprentissages (experts). Les données recueillies via la passation orale d'un TJPEA ont permis de (1) contribuer à identifier les différents facteurs susceptibles d'influencer le jugement professionnel en évaluation des répondants, (2) de comprendre davantage la dynamique de progression du développement de la compétence à évaluer, notamment en identifiant les différents schèmes de pensée entre les profils de répondants et (3) ressortir les différents obstacles à l'exercice du jugement évaluatif des sujets.

Même si nous sommes conscients que les résultats obtenus à cette étude ne sont pas généralisables en raison de son échantillon limité, il est tout de moins possible d'en faire découler de nouvelles avenues de recherche. Cela nous permettra éventuellement de proposer des solutions à la relève enseignante et de favoriser le développement d'un jugement en évaluation juste et équitable pour les élèves.

Remerciements

« La gratitude est la mémoire du cœur. »

Lao Tsé

Ce mémoire n'aurait pas vu le jour sans la présence, l'accompagnement et les précieux encouragements de mon directeur, monsieur Eric Dionne. Sa rigueur et sa passion pour la recherche m'ont inspirée et m'ont permis d'aller de l'avant dans cette grande aventure. Merci d'avoir investi temps et énergie pour me guider. Je t'en serai éternellement reconnaissante.

J'adresse mes vifs remerciements à monsieur Michel Laurier et à madame Diana Koszycki pour leurs bons conseils et pour avoir accepté de former mon comité. Cela a été un privilège de pouvoir échanger avec vous.

Je tiens également à remercier de façon particulière mes collègues de l'Université d'Ottawa qui m'ont permis de disposer de conditions propices à l'accomplissement de ce projet. Je remercie particulièrement monsieur Jean A. Roy, madame Diane Bouchard-Lamothe, madame Carole Fleuret, Maud Mediell, Joël Thibault et Marianne Dancause pour leurs encouragements et leurs amitiés.

Je dédie cette thèse à ma mère et à ma famille qui m'ont encouragée à aller au bout de mes ambitions, ainsi qu'à mon mari Simon pour avoir été un soutien indéfectible de par sa patience et son amour.

Table des matières

REMERCIEMENTS	III
TABLE DES MATIÈRES	IV
LISTES DES TABLEAUX	IX
LISTE DES FIGURES	X

Chapitre 1 : Problématique

1.1 L'émergence de l'approche par compétences au Québec	1
1.2 L'évaluation des apprentissages à la lueur des changements	2
1.3 Le rôle du jugement professionnel au sein de l'approche par compétences	4
1.4 L'insuffisance de formation au développement du jugement en évaluation	5
1.5 L'adaptation d'un dispositif employé en pédagogie des sciences de la santé	7
1.6 La question générale de recherche	8

Chapitre 2 : Cadre conceptuel

2.1 L'ascension des compétences en sein des programmes scolaires québécois	9
2.2 La démarche évaluative en contexte scolaire	12
2.3 Le jugement en sein de la démarche évaluative	14
2.4 Qualité du jugement en évaluation des apprentissages	15
2.5 Insuffisance d'outils de formation au développement du jugement en évaluation	16
2.6 Similitudes entre le jugement en évaluation et le jugement clinique	17
2.7 L'adaptation d'un instrument issu de la pédagogie des sciences de la santé	20
2.7.1 Les composantes du test de concordance de script	20
2.7.2 Les composantes du TJPEA	23
2.8 La question spécifique de recherche	24

Chapitre 3 : Méthodologie

3.1 Type de recherche	26
3.2 Participants	27
3.3 Sélection des participants	27
3.4 Étapes de réalisation du TJPEA	28
3.4.1 Détermination de l'utilisation prévue du test et définition de ce qu'il va mesurer	29
3.4.2 Construction d'un tableau de spécifications et sélection des thèmes des vignettes	29
3.4.3 Le format des vignettes	31
3.4.4 Rédaction des vignettes composantes du TJPEA	32
3.5 Administration du TJPEA	33
3.6 Description du questionnaire d'entrevue	33
3.7 Validation du questionnaire d'entrevue	34
3.8 Déroulement de l'entrevue	35
3.9 Analyse des données	35
3.10 Codage	36

Chapitre 4 : Résultats et interprétations

4.1 Vignette 1

4.1.1 Authenticité de la vignette 1	40
4.1.2 Identification des concepts	40
4.1.3 Données manquantes à la vignette 1	42
4.1.3.1 Données manquantes par les novices	43
4.1.3.2 Données manquantes par les praticiens	45
4.1.3.3 Données manquantes par les experts	46
4.1.3.4 Synthèse comparative des données manquantes par profil de participants	47

4.1.4 Item 1 de la première vignette	49
4.1.5 Item 2 de la première vignette	52
4.1.6 Item 3 de la première vignette	54
4.1.7 Synthèse de la vignette 1	57
4.2 Vignette 2	
4.2.1 Authenticité de la vignette 2	58
4.2.2 Identification des concepts	58
4.2.3 Données manquantes à la vignette 2	59
4.2.3.1 Données manquantes par les novices	60
4.2.3.2 Données manquantes par les praticiens	61
4.2.3.3 Données manquantes par les experts	62
4.2.3.4 Synthèse comparative des données manquantes par profil de participants	63
4.2.4 Item 1 de la deuxième vignette	64
4.2.5 Item 2 de la deuxième vignette	68
4.2.6 Item 3 de la deuxième vignette	72
4.2.7 Synthèse de la vignette 2	74
4.3 Vignette 3	
4.3.1 Authenticité de la vignette 3	75
4.3.2 Identification des concepts	75
4.3.3 Données manquantes à la vignette 3	77
4.3.3.1 Données manquantes par les novices	77
4.3.3.2 Données manquantes par les praticiens	78
4.3.3.3 Données manquantes par les experts	79
4.3.3.4 Synthèse comparative des données manquantes par profil de participants	80
4.3.4 Item 1 de la troisième vignette	81

4.3.5 Item 2 de la troisième vignette	84
4.3.6 Item 3 de la troisième vignette	87
4.3.7 Synthèse de la vignette 3	90
4.4 Vignette 4	
4.4.1 Authenticité de la vignette 4	91
4.4.2 Identification des concepts	91
4.4.3 Données manquantes à la vignette 4	92
4.4.3.1 Données manquantes par les novices	93
4.4.3.2 Données manquantes par les praticiens	94
4.4.3.3 Données manquantes par les experts	95
4.4.3.4 Synthèse comparative des données manquantes par profil de participants	95
4.4.4 Item 1 de la quatrième vignette	97
4.4.5 Item 2 de la quatrième vignette	100
4.4.6 Item 3 de la quatrième vignette	103
4.4.7 Synthèse de la vignette 4	106
4.5 Synthèse des résultats	107
Conclusion	
5.1 Bilan du travail de recherche	110
5.2 Limites à l'étude	111
5.3 Pistes de recherches éventuelles	112
RÉFÉRENCES	114

ANNEXES	123
ANNEXE A : Courriel d'invitation aux participants	126
ANNEXE B : Formulaire de consentement	128
ANNEXE C : Liste des thèmes abordés	131
ANNEXE D : Listes des ouvrages utilisés à la conception du TJPEA	132
ANNEXE E : Vignettes composant le TJPEA	133
ANNEXE F : Questionnaire d'entrevue	137

Liste des tableaux

Tableau 1. <i>Composantes liées à la compétence en évaluation chez les futurs enseignants québécois (MEQ, 2001)</i>	6
Tableau 2. <i>Présentation des définitions du concept de la « compétence »</i>	10
Tableau 3. <i>La compétence : principales caractéristiques</i>	11
Tableau 4. <i>Tableau illustrant les étapes de la démarche en évaluation des apprentissages (MEQ, 2003)</i>	13
Tableau 5. <i>Les caractéristiques de la modélisation synthétique du jugement évaluatif (Allal et Mottier Lopez, 2008)</i>	15
Tableau 6. <i>Caractéristiques de l'échantillon selon les profils</i>	29
Tableau 7. <i>Grille d'analyse de données</i>	36
Tableau 8. <i>Liste des codes finaux</i>	37
Tableau 9. <i>Tableau comparatif entre les concepts anticipés par le concepteur de la vignette 1 et ceux relevés par les trois groupes de participants (experts, praticiens et novices)</i>	40
Tableau 10. <i>Liste des codes et des sous-codes utilisés en lien avec les données manquantes de la vignette 1 du TJPEA</i>	42
Tableau 11. <i>Tableau comparatif entre les concepts anticipés par le concepteur de la vignette 2 et ceux relevés par les trois groupes de participants (experts, praticiens et novices)</i>	58
Tableau 12. <i>Liste des codes et des sous-codes utilisés en lien avec les données manquantes de la vignette 2 du TJPEA</i>	60
Tableau 13. <i>Tableau comparatif entre les concepts anticipés par le concepteur de la vignette 3 et ceux relevés par les trois groupes de participants (experts, praticiens et novices)</i>	76
Tableau 14. <i>Liste des codes et des sous-codes utilisés en lien avec les données manquantes de la vignette 3 du TJPEA</i>	77
Tableau 15. <i>Tableau comparatif entre les concepts anticipés par le concepteur de la vignette 4 et ceux relevés par les trois groupes de participants (experts, praticiens et novices)</i>	92
Tableau 16. <i>Liste des codes et des sous-codes utilisés en lien avec les données manquantes de la vignette 4 du TJPEA</i>	93

Liste des figures

Figure 1. Exemple d'une vignette de test de concordance de script	21
Figure 2. Exemple d'une vignette pour une dimension d'interprétation des résultats	24
Figure 3. Descriptif de la compétence 5 du <i>Référentiel de la formation en enseignement</i>	30
Figure 4. Exemple d'une vignette de TJPEA	32
Figure 5. Cas présenté dans le cadre de la vignette 1 du TJPEA	39
Figure 6. Exemple d'une codification de concepts	41
Figure 7. Données manquantes identifiées par les novices (n=4)	43
Figure 8. Données manquantes identifiées par les praticiens (n=4)	45
Figure 9. Données manquantes identifiées par les experts (n=4)	46
Figure 10. Synthèse des données manquantes identifiées par tous les participants (n=12)	48
Figure 11. Répartition des réponses des profils de répondants (n=12) à l'échelle A	50
Figure 12. Répartition des réponses des répondants (n=12) à l'échelle B par profil	51
Figure 13. Répartition des réponses des profils de répondants (n=12) à l'échelle A	53
Figure 14. Répartition des réponses des profils de répondants (n=12) à l'échelle A	54
Figure 15. Répartition des réponses des répondants (n=12) à l'échelle B par profil	56
Figure 16. Cas présenté dans le cadre de la vignette 2 du TJPEA	57
Figure 17. Données manquantes identifiées par les novices (n=4)	60
Figure 18. Données manquantes identifiées par les praticiens (n=4)	61
Figure 19. Données manquantes identifiées par les experts (n=4)	62
Figure 20. Synthèse des données manquantes identifiées par tous les participants (n=12)	63
Figure 21. Répartition des réponses des profils de répondants (n=12) à l'échelle A	65
Figure 22. Répartition des réponses des répondants (n=12) à l'échelle B par profil	67
Figure 23. Répartition des réponses des profils de répondants (n=12) à l'échelle A	69
Figure 24. Répartition des réponses des répondants (n=12) à l'échelle B par profil	70
Figure 25. Répartition des réponses des profils de répondants (n=12) à l'échelle A	73
Figure 26. Répartition des réponses des répondants (n=12) à l'échelle B par profil	74
Figure 27. Cas présenté dans le cadre de la vignette 3 du TJPEA	75
Figure 28. Données manquantes identifiées par les novices (n=4)	78
Figure 29. Données manquantes identifiées par les praticiens (n=4)	79
Figure 30. Données manquantes identifiées par les experts (n=4)	79
Figure 31. Synthèse des données manquantes identifiées par tous les participants (n=12)	80
Figure 32. Répartition des réponses des profils de répondants (n=12) à l'échelle A	82
Figure 33. Répartition des réponses des répondants (n=12) à l'échelle B par profil	83
Figure 34. Répartition des réponses des profils de répondants (n=12) à l'échelle A	84
Figure 35. Répartition des réponses des répondants (n=12) à l'échelle B par profil	86
Figure 36. Répartition des réponses des profils de répondants (n=12) à l'échelle A	88
Figure 37. Répartition des réponses des répondants (n=12) à l'échelle B par profil	90

Figure 38. Cas présenté dans le cadre de la vignette 4 du TJPEA	91
Figure 39. Données manquantes identifiées par les novices (n=4)	94
Figure 40. Données manquantes identifiées par les praticiens (n=4)	94
Figure 41. Données manquantes identifiées par les experts (n=4)	95
Figure 42. Synthèse des données manquantes identifiées par tous les participants (n=12)	96
Figure 43. Répartition des réponses des profils de répondants (n=12) à l'échelle A	98
Figure 44. Répartition des réponses des répondants (n=12) à l'échelle B par profil	99
Figure 45. Répartition des réponses des profils de répondants (n=12) à l'échelle A	101
Figure 46. Répartition des réponses des répondants (n=12) à l'échelle B par profil	103
Figure 47. Répartition des réponses des profils de répondants (n=12) à l'échelle A	104
Figure 48. Répartition des réponses des répondants (n=12) à l'échelle B par profil	105

Chapitre 1

Problématique

« Le vrai génie réside dans l'aptitude à évaluer l'incertain, le hasardeux, les informations conflictuelles. »
Churchill

Pour comprendre le contexte dans lequel s'inscrit l'objet d'étude de ce mémoire de maîtrise, il convient d'aborder les différents enjeux liés à l'évaluation en contexte scolaire québécois. Dans un premier temps, nous dressons un court historique de l'avènement des programmes par compétences afin de mieux saisir l'influence sur les modalités d'évaluation des enseignants. Dans un deuxième temps, nous soulignons l'insuffisance de dispositifs de formation disponibles à la formation des maîtres favorables au développement de la compétence à évaluer les élèves. Dans un troisième temps, nous décrivons comment le test de concordance pourrait mériter davantage d'attention en étant utilisé en situation de développement au jugement professionnel en évaluation. Dans un dernier temps, nous concluons ce chapitre par la formulation de la question générale de recherche sur laquelle repose notre étude.

1.1 L'émergence de l'approche par compétences au Québec

En 1996, le Rapport Delors intitulé « *L'éducation : un trésor caché dedans* » avait incontestablement eu un impact tangible sur les politiques de l'éducation à l'échelle mondiale, en servant de cadre philosophique aux réformes de plusieurs pays autant en Europe qu'en Amérique. L'un des objectifs visés par le Rapport était de rendre les systèmes d'éducation « plus pertinents, souples et justes, fondés sur les quatre piliers suivants : apprendre à connaître, apprendre à faire, apprendre à être et apprendre à vivre ensemble » (Draxler, 2010, p. 27). Cette version renouvelée de l'apprentissage a mené de nombreux systèmes scolaires (Canada, Suisse, Australie, Grande-Bretagne, etc.) à apporter de profonds changements administratifs et pédagogiques en vue de mieux préparer les jeunes aux défis qu'amène le 21^e siècle (Dionne, 2014). Les approches contemporaines de l'éducation ont ainsi voulu faire face « à un monde en transformation rapide » et répondre à un nombre élevé de tensions suscitées par les changements technologiques, sociaux et économiques. Les programmes formulés par compétences représentent, à cet effet, des

manifestations tangibles de ces volontés pédagogiques et politiques qui dominent encore aujourd'hui le milieu scolaire québécois.

Les questionnements qui ont suivis cette période de l'histoire scolaire ont favorisé le passage d'une pédagogie axée sur l'acquisition des connaissances vers l'une favorisant plutôt leur traitement. Généralement associée au modèle traditionnel de référence behavioriste de la pédagogie et de l'évaluation, la notion d'objectif a longtemps été destinée à juger des habiletés et des connaissances des élèves en quantifiant les apprentissages réalisés à un moment précis de leur parcours (Tardif, 2006). Au début des années 2000, l'adoption de l'approche par compétences a mis en branle de nombreux changements au sein du système scolaire. Inspirée des courants cognitivistes, constructivistes et socioconstructivistes associant la conception de l'apprentissage centrée sur l'élève à la manière dont ce dernier construit activement son savoir, la notion de compétence a progressivement pris son envol (Legendre, 2004). Tardif (2006, p. 22) définit la compétence comme étant « un savoir-agir complexe prenant appui sur la mobilisation et la combinaison efficaces d'une variété de ressources internes et externes à l'intérieur d'une famille de situations ». Enfin, c'est à la même époque que le ministère de l'Éducation du Québec (MEQ) réitère l'importance de l'évaluation formative, c'est-à-dire de la fonction d'aide à l'apprentissage visant à soutenir la progression des élèves dans une démarche d'enseignement/apprentissage.

1.2 L'évaluation des apprentissages à la lueur des changements

Ce changement paradigmatique issu du monde de l'éducation a nécessairement impliqué des ajustements majeurs, et ce, particulièrement sur le plan de l'évaluation. En effet, le monde de l'évaluation des apprentissages était en pleine effervescence et le corps enseignant se devait de revoir ses pratiques à cette lumière. Scallon (2004, p. 105) explique que « la méthodologie de l'évaluation a changé », puisqu'elle se caractérise maintenant par des tâches complexes, contextualisées et signifiantes permettant de mettre en œuvre plusieurs dimensions d'une compétence visée (De Ketele, 2000; De Ketele et Gérard, 2005; Figari, 2001; Paquay, 2002; Scallon, 2004; Roegiers, 2000; Tardif, 2006; Wiggins, 1998). Pour continuer, l'approche évaluative ciblée préconise une collecte de données représentatives de l'évolution des apprentissages témoignant de la compétence de chaque individu. Plus précisément, cette conception de l'évaluation rejette le principe de décomposition proposant de mesurer les

composantes d'une compétence individuellement ainsi que la décontextualisation, soit d'évaluer hors contexte (De Ketele, 2000; Figari, 2001; Jonnaert, 2002; Paquay, 2002; Roegiers, 2000; Scallon, 2004; Tardif, 2006). Comme le suggèrent Laurier, Tousignant et Morissette (2005), l'on doit idéalement compter sur des situations qui s'apparentent à des situations réelles qui permettent de solliciter les compétences par le biais de tâches qui requièrent la mise en œuvre simultanée de plusieurs connaissances dont des habiletés, des techniques ou des attitudes. Cette constatation nous rappelle d'ailleurs les travaux de Wiggins (1993, p. 78) dans lesquels nous pouvions lire que « l'utilisation de tâches contextualisées devrait se traduire par des situations signifiantes qui engagent et motivent l'élève ». Bref, il ne suffit plus d'apprendre pour apprendre, mais d'apprendre en étant conscient de la pertinence de ces apprentissages en situation réelle.

Comme la collecte d'informations sur l'objet à évaluer est l'une des étapes implorantes de la démarche évaluative, nombre de chercheurs se sont questionnés sur l'instrumentation nécessaire à l'évaluation par compétences en contexte scolaire (Dionne, 2005, 2008; Rey, Carette, Defrance et Kahn, 2003; Scallon, 2004). En effet, il devenait de plus en plus évident que des instruments de collecte de données adaptés aux apprentissages variés et complexes se devaient d'être développés afin de pallier aux limites des outils traditionnels disponibles¹ (Laurier, Tousignant et Morissette, 2005). Certes, ces derniers assuraient une assimilation de procédures automatisables reproductibles peu importe la situation présentée, comme dans le cas d'applications de règles grammaticales ou de calculs mathématiques (Rey, 2012, p. 18). Par contre, l'évaluation des compétences exigeait maintenant à l'élève d'être confronté à une tâche authentique qui pouvait changer de nature selon les situations présentées (ex. adapter un écrit selon le destinataire). Il s'agissait, de ce fait, de maîtriser une démarche, mais aussi d'interpréter une tâche afin d'adapter des procédures au contexte (*idem*, p. 18).

Par conséquent, le MEQ (2003) a proposé de recourir à certains outils d'évaluation, notamment la grille d'observation, d'accompagnement ou d'appréciation, l'autoévaluation, le portfolio, les entrevues, les cartes conceptuelles, les listes de vérification, les échelles descriptives, etc. Les enseignants qui avaient l'habitude d'évaluer à l'aide d'exams, de tests ou de travaux

¹ Les formes d'évaluation dites « traditionnelles » visent à questionner dans le but de rappeler des faits, vérifier la maîtrise d'habiletés de base et la transmission de valeurs traditionnelles. Parmi les formats les plus utilisés, on compte les tests, les récitations, les travaux et les instructions des enseignants (Legendre, 2005).

comme le proposait alors la *Politique générale d'évaluation pédagogique* (1981) ont dû se familiariser avec de nouveaux instruments, et ce, afin de répondre aux exigences du *Programme de formation de l'école québécoise* (2001), de la *Politique d'évaluation des apprentissages* (2003) et, plus récemment, des cadres d'évaluation (MELS, 2011).

1.3 Le rôle du jugement professionnel au sein de l'approche par compétences

Le jugement professionnel n'est pas un concept nouveau; les enseignants interprètent des données témoignant des apprentissages des élèves depuis longtemps (Laveault, 2008). Par contre, comme le souligne Leroux (2009), les enseignants basent parfois leur jugement sur la maîtrise d'une compétence en se référant à des évaluations de type sommatif où il y a une addition de notes. Bélair (Bélair, dans Ménard et St-Pierre, 2014, p. 360) arrive à la même conclusion : « Encore aujourd'hui, les enseignants corrigent, notent et calculent les productions des élèves avant de transférer le résultat dans un tableau de compilation permettant de donner une note finale sans avoir recours à une interprétation des données ». Dans un tel contexte, quelle place occupe le jugement dans la démarche évaluative de l'enseignant? Celui-ci ne devrait-il pas apparaître en filigrane dans l'ensemble du processus d'évaluation comme le propose la *Politique d'évaluation des apprentissages* de 2003?

Cette pratique est clairement incompatible avec l'approche par compétences visant plutôt le développement d'une compétence par sa progression (Scallon, 2004). Cet état de réalisation des apprentissages doit se faire à la lumière de l'interprétation des diverses informations recueillies lors de la réalisation de tâches complexes (Leroux, 2016). Loin de l'exécution « mécanique », on dénote donc que l'exercice du jugement en évaluation engendre parfois la peur de la subjectivité chez le corps enseignant. C'est pourquoi on lui rappelle de prendre appui sur des observations de qualité et en nombre suffisant afin d'en découler des décisions qu'il arrivera à argumenter si cela s'avère nécessaire (Morrissette et Legendre, 2011).

À l'ordre du secondaire, un manquement à cet égard peut avoir des répercussions significatives sur le parcours scolaire et la carrière professionnelle de l'élève (Lemay, 2000; Laveault, 2008; Lecoindre, 1997; Merle, 1996). Sur le plan affectif, des conséquences comme la perte d'estime de soi, l'angoisse, la démotivation, l'abandon et le décrochage scolaire ont été, par ailleurs, répertoriées (Bélair, 1999; Black et William, 2006; Crooks, 1988; Hobbs, 1975). Malgré

des précautions de validité et de fidélité pour s'assurer de limiter la subjectivité, il faut reconnaître que le jugement reste toujours tributaire à certains obstacles. Comme l'affirme Lafortune (2006, p. 27) : « l'exercice et le développement du jugement professionnel supposent une acceptation de cette subjectivité, une conscience de ses croyances et de ses pratiques, un éclairage de la prise de décisions à l'aide d'outils pertinents et suffisants ainsi qu'une argumentation ». Jorro (2000), dans ses travaux, tient aussi compte de cette réalité en s'intéressant au rapport de l'évaluateur à l'évaluation. Pour l'illustrer, il propose quatre imaginaires: (1) performance, (2) maîtrise, (3) construction et (4) compréhension. Ces imaginaires regroupent aussi des ensembles de valeurs, croyances, goûts et convictions à l'origine des pratiques des enseignants et se manifestent au sein de leurs choix pédagogiques. Enfin, la littérature (Jorro, 2000; Scallon, 2004) répertorie l'influence de certains biais pouvant influencer la bonne prise de décisions: (1) la stéréotypie favorable ou défavorable à l'égard d'un élève, (2) l'effet de halo, (3) l'ordre des copies, (4) l'effet de contraste entre les productions, (5) la contamination ou l'influence des notes antérieures et (6) la posture de l'évaluateur au regard de son niveau de sévérité. En somme, l'exercice du jugement en contexte de l'évaluation des compétences est une entreprise complexe sensible à de nombreuses menaces. Il convient donc aux milieux scientifiques et pratiques de développer des dispositifs d'évaluation qui limitent ces effets indésirables.

1.4 L'insuffisance de formation au développement du jugement en évaluation

Dans une visée de professionnalisation des enseignants, un référentiel de compétences professionnelles destiné aux enseignants et à leur relève a été élaboré (MEQ, 2001). Les compétences le constituant sont formulées autour de quatre axes principaux : les fondements, l'acte d'enseigner, le contexte social et scolaire et l'identité professionnelle. En ce qui a trait à la professionnalité à évaluer des enseignants, cela se traduit par le développement de la compétence 5, soit celle « d'évaluer la progression des apprentissages et le degré d'acquisition des compétences des élèves pour les contenus à apprendre » (MEQ, 2001, p. 96).

À l'issue de leur formation de quatre ans, il est ainsi attendu des étudiants inscrits à la formation des maîtres qu'ils aient développé durant leurs stages et leurs cours universitaires ceci :

Tableau 1

Composantes liées à la compétence en évaluation chez les futurs enseignants québécois (MEQ, 2001, p.96)

Au terme de la formation initiale, l'étudiante ou l'étudiant doit être en mesure :
<ul style="list-style-type: none">• de détecter, en situation d'apprentissage, les forces et les difficultés des élèves;• de préciser, de façon autonome, des correctifs à apporter à son enseignement;• de contribuer avec ses pairs à la préparation du matériel d'évaluation, à l'interprétation des productions des élèves au regard du développement des compétences et à l'élaboration d'outils de communication destinés aux parents;• de communiquer aux élèves les résultats d'un processus d'évaluation diagnostique et d'indiquer aux parents et aux membres de l'équipe pédagogique les éléments des stratégies d'intervention corrective envisagés.

Inévitablement, bien former les enseignants au développement de la compétence à évaluer dans le cadre de la formation des maîtres est essentiel afin de les préparer pour l'avenir. Mais, comment y parvenir? À ce sujet, Vanhulle (2009, p. 166) explique « que la formation des maîtres oscille entre la logique académique s'intéressant à la construction des savoirs référentiels scientifiques préparant le futur enseignant à penser à ses pratiques professionnelles à partir de modèles d'intelligibilité et la logique des compétences pouvant se manifester qu'en contexte de travail et donnant lieu aux savoirs expérientiels ». Au regard de cette affirmation, nous constatons actuellement que la plupart des étudiants formés dans les universités du Québec offrant un programme à l'enseignement ne suivent qu'un seul cours de trois crédits en évaluation des apprentissages (Fontaine et al., 2013). Au cours de la formation initiale d'une durée totale de quatre années, quatre stages en milieu scolaire sont également exigés. Malgré ces efforts curriculaires, très peu de ponts entre la pratique (stages) et la théorie (cours universitaires) permettent aux étudiants en cours de formation de développer leur jugement professionnel. La littérature en lien avec l'insertion professionnelle des jeunes enseignants (Cattonar, 2008; Hudson, Beutel et Hudson,

2008; Mukamurera, 2011) relève entre autres le choc de la réalité, le sentiment d'incompétence et l'inconfort psychologique vécu lors de la période de transition entre le terrain et la formation universitaire. Malgré l'urgence de la situation soulignée par Nizet en 2014, à ce jour, peu d'études se sont intéressées au sujet. Or, le défi n'est pas si simple. La rareté des outils disponibles permettant d'exprimer le jugement professionnel de l'enseignant pour en reconnaître le caractère complexe et non mécanique en est qu'un exemple (Mottier Lopez et Allal, 2010).

1.5 L'adaptation d'un dispositif employé en pédagogie des sciences de la santé

À la lumière de la problématique que nous venons de décrire, nous croyons qu'en s'inspirant de d'autres disciplines où le jugement professionnel est développé tout au long de la formation, les sciences de l'éducation pourraient bonifier le développement du jugement évaluatif par l'entremise d'outils formateurs mieux adaptés. Pour y arriver, il devient impératif de se pencher sur les dispositifs de formation au jugement afin d'en avoir une meilleure compréhension et ultimement d'arriver à les améliorer. De nombreux points de convergence ont été identifiés entre la profession médicale et celle du pédagogue, plus précisément entre le développement du jugement clinique et celui du jugement en évaluation. L'article de Jouquan, Parent et Audétat, publié en 2013, fait référence aux finalités centrées sur le développement de l'autonomie de la personne, à la nature diagnostique de l'approche et de leurs objectifs ainsi qu'au parallélisme possible entre l'approche éducationnelle centrée sur l'apprenant et l'approche en santé centrée sur le patient. L'éducation médicale fait appel, de plus en plus, à un instrument appelé le test de concordance de script (TCS) pour développer le processus de raisonnement clinique des futurs médecins. Les principes inhérents au TCS sont issus de la psychologie cognitive et consistent à évaluer le jugement dans une situation complexe, incertaine et authentique (Charlin, Gagnon, Kazi-Tani et Thivierge, 2005). Plus précisément, le TCS est composé de différentes vignettes cliniques comportant des informations ambiguës, incomplètes ou superflues. Ces vignettes se caractérisent également par leur authenticité au sens où elles sont évocatrices de situations réelles vécues par les enseignants du secondaire. Une série de questions est associée à chacun des scénarios proposés. L'objectif de cet outil d'évaluation, principalement utilisé en contexte d'évaluation formative, est de comparer la façon dont les étudiants organisent leurs connaissances (scripts) à celles des experts du domaine (Charlin, Gagnon, Sibert et Van der Vleuten, 2002). Selon les études, le TCS permettrait d'évaluer fidèlement le niveau de raisonnement clinique en vue de

classer les répondants selon un niveau d'expertise (Lambert, 2005). Tout comme l'enseignant dans sa salle de classe, « le médecin doit ainsi tenir compte de procédures et de règles imposées par l'institution sans que celles-ci nuisent à l'autonomie de son jugement » (Allal et Lafortune, 2008, p. 3).

C'est en se basant sur cette idée de similitudes existantes entre le jugement clinique et le jugement en évaluation des apprentissages que Dionne a développé un format d'outil visant le développement de la compétence à évaluer chez les futurs enseignants. En 2012, il a mené une première expérimentation du test de jugement professionnel en évaluation des apprentissages (TJPEA). L'instrument en question visait alors à recueillir des données sur la façon dont les étudiants (n=72) composaient avec l'incertitude et arrivaient à résoudre des problèmes mal définis en lien avec des dilemmes pédagogiques en évaluation. Les résultats positifs présentés nous poussent à croire qu'il pourrait être pertinent de faire une expérimentation plus complète du TJPEA dans le cadre de cette recherche. Cette étude pourrait également répondre au besoin exprimé par maints chercheurs (Baker, 1997; Hogan et Murphy, 2007; Scallon, 2004; Solano-Flores, Shavelson et Schneider, 2001), soit celui de concevoir des situations d'évaluation authentiques visant à mesurer des performances représentatives de construits complexes comme celui des compétences.

1.6 La question générale de recherche

En somme, il existe un besoin d'explorer le jugement en évaluation en considérant sa nature multidimensionnelle et longitudinale de développement. En d'autres mots, il s'agit de mieux comprendre un apprentissage complexe. C'est précisément dans un contexte d'exploration que nous souhaitons arriver à préciser les caractéristiques du jugement en évaluation des enseignants à partir de l'instrument du TJPEA. Au final, les vignettes présentées par le TJPEA permettent-elles de rendre compte du jugement en évaluation des enseignants ? Il s'agit de la question générale de recherche dont cette étude fait l'objet.

Chapitre 2

Cadre conceptuel

*« Quand on lit pour s'instruire, on voit tout ce qui nous a échappé lorsqu'on ne lisait qu'avec les yeux. »
Voltaire*

Le chapitre précédent nous a permis de présenter la problématique entourant la complexité à développer le jugement professionnel en évaluation dans le cadre de la formation des maîtres et d'articuler notre question générale de recherche. Le présent chapitre vise à exposer les concepts sur lesquels nous comptons nous appuyer pour mener à bien cette étude, ainsi qu'à tisser les liens entre la littérature scientifique issue de deux disciplines : (1) l'évaluation des apprentissages et (2) la pédagogie des sciences de la santé. Tout d'abord, les concepts suivants seront abordés : la démarche évaluative en contexte des programmes par compétences, la place du jugement en évaluation dans le cadre des pratiques enseignantes, les similitudes existantes entre le jugement en évaluation et le jugement clinique ainsi que les composantes instrumentales du TJPEA. Enfin, en cohérence avec la question générale, la présentation de la question spécifique clora ce chapitre.

2.1 L'ascension des compétences en sein des programmes scolaires québécois

L'émergence du concept de la « compétence » dans le contexte scolaire québécois s'inscrit dans la volonté « de doter l'élève de savoirs vivants susceptibles d'être mobilisés pour la résolution de problèmes dans le monde qui l'entoure » (Romainville, 1996, p. 137). Au fil des ans, plusieurs auteurs ont proposé des définitions à cette base théorique des nouveaux programmes d'enseignement. Malgré quelques nuances sur le plan sémantique, une certaine convergence semble se dégager d'entre elles. Le tableau suivant en présente quelques-unes.

Tableau 2. *Présentation des définitions du concept de la « compétence »*

Le Boterf (1994, p. 21)	Jonnaert (2002, p. 38)	MELS (2006, p. 7)	Roegiers (2000, p. 66)	Tardif (2006, p. 22)
« Capacité d'intégrer des savoirs divers et hétérogènes pour les finaliser sur une réalisation d'activités. »	« Mise en œuvre, par une personne particulière ou par un groupe, de savoirs, de savoir-être, de savoir-faire ou de savoir-devenir dans une situation donnée. »	« Savoir-agir fondé sur la mobilisation et l'utilisation efficaces d'un ensemble de ressources. »	« Possibilité, pour un individu, de mobiliser de manière intériorisée un ensemble intégré de ressources en vue de résoudre une famille de situations-problèmes. »	« Savoir-agir complexe prenant appui sur la mobilisation et la combinaison efficaces d'une variété de ressources internes et externes à l'intérieur d'une famille de situations. »

À la lecture de ces dernières, certaines similitudes peuvent être soulignées. Plus spécifiquement, toutes les définitions présentées mettent en évidence la capacité de l'élève à mettre en action ses connaissances, ses habiletés et ses aptitudes dans une situation précise. Toutefois, on peut constater que l'appellation « capacité » présentée par Le Boterf (1994) a rapidement été mise de côté et remplacée par l'idée de « mise en œuvre » et de « mobilisation » préconisée par les quatre autres auteurs. En effet, le développement de la compétence passe par une « mobilisation de multiples ressources dont les savoirs théoriques, professionnels et expérientiels » (Perrenoud, 1998, p. 17). On distingue l'ajout caractéristique du processus intériorisé des définitions formulées et du concept de « famille de situations » par Roegiers (2000) et Tardif (2006). Roegiers (2000, p. 130) définit une famille de situations comme « un ensemble de situations proches l'une de l'autre ». Ce concept est important, car il permet de circonscrire le type de situations auquel seront confrontés les élèves.

Dans le même ordre d'idées, Jonnaert explique que « l'organisation en réseau des ressources mobilisées par l'élève est importante, car il ne s'agit pas d'une somme qu'il empilerait, mais bien d'une articulation de ressources pertinentes articulées entre elles permettant un traitement efficace d'une situation complexe » (2002, p. 36). L'enseignant place donc l'élève face à une tâche qui l'amène à « choisir la voie » qui lui permettra de résoudre le travail demandé pour lequel il n'existe pas de réponse unique, mais plutôt un faisceau de solutions. Pour continuer, la définition de Jonnaert (2002) est l'unique à faire référence au caractère « partagé » de la compétence, et ce, en y introduisant les mots « personne et groupe ». En s'inscrivant dans le courant socioconstructiviste, cet auteur postule qu'il n'existe pas de réalité indépendante du sujet. Ainsi, c'est par ses expériences et ses interactions qu'une personne se construit. Le même auteur ajoute que la compétence est

toujours contextualisée dans une situation précise et est toujours dépendante de la représentation que la personne se fait de la situation (Jonnaert, 2011, p. 27).

Enfin, comme la définition de Tardif (2006) met en évidence les concepts de mobilisation, de combinaison des ressources ainsi que le caractère complexe commun aux autres définitions, c'est elle qui servira à illustrer nos propos pour la suite de cet écrit. Pour continuer, la littérature entourant le concept de « compétence » lui attribue six principales caractéristiques : (1) complexe, (2) contextualisée, (3) évolutive, (4) finalisée, (5) intégrative et (6) intériorisée. Le tableau synthèse tiré d'Hébert (2013), présente un portrait des principaux écrits décrivant chacune d'entre elles.

Tableau 3. *La compétence : principales caractéristiques*

La compétence est...	Dans la mesure où...	Auteurs
Complexe	elle « est généralement associée à des tâches d'un certain niveau de complexité qui font appel à de la résolution de problèmes et nécessitent le recours à des ressources variées » (Legendre, 2006, p. 23)	Legendre, 2001a, 2001b, 2004a, 2004b, 2006; MEQ, 2002.
Contextualisée	elle « ne s'exerce pas à vide » (Lasnier, 2000, p. 230)	Beckers, 2002; Gilbert & Parlier, 1992; Jonnaert, 2002; Lasnier, 2000; Louis, 2004; Louis & Hensler, 2003; Roegiers, 2000; Tardif, 2006; Wittorski, 1998.
Évolutive	« elle demande du temps pour se développer et [...] il est toujours possible de progresser » (MEQ, 2002, p. 6)	Fourez, 1999; Laurier, 2005; Legendre, 2001a, 2001b, 2004a, 2004b; MEQ, 2002; Tardif, 2006.
Finalisée	« elle n'a de sens que par rapport à l'action - elle est toujours "compétence à agir" - et au but que poursuit cette action » (Gilbert & Parlier, 1992, p. 15)	Detheux, 2005; Fourez, 1999; Gilbert & Parlier, 1992; Jonnaert, 2002; Rey, 2004; Rey et al., 2003; Roegiers, 2000; Wittorski, 1998.
Intégrative	« elle fait appel à une diversité de ressources, aussi bien externes qu'internes, et [...] réside dans la manière dont le sujet orchestre l'utilisation dans une situation donnée » (Legendre, 2001a, p. 17)	Gilbert & Parlier, 1992; Laurier, 2005; Legendre, 2001a, 2001b, 2004a; Louis, 2004; Louis & Hensler, 2003; MEQ, 2002; Tardif, 2006.
Intériorisée	elle « appartient en propre à un individu » (Rey et al., 2003, p. 13)	Louis, 2004; Rey et al., 2003; Roegiers, 2000.

Les caractéristiques répertoriées par les auteurs ci-haut inscrivent, en résumé, la compétence dans un contexte de complexité où les individus sont confrontés à des situations ciblées leur permettant de démontrer ce dont ils sont capables par la mobilisation d'un nombre important de ressources diverses. Plus précisément, Legendre (2006, p. 24) qualifie la compétence d'intégratrice, puisqu'elle « fait référence à une diversité de ressources internes et externes orchestrées par le sujet ». Elle est aussi complexe, car il ne suffit pas d'empiler, d'additionner ou de juxtaposer des ressources existantes pour constituer une compétence, mais il faut pouvoir les mobiliser et les combiner. Elle possède un potentiel d'action finalisé, car elle prend tout son sens dans l'action en quête du but qu'elle poursuit. Elle est également contextualisée, car elle se rapporte à des situations précises. Par conséquent, ce n'est pas « une capacité abstraite isolée de tout contexte » (Scallon, 2004, p. 103). Enfin, la compétence est évolutive et intériorisée, car elle est le résultat d'une construction personnelle résultant des apprentissages réalisés au fil du temps.

2.2 La démarche évaluative en contexte scolaire

L'adoption d'une approche par compétences en territoire québécois a mis en branle de nombreux changements au sein des pratiques évaluatives du corps enseignant (Tardif, 2006). Par conséquent, les prochains paragraphes exposent plus particulièrement cette transformation en abordant la démarche évaluative en contexte scolaire.

L'évaluation est définie comme « processus qui consiste à porter un jugement sur les apprentissages, soit des connaissances et des compétences disciplinaires, à partir des données recueillies, analysées et interprétées, en vue de décisions pédagogiques et administratives » (MEQ, 2003, p. 2). Ce « processus » s'inscrit dans une démarche d'évaluation qui permet d'analyser le cheminement (Scallon, 2004) et les difficultés d'apprentissage (Legendre, 2005) des élèves par des méthodes scientifiques (Champagne, Contradriopoulos, Brousselle, Hartz et Denis, 2009). La démarche évaluative se caractérise par différentes étapes. Traditionnellement, elle se présentait en trois temps : mesure, jugement et décision (MEQ, 1983). Par la suite, se sont greffés l'intention évaluative et l'objet (MEQ, 1988). D'autres auteurs comme Tousignant et Morissette (1990) la présentent plutôt en deux étapes distinctes : la mesure (incluant la collecte, l'organisation et l'interprétation des données) et l'évaluation (incluant le jugement et la décision). De leur côté,

Laurier et *al.* (2005) décrivent le tout en cinq temps : (1) la planification, (2) la prise d'information, (3) l'interprétation, (4) le jugement et (5) la communication. Enfin, Durand et Chouinard (2012) favorisent également cette approche en y ajoutant une légère précision : la planification (globale et spécifique), la prise d'information, l'interprétation, le jugement et la décision ainsi que la communication.

Dans le cadre de cette étude, le découpage proposé par le MEQ en 2003 est retenu, car il s'agit du cadre de référence officiel en conformité avec la réalité des établissements scolaires québécois où œuvrent les participants ciblés. À titre informatif, le tableau suivant présente plus précisément chacune de ces étapes telles que décrites à la page 32 de la *Politique d'évaluation des apprentissages*.

Tableau 4. *Tableau illustrant les étapes de la démarche en évaluation des apprentissages (MEQ, 2003)*

Étape	Description
Planification	<ul style="list-style-type: none"> ○ Établir l'intention de l'évaluation ○ Choisir les moyens appropriés à l'évaluation selon l'intention retenue ○ S'assurer d'une planification souple, rigoureuse et adaptée aux diverses situations
Prise d'information et son interprétation	<ul style="list-style-type: none"> ○ Recueillir des données sur les apprentissages des élèves selon leurs besoins ○ S'assurer d'une instrumentation permettant une cueillette suffisante et pertinente essentielle à un bon jugement ○ Comparer les données recueillies avec les résultats attendus
Jugement	<ul style="list-style-type: none"> ○ Agir comme toile de fond du processus complet de la démarche en évaluation (de la planification à l'action) ○ Analyser et synthétiser les données recueillies
Décision-action	<ul style="list-style-type: none"> ○ Représente la finalité de l'évaluation ○ Représente la décision (pédagogique ou administrative) prise à la suite du jugement

2.3 Le jugement en sein de la démarche évaluative

Dans la cadre de la poursuite de ce projet, l'étude spécifique de l'étape du « jugement » est d'intérêt. L'appellation « jugement évaluatif » est d'ailleurs endossée par plusieurs auteurs (Allal et Mottier Lopez, 2008; Durand, 2010; Legendre, 2005; Leroux, 2014 et Mottier Lopez, 2008). Legendre (2005) le définit comme étant « une étape centrale de la démarche d'évaluation pédagogique qui consiste à donner un avis sur la progression ou l'état de réalisation des apprentissages, et ce, à la lumière des diverses informations recueillies ». Leroux (2014, p. 80) souligne également le rôle primordial qu'il occupe au sein de la démarche évaluative, car « il donne un avis sur le développement d'une compétence en tenant compte de diverses informations recueillies lors de la réalisation de tâches complexes ». Dans un contexte d'évaluation de compétences articulé autour de tâches complexes et authentiques, le jugement de l'enseignant décrit en profondeur les performances observées afin de discriminer les niveaux de performance entre les élèves (Scallon, 2004; Wiggins, 1998). Pour ce faire, l'enseignant porte un jugement en ayant un regard global et évolutif sur la capacité de l'élève et sur sa progression, ce qui nécessite une appréciation instrumentalisée et bien documentée (Durand, 2010, p. 133). Ce processus complexe se fonde en grande partie sur un jugement balisé afin d'assurer la crédibilité des actions posées par le pédagogue. Pour recueillir les données nécessaires, l'enseignant place l'élève dans des tâches suffisamment nombreuses et diversifiées afin de lui permettre de montrer ses compétences (Durand, 2010). Enfin, Allal et Mottier Lopez (2008, p. 466) appuient également cette démarche en expliquant que « l'enseignant a recours au jugement évaluatif par son choix de l'objet à évaluer, ses méthodes et ses instruments de mesure, son interprétation et ses décisions émises ».

Dans la présente étude, nous nous sommes appuyés sur les caractéristiques du jugement évaluatif tirées d'une recherche menée par Allal et Mottier Lopez en 2008. Nous considérons que cette modélisation du jugement évaluatif met bien en évidence le caractère heuristique sur lequel s'appuient les raisonnements des enseignants lors de prises de décisions évaluatives, en plus des influences possibles du contexte sur cette démarche. Enfin, tout comme Scallon (2004) et Tardif (2006), les auteurs de cette étude s'entendent sur l'importance de l'aspect qualitatif de l'exercice du jugement, et ce, même si des valeurs quantitatives (ex. scores) sont parfois attribuées. Le tableau suivant précise les caractéristiques de la modélisation du jugement évaluatif sur lesquels nous basons la présente étude.

Tableau 5. *Les caractéristiques de la modélisation synthétique du jugement évaluatif (Allal et Mottier Lopez, 2008, p. 466)*

1	<p>Il comprend deux démarches de base :</p> <ul style="list-style-type: none"> ○ Réunir et confronter des sources d'information de natures diverses ○ Exécuter un raisonnement interprétatif pouvant inclure un algorithme de calculs, mais qui ne s'y réduit pas.
2	<p>Il s'appuie sur des procédures d'évaluation que l'enseignant élabore et applique à l'ensemble des élèves, mais une de ses fonctions essentielles est d'assurer l'adaptation des procédures aux situations singulières qui l'exigent.</p>
3	<p>Acte cognitif individuel, mais s'élaborant et s'intégrant à un réseau de communication au sein d'un contexte social et institutionnel.</p>
4	<p>Il implique une mise en relation entre les exigences du système scolaire, la culture de l'établissement, les valeurs personnelles et les stratégies pédagogiques de l'enseignement.</p>

2.4 Qualité du jugement en évaluation des apprentissages

La tension existante entre l'objectivité et la subjectivité en contexte évaluatif a été soulignée par plusieurs auteurs (Gérard, 2002; Lafortune, 2007; Leroux, 2014). Parmi eux, Leroux (2014, p. 5) suggère que « la subjectivité soit prise en compte dans l'exercice du jugement, mais que cette dernière soit soutenue par une démarche de cueillette de données pertinentes et suffisantes ainsi qu'une interprétation à l'aide d'instruments de jugement dont le but est de prendre

des décisions qui pourront être justifiées ». Savoie-Zajc (2013, p. 105) ajoute que le jugement de l'enseignant peut être influencé par les caractéristiques de la personne (expérience, formation et identité). Comme il est impossible de faire complète abstraction de ces éléments, « il est important de fonder son jugement sur des qualités de rigueur, de cohérence et de transparence (*idem*, 2013). Même si l'on reconnaît que la subjectivité est inévitable, des caractéristiques semblent favoriser une certaine objectivité. Les caractéristiques qui tendent à favoriser cette condition sont les suivantes : l'exercice d'un jugement instrumenté, évolutif, éthique et conforme à un ensemble de valeurs fondé sur l'autonomie (Bressoux et Pansu, 2003; Legendre, 2001; Durand et Chouinard, 2012). Par conséquent, il n'est pas si facile pour un enseignant d'agir de manière objective et juste. Les travaux en lien avec les biais de l'évaluateur (l'effet de la tendance centrale, la stéréotypie, l'effet de Halo, l'effet de contraste, etc.) en témoignent (Leclercq, 2004).

Dans une perspective visant à limiter l'influence de ces biais, Lafortune et Allal (2008) émettent quelques suggestions : expliciter le contenu à évaluer et la manière de les évaluer, travailler en équipe de professionnels lors de la construction des outils d'évaluation, confronter les résultats des évaluations à des évaluateurs externes et enrichir les interprétations en sollicitant d'autres avis de professionnels. Enfin, Mottier Lopez et Allal (2008, p. 478) affirment que « promouvoir le développement du jugement professionnel en évaluation est un moyen d'amener les enseignants à prendre conscience des biais évaluatifs et à adopter des démarches susceptibles de les réduire ». Malheureusement, peu de ressources existent. Les prochaines lignes s'attardent donc à l'insuffisance d'outils de formation au développement du jugement professionnel en évaluation chez les enseignants et la relève.

2.5 Insuffisance d'outils de formation au développement du jugement en évaluation

Des études menées au cours d'une récente période exposent le défi entourant le manque de dispositifs de formation à l'acte d'évaluer chez le corps enseignant (Baribeau, 2009; Dionne, 2005; Proulx, 2005; Maray, 2002 ; Tardif et Lessard, 2004). En effet, de nombreux enseignants affirment avoir peu d'accompagnement et de temps pour échanger avec leurs collègues (Baribeau, 2009). Historiquement, Dionne (2005) explique comment le soutien au corps professoral dans le cadre de l'implantation du Renouveau pédagogique n'a pas été suffisant pour contrer l'incompréhensibilité du personnel. Malgré les pistes proposées par le MELS (2008) à cette époque pour encadrer les

pratiques évaluatives des enseignants, plusieurs zones grises persistent toujours. Malgré son importance, l'évaluation dans le contexte scolaire apparaît encore, pour plusieurs acteurs, comme une contrainte inutile qui les stresse et leur fait perdre temps et énergie (Van der Maren et Loye, 2011). Cette situation s'avère étonnante dans la mesure où l'évaluation est présente dans chacune des activités des enseignants visant l'articulation des actions d'enseignement avec les processus d'apprentissage des élèves (Allal, 2007).

À la lumière de ces réalités, il devient essentiel de mieux encadrer le corps enseignant, dès leur entrée à la formation des maîtres, au développement d'un jugement professionnel adéquat en évaluation. Actuellement, la plupart des enseignants issus des universités du Québec offrant un programme à l'enseignement ne suivent qu'un seul cours de trois crédits en évaluation des apprentissages (Fontaine et *al.*, 2013). En nous inspirant de d'autres disciplines où le jugement professionnel est développé tout au long de la formation, nous croyons que les sciences de l'éducation pourraient bonifier le développement du jugement évaluatif de la relève par l'entremise d'outils formateurs adaptés. Pour y arriver, il devient impératif de se pencher sur les dispositifs de formation au jugement disponibles afin d'en avoir une meilleure compréhension et ultimement d'arriver à les bonifier.

2.6 Similitudes entre le jugement en évaluation et le jugement clinique

Lafortune (2006, p. 4) nous présente le jugement en évaluation comme « un processus qui mène à une prise de décisions, laquelle prend en compte différentes considérations issues de son expertise (expérience et formation) professionnelle. Ce processus exige rigueur, cohérence et transparence ». Tel que mentionné, le contexte scolaire incite souvent le jugement de l'enseignant à être basé sur la confrontation de données de natures différentes et l'élaboration d'une combinaison qui ne peut se limiter à l'application d'un algorithme. En effet, plusieurs facteurs doivent être considérés dans un processus où le traitement de l'information doit être réalisé à la lumière de contraintes administratives et institutionnelles. Le jugement de l'enseignant est ainsi un processus cognitif individuel et une pratique sociale située au cœur de la communauté professionnelle (Lafortune, 2006; Laveault, 2005). Plus précisément, les écrits des chercheurs intéressés par le jugement en évaluation (Gohier, Bednarz, Gaudreau, Pallascio et Parent, 1999; Lafortune et Allal, 2008; Laveault, 2005; Paquay, Altet, Charlier et Perrenoud, 1996) nous

expliquent que celui-ci se développe au regard de quelques conditions propres à la profession enseignante. Parmi ses caractéristiques spécifiques, on remarque notamment les suivantes :

- le jugement évaluatif s'exerce dans des situations de **problèmes complexes** pour lesquelles l'application de règles standardisées ne convient pas;
- la résolution nécessite des **savoirs d'expérience** et la mobilisation de **connaissances professionnelles**;
- des prises de décisions qui relèvent de la **responsabilité** et d'une certaine **autonomie** de l'enseignant;
- une analyse des ressources et des **contraintes liées aux contextes** concernés est impliquée;
- une **réflexion critique** tant individuelle que collective est suscitée;
- des formes **d'ajustements créatifs** au regard des situations singulières sont engagées.

De manière similaire, une composante du développement de la compétence professionnelle du médecin repose également sur sa capacité à résoudre des problèmes complexes dans des contextes d'incertitude (Shön, 1983). C'est le « jugement clinique » qui devient alors nécessaire afin de régler des situations qui sont caractérisées comme étant indéterminées, ambiguës et impliquant parfois des conflits éthiques, d'intérêts ou de valeurs (Tanner, 2006). Le paragraphe suivant définit plus spécifiquement le rôle du jugement clinique en contexte de soins.

La littérature issue de la pédagogie des sciences de la santé place au cœur de la démarche clinique le raisonnement médical (Jouquan, Parent et Audétat, 2013). Plus spécifiquement, les termes de « jugement clinique » et « raisonnement clinique » sont fréquemment utilisés de manière interchangeable. Nendaz et *al.* (2005, p. 236) désignent le « raisonnement clinique » comme étant un « processus de résolution de problèmes et de prise de décisions que développe un clinicien lorsque, à partir de la plainte formulée par patient, il catégorise le problème de santé auquel est confronté ce dernier en termes suffisamment opérationnels pour qu'un plan d'action efficient puisse être engagé ». Le mécanisme décisionnel impliqué lors d'un jugement clinique peut être de deux ordres : analytique ou intuitif (King et Mcleod, 2002). Les raisonnements non analytiques relèvent d'une pensée intuitive, réflexe ou automatique s'appuyant le plus souvent sur la reconnaissance d'une similitude entre le problème posé et les expériences antérieures du professionnel. Les raisonnements non analytiques sont surtout retrouvés chez les professionnels expérimentés, ayant emmagasiné en mémoire à long terme de nombreux cas et raisonnements. C'est un processus de raisonnement qui se développe avec l'expérience clinique (Paucart-Dupont

et Marchand, 2014). La recherche associée au raisonnement clinique s'est d'ailleurs intéressée à cette aptitude cognitive qui permet aux professionnels de la santé d'organiser leurs connaissances de manière à générer des hypothèses les menant à l'établissement d'un diagnostic et à l'élaboration d'un plan de traitement (Charlin, Tardif et Boshuizen, 2000).

Pour continuer, une analyse menée par Dionne et Forte (2012) a permis de recenser une série de similitudes entre le jugement exercé en contexte d'évaluation des apprentissages et celui exercé en milieu clinique. Parmi celles-ci, on peut compter :

- 1) la prise en compte de réseaux de **connaissances préalables** ou connexes (Charlin et *al.*, 2005; Charlin, Boshuizen, Custers et Feltovich, 2007; Charlin, Tardif et Boshuizen, 2000);
- 2) l'identification d'informations **manquantes ou ambiguës** (Charlin et *al.*, 2000, 2005);
- 3) la présentation d'éléments qui **confirment** la réponse (Charlin et *al.*, 2005);
- 4) la **recherche d'indices** qui confirment l'opinion ou la décision (Charlin et *al.*, 2007) et la **reconnaissance d'un modèle** (ou catégorisation de tâches) (Charlin et *al.*, 2000, 2007);
- 5) le raisonnement **hypothético-déductif** (Charlin et *al.*, 2000, 2007);
- 6) la **prédiction** d'évènements (Charlin et *al.*, 2007);
- 7) la prise de décisions à partir de la **résolution de problèmes** (Ramaekers et *al.*, 2010).

Les similitudes existantes entre les deux types de jugement (évaluatif et clinique) nous ont semblé notables. Sans être identiques, on a toutefois constaté que les jugements ciblés s'exercent dans des contextes similaires où l'incertitude et l'ambiguïté sont bien présentes. Dans les deux cas, on a également remarqué l'influence possible du processus de raisonnement basé sur les connaissances et les expériences préalables de l'évaluateur sur sa prise de décisions. Pour ces raisons, dans le cadre de cette étude, nous postulons qu'il existe suffisamment de similarités entre le raisonnement clinique et le jugement en évaluation des apprentissages pour justifier l'adaptation d'un test utilisé au développement du jugement clinique à l'application du développement du jugement évaluatif de l'enseignant.

2.7 L'adaptation d'un instrument issu de la pédagogie des sciences de la santé

2.7.1 Les composantes du test de concordance de script

Parmi les méthodes d'évaluation du raisonnement clinique, on accorde au test de concordance de script (TCS) une bonne crédibilité (Carrière et *al.*, 2009; Charlin, 1998; Fournier, Thiercelin, Pulcini, Alunni-Perret, Gilbert, Minget et Bertrand, 2006; Gagnon et *al.*, 2005; Lambert, 2005; Marie et *al.*, 2005; Sibert et *al.*, 2001). En effet, les résultats d'études s'intéressant au TCS ont permis de conclure qu'il évaluait fidèlement le raisonnement clinique en vue de classer les répondants selon un niveau d'expertise tel que novice, compétent ou expert (Lambert, 2005). Basé sur la théorie cognitive des scripts, c'est-à-dire sur la manière dont sont organisées les connaissances acquises des individus en situations de résolution de problèmes complexes, cet instrument expose le sujet à une série de cas cliniques aussi appelés *vignettes*. À la suite de la lecture des situations présentées, l'instrument propose au répondant des hypothèses solutionnaires suivies d'informations nouvelles (symptômes physiques, d'antécédents de santé, de fausses données scientifiques ou encore de résultats de tests de laboratoire). Pour chacune des hypothèses données, le participant est invité à qualifier l'effet de cette information nouvelle sur son hypothèse initiale, et ce, à partir d'une échelle de type Likert (-2 à +2., par exemple). Le score attribué au sujet est déterminé à partir de la fréquence de concordance entre ses réponses et celles obtenues par un panel d'experts ayant répondu aux mêmes questions. La figure suivante présente un exemple de vignette de TCS (Charlin, Kazi-Tani, Gagnon et Thivierge, 2005).

- Une patiente de 65 ans est adressée aux urgences pour une suspicion d'embolie pulmonaire.
- Elle a des antécédents de diabète, d'insuffisance cardiaque et de BPCO.

► Une réponse pour chaque ligne.

Hypothèse	Nouvelle information	Conséquence sur l'hypothèse initiale				
Si vous pensiez faire	Et qu'alors vous trouvez...	L'effet sur la nécessité de demander ce test est le suivant... (cercler votre réponse)				
Un angioscan thoracique	Un traitement par metformine	- 2	- 1	0	+ 1	+ 2
Un dosage des D-dimères	Un cancer de l'ovaire en cours de traitement	- 2	- 1	0	+ 1	+ 2
Un écho-Doppler veineux	Un signe de Homans	- 2	- 1	0	+ 1	+ 2

► Réponses possibles

- 2 absolument contre-indiqué
- 1 peu utile
- 0 non pertinent dans cette situation
- + 1 utile et souhaitable
- + 2 indispensable

Figure 1. Exemple d'une vignette de test de concordance de script (Charlin, Kazi-Tani, Gagnon et Thivierge, 2005)

La littérature entourant les règles de conception d'un TCS nous informe tout d'abord sur l'objectif principal de l'outil : mesurer le niveau de raisonnement clinique tout en s'attardant à la qualité de l'organisation des connaissances du répondant (Dumont, 2013). Plus précisément, le test permet de recueillir et d'interpréter des données dans un contexte de la résolution d'un problème médical mal défini (Charlin, Kazi-Tani, Gagnon et Thivierge, 2005). Fournier, Demeester et Charlin (2008) ont d'ailleurs publié un article référentiel intitulé *Script Concordance Tests: Guidelines for Construction*, portant essentiellement sur les principales caractéristiques de la vignette clinique du TCS : présence importante de l'incertitude et de l'ambiguïté aux deux

différentes composantes du test : le scénario et les questions. Contrairement aux questions à choix multiple, le TCS vise un certain niveau de variabilité dans les réponses des membres du panel d'experts (Charlin, Gagnon, Sibert et Van der Vleuten, 2002). Ainsi, l'expert confronté au problème présenté n'a pas entre les mains tous les éléments essentiels à la résolution du problème, tout comme cela pourrait être le cas dans son cabinet médical.

Nous soulignons que l'utilisation d'un panel d'experts est un élément essentiel à l'élaboration d'un TCS, car c'est à leurs réponses que sont comparées celles des participants, et ce, pour les mêmes vignettes. Dans le but de renforcer la validité de cet outil d'évaluation, de nombreux chercheurs se sont intéressés aux caractéristiques de composition de ce dernier. Comme le souligne Norman (2005), la complexité de la formation d'un tel panel est justifiée notamment par le manque de consensus entourant la définition même que ce qu'est un expert dans un domaine. À ce jour, dans un contexte de TCS, un expert est défini comme un clinicien expérimenté dont la présence dans un jury est légitime, et ce, considérant le niveau de formation des candidats qui auront à passer le test (Blais, Charlin, Grodin, Lambert, Loye et Gagnon, 2011). À cette définition ajoutent Lubarsky, Dory, Duggan, Gagnon et Charlin (2013) d'opter pour des cliniciens expérimentés dont le domaine de spécialisation est identique à celui abordé dans les vignettes cliniques composant le TCS. Dory, Gagnon, Vanpee et Charlin (2012) renchérissent cette affirmation en soulignant que les panels d'experts, composés de médecins issus de la même culture clinique que la matière abordée ainsi que la présence de cliniciens-enseignants, augmenteraient généralement la validité des scores obtenus. Enfin, une étude de Charlin, Gagnon, Sauve et Coletti (2007) en est venue à des conclusions similaires en comparant deux panels d'experts dont l'un était composé de médecins de famille ayant des tâches d'enseignement ou d'un autre qui n'en avait aucune. Les résultats ont démontré une forte corrélation entre la fonction d'enseignant (coefficient de 0.98) et les données obtenues.

Pour continuer, Gagnon, Charlin, Coletti, Sauve et Van der Vleuten (2005) se sont penchés sur le nombre idéal de membres que devrait contenir le groupe d'experts. Leur étude a démontré que le panel composé d'au moins dix experts donnait des résultats suffisants et qu'un groupe de plus de vingt ne démontrait pas de réel gain en termes de cohérence interne. Par contre, comme cette étude s'intéressait qu'à un seul TCS, d'autres résultats seraient nécessaires pour faire de cette

affirmation une règle d'application plus formelle. D'autres chercheurs se sont interrogés sur la validité conceptuelle en lien avec les différents niveaux de spécialisation des médecins et le type de pratique composant le panel d'experts. Plus précisément, les travaux de Bursztejen et *al.* (2011) ont permis de conclure que les résultats obtenus aux TCS étaient plutôt influencés par le type de pratique (hôpital par opposition à la pratique clinique) plutôt que par la spécialité des membres experts (dermatologie opposée à la médecine familiale). Malgré tout, certains auteurs ont tout de même relevé certaines menaces à la validité des inférences à partir des scores obtenus avec cet instrument en contestant le processus de sélection ainsi que les caractéristiques de composition du panel d'experts (Lineberry, Kreiter et Bordage, 2013). En effet, il peut s'agir-là d'une étape névralgique puisque la structure de détermination des scores en dépend. Malgré les balises de critères de sélection (nombre, spécialité, nombre d'années d'expérience, etc.) proposées par la littérature, il en résulte qu'il s'agit encore d'un aspect influant dont l'impact est encore mal estimé. C'est pourquoi Lubarsky, Gagnon et Charlin (2013) encouragent la communauté scientifique à continuer de se pencher sur le sujet.

2.7.2 Les composantes du test de jugement professionnel en évaluation des apprentissages (TJPEA)

Le format de vignette proposé par le TJPEA s'inspire directement du gabarit et des règles rédactionnelles du test de concordance de script élaboré par Charlin et *al.* (2002). Dans un premier temps, chaque canevas expose un scénario pédagogique offrant une courte description d'une situation authentique présentant des données ambiguës, incomplètes ou incertaines. Dans un deuxième temps, tout comme dans le format proposé par le TCS, le TJPEA suggère au répondant une intervention réalisée suivie de sa conséquence. Dans un troisième temps, les scores des répondants sont colligés au moyen d'items de type Likert à quatre échelons (échelle de pertinence). Dans un quatrième temps, contrairement au TCS, l'instrument du TJPEA propose une deuxième échelle de type Likert à quatre échelons, soit celle en lien avec la probabilité à réaliser l'item. Le concepteur de l'outil, Dionne (2012), a jugé intéressant de faire cet ajout, et ce, afin de mieux comprendre l'influence possible de la désirabilité sociale sur les réponses des sujets. La Figure 2 présente un exemple de vignette pouvant constituer le TJPEA.

Directives : À la suite de la lecture du cas, une intervention hypothétique est présentée au répondant à laquelle on associe une conséquence probable. Le participant est ensuite appelé à répondre à deux questions déterminant s'il a jugé utile l'intervention proposée et s'il compte reproduire cette action une fois en salle de classe.

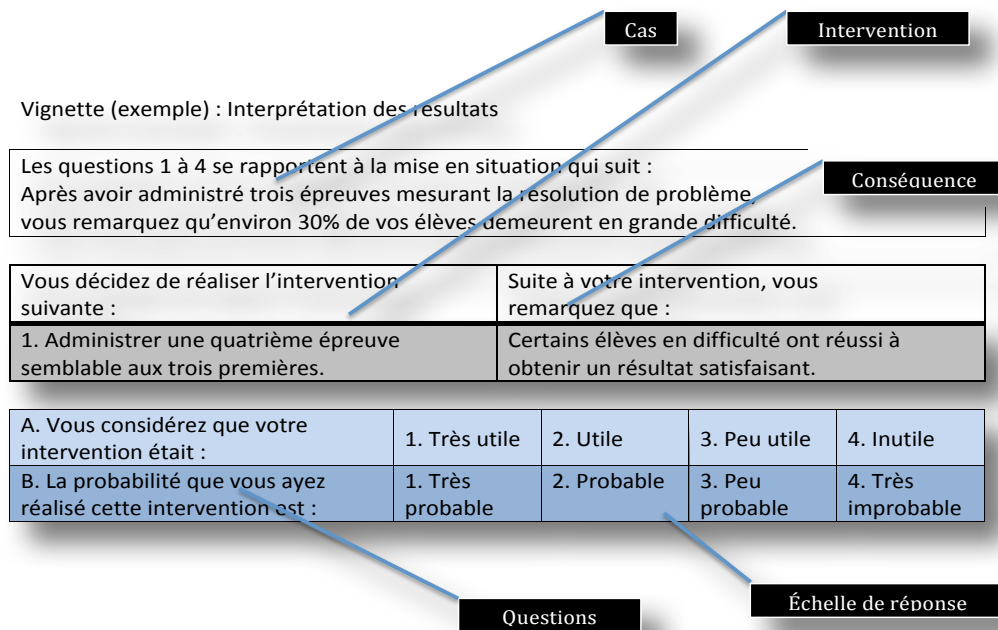


Figure 2. Exemple d'une vignette pour une dimension d'interprétation des résultats (Dionne, 2012)

Étant donné le caractère novateur de l'adaptation de l'instrument qu'est le TJPEA, une présentation détaillée de chacune des étapes de réalisation est présentée au prochain chapitre de ce mémoire. De plus, la démarche méthodologique choisie dans le cadre de cette étude, soit l'entrevue par la verbalisation inspirée des travaux d'Ericsson et Simon (1980), mettra en évidence comment la passation du TJPEA peut, selon notre hypothèse, rendre compte du jugement professionnel en évaluation des différents profils de répondants.

2.8 La question spécifique de recherche

En conclusion, dans le cadre de ce présent chapitre, nous avons défini nos principaux concepts, soit celui du jugement professionnel en évaluation des apprentissages et du jugement clinique. Basé sur la littérature issue des disciplines de l'évaluation des apprentissages et de la pédagogie des sciences de la santé, nous avons ensuite proposé l'application d'un instrument de formation destiné au développement du jugement clinique à celui du jugement évaluatif. Plus précisément, nous nous questionnons quant au potentiel pédagogique possible en lien avec la passation de ce type de test : le test de jugement professionnel en évaluation des apprentissages (TJPEA). Notre question de recherche principale était : « Les vignettes présentées par le TJPEA permettent-elles de rendre compte du jugement en évaluation des enseignants ? »

Afin de mieux comprendre le jugement professionnel qu'exercent les enseignants et d'être cohérents avec la question de recherche principale, notre question spécifique s'articule ainsi : « À partir des réponses des participants aux vignettes du TJPEA, pourrions-nous démontrer des différences entre le jugement en évaluation des novices, des praticiens et celui des experts ? »

Chapitre 3

Méthodologie

« Une bonne méthode est un télescope avec lequel on voit ce qui échappe à l'œil nu. »
Condillac

Cette partie du présent travail dévoile les éléments méthodologiques qui ont permis de réaliser cette étude. Elle expose le type de recherche, les étapes de réalisation et de validation de l'outil d'évaluation, l'échantillon ainsi que les conditions d'administration du test de jugement en évaluation des apprentissages (TJPEA) qui ont permis de répondre à la question générale de recherche : les vignettes présentées par le TJPEA permettent-elles de rendre compte du jugement en évaluation des enseignants ?

3.1 Type de recherche

L'objectif de cette étude exploratoire qualitative interprétative est de mieux comprendre le processus de développement du jugement en évaluation des apprentissages des futurs enseignants par la passation d'un TJPEA. Comme le soulignent Guba et Lincoln (1989), la recherche qualitative interprétative appelle fondamentalement des outils qualitatifs visant à mesurer des données difficilement mesurables. Savoie-Zajc (2004) ajoute que deux facteurs peuvent justifier ce type de démarche : l'accessibilité des données et des connaissances produites par la recherche ainsi que la présence essentielle du caractère de l'interactivité. Les caractéristiques de l'étude nous ont donc amenés à adopter cette approche, car au-delà d'une simple description, elle permet de comprendre et de considérer les significations que les individus donnent à leurs actions (Huberman et Miles, 1991; Lessard-Hebert, Goyette et Boutin, 1995). Non seulement on vise à comprendre le processus de développement du jugement en évaluation des apprentissages à partir des impressions des participants face aux vignettes présentées par le TJPEA, mais la présente étude propose également d'explorer un domaine d'intérêt en évaluation pour lequel il n'existe peu ou pas de données disponibles, et ce, malgré le fait que la compétence ciblée soit partie intégrante du *Cadre de référence québécois de la profession enseignante du Québec* (MEQ, 2001).

3.2 Participants

Dans une perspective de description du processus de jugement en évaluation, nous avons choisi de constituer ce que Deslauriers (1991) nomme un échantillon intentionnel. Ce type de recrutement privilégie des cas typiques en lien avec les objectifs de la recherche (Mayer et Ouellet, 1991) et vise à produire le maximum d'information par la composition de cas qui, par rapport aux critères de sélection, se différencient et représentent les divers aspects d'un problème à l'étude (saturation de l'échantillon). Savoie-Zajc (2004) suggère les quatre critères suivants pour délimiter un échantillon rigoureux : (1) il est intentionnel et constitué en fonction des buts de la recherche, (2) il est balisé et justifié, (3) il est cohérent avec les postures épistémologiques et méthodologiques et (4) il se traduit un souci éthique. L'échantillon constitué dans le cadre de cette étude satisfait à ces critères.

Finalement, ce projet de recherche vise trois différentes populations : (1) quatre étudiants de première année inscrits au programme de formation des maîtres (profil secondaire) dans une université québécoise, (2) quatre enseignants actifs dans une école secondaire québécoise et (3) quatre experts issus du domaine de l'évaluation des apprentissages.

3.3 Sélection des participants

Comme cette étude s'inscrit dans un contexte québécois, la sélection des participants tiendra compte de cette caractéristique dans ses critères de sélection.

○ *Novices*

Le recrutement de quatre participants ainsi que la sélection des sujets de la catégorie novice se sont réalisés selon les critères suivants :

- être inscrit à un programme de formation des maîtres (volet secondaire) offert en français au Québec;
- ne jamais avoir enseigné dans un établissement scolaire;
- avoir le français comme langue de scolarisation depuis la première secondaire, et ce, afin de s'assurer de la pleine compréhension des vignettes dans le cadre de la passation du TJPEA.

○ *Enseignants (Praticiens)*

Les quatre participants composant le groupe de praticiens devaient répondre aux critères prédéterminés suivants :

- être enseignant dans une école secondaire publique québécoise depuis plus de 10 ans;
- détenir un brevet d'enseignement valide dans la province de pratique, soit le Québec;
- avoir le français comme langue de scolarisation depuis la première secondaire, et ce, afin de s'assurer de la pleine compréhension des vignettes dans le cadre de la passation du TJPEA.

○ *Experts*

Les quatre participants composant le groupe d'experts devaient également répondre aux critères prédéterminés suivants :

- avoir le français comme langue de scolarisation depuis la première secondaire, et ce, afin de s'assurer de la pleine compréhension des vignettes dans le cadre de la passation du TJPEA;
- occuper un poste de professeur dans une université canadienne francophone;
- avoir enseigné un cours d'évaluation des apprentissages au niveau universitaire;
- mener des travaux de recherche en lien avec la discipline ciblée (évaluation des apprentissages);
- détenir une expérience d'enseignement d'au moins 10 ans.

Tous les participants qui ont accepté de participer à cette recherche ont été recrutés sur une base volontaire par l'entremise d'un courriel d'invitation (Annexe A) parvenu par l'étudiante-chercheuse du présent projet. Tous ont choisi de signer le formulaire de consentement (Annexe B) nous autorisant à mener cette étude auprès d'eux par la passation d'un test de jugement professionnel en évaluation des apprentissages. À titre informatif, le tableau suivant présente certaines caractéristiques de l'échantillon.

Tableau 6*Caractéristiques de l'échantillon selon les profils*

	Sexe (Féminin et masculin)	Palier d'enseignement (secondaire ou universitaire)	Milieu d'enseignement (Privé ou public)	Nombre d'années d'expérience
Novices				
A	Féminin	Non applicable	Non applicable	0
B	Féminin	Non applicable	Non applicable	0
C	Féminin	Non applicable	Non applicable	0
D	Féminin	Non applicable	Non applicable	0
Praticiens				
A	Féminin	Secondaire	Public et privé	Entre 10 et 15 ans
B	Masculin	Secondaire	Public	Entre 10 et 15 ans
C	Féminin	Secondaire	Public	Plus de 20 ans
D	Féminin	Secondaire	Public	Entre 10 et 15 ans
Experts				
A	Masculin	Universitaire	Public	Plus de 20 ans
B	Masculin	Universitaire	Public	Plus de 20 ans
C	Féminin	Universitaire	Public	Plus de 20 ans
D	Féminin	Universitaire	Public	Plus de 20 ans

3.4 Étapes de réalisation du TJPEA

Cette section aborde les différentes étapes procédurales qui ont permis la réalisation du TJPEA. Nous nous sommes référés aux indications de l'auteur (Dionne) qui s'est lui-même basé sur les propositions de Downing et Haladyna (2006) afin de développer le TJPEA.

3.4.1 Détermination de l'utilisation prévue du test et définition de ce qu'il va mesurer

Le TJPEA développé vise à explorer le niveau de jugement professionnel en évaluation des étudiants inscrits à la formation des maîtres à l'ordre du secondaire. Afin d'y arriver, l'organisation de leurs connaissances (scripts) a été comparée à celle d'un groupe d'experts en évaluation des apprentissages ainsi qu'à un groupe d'enseignants d'expérience au secondaire. Il est à noter que la passation du ce test est destinée à un contexte formatif et que le contenu présenté est uniquement en français.

3.4.2 Construction d'un tableau de spécifications et sélection des thèmes des vignettes

Un tableau de spécifications a été construit conformément aux suggestions de Downing et Haladyna (2006). Ce dernier a permis d'établir la validité de contenu du test élaboré grâce à la

représentativité et à l'exhaustivité des éléments présentés. Comme le soulignent Laveault et Grégoire, 2002, p. 24) : « Le tableau de spécifications est un moyen utilisé depuis longtemps pour s'assurer que l'échantillonnage des questions à un examen est véritablement représentatif de la situation qui a prévalu en salle de classe ».

Plus spécifiquement, le tableau de spécifications a été créé en deux temps. Tout d'abord, une liste de thèmes (Annexe C) a été réalisée à partir d'une synthèse des contenus abordés par les principaux ouvrages (Annexe D) utilisés par les professeurs enseignant la discipline ciblée en milieu universitaire canadien, c'est-à-dire l'évaluation des apprentissages. Dans un deuxième temps, par souci de validité, le choix des contenus des vignettes a été fait à la lumière des attentes de la compétence en évaluation chez les futurs maîtres québécois présenté à la figure suivante.

**Compétence, composantes et attentes liées à la compétence
en évaluation chez les futurs maîtres québécois**

Compétence	Évaluer la progression des apprentissages et le degré d'acquisition des compétences des élèves pour les contenus à faire apprendre.
Composantes	<ol style="list-style-type: none"> 1. En situation d'apprentissage, prendre des informations afin de repérer les forces et les difficultés des élèves ainsi que de revoir et d'adapter l'enseignement en vue de favoriser la progression des apprentissages. 2. Établir un bilan des acquis afin de porter un jugement sur le degré d'acquisition des compétences. 3. Construire ou employer des outils permettant d'évaluer la progression et l'acquisition des compétences. 4. Communiquer aux élèves et aux parents, de façon claire et explicite, les résultats attendus ainsi que les rétroactions au regard de la progression des apprentissages et de l'acquisition des compétences. 5. Collaborer avec l'équipe pédagogique à la détermination du rythme et des étapes de progression souhaitées à l'intérieur du cycle de formation.
Attentes	<p>L'étudiant devrait être capable de :</p> <ul style="list-style-type: none"> - détecter, en situation d'apprentissage, les forces et les difficultés des élèves. - préciser, de façon autonome, des correctifs à apporter à son enseignement. - contribuer avec ses pairs à la préparation du matériel d'évaluation, à l'interprétation des productions des élèves au regard du développement des compétences et à l'élaboration d'outils de communication destinés aux parents. - communiquer à l'élève les résultats d'un processus d'évaluation diagnostique et d'indiquer aux parents et aux membres de l'équipe pédagogique les éléments des stratégies d'intervention corrective envisagés.

Source : Ministère de l'Éducation du Québec, 2001.

Figure 3. Descriptif de la compétence 5 du *Référentiel de la formation en enseignement* (MEQ, 2001)

Enfin, les vignettes sélectionnées dans le cadre de cette étude abordaient les contenus théoriques suivants : la démarche en évaluation, la rédaction des items, la progression et la structure de notation ainsi que la communication des résultats.

3.4.3 Le format des vignettes

Le format des vignettes proposées par le TJPEA s'inspire directement du test de concordance de script élaboré par Charlin et *al.* (2002). Chaque vignette comporte quatre parties distinctes. La première partie expose un scénario pédagogique sous la forme d'une courte description de situation authentique où le contexte présenté est mal défini. Les informations dévoilées sont donc incomplètes, ambiguës, incertaines et peuvent faire l'objet de différentes interprétations selon le répondant. La deuxième section nommée « intervention » propose une série de trois items indépendants des uns des autres. Par conséquent, ces items ne peuvent pas être traités comme étant cumulatifs, et ce, même s'ils relèvent tous du même scénario, de la même vignette (Caire et *al.*, 2004; Fournier et *al.*, 2008). Pour chaque intervention présentée, la conséquence sur les intervenants impliqués est donnée. La troisième section impose au participant de se prononcer sur la pertinence de l'intervention proposée sur une échelle de Likert à quatre encrages (très pertinent, pertinent, peu pertinent, non pertinent). Enfin, la dernière section demande au participant de situer la probabilité qu'il décide d'agir comme l'intervention proposée sur une autre échelle de Likert (très probable, probable, peu probable et improbable). Cette deuxième échelle, ne faisant pas partie du modèle original du format TCS, a été introduite au TJPEA afin de questionner l'influence de la désirabilité sociale chez les sujets. Par exemple, un participant pourrait juger une intervention « très pertinente » à l'échelle A, mais opter pour « peu probable » à l'échelle B, car il ne croit pas qu'il agirait de la sorte une fois en pratique.

Afin de faciliter la compréhension de la structure du format de l'instrument, la figure suivante présente une vignette tirée du TJPEA ainsi qu'un exemple d'item.

Directives : À la suite de la lecture du cas, une intervention hypothétique est présentée au répondant à laquelle on associe une conséquence probable. Le participant est ensuite appelé à répondre à deux questions déterminant s'il a jugé utile l'intervention proposée et s'il compte reproduire cette action une fois en salle de classe.

Vignette 1

Les questions se rapportent à la mise en situation qui suit :

Lors de la comptabilisation des notes pour le bulletin, vous remarquez qu'un de vos élèves est en échec de peu en raison des notes extrêmement basses qu'il a obtenues au début de l'étape. Par contre, si vous prenez en compte ses notes plus récentes, vous jugez qu'il a fait un progrès exceptionnel.

Option 1

Vous décidez de prendre la décision suivante :	À la suite de votre décision, vous remarquez que :
Vous décidez de modifier la pondération des évaluations de cet élève afin d'augmenter le poids des plus récentes.	L'élève termine son étape avec une bien meilleure note et sa confiance augmente drastiquement.

A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Figure 4. Exemple d'une vignette de TJPEA

3.4.4 Rédaction des vignettes composantes du TJPEA

Une fois le tableau de spécifications établi, les contenus des vignettes ont été rédigés par l'auteur du TJPEA et un étudiant en sciences de l'éducation. Dans un souci de qualité, une validation auprès de deux experts sélectionnés, selon des critères d'expérience et de connaissance du contexte évaluatif des enseignants, a permis de s'assurer de la clarté, de la précision et de la compréhension des vignettes.

3.5 Administration du TJPEA

L'objectif principal de cette collecte de données qualitatives a été de mieux comprendre le processus de jugement en évaluation d'experts, de praticiens et de novices par une passation orale d'un TJPEA. Le tout a été réalisé dans le cadre d'entrevues semi-dirigées d'une durée de 60 à 90 minutes enregistrées sur une bande audio. Comme l'explique Van der Maren (1996, p.312), l'entrevue est un instrument qui permet au chercheur « d'obtenir des informations sur les perceptions, les états affectifs, les jugements, les opinions, les représentations des individus à partir de leur cadre personnel de référence et par rapport à la situation présentée ». Malgré l'enregistrement audio nécessaire à cette collecte, nous avons eu le souci d'établir une relation de confiance et de mener l'entrevue dans un cadre familier en démontrant à la personne interviewée que l'on était à son écoute (Poupart, 1997). Les experts, les praticiens et les novices se sont vu présenter individuellement les quatre mêmes vignettes composées de trois items (Annexe E). À chaque item, le participant interrogé a partagé oralement son opinion face aux contenus des situations présentées et a justifié ses actions. Enfin, à la lumière de la technique d'entrevue semi-dirigée, nous avons approfondi certains aspects de réponses lorsque cela s'avérait nécessaire.

3.6 Description du questionnaire d'entrevue

Le questionnaire d'entrevue (Annexe F) a été construit à partir de notre cadre théorique portant principalement sur les caractéristiques de l'instrument ainsi que sur les pratiques évaluatives des enseignants en contexte scolaire du niveau secondaire. La période de questions comportait deux parties distinctes : (1) l'exploration du scénario pédagogique présenté et (2) la réponse aux items en lien avec ce dernier. L'ordre des questions nous a permis, dans un premier temps, d'avoir une bonne compréhension des aspects constitutifs de la vignette présentée : l'authenticité, les données manquantes, les données superflues, les données ambiguës et l'identification des concepts clés théoriques abordés.

Dans un deuxième temps, les participants ont répondu aux items en se situant sur les deux échelles de Likert proposées par l'instrument (TJPEA) et en justifiant leurs choix par la technique du « think-aloud », aussi appelée verbalisation. Inspirée des travaux de Ericsson et Simon (1980) visant à expliciter à voix haute chronologiquement les actions mentales face à une situation, chaque participant a donc été invité à commenter à des moments prédéterminés afin de permettre

l'explicitation du jugement en évaluation et optimiser le recours à la mémoire à court terme. Cela aura eu pour effet d'accroître la validité des informations recueillies (Tochon, 1996). Comme nous reconnaissons que les actions cognitives sont souvent restées inconscientes et constituent une source privilégiée d'information, car elles représentent « la seule source fiable d'inférences mettant en évidence les raisonnements effectivement mis en œuvre » (Vermersch, 2006), la prise de conscience des actions au moyen de la verbalisation a permis de les rendre explicites en vue de les conceptualiser. Étant donné que l'objectif principal de cette étude consistait à illustrer un construit intangible et complexe, nous avons pensé pertinent de se tourner vers cette méthodologie contrairement aux études basées sur le test de concordance de script s'intéressant plutôt exclusivement à la finalité du score obtenu par les participants. Ainsi, il nous serait possible, par exemple, de déceler et de comprendre si deux experts qui ont opté pour la même réponse l'ont fait pour des raisons différentes.

3.7 Validation du questionnaire d'entrevue

Une validation du questionnaire d'entrevue a été réalisée en deux temps (deux blocs de trois heures) auprès du concepteur de l'instrument du TJPEA, aussi spécialiste en évaluation des apprentissages. Ce dernier s'est prêté au même exercice que les participants. Plus spécifiquement, à chaque item, le concepteur a partagé oralement son opinion face aux contenus des situations présentées et justifié ses actions en commentant les aspects suivants : l'authenticité de la vignette, les données manquantes, les données superflues, les données ambiguës ainsi que les concepts clés théoriques abordés. À la lumière de cette première expérimentation, deux des six vignettes mises à l'essai ont été retirées. En effet, ces dernières présentaient des scénarios jugés trop simples par l'étudiante-chercheuse. De plus, comme la profondeur de la réflexion obtenue était moindre que dans le cas des autres vignettes, nous craignons ne pas atteindre l'objectif de cette recherche, soit de rendre compte du jugement évaluatif des participants.

Pour continuer, un nombre d'ajustements terminologiques a été apporté à certaines vignettes, afin de clarifier ou d'adapter la formulation employée (ex. « fin d'étape » plutôt que « fin de session »). Ces changements ont permis de s'assurer d'une bonne fluidité tout au long des entrevues à mener. Enfin, nous avons considéré cette étape de l'étude comme étant essentielle, car elle a

permis de suivre la recommandation de Van der Maren (1996) soulignant l'importance de s'entraîner aux techniques de l'entrevue pour éviter de contaminer les données.

3.8 Déroulement de l'entrevue

Toutes les entrevues ont été menées de manière individuelle à un moment et à un endroit choisi par le participant (bureau, résidence du participant ou de l'étudiante-chercheuse). La durée de chaque entrevue a oscillé entre 60 et 90 minutes. À chaque fois, nous avons utilisé un logiciel d'enregistrement audio disponible sur ordinateur afin de garder des traces des entretiens. Nous avons également pris en note les réponses des répondants à chaque item du TJPEA sur une version papier imprimée de l'instrument.

3.9 Analyse des données

La collecte de données a été réalisée par le biais d'entrevues qualitatives conduites à l'aide d'un guide d'entrevue dont les thèmes ont été délimités par les objectifs de l'étude. Une transcription partielle des enregistrements a été effectuée par traitement de texte de type *Word*. Plus précisément, la retranscription a été ciblée aux similitudes et aux différences de jugement déterminées à partir des réponses des répondants. Bien qu'une retranscription complète du *verbatim* aurait été idéale, notre choix méthodologique justifié par le cadre exploratoire de cette étude nous a tout de même permis de dégager les premières observations en lien avec notre question de recherche.

Le tableau suivant présente la grille de consignation des données recueillies. Cette dernière a été élaborée à la suite de la première expérimentation des vignettes auprès du concepteur du TJPEA.

Tableau 7
Grille de consignation des données

Lecture du scénario de la vignette 1	<u>Authenticité de la vignette</u>	<u>Données manquantes mentionnées</u>	<u>Données superflues mentionnées</u>	<u>Concepts abordés par la vignette</u>
Item 1	Réponse sur l'échelle A			Justification donnée
Item 1	Réponse sur l'échelle B			Justification donnée
Item 2	Réponse sur l'échelle A			Justification donnée
Item 2	Réponse sur l'échelle B			Justification donnée
Item 3	Réponse sur l'échelle A			Justification donnée
Item 3	Réponse sur l'échelle B			Justification donnée

3.10 Codage

Le *verbatim* des entretiens a fait l'objet d'une analyse de contenu thématique à l'aide du logiciel d'analyse de données N'Vivo 10. Comme le proposent les démarches de Van der Maren (2003) et de Miles et Huberman (2002) pour l'analyse du contenu écrit, nous avons suivi les étapes suivantes : sélection des données pertinentes pour chaque entrevue, identification des unités de sens, classement des unités de sens dans chaque catégorie de stratégies sous la stratégie qu'elle mobilise et regroupement des unités de sens selon des thèmes communs. Une logique inductive qui consiste à utiliser le cadre théorique comme guide du processus d'analyse (Savoie-Zajc, 2004) a été préconisée. L'analyse de contenu à partir des réponses des répondants a porté sur la nature des points de vue recueillis, et ce, afin de dégager la diversité des perceptions des participants sur les aspects ciblés. Une analyse verticale a ensuite été réalisée pour chacune des entretiens. Enfin, une analyse transversale a été effectuée afin de mettre en évidence les différentes données récurrentes ou divergentes au travers de l'ensemble des entretiens entre les novices, les praticiens et les experts.

Le codage des entrevues a donné lieu à quatre métacodes qui nous ont permis de travailler à la question de recherche : « À partir des réponses des participants aux vignettes du TJPEA, pourrions-nous démontrer des différences entre le jugement en évaluation des étudiants, des praticiens et celui des experts ? » La liste des codes et sous codes finaux est présentée au tableau suivant.

Tableau 8
Liste des codes finaux

Métacodes	Sous-codes
Profil de l'élève	<ul style="list-style-type: none"> • Niveau d'effort • Notes antérieures • Situation familiale • Troubles d'apprentissage • Santé de l'élève
Profil du groupe	<ul style="list-style-type: none"> • Notes antérieures du groupe • Troubles d'apprentissage • Niveau du groupe (faible, moyen ou fort)
Tâche	<ul style="list-style-type: none"> • Pondération de la tâche • Pondération des questions • Type de tâche • Type de questions • Niveau de difficulté de l'épreuve • Matière vue • Connaissances préalables • Note obtenue • Validation de la tâche • Amplitude des scores • Biais possibles • Clarté de la tâche • Qualité de la clé de correction
Contexte	<ul style="list-style-type: none"> • Moment dans l'année • Notes des autres élèves • Qualité de l'enseignement • Durée de l'épreuve • Historique comportemental du parent • Attentes formulées par le parent

L'étudiante a effectué le codage de l'ensemble des données. Compte-tenu de la quantité importante des données recueillies, un contre-codage a été effectué par la même personne pour l'ensemble des données (CCI=0,93). À ceci s'ajoutent, tout au long de l'analyse, des rencontres avec le directeur de ce mémoire qui ont permis de valider les codes et le codage. La première vague d'analyses a permis d'établir la liste des métacodes et des sous codes (Miles et Huberman, 2005). Plus précisément, des quatre métacodes se sont vus attribués des sous-codes aussi appelés « extensions ». Ces derniers ont été ajoutés afin d'illustrer les nuances possibles du métacode initial. Par exemple, le métacode « profil de l'élève » s'est vu ajouter les sous-codes suivants : niveau d'effort, situation familiale, notes antérieures, santé de l'élève, etc. Au fil du processus, d'autres opérations se sont également réalisées. Qu'il s'agisse de l'insertion (ajout d'un nouveau sous-code), de l'élimination (retrait d'un sous-code) ou de la combinaison (jumelage de deux sous-codes afin de garder le plus représentatif); elles étaient nécessaires à la bonne conduite de cette analyse.

Chapitre 4

Résultats et interprétations

« Ce que nous appelons le hasard n'est et ne peut être que la cause ignorée d'un effet connu. »
Voltaire

Cette partie du présent mémoire est consacrée à la présentation et à l'analyse des résultats. Dans le but d'illustrer le panorama des thèmes qui ressortent de l'analyse des quatre vignettes composant le test de jugement professionnel en évaluation des apprentissages (TJPEA), nous détaillerons dans quelle mesure les réponses ont permis de répondre à notre question spécifique de recherche : *À partir des réponses des participants aux vignettes du TJPEA, pourrions-nous démontrer des différences entre le jugement en évaluation des novices, des praticiens et celui des experts ?* Afin de faciliter la présentation des données, ces dernières seront organisées par vignette, puis par groupe de participants (novices, praticiens et experts). Nous pensons que ce choix facilitera la compréhension du lecteur et mettra en valeur la richesse des résultats. Nous avons également fait le choix de présenter simultanément les résultats et leurs interprétations par souci d'efficacité afin d'éviter les longueurs et les redites.

4.1 Vignette 1

Afin de mieux contextualiser les résultats qui suivront, la Figure 5 présente la première vignette composant le test de jugement professionnel en évaluation des apprentissages. Cette dernière a été lue à voix haute en plus d'être accessible à partir du support papier remis à chaque participant.

Les questions se rapportent à la mise en situation qui suit :

Lors de la comptabilisation des notes pour le bulletin, vous remarquez qu'un de vos élèves est en échec de peu en raison des notes extrêmement basses qu'il a obtenues au début de l'étape. Par contre, si vous prenez en compte ses notes plus récentes, vous jugez qu'il a fait un progrès exceptionnel.

Figure 5. Cas présenté dans le cadre de la vignette 1 du TJPEA

4.1.1 Authenticité de la vignette 1

À la question « *Trouvez-vous que la vignette présentée représente une situation possible de se dérouler en salle de classe ou dans une école ?* », tous les participants (n=12) ont répondu par l'affirmative.

4.1.2 Identification des concepts

Dans un souci de validité, nous avons posé la question suivante à chaque sujet : « Quels concepts sont abordés par le cas présenté par la vignette ? ». Dans le tableau ci-dessous, une présentation comparative permet d'illustrer les concepts disciplinaires relevés par les participants et ceux anticipés par le concepteur de l'instrument, soit un expert en évaluation des apprentissages.

Tableau 9

Tableau comparatif entre les concepts anticipés par le concepteur de la vignette 1 et ceux relevés par les trois groupes de participants (experts, praticiens et novices)

Profils	Concepteur N=1	Experts N=4	Praticiens N=4	Novices N=4
Concepts				
Planification de l'évaluation	Oui 1/1	Oui 1/4	Oui 1/4	Non
Évaluation par compétences	Oui 1/1	Oui 3/4	Oui 3/4	Non
Fonctions de l'évaluation	Oui 1/1	Oui 1/4	Non	Non
Rythme d'apprentissage	Oui 1/1	Oui 2/4	Non	Non
Valeurs	Non	Oui 3/4	Oui 1/4	Oui 2/4
Niveau de difficulté de l'épreuve	Non	Oui 3/4	Non	Non

Selon le point de vue du concepteur, quatre principaux concepts liés à l'évaluation des apprentissages émergent de la première vignette du TJPEA : la planification de l'évaluation, la pondération évolutive préconisée par l'approche par compétences, les fonctions de l'évaluation et le rythme d'apprentissage. En jetant un coup d'œil général au tableau, on dénote une plus grande

facilité pour les experts que les praticiens et les novices à répondre à la question d'identification des concepts touchés par la vignette. En effet, le groupe d'experts a réussi à relever six contenus disciplinaires différents contre trois par les praticiens et un seul par les novices. Nous jugeons que cette tendance pourrait être expliquée par le fait que ce groupe possède une meilleure connaissance des contenus théoriques (ex. savoirs). Plus spécifiquement, ce dernier doit les maîtriser, et ce, afin d'être en capacité de livrer leur enseignement de cours en évaluation des apprentissages en contexte universitaire. Aussi, tous les experts mènent des recherches scientifiques en lien avec la discipline ciblée, ce qui exige de leur part une meilleure compréhension des enjeux et des liens entre les construits de la discipline.

De leur côté, les praticiens œuvrant dans les écoles appliquent aussi des théories en évaluation, mais sans nécessairement arriver à nommer les concepts auxquels leurs pratiques sont associées. Dans le cadre de nos entrevues, ceci s'est présenté par une incompréhension du mot « concept » de la part des praticiens, puis d'une absence de réponses spontanées, lorsque confrontés à cette question. Nous sommes tout de même parvenus à recueillir des éléments de réponse en associant une vulgarisation (la description d'un concept « dans leurs mots ») à un concept en évaluation (code). La Figure 6 présente un exemple de codification réalisée à partir de vulgarisation de praticiens.

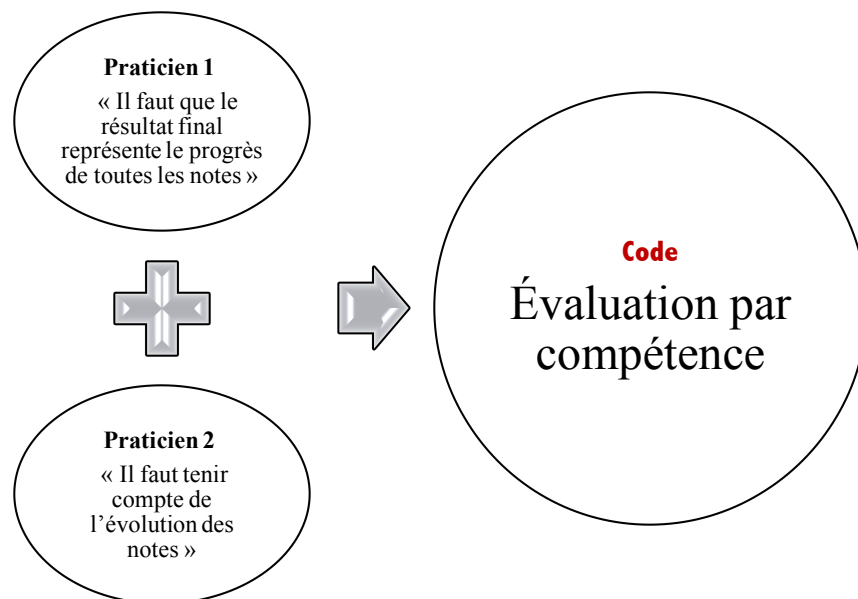


Figure 6. Exemple d'une codification de concepts

En conclusion, toujours en lien avec la première vignette, le concept identifié comme prédominant par les trois profils de participants est celui des valeurs. En effet, trois experts, un praticien et deux novices ont réussi à l'identifier. L'enjeu de la pondération évolutive par compétences arrive en second rang, après avoir été relevé par trois experts et trois praticiens. Enfin, la présence de la planification en matière de démarche en évaluation a été dénotée par un expert et un praticien.

4.1.3 Données manquantes à la vignette 1

Les figures suivantes présentent les réponses obtenues à la question « *Quelles informations non présentées par la vignette pourraient vous être utiles ?* » par groupe de répondants. Dans un premier temps, nous présentons une liste des trois principaux codes et de leurs sous-codes utilisés dans le cadre de cette partie de notre analyse. Dans un deuxième temps, nous présentons les réponses des participants à cette question par profil de répondants (novices, praticiens et experts). Dans un troisième temps, nous illustrons par une figure synthèse les différences entre les trois, et ainsi, répondons à notre question de recherche s'intéressant aux similitudes et aux différences de jugement de nos sujets.

Tableau 10

Liste des codes et des sous-codes utilisés en lien avec les données manquantes de la vignette 1 du TJPEA

<p align="center">CODE 1 Profil de l'élève</p>	<p align="center">CODE 2 Tâche finale</p>	<p align="center">CODE 3 Contexte</p>
<p>•Sous-codes 1</p> <ul style="list-style-type: none"> • Niveau d'effort • Notes antérieures • Situation familiale • Trouble d'apprentissage • Santé de l'élève 	<p>• Sous-codes 2</p> <ul style="list-style-type: none"> • Pondération de la tâche <ul style="list-style-type: none"> • Type de tâche • Niveau de difficulté de l'épreuve <ul style="list-style-type: none"> • Matière vue • Connaissances préalables <ul style="list-style-type: none"> • Note obtenue 	<p>• Sous-codes 3</p> <ul style="list-style-type: none"> • Moment dans l'année • Notes des autres élèves

À partir des réponses obtenues à cette question par tous les répondants, trois grandes catégories de données manquantes (codes principaux) se sont créées : celles liées au profil de l'élève, celles liées à la tâche et celles liées au contexte. Pour chacune de ces catégories, des sous-codes ont été identifiés. Les figures suivantes illustrent cette catégorisation par profil de répondants.

4.1.3.1 Données manquantes par les novices

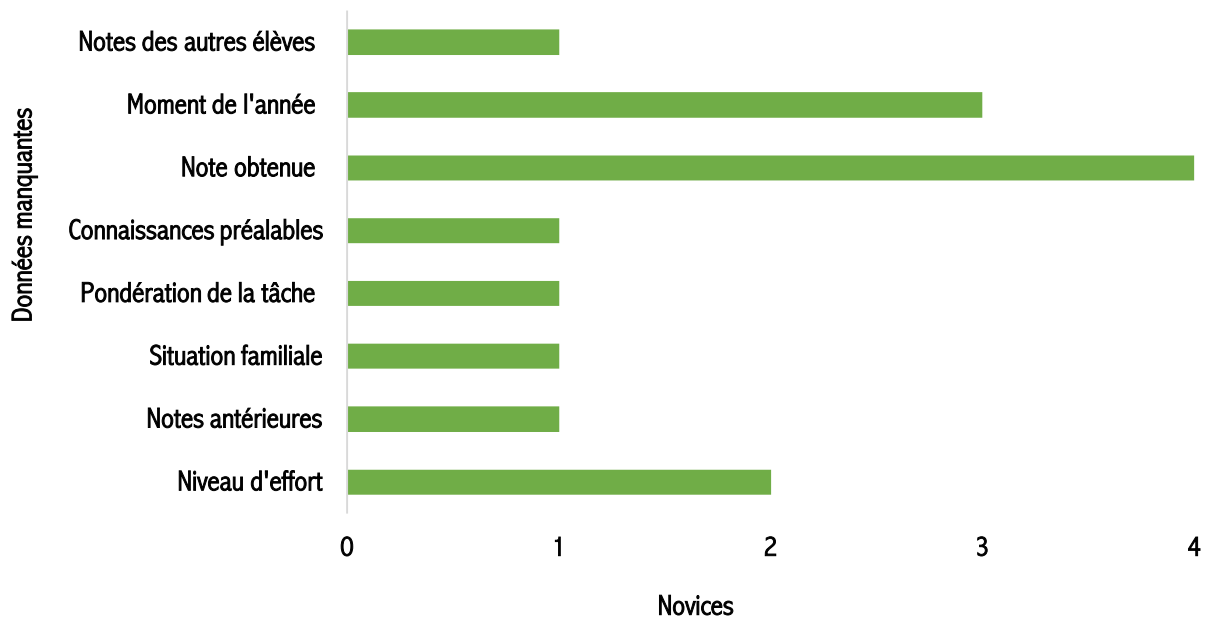


Figure 7. Données manquantes identifiées par les novices (n=4)

Parmi les données manquantes relevées par les participants novices, la note obtenue en fin d'étape (n=4), le moment de l'année où cela se déroule (n=3) et le niveau général d'effort de l'élève (n=2) ont été les plus souvent mentionnées. En effet, il semblerait que ces éléments issus de trois catégories de codes principaux différentes (tâche, contexte et profil de l'élève) pourraient influencer l'exercice du jugement de l'enseignant dans le cadre de la situation présentée par la première vignette de notre instrument.

Code 1 - Profil de l'élève

Verbatim du novice 1 en lien avec le niveau d'effort de l'élève

« J'aimerais savoir si c'est un élève travaillant... Est-ce qu'il le mérite? S'il n'a rien fait tout au long de l'étape, puis dans le cadre des deux derniers examens il a eu des bonnes notes, ça joue sur le fait qu'il mérite ou non son ajustement de notes. »

Code 2 - Tâche finale

Verbatim du novice 2 en lien avec la note obtenue par l'élève

« J'aimerais connaître la note de l'élève. Est-ce que l'élève a eu 59% et je décide de lui mettre 60%? ou s'il a eu 50% et je dois lui donner 10% de plus pour qu'il puisse passer? »

Code 3 – Contexte

Verbatim du novice 3 en lien avec le moment de l'année

« Est-ce le bulletin final? Est-ce la première étape? La vignette fait référence à une étape, mais on ne sait pas laquelle. Si c'est la première étape, c'est moins pire que l'élève vive un échec que si on parle de la dernière... et que là ça vient jouer tout le cours. Ça influence ma décision. Des fois, avoir un échec à la première étape peut motiver l'élève... »

Enfin, d'autres données manquantes ont été soulevées par une minorité de novices (n=1). Parmi celles liées au profil de l'élève (code 1), la situation familiale et les notes antérieures de l'élève ont été soulevées. Parmi celles liées à la tâche finale (code 2), l'on peut ajouter les connaissances préalables à l'exécution de la tâche finale et la pondération de cette dernière. Finalement, parmi les données manquantes souhaitées en lien avec le contexte (code 3), seules les notes des autres élèves ont été relevées.

4.1.3.2 Données manquantes par les praticiens

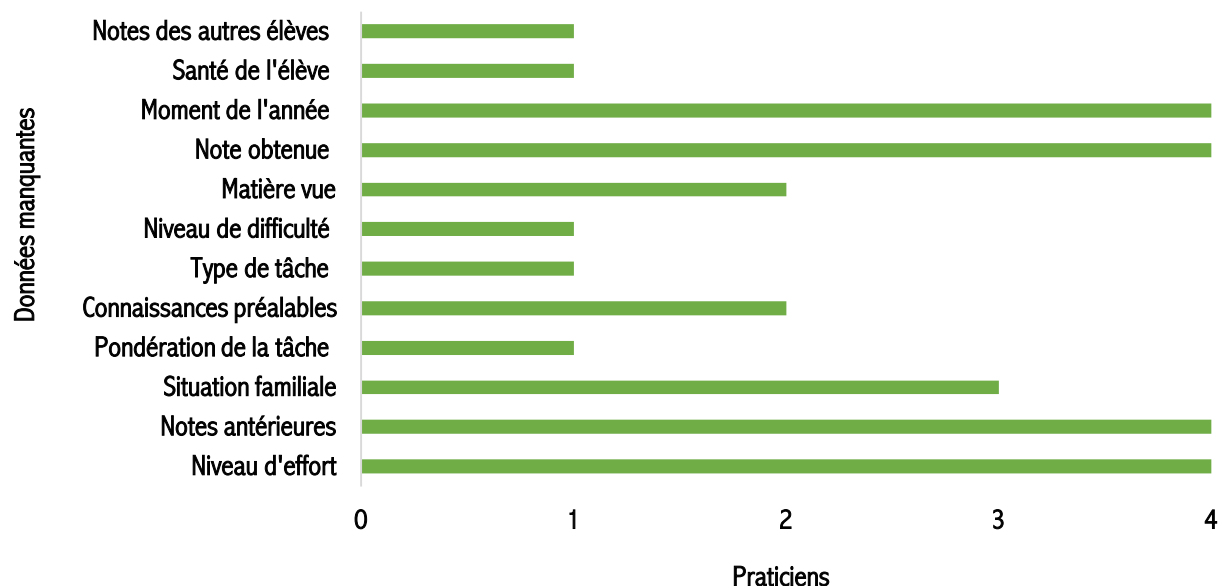


Figure 8. Données manquantes identifiées par les praticiens (n=4)

Parmi les réponses obtenues chez les praticiens concernant les données manquantes à la vignette 1, les plus fréquentes sont issues des trois grandes catégories de codes. Pour les données manquantes en lien avec le profil de l'élève, la situation familiale (n=3), le niveau général d'effort (n=4) et les notes antérieures de l'élève (n=4) sont prédominantes. Sous la catégorie des données liées à la tâche finale, la note obtenue par l'élève (n=4), la matière vue (n=2) et les connaissances préalables nécessaires à l'exécution de la tâche (n=2) sont les plus répertoriées. Enfin, parmi les données identifiées comme manquantes en lien avec le contexte, les notes des autres élèves (n=1) et le moment de l'année (n=4) prévalent.

Code 1 - Profil de l'élève

Verbatim du praticien 3 en lien avec le niveau d'effort de l'élève

« Il faudrait vérifier le comportement de l'élève durant toute l'année. C'est sûr que ça peut être choquant comme enseignant quand tu sais que l'élève a travaillé seulement aux mois de mai et de juin et que durant tout le reste de l'année il a eu un comportement désagréable. Il a peut-être nui à d'autres élèves... Ça peut aussi être un élève qui a été super gentil toute l'année... Il faut que ça soit comptabilisé. Je ne dis pas que j'agis ainsi, mais ça peut arriver. »

Code 2 - Tâche finale

Verbatim du praticien 1 en lien avec la tâche

« Ça dépend... Les notes basses sont en lien avec quels types de tâches : des projets faits en équipes ou à la maison ou encore un examen de lecture qui regroupait tout? Il faut donc analyser les travaux de l'élève. »

Code 3 – Contexte

Verbatim du praticien 3 en lien avec le moment de l'année

« Si je suis à la troisième étape, cela a plus grosse influence sur mon jugement. À la première étape, je n'aurais pas changé les notes, mais à la troisième étape, je regarde tout ce qui a été accompli pendant l'année et je me garde une petite réserve. Le temps dans l'année fait une différence sur ma décision. »

Finalement, une minorité de praticiens (n=1) ont soulevé les données manquantes suivantes : le niveau de santé de l'élève, le niveau de difficulté, le type et la pondération de la tâche ainsi que les notes obtenues par les autres élèves à la même épreuve.

4.1.3.3 Données manquantes par les experts

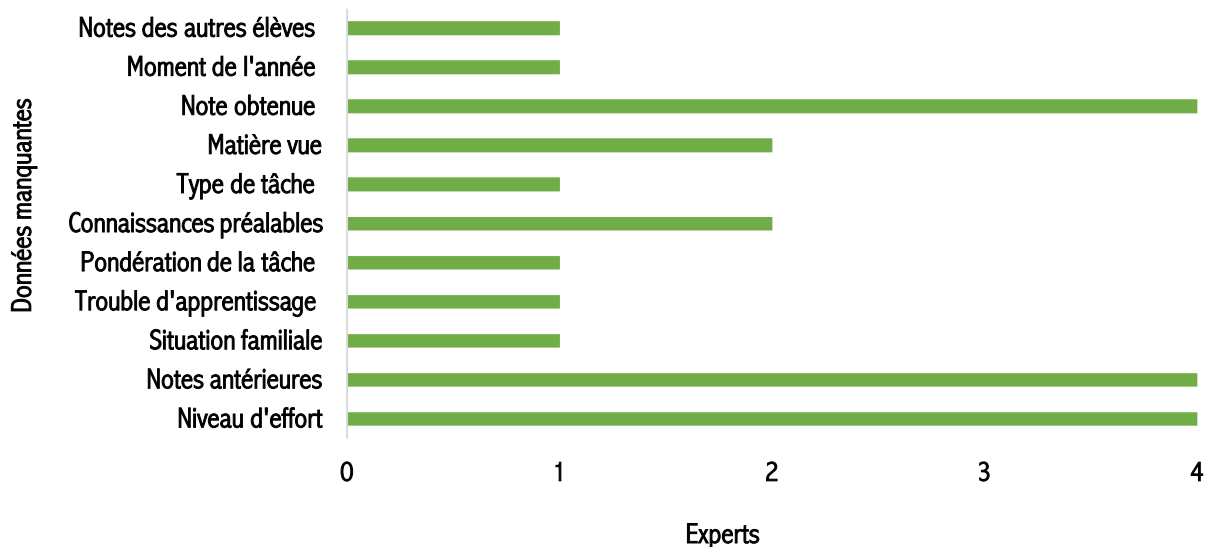


Figure 9. Données manquantes identifiées par les experts (n=4)

Parmi les données manquantes identifiées par les experts, les plus fréquentes touchent le profil de l'élève et la tâche finale. Plus précisément, les experts auraient voulu connaître le niveau d'effort général de l'élève (n=4) ainsi que ses notes antérieures (n=4). En lien avec la tâche de fin d'étape, ils auraient voulu spécifier la note obtenue par l'élève (n=4), la matière vue (n=2) ainsi que les connaissances préalables nécessaires à la réalisation de la tâche (n=2).

Code 1 - Profil de l'élève

Verbatim de l'expert 1 en lien avec les notes antérieures de l'élève

« Qu'est-ce que cela veut dire des *notes extrêmement basses*? Pour certains enseignants et certains élèves cela peut être 50%, alors que pour d'autres c'est 30%... Pour ce qui est du progrès, on n'a pas besoin de notes là-dessus. Dans les deux cas, c'est parlant. »

Code 2 - Tâche finale

Verbatim de l'expert 2 en lien avec la note obtenue par l'élève

« J'aurais aimé connaître le seuil de réussite. Est-ce 50%? Est-ce 60%? Aussi, est-ce que l'élève est en échec à 58% ou à 35%? Est-ce juste un élève qui est juste passé à côté ou est-ce un élève qui n'a vraiment pas compris? »

Enfin, d'autres données se sont qualifiées comme étant manquantes par une minorité d'experts (n=1) : la situation familiale de l'élève, la présence d'un trouble d'apprentissage, le type de tâche finale et sa pondération, les notes des autres élèves et le moment de l'année de la vignette.

4.1.3.4 Synthèse comparative des données manquantes par profil de participants

Afin de répondre à notre question de recherche « *À partir des réponses des participants aux vignettes du TJPEA, pourrions-nous démontrer des différences entre le jugement en évaluation des novices, des praticiens et celui des experts ?* », nous avons jugé pertinent de s'attarder à la synthèse des tableaux précédents, et ce, à la lumière des différences et des similitudes des données manquantes identifiées par les trois groupes de sujets.

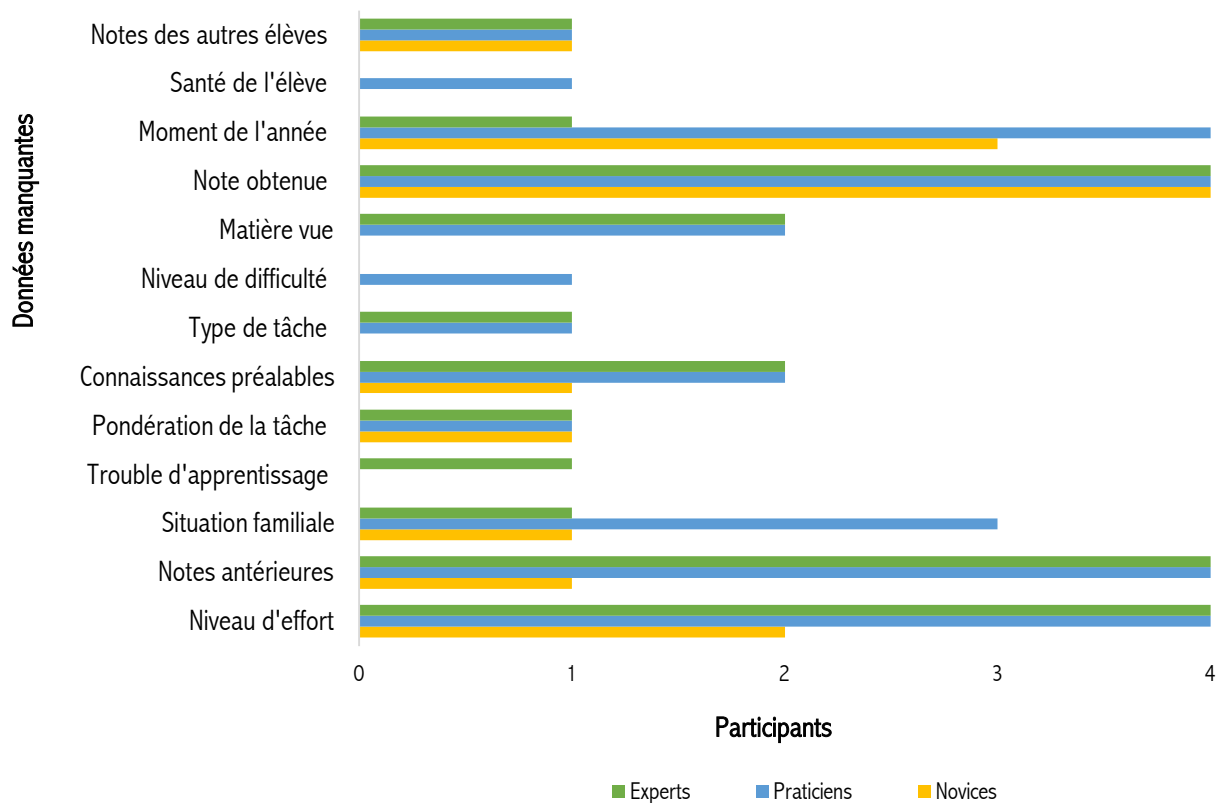


Figure 10. Synthèse des données manquantes identifiées par tous les participants (n=12)

Comme nous le témoigne la Figure 10, « la note obtenue de l'élève à la réalisation de la tâche finale », est une donnée manquante pour laquelle tous les participants (n=12) ont montré de l'intérêt. Malgré cette première régularité, de manière générale, les novices identifient un moins grand nombre de variables manquantes (n=8 données) que le groupe de praticiens (n=12 données) et d'experts (n=11 données). Bien qu'une meilleure capacité à identifier les concepts par les termes exacts par les experts ait été annoncée, le nombre de données relevées par les praticiens est quasi identique (n=12). Comparativement au groupe de novices pour qui le nombre de données repérées est plus faible (n=8), les premières observations de ces résultats nous poussent à croire qu'il existe une possible différence du niveau de prise en compte des variables susceptibles d'influencer le jugement professionnel de l'enseignant entre les novices et les deux autres groupes (praticiens et experts).

Pour continuer, « le niveau d'effort de l'élève » a aussi été relevé par un plus grand nombre d'experts (n=4) et de praticiens (n=4) que de novices (n=2). Les « notes antérieures de l'élève » suivent la même logique avec une identification opérée par un seul novice, quatre praticiens et quatre experts. Avec des distributions de fréquences pratiquement identiques, « la situation familiale de l'élève » (un novice, trois praticiens et un expert) et les « connaissances préalables requises » (un novice, deux praticiens et deux experts) démontrent également une distinction entre les novices et les deux autres groupes.

En contrepartie, qu'en est-il des différences possibles entre les praticiens et les experts? Se manifestent-elles dans le cadre de cette identification de données manquantes issues de la première vignette de notre instrument? Au premier regard, non. En effet, à l'exception de la « présence d'un trouble d'apprentissage », toutes les données manquantes relevées par les experts l'ont aussi été par les praticiens. Enfin, deux données ont été exclusivement signalées par les praticiens (« niveau de santé de l'élève » et « niveau de difficulté de la tâche finale »).

4.1.4 Item 1 de la première vignette

Toujours en lien avec la première vignette du TJPEA, voici le premier item qui a été présenté aux participants.

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Vous décidez de modifier la pondération des évaluations de cet élève afin d'augmenter le poids des plus récentes.		L'élève termine son étape avec une bien meilleure note et sa confiance augmente drastiquement.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Dans un premier temps, les sujets devaient se situer sur l'échelle de type Likert présentée en lien avec la pertinence de l'intervention indiquée. La figure suivante illustre bien la distribution des réponses obtenues. Plus spécifiquement, une majorité de novices (3/4 novices) ont jugé « très pertinente » l'intervention présentée contre un seul répondant novice la jugeant plutôt de « non pertinente ». De leur côté, les praticiens se sont départagés entre « pertinente » (2/4 praticiens) et « peu pertinente » (2/4 praticiens). Enfin, un expert (1/4) a jugé l'intervention « très pertinente », un autre « pertinente » (1/4) et deux autres (2/4) « non pertinente ».

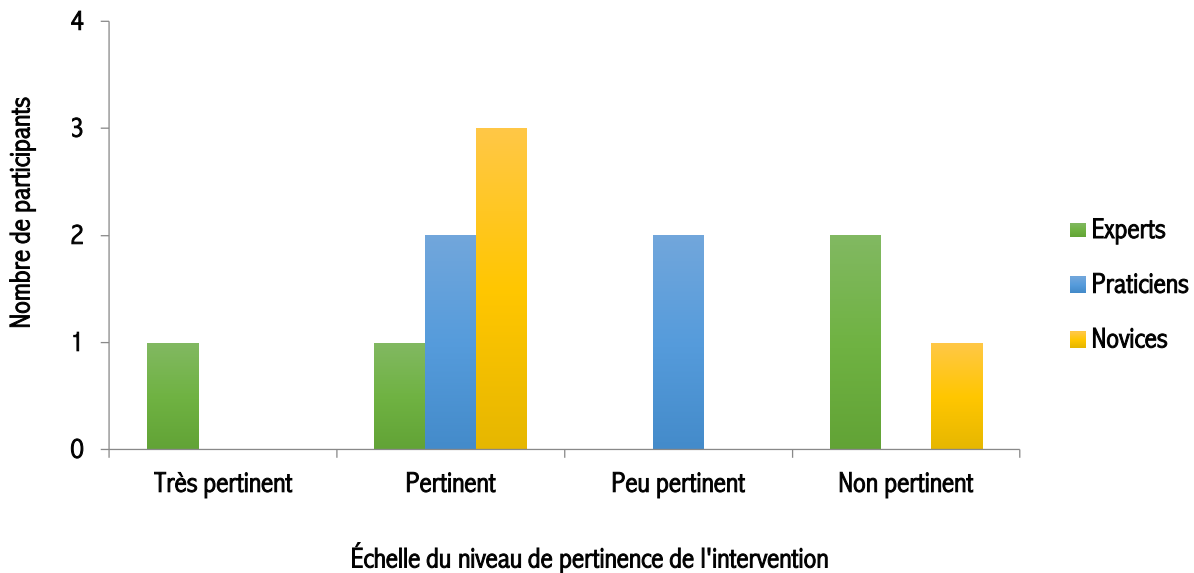


Figure 11. Répartition des réponses des profils de répondants (n=12) à l'échelle A

Au premier regard, nous remarquons surtout qu'il y existe une grande variété de réponses chez les participants. Le groupe d'experts a opté pour trois différents niveaux de réponses, tandis que les novices se rapprochent plus du consensus (3/4 novices optent pour l'option « pertinent »). Pour leur part, les praticiens se sont placés plus au centre du continuum que propose l'échelle de Likert de l'outil.

Les extraits suivants démontrent plus spécifiquement ce qui a accroché les participants. Il semblerait qu'une majorité d'experts et de praticiens considèrent l'intervention proposée comme « injuste » pour les autres élèves du groupe.

Verbatim du praticien 2 en lien avec l'item 1

« Si je le fais pour lui, je le fais pour tout le monde. Je ne pense pas que je le ferais pour un seul élève. Je lui demanderais un travail supplémentaire plutôt. »

Verbatim de l'expert 1 en lien avec l'item 1

« Les valeurs de la *Politique d'évaluation* font référence à l'équité. Si je fais ça, il faut que je le fasse pour tout le monde. Sinon, il faut que je trouve quelque chose qui me dit que cet élève-là avait des caractéristiques tellement

différentes de tout le monde que cela peut justifier que c'est un cas unique dans lequel je peux le faire. »

Dans un deuxième temps, les répondants devaient également se situer sur la deuxième échelle en lien avec la probabilité à réaliser l'action proposée par l'item. Cette deuxième question nous a semblé pertinente, et ce, afin d'éviter l'influence de désirabilité sociale.

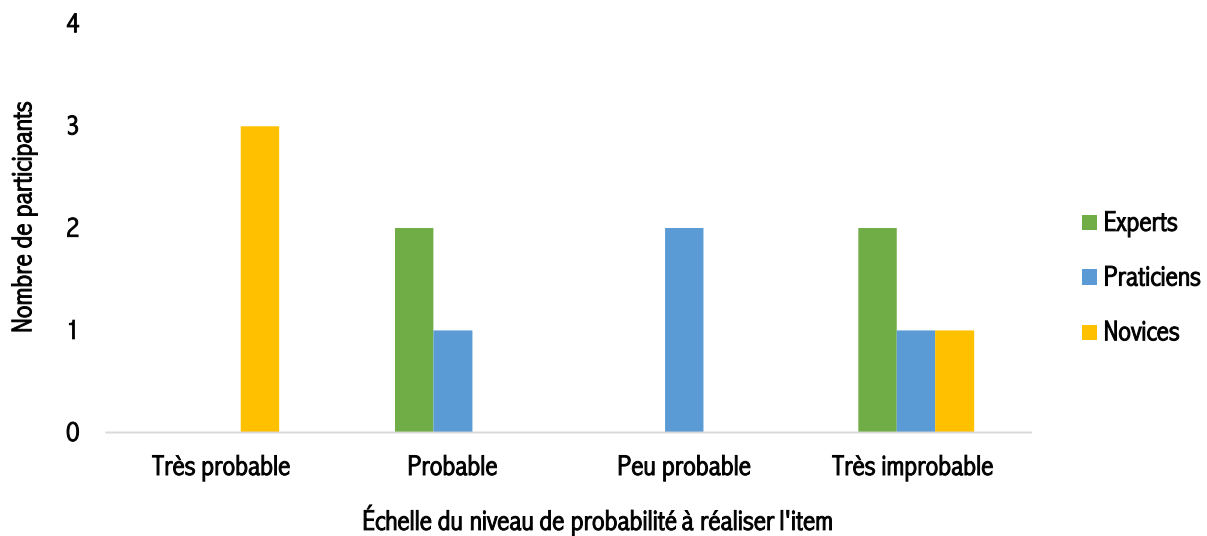


Figure 12. Répartition des réponses des répondants (n=12) à l'échelle B par profil

La répartition des réponses des répondants démontre que trois novices ont affirmé « très probable » qu'ils décident de modifier la pondération comme le propose l'item 1, contre un seul « très improbable ». Les praticiens sont quant à eux plus partagés : un a jugé « probable », deux « peu probable » et un « très improbable ». Chez les experts, deux clans se sont formés : la moitié d'entre eux (n=2) a dit « probable » et l'autre (n=2) « très improbable ». Encore une fois, les répondants qui se sont sentis moins interpellés par l'intervention de l'item 1 justifiaient leur choix par un manque d'équité entre les élèves. Ainsi, cette valeur semble moins forte chez les novices que chez les deux autres groupes qui ont été plus déchirés.

Verbatim du novice 4 en lien avec l'item 1

« Quelqu'un qui fait des efforts, je trouve ça encore plus beau qu'un élève qui traîne. Je trouve qu'il mérite qu'on change sa note, même si ce n'est pas évident... »

Verbatim de l'expert 1 en lien avec l'item 1

« Il y a toujours ces deux principes d'égalité et d'équité... Il faudrait le faire pour l'ensemble du groupe d'élèves, à moins que l'élève en question ait un plan d'intervention. »

4.1.5 Item 2 de la première vignette

Voici le deuxième item présenté aux participants, et ce, en lien avec la première vignette. Vous remarquerez que la première partie en lien avec l'action proposée est identique au premier item. Par contre, ici, la conséquence présentée est différente. Ainsi, cette répercussion nouvelle présentée aux répondants pourrait (ou non) influencer le jugement exercé. Voyons de plus près les résultats obtenus.

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Vous décidez de modifier la pondération des évaluations de cet élève afin d'augmenter le poids des plus récentes.		La situation se fait apprendre par des élèves et des parents de la classe réclament des traitements similaires.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

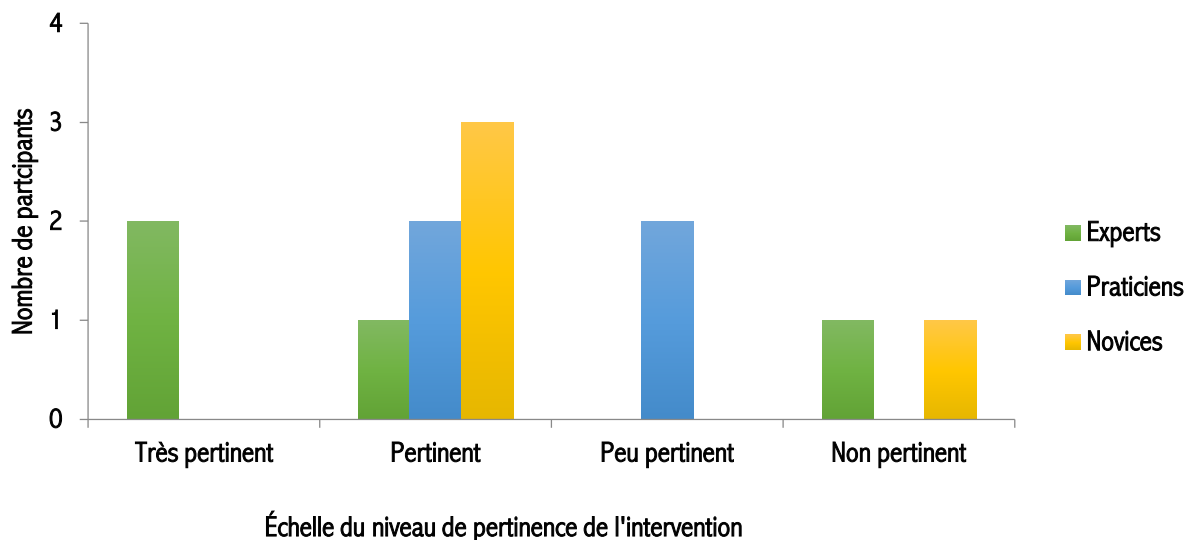


Figure 13. Répartition des réponses des répondants (n=12) à l'échelle A par profil

La répartition des réponses obtenues par le groupe d'experts est la suivante : deux experts considèrent toujours l'intervention proposée par l'item 2 « très pertinente », un expert la considère « pertinente » et un autre « non pertinente ». Dans une perspective comparative entre l'item 1 et l'item 2 dont l'action proposée en première partie de l'item est la même, mais dont la conséquence est différente (2^e partie), il semblerait que l'item 2 « gagne » un expert sur l'échelle de pertinence en proposant d'offrir un traitement similaire aux autres élèves de la classe. Ainsi, un expert est passé de l'option « très impertinent » à celle de « très pertinent » en se basant sur la valeur de l'équité des traitements offerts à l'ensemble des élèves. Enfin, pour tous les autres répondants, la distinction présentée par l'item 2 n'a aucunement changé leur livraison de réponses. Ainsi, la distribution reste identique à celle de l'item 1, à l'exception de la réponse de l'expert dont nous venons de faire mention.

4.1.6 Item 3 de la première vignette

Enfin, le troisième item en lien avec la première vignette traite entre autres de l'influence de la direction d'école dans l'octroi des résultats des élèves.

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Votre direction vous indique que ce genre de modifications ne sont pas acceptables et que vous ne pouvez rien changer aux résultats. Vous suivez les directives obtenues.		L'élève termine son étape avec un échec et semble avoir un déclin d'intérêt et de motivation face à l'école.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

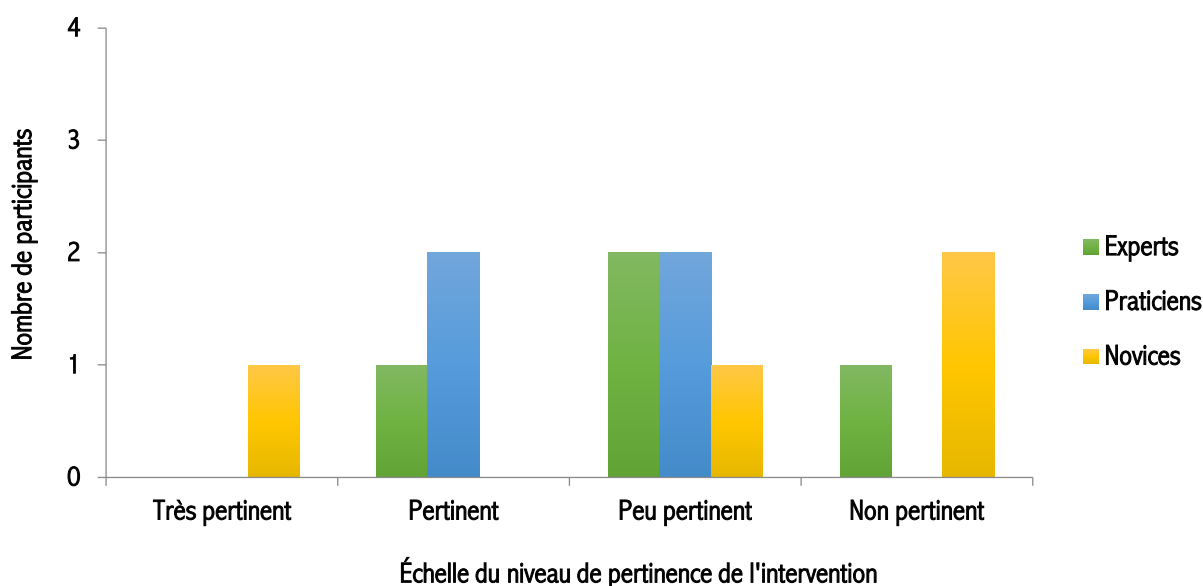


Figure 14. Répartition des réponses des répondants (n=12) à l'échelle A par profil

Au premier regard, la distribution des réponses des participants semble assez inégale. Chez les novices, la majorité (n=2) considère l'item « non pertinent » contre un novice qui le juge « peu pertinent » et un autre « très pertinent ». Chez les experts, la même disparité se fait remarquer : deux experts jugent l'item « peu pertinent » contre un « pertinent » et un autre « non pertinent ». Enfin, la moitié (n=2) des praticiens affirment que l'action présentée par l'item 1 est « pertinente » contre l'autre moitié (n=2) qui la juge « peu pertinente ».

Pour continuer, toutes catégories confondues, le plus grand nombre de répondants (deux experts, deux praticiens et un novice) jugent l'intervention « peu pertinente ». Les propos justificatifs recueillis semblent indiquer que ces participants reconnaissent l'influence du pouvoir décisionnel de la direction dans un contexte de jugement « certificatif », et ce, même si cela va à l'encontre de ce que l'enseignant pourrait juger favorable pour l'élève.

Verbatim de l'expert 4 en lien avec l'item 3

« La réaction de la direction est pertinente si l'élève n'a pas de plan d'intervention. Malheureusement, il n'a pas le choix. Une fois que la direction donne la ligne de conduite, l'enseignant doit agir tel que demandé. »

Verbatim de l'expert 2 en lien avec l'item 3

« Ce n'est pas à la direction de se substituer à l'enseignant. Principalement, c'est l'enseignant qui est responsable de l'évaluation des élèves et des moyens qui sont mis en œuvre. C'est lui qui observe. À mon avis, ce n'est pas une intervention pertinente de la part de la direction. J'expliquerais ma position au directeur, mais je ne suis pas certain que j'irais jusqu'à l'insubordination. »

Verbatim de l'expert 3 en lien avec l'item 3

« Quand le prof n'est pas capable d'argumenter, cela peut être en lien avec n'importe quels biais... Je ne considère pas que la direction possède un niveau professionnel plus haut que l'enseignant. Elle a des pouvoirs comme gestionnaire, mais elle n'a pas la compétence plus élevée que l'enseignant en évaluation. »

Enfin, la Figure 15 présente la distribution des réponses des répondants selon le niveau de probabilité à réaliser l'item 3, soit de ne rien faire face au résultat de l'élève à la suite des directives reçues par la direction de l'école. Dans un premier temps, la moitié des experts (n=2) jugent « très improbable » d'agir de la sorte, contre un expert qui considère le tout « peu probable » et un autre « très improbable ». Dans un deuxième temps, la majorité des praticiens (n=3) affirme « très probable » qu'ils agissent ainsi, contre un seul (n=1) « peu probable ». Dans un dernier temps, la

majorité des novices (n=3) indique leur réponse comme étant « peu probable » contre un seul « très probable ».

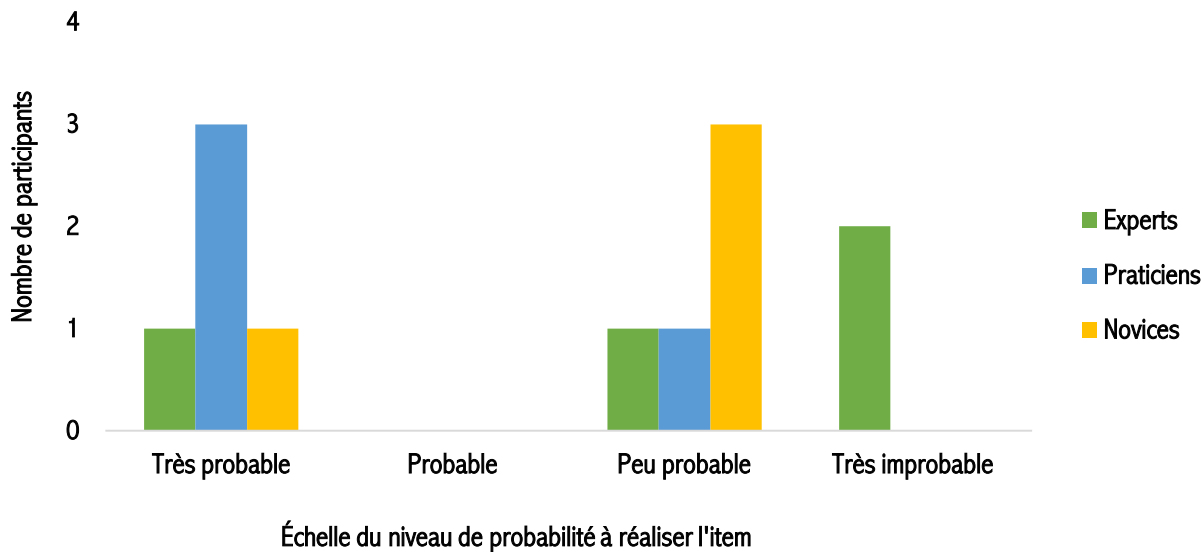


Figure 15. Répartition des réponses des répondants (n=12) à l'échelle B par profil

Il est intéressant de constater que les deux profils de répondants « hors terrain », c'est-à-dire n'œuvrant pas dans les écoles (experts et novices) semblent plus contestataires face aux directives de la direction. Plus précisément, trois novices, un praticien et trois experts soulignent l'importance que la décision face à la performance d'un élève revienne à l'enseignant, et ce, peu importe les contraintes administratives auxquelles sont confrontées les équipes de direction. À l'inverse, les praticiens semblent plus « refroidis » à l'idée de confronter leur supérieur hiérarchique. En effet, trois praticiens sur quatre disent vouloir se conformer à la décision rendue par l'administration. De ce fait, la variable « pouvoir administratif » semble influencer ce profil de répondants. Notons que nos praticiens étaient tous issus d'écoles secondaires différentes.

Verbatim du praticien 2 en lien avec l'item 3

« Je ne serais pas content que la direction prenne une décision dans ma classe, mais je ne me serais pas obstiné. Je suis comme ça. Quand la direction prend une décision et que me l'on explique... que je sois d'accord ou pas ça ne changera rien de toute façon.

« Ça prend des motifs exceptionnels pour changer une pondération. La direction cherche à uniformiser. En contexte de force majeure... quand c'est justifié comme à la fin de l'année, ça pourrait arriver, mais c'est des cas uniques. La direction fait le bon choix; c'est très pertinent. »

4.1.7 Synthèse de la vignette 1

En conclusion, la première vignette permet de déceler une meilleure aptitude chez les experts que chez les novices et les praticiens à identifier les concepts abordés par le scénario ainsi que les données manquantes susceptibles d'influencer l'exercice du jugement en évaluation. À la lecture des résultats obtenus aux deux premiers items, une correspondance de raisonnement est observée chez les praticiens et les experts par le soulignement du concept de l'équité et de son rôle à jouer dans la cadre de cette vignette. Cette observation n'a pas été repérée par les novices. Le troisième item présente une similitude entre les experts et les novices de « contestation face à l'équipe de direction » quant au rôle de l'enseignant dans l'exercice du jugement professionnel. De leur côté, les praticiens semblent moins susceptibles de critiquer la décision de leur patron, et ce, même s'ils laissent sous-entendre qu'ils sont les mieux placés pour évaluer. Enfin, la première vignette permet de rendre compte du jugement professionnel en évaluation des sujets et d'en déceler certaines similitudes ou différences entre les profils de répondants.

4.2 Vignette 2

La Figure 16 présente la deuxième vignette composant notre test de jugement professionnel en évaluation des apprentissages. Nous rappelons que cette dernière a été lue à voix haute en plus d'être accessible à partir du support papier remis à chaque participant.

Les questions se rapportent à la mise en situation qui suit :

À la suite de la correction d'une évaluation sommative, la distribution des notes d'un groupe d'élèves s'éloigne substantiellement de ce à quoi vous êtes habitué.

Figure 16. Cas présenté dans le cadre de la vignette 2 du TJPEA

4.2.1 Authenticité de la vignette 2

À la question « *Trouvez-vous que la vignette présentée représente une situation possible de se dérouler en salle de classe ou dans une école ?* », tous les participants (n=12) ont répondu par l'affirmative.

4.2.2 Identification des concepts

Dans un souci de validité, nous avons posé la question suivante à chaque sujet : « *Quels concepts sont abordés par le cas présenté par la vignette ?* ». Le prochain tableau dresse les concepts disciplinaires relevés par les participants et ceux anticipés par le concepteur de l'instrument, soit un expert en évaluation des apprentissages.

Tableau 11

Tableau comparatif entre les concepts anticipés par le concepteur de la vignette 2 et ceux relevés par les trois groupes de participants (experts, praticiens et novices)

Profils	Concepteur N=1	Experts N=4	Praticiens N=4	Novices N=4
Concepts				
Analyse d'items	Oui 1/1	Oui 3/4	Oui 1/4	Oui 1/4
Conception d'items	Oui 1/1	Oui 2/4	Non	Oui 1/4
Contexte de passation de l'épreuve	Oui 1/1	Oui 1/4	Oui 2/4	Non
Validité de l'épreuve	Oui 1/1	Oui 2/4	Oui 3/4	Non
Intérêt pour la matière visée	Non	Non	Non	Oui 1/4

Selon le point de vue du concepteur, quatre principaux concepts liés à l'évaluation des apprentissages émergent de la deuxième vignette du TJPEA : l'analyse d'items, la conception d'items, la validité de l'épreuve et son contexte de passation. Dans une perspective comparative, le Tableau 10 présente la fréquence des concepts qui ont aussi été relevés par chacun des groupes (novices, praticiens et experts). Encore une fois, on dénote une plus grande facilité pour les experts

que les praticiens et les novices à répondre à la question d'identification des concepts touchés par la vignette. En effet, le groupe d'experts a réussi à relever quatre contenus disciplinaires différents contre trois pour les praticiens et les novices. De surcroît, l'analyse des réponses obtenues a permis de déterminer que le groupe d'experts a associé à ces quatre concepts généraux sept sous-concepts (biais, degré de difficulté de l'item, distribution des scores, prévisibilité de l'épreuve, types de questions, clarté des items et la qualité de la clé de correction). Les praticiens en ont identifié qu'un seul (biais), tandis que les novices en ont relevé trois (degré de difficulté de l'item, types de questions et clarté des items).

Pour continuer, le concept reconnu par les trois profils de participants est celui de l'analyse d'items. En effet, trois experts, un praticien et un novice ont réussi à l'identifier. L'enjeu de la validité arrive en second rang, et ce, après avoir été relevé par trois praticiens et deux experts. Non loin derrière, la conception d'items semble influencer trois experts et un novice. Le contexte de la passation de l'épreuve, quant à lui, a été une réponse émergente du discours de deux praticiens et d'un expert. Enfin, un seul novice a fait la mention de l'intérêt des élèves pour la matière traitée par l'épreuve.

4.2.3 Données manquantes à la vignette 2

Les figures suivantes présentent les réponses obtenues à la question « *Quelles informations non présentées par la vignette pourraient vous être utiles ?* » par groupe de répondants. Dans un premier temps, nous présentons une liste des trois principaux codes et de leurs sous-codes utilisés dans le cadre de cette partie de notre analyse. Dans un deuxième temps, nous présentons les réponses des participants à cette question par profil de répondants (novices, praticiens et experts). Dans un dernier temps, nous illustrons par une figure synthèse les différences entre les trois, et ainsi, répondons à notre question de recherche s'intéressant aux similitudes et aux différences de jugement de nos sujets.

Premièrement, à partir des réponses obtenues par les répondants, trois grandes catégories de données manquantes (codes principaux) se sont créées : celles liées au profil du groupe d'élèves, celles liées à la tâche et celles liées au contexte. Pour chacune de ces catégories, des sous-codes ont été identifiés. Le tableau suivant présente ces listes.

Tableau 12

Liste des codes et des sous-codes utilisés en lien avec les données manquantes de la vignette 2 du TJPEA

CODE 1 Profil du groupe	CODE 2 Tâche	CODE 3 Contexte
<ul style="list-style-type: none">• Sous-codes 1• Notes antérieures• Troubles d'apprentissage	<ul style="list-style-type: none">• Sous-codes 2• Validation de la tâche• Types de questions• Amplitude des scores• Niveau de difficulté• Biais possibles	<ul style="list-style-type: none">• Sous-codes 3• Qualité de l'enseignement• Moment de l'année

4.2.3.1 Données manquantes par les novices

Parmi les données manquantes souhaitées par les participants novices, la qualité de l'enseignement offert ainsi que le moment dans l'année scolaire semblent les plus fréquentes (n=2/donnée manquante). Pour continuer, d'autres données ont aussi été identifiées comme étant manquantes par un novice, soit le niveau de difficulté de la tâche, l'amplitude des scores obtenus par les élèves, le processus de validation de la tâche ainsi que l'historique des notes antérieures des élèves. Enfin, malgré une bonne variété de données manquantes ciblées (n=6) par ce groupe de répondants, l'on peut remarquer qu'aucune d'entre elles n'a été relevée par plus de la moitié des novices (n=2).

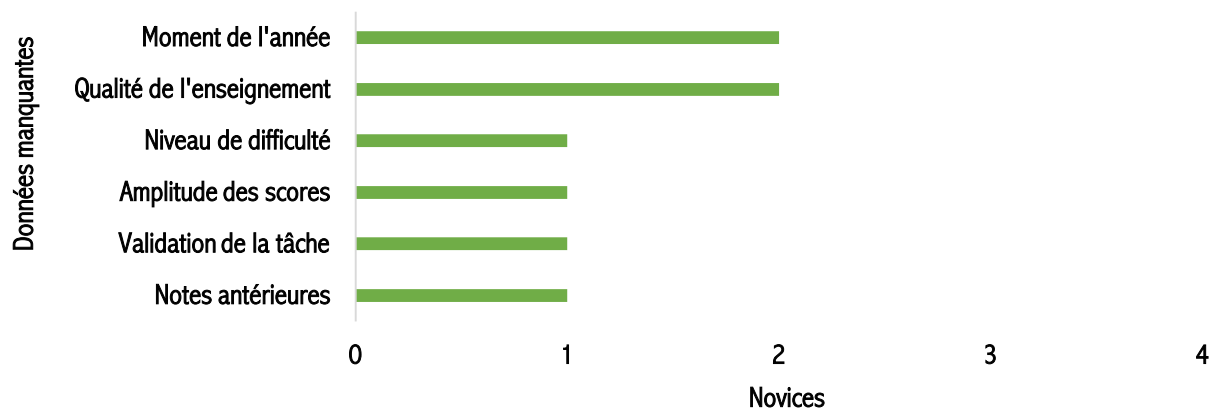


Figure 17. Données manquantes identifiées par les novices (n=4)

4.2.3.2 Données manquantes par les praticiens

Parmi les réponses obtenues chez les praticiens concernant les données manquantes à la vignette 2, les plus fréquents sont issues des trois grandes catégories de codes. Plus précisément, les notes antérieures des élèves (n=4), l'amplitude des scores obtenus (n=4), la qualité de l'enseignement (n=3), les biais possibles (n=3), les types de questions proposées par la tâche (n=3) et le processus de validation de l'épreuve sont les plus prédominants. En complémentarité, deux praticiens auraient voulu en connaître davantage sur le niveau de difficulté de l'épreuve ainsi que sur la présence (ou non) de troubles d'apprentissage parmi les élèves de la classe. Enfin, un seul praticien aurait souhaité connaître le moment dans l'année durant lequel se déroule la situation présentée par la vignette 2.

Au total, neuf données manquantes ont été relevées par le groupe de praticiens, dont la majorité d'entre elles l'ont été par trois ou quatre praticiens du groupe.

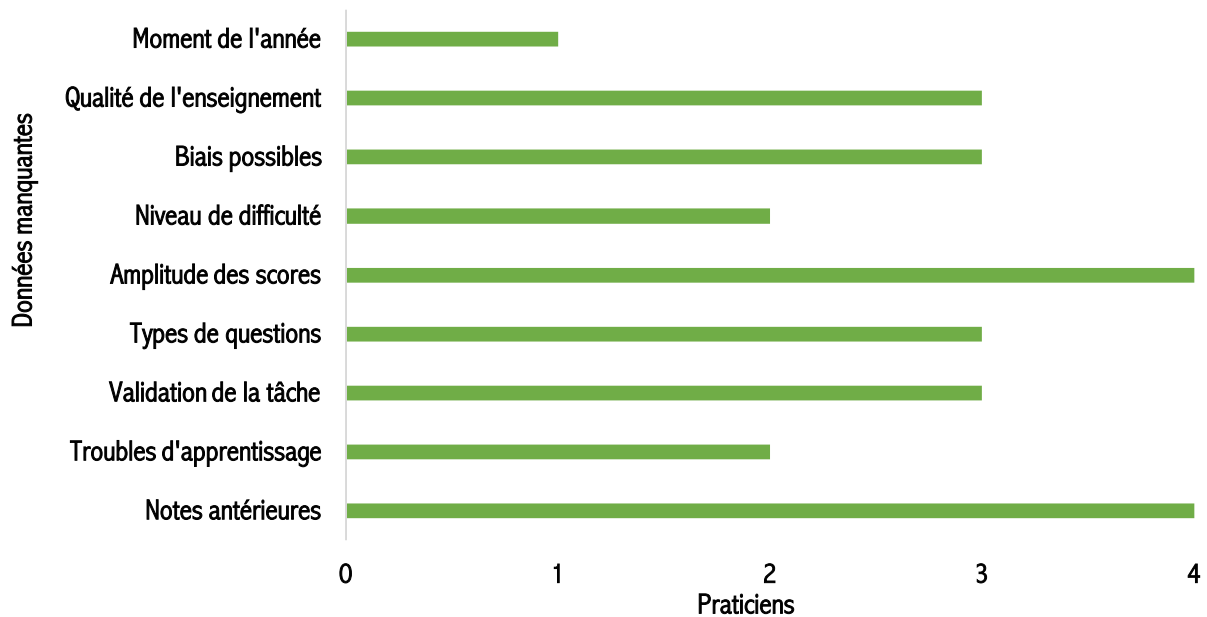


Figure 18. Données manquantes identifiées par les praticiens (n=4)

4.2.3.3 Données manquantes par les experts

Parmi les données manquantes identifiées par les experts, les plus fréquentes touchent le niveau de difficulté de la tâche (n=4), la qualité de l'enseignement (n=3) ainsi le processus de validation de la tâche présentée aux élèves (n=3). Pour continuer, la moitié des répondants de ce groupe (n=2) a mentionné l'influence possible de l'amplitude des scores obtenus par les autres élèves et les types de questions qui composent l'épreuve. Enfin, un seul expert a fait la mention du moment de l'année ainsi que les notes antérieures des élèves concernés.

Au total, sept données manquantes ont été relevées par ce groupe dont la majorité d'entre elles ont été mentionnées par plus de la moitié des répondants experts.

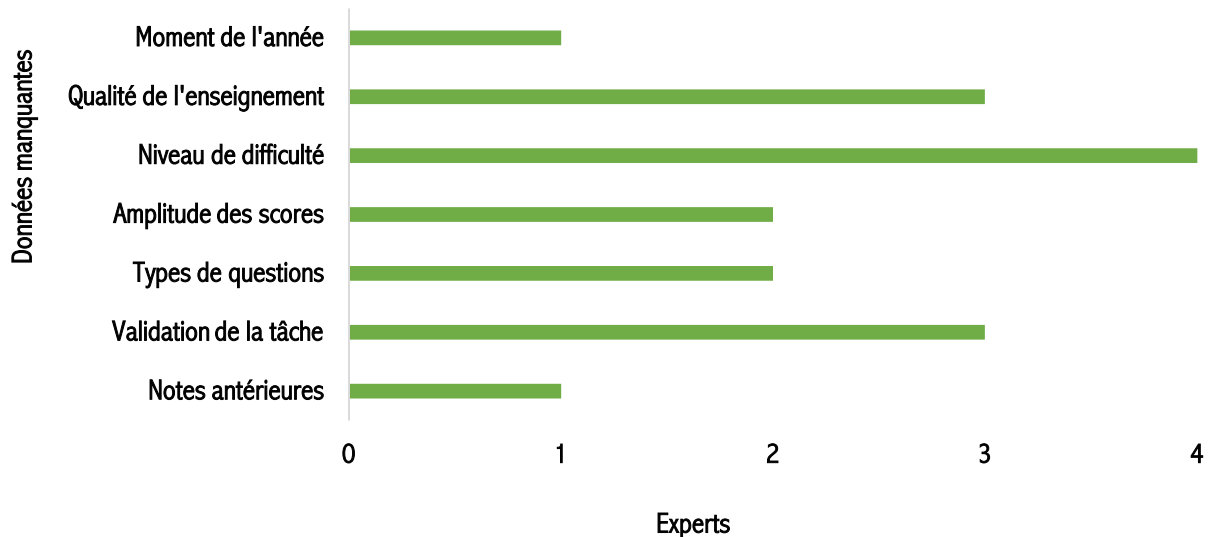


Figure 19. Données manquantes identifiées par les experts (n=4)

4.2.3.4 Synthèse comparative des données manquantes par profil de participants

Afin de répondre à notre question de recherche « À partir des réponses des participants aux vignettes du TJPEA, pourrions-nous démontrer des différences entre le jugement en évaluation des novices, des praticiens et celui des experts ? », nous présentons la synthèse des tableaux précédents, et ce, à la lumière des différences et des similitudes des données manquantes identifiées par les trois groupes de sujets.

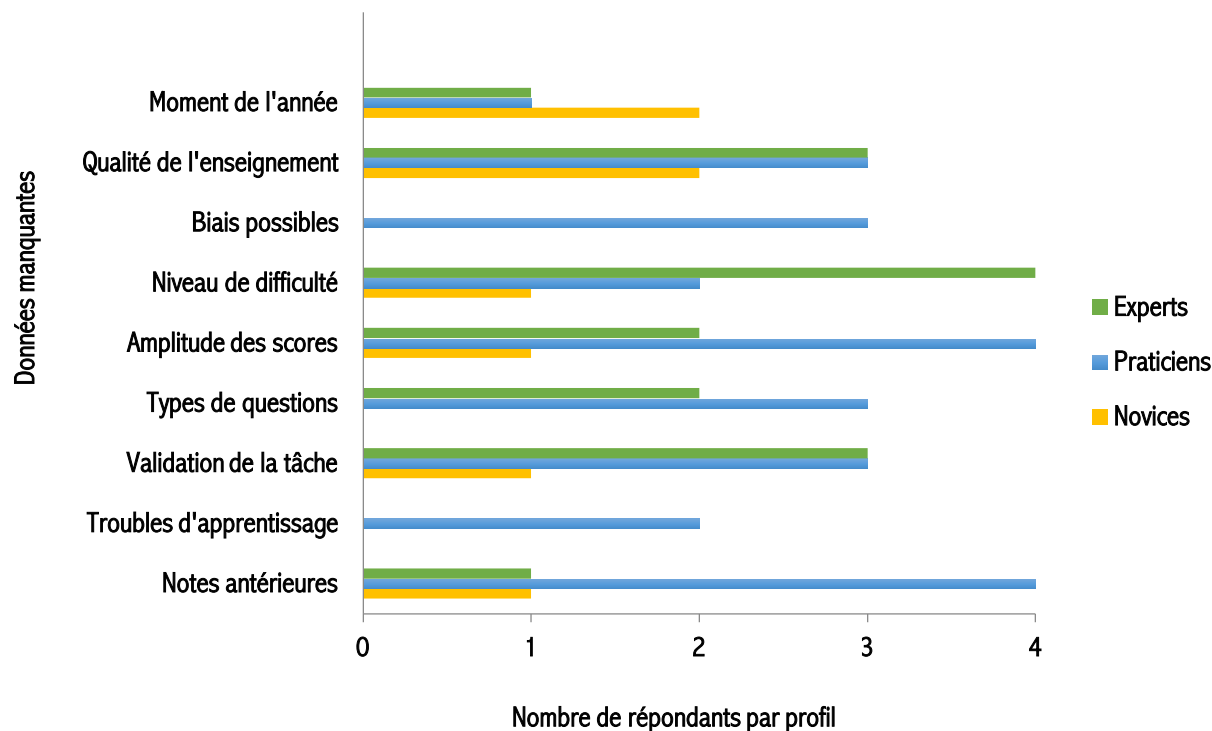


Figure 20. Synthèse des données manquantes identifiées par tous les participants (n=12)

Comme nous le témoigne la Figure 20, certaines données manquantes ont été relevées par les trois profils de répondants : le moment de l'année, la qualité de l'enseignement, le niveau de difficulté de l'épreuve, l'amplitude des scores, le processus de validation de la tâche et les notes antérieures des élèves. Malgré cette constatation, nous remarquons qu'à l'exception « du moment de l'année », les novices sont toujours moins nombreux que les deux autres profils de répondants à avoir relevé l'information. Par exemple, pour la donnée manquante « niveau de difficulté de la tâche », quatre experts l'ont soulignée contre deux praticiens et un seul novice. Le même scénario se répète avec la « validation de la tâche » pour laquelle trois experts et trois praticiens l'ont identifiée, contre un seul novice.

Pour continuer, malgré le fait que les trois profils de participants ont identifié un nombre total similaire de données manquantes (six pour les novices, neuf pour les praticiens et sept pour les experts), la fréquence relevée par participant nous précise un détail important. Sous cet angle, les novices signalent l'absence d'information huit fois, les praticiens 25 et les experts 16. Ainsi, les praticiens semblent faire partie du groupe de répondants capables de nommer un plus grand

nombre de facteurs susceptibles d'influencer leur jugement face à la situation présentée à la vignette deux. Les experts sont non loin derrière avec neuf fréquences de moins. Finalement, les novices ont 17 fréquences de moins que le groupe prédominant, soit les praticiens.

Tout comme dans le cas de la première vignette de ce TJPEA, les résultats obtenus aux données manquantes de la deuxième vignette nous poussent à croire qu'il existe bien une différence de niveau d'habileté à identifier les facteurs susceptibles d'influencer le jugement professionnel de l'enseignant entre les novices et les deux autres groupes (praticiens et experts). Toutefois, aucun résultat ne semble révéler une différence notable entre ces deux derniers groupes de sujets. De ce fait, cette observation concorde avec celle relevée dans le cadre de l'interprétation de la première vignette de l'instrument.

4.2.4 Item 1 de la deuxième vignette

Toujours en lien avec la deuxième vignette du TJPEA, voici le premier item qui a été présenté aux participants.

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Vous ne faites rien de particulier puisqu'il s'agit probablement de l'effet du hasard.		Les notes reviennent à la normale lors de l'évaluation suivante.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Dans un premier temps, les sujets devaient se situer sur l'échelle de type Likert présentée en lien avec la pertinence de l'intervention indiquée. La figure suivante illustre bien la distribution des réponses obtenues.

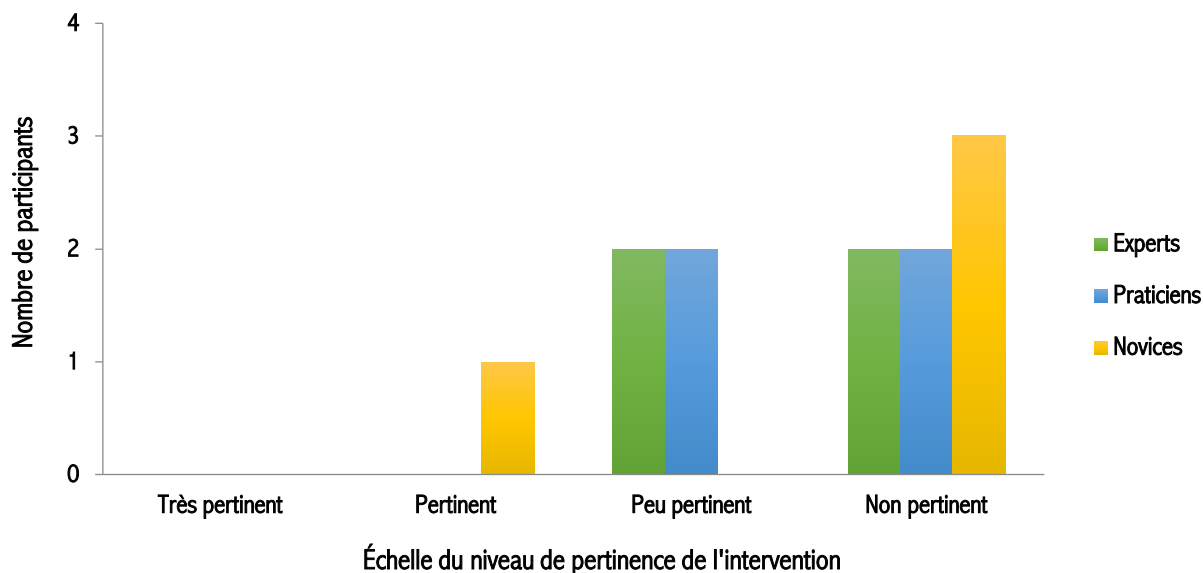


Figure 21. Répartition des réponses des profils de répondants (n=12) à l'échelle A

Au premier regard, nous pouvons remarquer que les praticiens et les experts ont exactement le même patron de réponses : la moitié du groupe (n=2) a considéré l'intervention présentée par l'item 1 comme étant « peu pertinente » et l'autre moitié (n=2) « non pertinente ». Du côté des novices, un seul d'entre eux a jugé l'intervention « pertinente », tandis que les trois autres ont jugé le tout « non pertinent ». Malgré une légère différence parmi les réponses obtenues au sein de l'ensemble des participants, l'item 1 ne semble pas démontrer de différence notable entre les trois profils. Ainsi, l'item concerné ne permet pas de déceler de différences de jugement entre nos trois groupes de répondants. Par contre, l'analyse du *verbatim* nous permet tout de même de déceler certaines nuances de profondeur entre les justifications apportées par les trois profils de répondants.

Verbatim du praticien 3 en lien avec l'item 1

« Tu ne peux pas juste laisser passer. Il faut que tu reviennes sur ce que tu as fait comme prof et que tu réajustes le tir pour l'année suivante. Dans mon livre à moi, tu ne peux pas faire ça. Tu n'es pas un bon prof quand tu fais ça. »

Verbatim de l'expert 2 en lien avec l'item 1

« Il faut regarder ce qui s'est passé. On regarde toutes les explications que l'on peut voir. Est-ce que les élèves n'ont pas fait tous les apprentissages? Est-ce qu'il y a un thème biaisé? Est-ce qu'il y avait une question qui abordait quelque chose qui était complètement en dehors du cours où les élèves parlaient de leurs connaissances antérieures? Il faut se poser la question pourquoi les résultats ne correspondent pas à ce qui était prévu. »

Verbatim du novice 2 en lien avec l'item 1

« Ça dépend de l'impact que cela a sur les notes des élèves. Si l'examen est anormalement fort, c'est tant mieux pour les élèves. Je ferais un retour sur l'épreuve avec eux pour savoir ce qui était plus clair pour eux... Si c'est négatif, je trouve ça dommage. Ça se peut que cela soit mon erreur. »

On remarque que l'expert est capable de déceler les facteurs possibles d'avoir influencé la situation présentée par l'item sans hésitation. Son discours donne d'ailleurs de nombreux exemples du questionnement que pourrait vivre l'enseignant dans une telle situation. Du côté du praticien, encore une fois, le désir d'agir et de se questionner est aussi relevé. Par contre, aucune précision quant aux facteurs explicatifs n'a été donnée. Enfin, le discours du novice se centre plutôt sur l'impact des résultats des élèves sur leur réussite globale. Ainsi, il semble s'inquiéter davantage face à la réaction des élèves (positive ou négative) plutôt que d'essayer d'identifier la source de la variation des scores.

La figure suivante démontre la répartition des réponses des répondants en lien avec la deuxième échelle de l'item1, soit celle de la probabilité à réaliser l'action proposée.

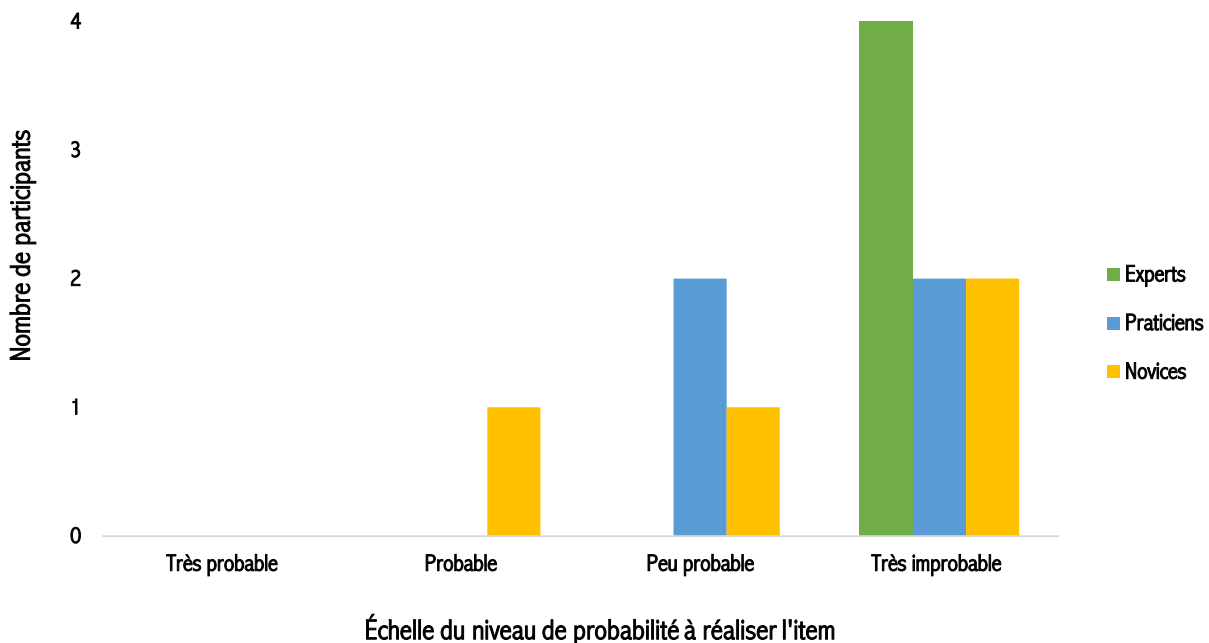


Figure 22. Répartition des réponses des répondants (n=12) à l'échelle B par profil

À première vue, la répartition des réponses des répondants est relativement différente selon les profils. Premièrement, tous les experts ont jugé « très improbable » pour eux de réaliser l'item 1, soit de ne rien faire de particulier. Chez les praticiens, les réponses se partagent en deux moitiés : deux praticiens jugent « très improbable » et deux autres « peu probable ». Chez les novices, les réponses sont relativement très partagées : deux novices ont jugé « très improbable » la probabilité de réaliser l'action, un « peu probable » et un « probable ». Enfin, en jetant un coup d'œil aux trois profils de répondants, l'on peut remarquer que la fluctuation des réponses est plus présente chez les novices et les praticiens que chez les experts. L'analyse des verbatim présentés un peu plus bas met en évidence la limite des scores de l'instrument. Par conséquent, des répondants peuvent opter pour la même réponse, mais pas pour les mêmes raisons.

Ainsi, contrairement à l'échelle A, la distribution des réponses à l'échelle B permet une lecture plus représentative des justifications données par les trois groupes de sujets dans le cadre des entrevues. Les experts ont, à l'unanimité, souligné l'importance d'agir et de se questionner quant aux facteurs explicatifs de la variation des notes obtenues par les élèves. Les praticiens, quant à eux, l'ont aussi mentionné, sans toutefois aller en profondeur quant aux éléments possibles

d'avoir influencé les résultats. Enfin, les novices sont plus partagés : la situation avantage t'elle les élèves? Si c'est le cas, ils pourraient décider de fermer les yeux.

Verbatim du praticien 2 en lien avec l'item 1

« C'est sûr qu'il s'est passé quelque chose. Au minimum, tu vas te questionner. C'est clair. Je ferais une rétroaction avec les élèves. J'irais peut-être voir un autre prof pour savoir si mon examen fait du sens... Oui, non et pourquoi. »

Verbatim de l'expert 4 en lien avec l'item 1

« Dès qu'il y a un changement, il faut se questionner. Je regarderais s'il s'est passé quelque chose durant cette évaluation-là? Est-ce que j'ai moins enseigné certains concepts? Je retourne à l'examen et j'essaie de voir si cela touche des concepts en particulier. Est-ce que c'est l'ensemble de l'épreuve? Si c'est l'ensemble, est-ce que mes élèves avaient vécu quelque chose? Il faut connaître le contexte de la passation de l'examen et mieux analyser l'outil. Il faut aussi retourner dans son enseignement. »

Verbatim du novice 3 en lien avec l'item 1

« J'irais chercher de l'information chez les élèves. Je ferais un autre examen pour voir si on peut augmenter leurs notes. Si c'est bon, je suis contente pour eux. »

4.2.5 Item 2 de la deuxième vignette

Voici le deuxième item présenté aux participants, et ce, en lien avec la deuxième vignette.

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Vous modifiez les résultats afin d'attribuer à ces élèves les notes qu'ils atteindraient en temps normal.		La distribution des notes ressemble aux précédentes.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

La répartition des réponses obtenues par les experts à cet item est assez variée. Comme le témoigne la Figure 23, deux experts ont considéré l'intervention de l'item 2 comme étant « pertinente », un expert « peu pertinente » et un autre « non pertinente ». Les praticiens, de leur côté, sont unanimes en considérant l'item 2 comme étant « non pertinent ». Enfin, les novices semblent plutôt aller dans le même sens que les praticiens. Plus précisément, trois novices ont opté pour l'option « non pertinent » et un seul « peu pertinent ».

Dans une perspective comparative, l'item 2 de la deuxième vignette semble démontrer une différence entre les réponses obtenues par les experts : une plus grande variété est observée que pour les deux autres profils de répondants. En effet, les novices et les praticiens sont, quant à eux, plus unanimes. Ceci étant dit, cette similitude de réponses entre ces deux groupes démontre que l'item 2 a permis de déceler des différences entre les experts et les praticiens et les novices. Par contre, aucune différence notable n'a pu être dénotée entre les praticiens et les novices.

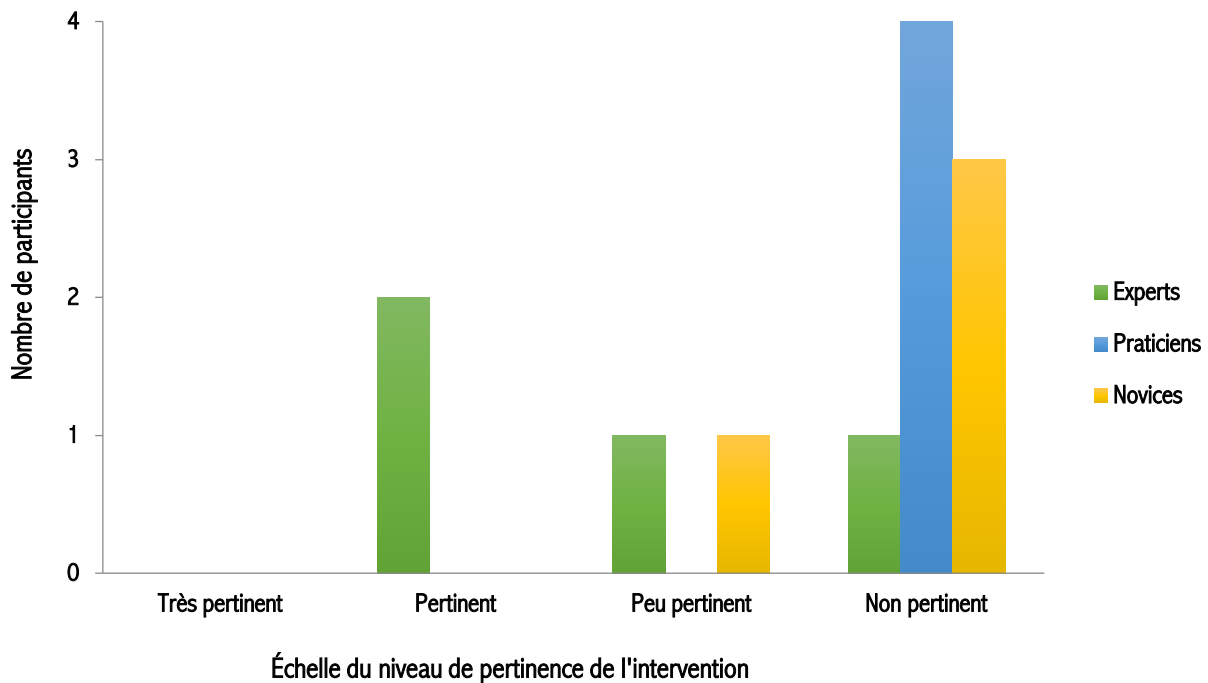


Figure 23. Répartition des réponses des répondants (n=12) à l'échelle A par profil

La Figure 24 présente la probabilité pour les répondants des différents profils à réaliser l'item 2.

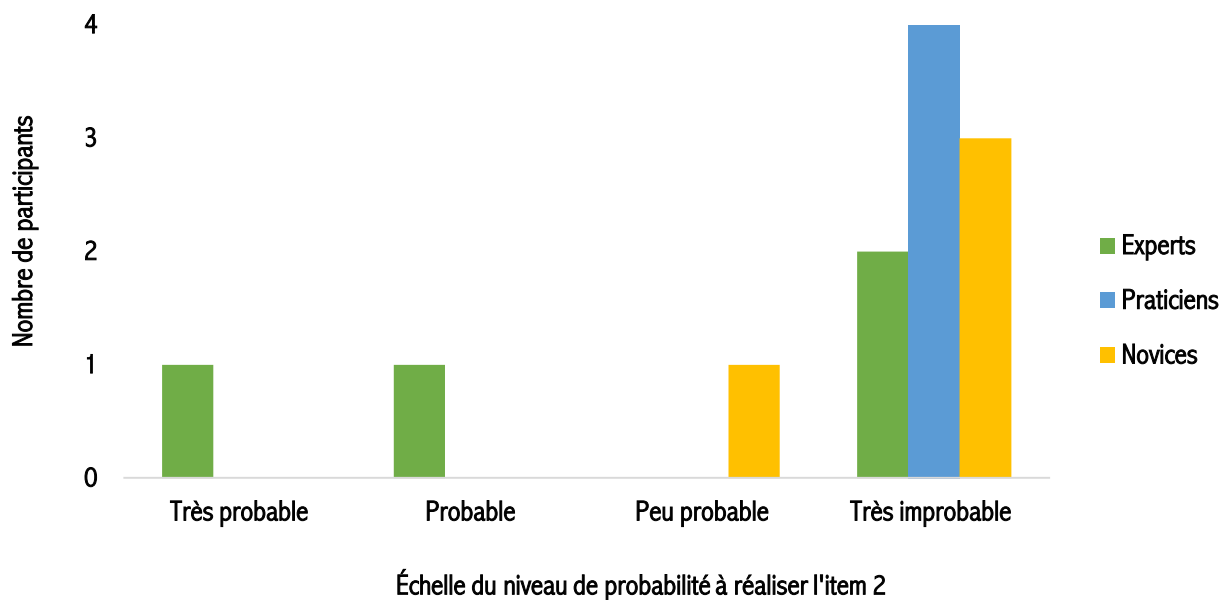


Figure 24. Répartition des réponses des répondants (n=12) à l'échelle B par profil

Tous profils confondus, la majorité des répondants ont considéré « très improbable » de réaliser l'intervention proposée par l'item 2, soit de « modifier les résultats afin d'attribuer à ces élèves les notes qu'ils atteindraient en temps normal ». Parmi les répondants ayant opté pour cette option, l'on peut dénombrer deux experts contre quatre praticiens et trois novices. Les experts sont, encore une fois, plus variés dans leurs réponses sur ce continuum : un ayant opté pour « très probable » et un autre pour « probable ». Ceci étant dit, il semblerait que nos novices et nos praticiens s'entendent davantage quant à la probabilité à réaliser l'intervention proposée. Nos experts, de leur côté, sont plus partagés. Les extraits de *verbatim* suivants précisent les raisons qui ont partagé les réponses des experts.

Verbatim de l'expert 2 en lien avec l'item 2

« Je serais très hésitant à faire ce genre de chose... cela dépend de ce que tu trouves quand tu as essayé de voir pourquoi c'est comme ça. Mais, il y a un certain nombre de modifications que tu peux faire pour être équitable. Par exemple, si c'est un problème avec le niveau de

difficulté de la tâche... car elle était beaucoup plus difficile ou plus facile que ce qui était anticipé... tu peux faire un ajustement linéaire uniforme pour tout le monde. Ça c'est faisable. Le MEQ le fait régulièrement. »

Verbatim de l'expert 4 en lien avec l'item 2

« Ce n'est pas quelque chose que je trouve pertinent du tout. Par contre, c'est étonnant tout le travail qui est fait avec les examens du Ministère comme la modération et les ajustements de notes... Maintenant, ce sont des examens externes. Quand on est dans sa propre classe, ce n'est pas du tout pertinent parce qu'il s'est passé quelque chose. Ce n'est pas correct de faire ça... C'est de cacher le fait qu'il y a une problématique dans son enseignement ou dans son outil. »

Verbatim de l'expert 1 en lien avec l'item 2

« Tant et aussi longtemps que je ne comprends pas le problème, de faire des ajustements... je vais peut-être les pénaliser ou les faire gagner... Je n'ai aucune idée. De revenir à la normalité n'est peut-être pas la meilleure solution. L'évaluation doit nous donner leur potentiel et elle ne semble pas le donner. Je me vois mal de pondérer... Je trouve cela difficile de savoir quoi faire, mais je devrais fouiller pour trouver de l'information. »

Malgré la disparité des réponses obtenues à l'échelle A, les justificatifs donnés par le groupe d'experts révèlent tout de même un consensus. Tous expriment l'importance de trouver la raison de la fluctuation des notes des élèves. La majorité des experts font référence à certaines procédures mathématiques (modération, ajustement linéaire) qui pourraient concrétiser un ajustement des notes équitable, mais aucun d'entre eux se dit confiant quant à la pertinence d'aller de l'avant avec elle telle démarche dans un contexte de salle de classe.

En somme, les entrevues ont permis de déceler une similitude quant à la réflexion des trois profils. Plus précisément, la majorité des sujets ne se sentent pas à l'aise quant à l'intervention proposée. Par contre, seuls les experts font mention d'une démarche de calcul possible si la situation se présentait à plus grande échelle. C'est d'ailleurs cette éventualité qui pourrait expliquer le manque d'accord au sein de ce profil de sujets. Les deux autres groupes de participants ont aussi ressenti un malaise et ne supportent pas l'idée proposée par l'item. Les praticiens semblent vouloir comprendre ce qui s'est produit avant d'orienter leurs actions, tandis que les novices ne proposent pas de solutions spécifiques.

Verbatim du praticien 1 en lien avec l'item 2

« Il y a plein d'affaires qui ont pu arriver. Commencer à jouer avec les résultats, il faut que ça soit quelque chose de spécifique. Moi, je vais plus aller comprendre. Je ne réajuste pas, mais je peux donner un autre examen ou un autre travail. »

Verbatim du novice 2 en lien avec l'item 2

« Je ne modifierais pas les résultats. Je trouve cela pas équitable. J'ai juste l'impression que c'est injuste. »

4.2.6 Item 3 de la deuxième vignette

Enfin, le troisième item en lien avec la deuxième vignette souligne l'influence possible de l'alignement entre l'enseignement offert aux élèves et les contenus évalués.

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Vous réévaluez votre enseignement afin de déterminer la source de la variation.		La matière vue était beaucoup moins dynamique et stimulante que celle vue précédemment.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

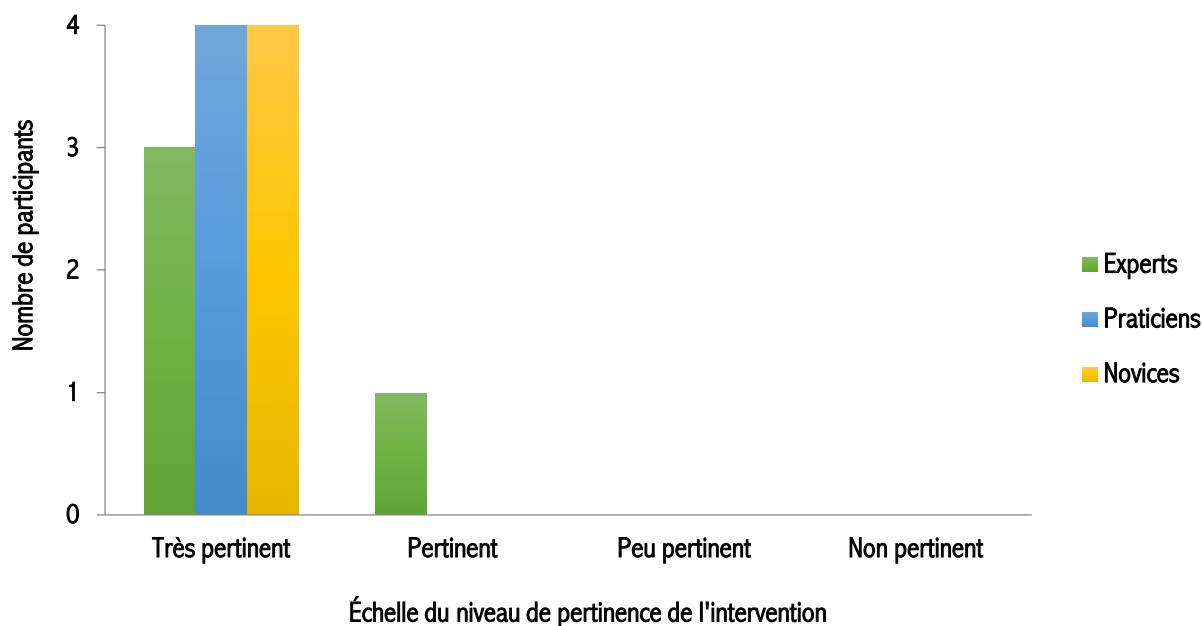


Figure 25. Répartition des réponses des répondants (n=12) à l'échelle A par profil

Au premier regard, nos trois profils de répondants semblent unanimes, et ce, à l'exception d'un seul expert. En effet, trois experts, quatre praticiens et quatre novices ont considéré l'intervention proposée par l'item 3 comme étant « très pertinente ». Ainsi, peu importe le profil, tous s'entendent pour relever l'importance de se questionner face à la manière dont l'enseignant livre les contenus qui seront évalués.

Verbatim de l'expert 4 en lien avec l'item 3

« Ce n'est pas suffisant, mais je pense que cela est très pertinent. Il faut aussi analyser l'outil, c'est clair. Il faut aussi regarder la cohérence entre l'outil et ce qui a été enseigné, l'organisation de l'examen et le contexte. »

Verbatim du praticien 1 en lien avec l'item 3

« Il faut qu'on se questionne. Faut en parler aussi avec des collègues... Ça fait aussi partie de notre développement professionnel. »

Verbatim du novice 2 en lien avec l'item 3

« C'est justement ça qui va permettre à une situation comme celle-là de ne pas se reproduire. Je n'ai pas envie de décourager mes élèves. Si mes élèves n'ont pas aimé le sujet de mon

examen, c'est moi. C'est moi qui a influencé cet examen-là en ne m'assurant pas que cela soit intéressant... »

Ceci étant dit, l'item 3 ne permet de pas de déceler des différences entre les profils. Au contraire, il met bien en valeur la similitude commune entre les trois profils, soit celle de prioriser l'alignement entre la matière enseignée et celle questionnée dans le cadre de l'évaluation.

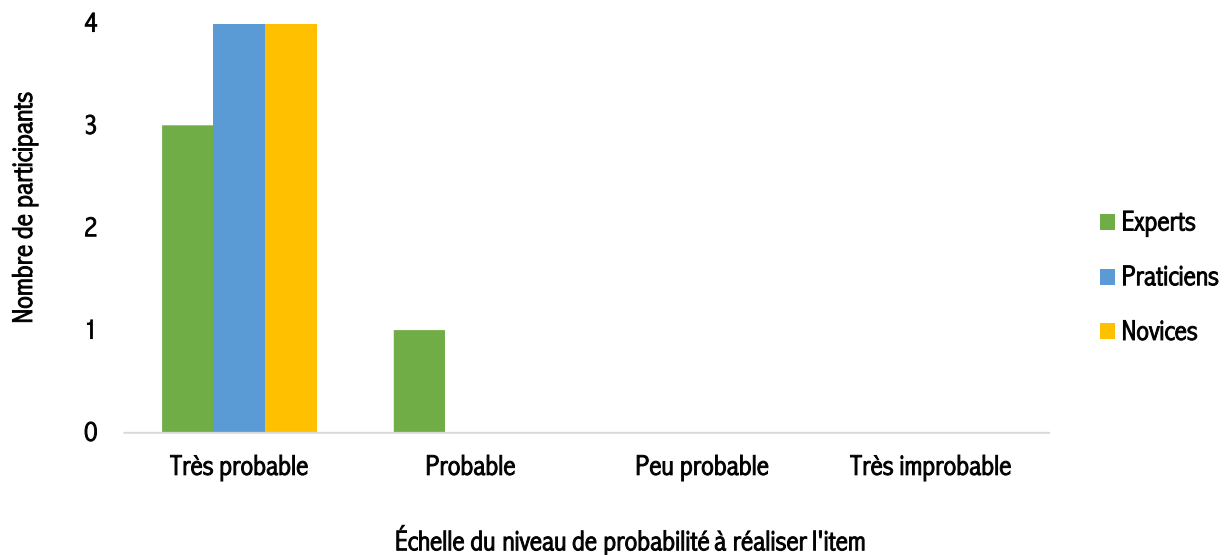


Figure 26. Répartition des réponses des répondants (n=12) à l'échelle B par profil

Pour continuer, la Figure 26 présente la distribution des réponses des répondants selon le niveau de probabilité à réaliser l'item 3, soit s'interroger sur son enseignement afin d'identifier la source de variation possible. Chez les praticiens et les experts, tous ont été unanimes : il était « très probable » qu'ils agissent ainsi. Une majorité d'experts (n=3/4) et de novices (n=4/4) ont opté pour la même réponse. Comme le témoignait l'échelle précédente de pertinence, une similitude de jugement est bien représentée grâce à cet item.

4.2.7 Synthèse de la vignette 2

En conclusion, la deuxième vignette présentée par le TJPEA permet de mettre en évidence une similitude quant à l'avis des sujets face à la pertinence des interventions proposées par les trois items. Malgré cette observation, l'analyse des extraits d'entrevues permet de déceler une

profondeur de raisonnement plus marquée chez les experts qui fournissent un ensemble de solutions ou de facteurs plus nombreux que les praticiens et les novices. Les praticiens sont, quant à eux, souvent aptes à mentionner l'action qu'ils privilégieraient, mais sans nécessairement pouvoir en mentionner d'autres. Enfin, les novices émettent souvent leur malaise ou leur accord, sans toutefois pouvoir justifier de manière précise pourquoi. De ce fait, la deuxième vignette permet de rendre compte du jugement professionnel en évaluation des sujets et d'en déceler certaines similitudes ou différences entre les profils de répondants.

4.3 Vignette 3

La Figure 27 présente la troisième vignette composant le test de jugement professionnel en évaluation des apprentissages. Nous rappelons que cette dernière a été lue à voix haute en plus d'être accessible à partir du support papier remis à chaque participant.

Les questions se rapportent à la mise en situation qui suit :

Lors de la correction d'une épreuve écrite une question a été comprise d'une toute autre façon que ce que vous aviez prévu initialement lors de la confection de cette épreuve.

Figure 27. Cas présenté dans le cadre de la vignette 3 du TJPEA

4.3.1 Authenticité de la vignette 3

À la question « *Trouvez-vous que la vignette présentée représente une situation possible de se dérouler en salle de classe ou dans une école ?* », tous les participants (n=12) ont répondu par l'affirmative.

4.3.2 Identification des concepts

Dans un souci de validité, nous avons posé la question suivante à chaque sujet : « Quels concepts sont abordés par le cas présenté par la vignette ? ». Le prochain tableau dresse les concepts disciplinaires relevés par les participants et ceux anticipés par le concepteur de l'instrument, soit un expert en évaluation des apprentissages.

Tableau 13

Tableau comparatif entre les concepts anticipés par le concepteur de la vignette 3 et ceux relevés par les trois groupes de participants (experts, praticiens et novices)

Profils	Concepteur N=1	Experts N=4	Praticiens N=4	Novices N=4
Concepts				
Validité	Oui 1/1	Oui 3/4	Oui 2/4	Non
Conception d'items	Oui 1/1	Oui 4/4	Oui 1/4	Oui 2/4
Analyse d'items	Oui 1/1	Non	Non 2/4	Non
Correction	Non	Oui 3/4	Oui 2/4	Oui 2/4

Selon le point de vue du concepteur, trois principaux concepts liés à l'évaluation des apprentissages émergent de la troisième vignette du TJPEA : la validité, la conception des items et l'analyse des items. Du côté des experts, deux de ces trois concepts sont ressortis : la validité (3/4 experts) et la conception d'items (4/4 experts). Ces deux mêmes concepts ont également été relevés par les praticiens (respectivement 2/4 praticiens et 1/4 praticiens). Les novices, quant à eux, ont identifié qu'un seul des concepts anticipés par le concepteur, soit celui de la conception des items (2/4 novices).

Pour continuer, pour les trois groupes de répondants, la correction des items semblait être un important facteur à considérer dans le cadre de la troisième vignette. Malgré que le concepteur n'y fit pas allusion dans son anticipation, les experts (3/4 experts), les praticiens (2/4 praticiens) et les novices (2/4 novices) s'y sont attardés. Plus précisément, les individus concernés se sont intéressés à la clarté des réponses attendues dans le cadre de l'étape de la correction de l'épreuve. Enfin, aucun participant n'a fait allusion à l'analyse des items, concept qui avait été suggéré par le concepteur.

4.3.3 Données manquantes à la vignette 3

Les figures suivantes présentent les réponses obtenues à la question « *Quelles informations non présentées par la vignette pourraient vous être utiles ?* » par groupe de répondants. La séquence de présentation des résultats reste la même que les deux précédentes.

Ainsi, à partir des réponses obtenues à cette question par tous les répondants, trois grandes catégories de données manquantes (codes principaux) se sont créées : celles liées au profil du groupe d'élèves, celles liées à la tâche et celles liées au contexte. Pour chacune de ces catégories, des sous-codes ont été identifiés. Ici, la majorité des questionnements émis par nos participants étaient en lien avec la tâche, code pour lequel se sont associés cinq sous-codes, et ce, contrairement aux deux autres catégories pour lesquelles un seul sous-code n'a été nécessaire. Le tableau suivant présente la structure des codes et des sous-codes utilisés.

Tableau 14

Liste des codes et des sous-codes utilisés en lien avec les données manquantes de la vignette 3 du TJPEA

CODE 1 Profil du groupe	CODE 2 Tâche	CODE 3 Contexte
•Sous-code 1 • Niveau du groupe (faible, moyen ou fort)	•Sous-codes 2 • Clarté de la tâche • Types de questions posées • Qualité de la clé de correction • Pondération des questions • Niveau de difficulté de la tâche	• Sous-code 3 • Durée de l'épreuve

4.3.3.1 Données manquantes par les novices

Parmi les données manquantes souhaitées par les participants novices, le niveau de difficulté et la clarté de la tâche semblent les plus populaires, car elles ont été demandées par 3/4 des novices. Un seul novice a fait la requête d'en savoir davantage sur les trois données suivantes : la pondération de la question, la qualité de la clé de correction et le niveau du groupe d'élèves

(faible, moyen ou fort). En somme, cinq données manquantes ont été soulevées par le profil des novices.

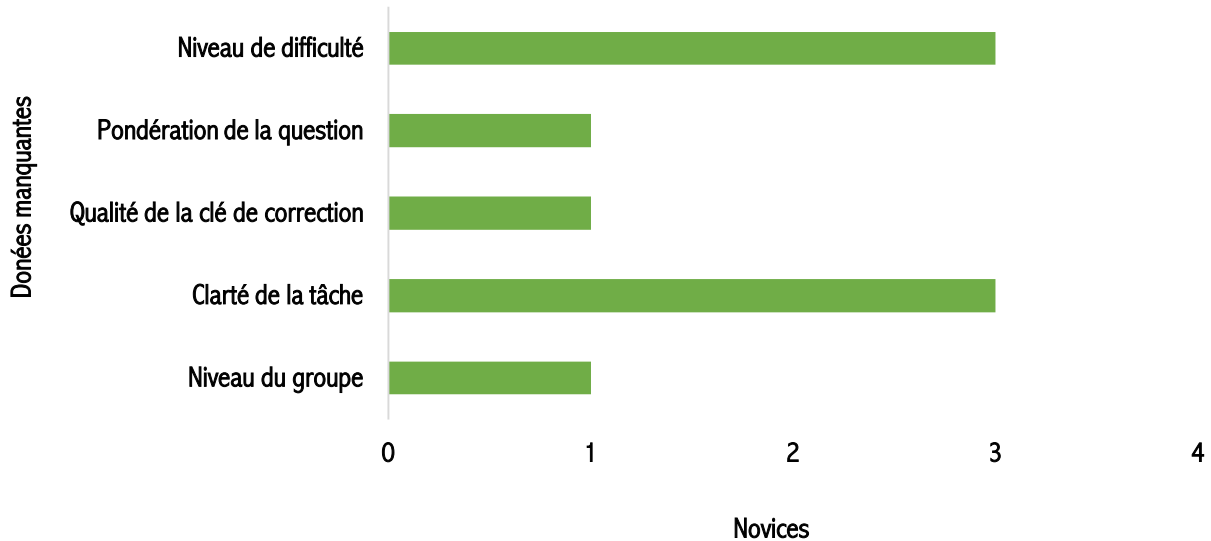


Figure 28. Données manquantes identifiées par les novices (n=4)

4.3.3.2 Données manquantes par les praticiens

Parmi les réponses obtenues chez les praticiens concernant les données manquantes à la vignette 3, les plus fréquentes (4/4 praticiens) sont les suivantes : le niveau de difficulté de l'épreuve, la clarté de la tâche et le niveau du groupe d'élèves (faible, moyen ou fort). Non loin derrière, la pondération de la question et la qualité de la clé de correction ont été identifiées par trois novices sur une possibilité de quatre. Un seul novice aurait désiré connaître la durée de l'épreuve dont il question de la vignette, et ce, afin de mieux exercer son jugement. Enfin, au total, six données manquantes ont été relevées par le groupe de praticiens, dont la majorité d'entre elles l'ont été par trois ou quatre praticiens du groupe.

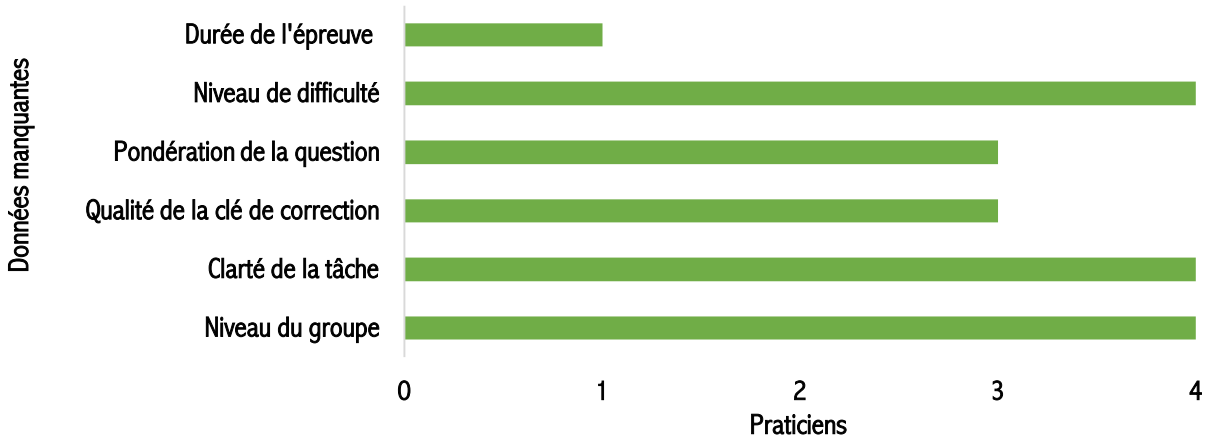


Figure 29. Données manquantes identifiées par les praticiens (n=4)

4.3.3.3 Données manquantes par les experts

Parmi les données manquantes identifiées par les experts, les plus fréquentes touchent le niveau de difficulté de la tâche et sa clarté (3/4 experts). Pour continuer, la moitié des répondants de ce groupe (n=2) a mentionné l'influence possible des caractéristiques de la clé de correction ainsi que les types de questions qui constituent l'examen. Enfin, un seul expert a fait la mention de la durée de l'épreuve comme étant une donnée manquante. Au total, cinq informations manquantes ont été relevées par ce groupe dont la majorité d'entre elles ont été mentionnées par plus de la moitié des répondants experts.

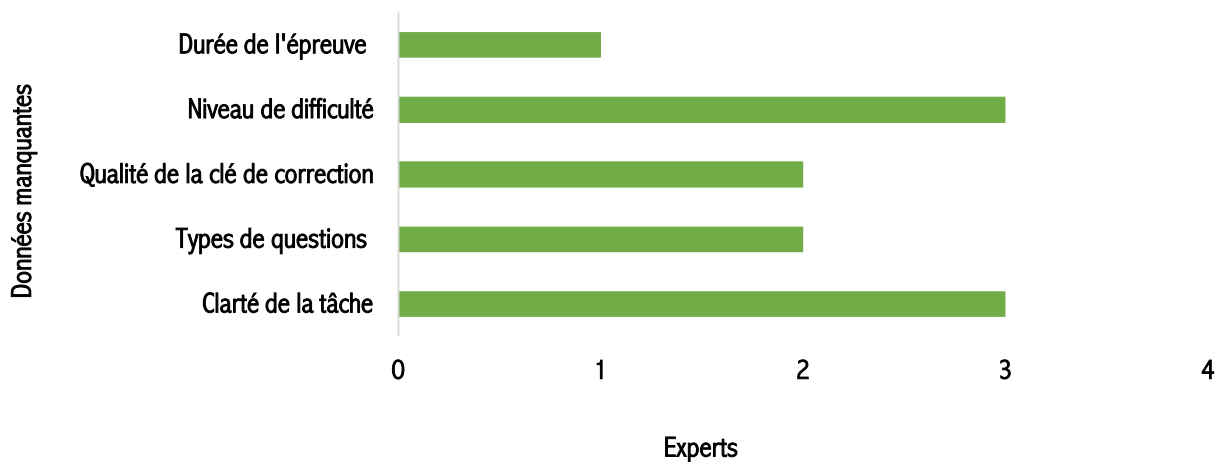


Figure 30. Données manquantes identifiées par les experts (n=4)

4.3.3.4 Synthèse comparative des données manquantes par profil de participants

Afin de répondre à notre question de recherche « À partir des réponses des participants aux vignettes du TJPEA, pourrions-nous démontrer des différences entre le jugement en évaluation des novices, des praticiens et celui des experts ? », nous présentons la synthèse des tableaux précédents, à la lumière des différences et des similitudes des données manquantes identifiées par les trois groupes de sujets.

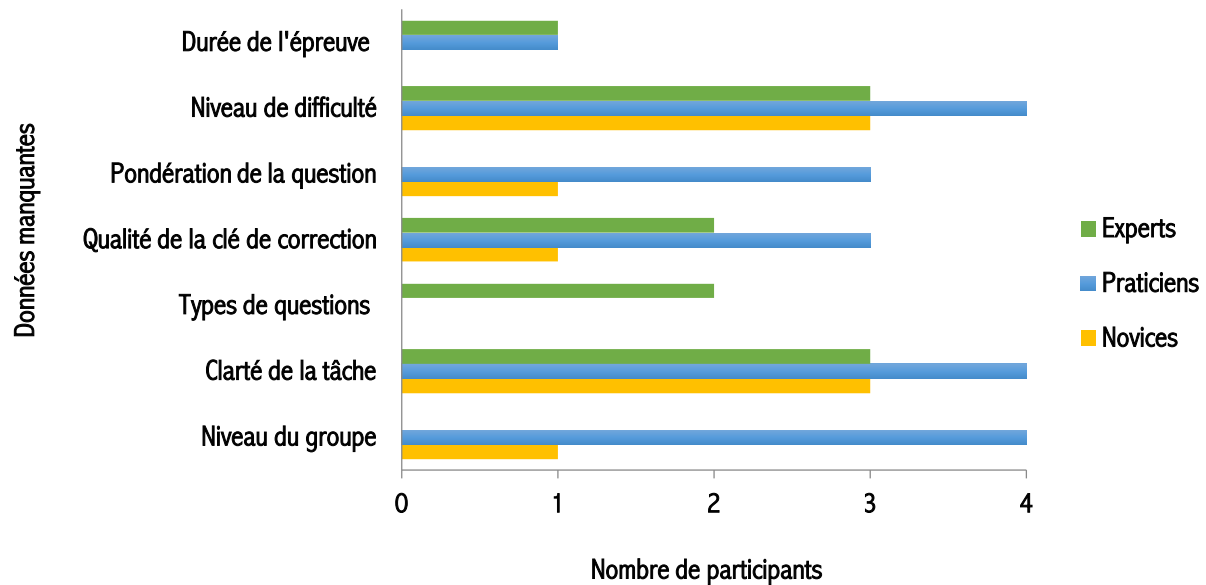


Figure 31. Synthèse des données manquantes identifiées par tous les participants (n=12)

Comme nous le témoigne la Figure 31, certaines données manquantes ont été relevées par les trois profils de répondants : le niveau de difficulté de la tâche, la qualité de la clé de correction et la clarté de la tâche. Pour ces trois cas, ce sont les praticiens qui ont été les plus nombreux à identifier les données manquantes (11 praticiens, contre huit experts et sept novices). Deux données ont été relevées uniquement par les praticiens et les novices, soit la pondération de la question (trois experts et un novice) et le niveau du groupe (quatre experts et un novice). Pour continuer, un praticien et un seul expert se sont questionnés face à la durée de l'épreuve, tandis que deux experts l'ont fait en lien avec les types questions.

Malgré le fait que les trois profils de participants aient identifié un nombre total similaire de données manquantes (cinq pour les novices, six pour les praticiens et cinq pour les experts), la fréquence par sujet n'est pas la même. Sous cet angle, les novices signalent l'absence de l'information neuf fois, les praticiens 19 et les experts 11. Ainsi, les praticiens semblent faire partie du groupe de répondants capables de nommer un plus grand nombre de facteurs susceptibles d'influencer leur jugement face à la situation présentée. Les experts sont, quant à eux, huit fréquences derrière ; les novices 10 répétitions derrière.

Tout comme dans le cadre de la première et de la deuxième vignette de ce TJPEA, les résultats obtenus dans le cadre de la question liée aux données manquantes de la troisième vignette nous poussent à croire qu'il existe bien une différence d'habileté à identifier les facteurs susceptibles d'influencer le jugement professionnel de l'enseignant entre les novices et les deux autres groupes (praticiens et experts). Toutefois, dans le cadre de la troisième vignette, l'écart existant entre les novices et les experts est moindre que dans le cas des deux premières vignettes.

4.3.4 Item 1 de la troisième vignette

Toujours en lien avec la troisième vignette du TJPEA, voici le premier item qui a été présenté aux participants.

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Vous conservez les résultats des élèves sans rien changer.		Les notes des étudiants réputés très performants sont plus faibles pour cette question.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Dans un premier temps, les sujets devaient se situer sur l'échelle de type Likert présentée en lien avec la pertinence de l'intervention indiquée. La figure suivante illustre bien la distribution des réponses obtenues.

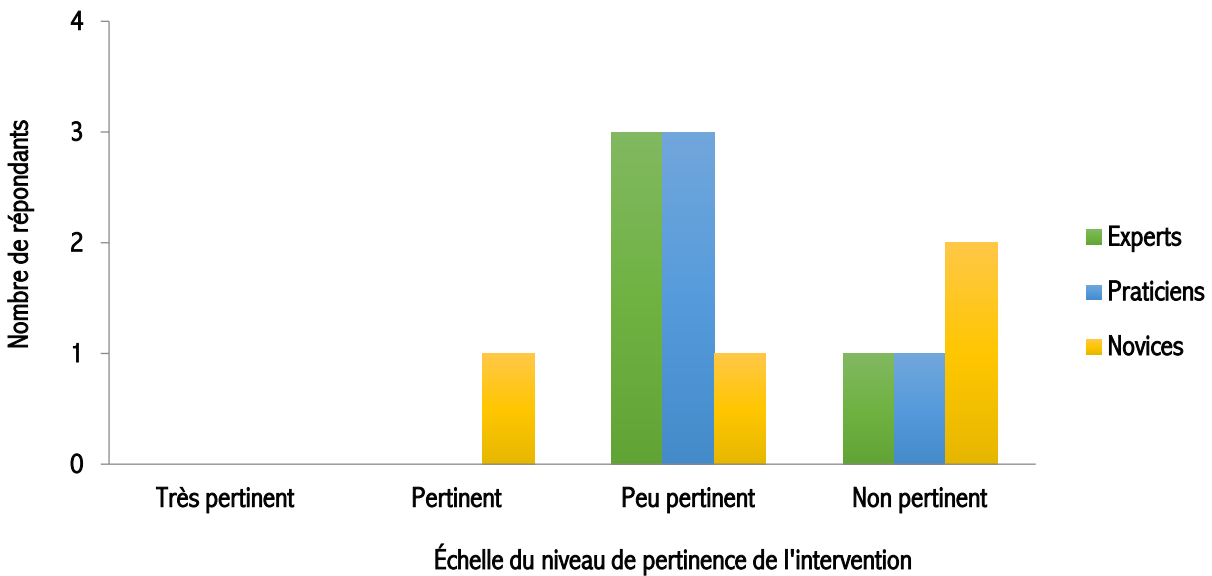


Figure 32. Répartition des réponses des profils de répondants (n=12) à l'échelle A

Au premier regard, nous pouvons remarquer que les praticiens et les experts ont exactement le même patron de réponses : trois répondants sur quatre ont considéré l'intervention présentée par l'item 1 comme étant « peu pertinente » et un seul « non pertinente ». Du côté des novices, la moitié du groupe (n=2) a considéré l'intervention « non pertinente », tandis que l'autre moitié la considère « peu pertinente » (n=1) et « pertinente » (n=1).

D'un point de vue comparatif, nous pouvons distinguer une plus grande variance dans les réponses obtenues par les novices que dans les deux autres groupes de répondants. En effet, certains d'entre eux se questionnent face à la situation, mais n'iront pas jusqu'à l'action si cela n'a pas d'impact significatif sur le score global des élèves à l'examen.

Verbatim du novice 3 en lien avec l'item 1

« Je changerais quelque chose, mais il faut que cela soit arrivé avec plus qu'un élève. Si c'est juste quelques-uns, je ne pense pas que je ferais quelque chose nécessairement. »

Verbatim du novice 1 en lien avec l'item 1

« Peut-être que mes élèves étaient moins concentrés. Ça dépend... Ça peut arriver. »

Chez les experts et les praticiens, le questionnement face à la situation présentée est plus détaillé. Les deux groupes soulignent l'importance d'agir quand une telle situation se produit. Par conséquent, les sujets ne permettent pas qu'une « mauvaise question » soit comptabilisée dans le score final des élèves concernés. Les extraits des *verbatim* suivants illustrent bien cette pensée.

Verbatim de l'expert 3 en lien avec l'item 1

« Quand on a un problème qui est aussi généralisé, on ne peut pas faire abstraction. Je suggérerais de retirer la question tout simplement. Il faut faire un retour en classe avec les élèves par rapport à ce qui n'a pas été bien compris, mais on ne la comptabilise pas dans les notes. »

Verbatim du praticien 2 en lien avec l'item 1

« C'est sûr que je ne conserve pas les résultats des élèves. Je dois corriger le tir. J'annulerais la question. La question n'est pas valide. »

Pour continuer, le graphique suivant démontre la répartition des réponses des répondants en lien avec la deuxième échelle de l'item 1, soit celle de la probabilité à réaliser l'action proposée.

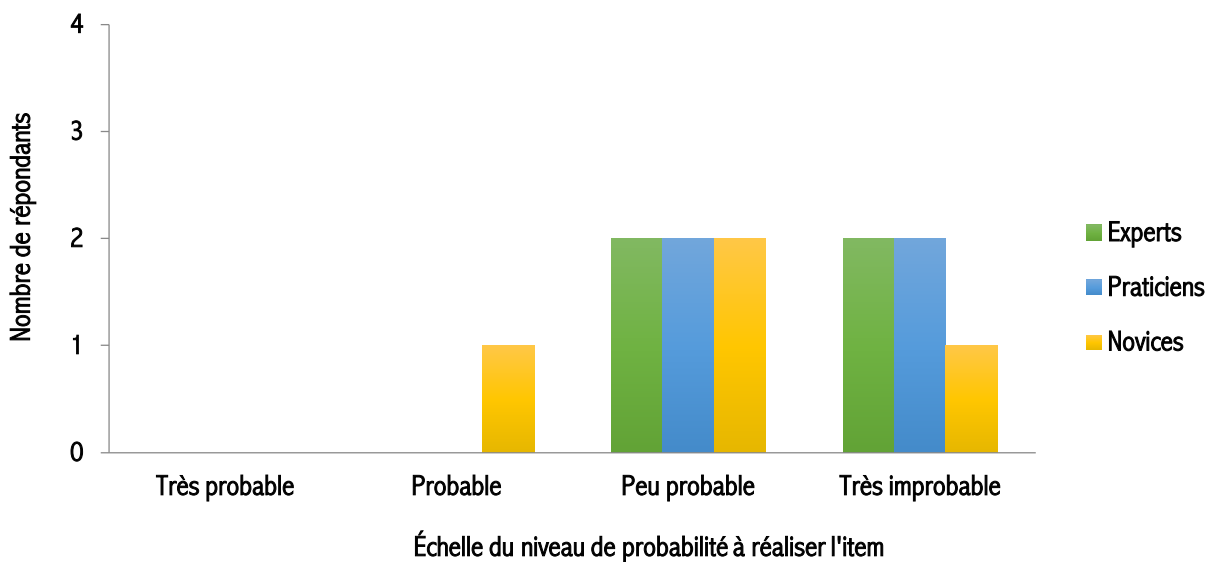


Figure 33. Répartition des réponses des répondants (n=12) à l'échelle B par profil

À première vue, la répartition des réponses des répondants est identique chez les experts et chez les praticiens : la moitié du groupe (n=2) a considéré « peu probable » de réaliser l’item 1, tandis que l’autre moitié (n=2) a jugé le tout « très improbable ». Chez les novices, les réponses sont relativement très diverses : la moitié (n=2) est allée dans le même sens que les experts et les praticiens en considérant la probabilité à réaliser l’item 1 comme étant « peu probable », tandis qu’un seul a indiqué « probable » et un autre « très improbable ». Globalement, nos répondants considèrent donc « très improbable » (n=5) ou « peu probable » (n=6) de « conserver les résultats des élèves sans rien changer ».

4.3.5 Item 2 de la troisième vignette

Voici le deuxième item présenté aux participants, et ce, en lien avec la troisième vignette.

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Vous administrez une nouvelle question au cours suivant afin remplacer la question problématique.		Les élèves sont mécontents de devoir répondre à la nouvelle question.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

La figure suivante présente la répartition des réponses des répondants à l’échelle A de l’item 2 de la troisième vignette.

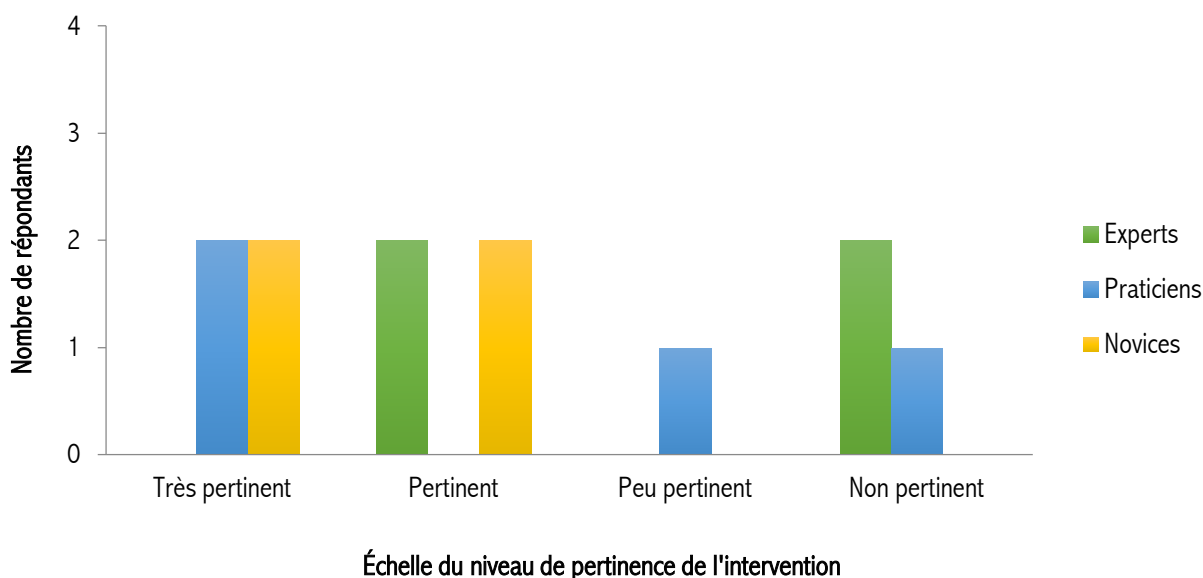


Figure 34. Répartition des réponses des répondants (n=12) à l’échelle A par profil

La répartition des réponses obtenues à l'item 2 est très variée. Par profil, les novices se sont divisés : la moitié d'entre eux (n=2) a considéré le fait « d'administrer une nouvelle question au cours suivant » comme étant « très pertinent ». L'autre moitié de groupe (n=2) a plutôt opté pour l'option « pertinent ». Chez les praticiens, nous remarquons très peu d'accord : deux ont jugé l'intervention « très pertinente », un « peu pertinente » et un autre « non pertinent ». Enfin, chez les experts, la moitié (n=2) a voulu répondre « pertinent » et l'autre moitié (n=2) « non pertinente ».

Dans une perspective comparative, l'item 2 de la troisième vignette, les praticiens et les novices (n=4) semblent en accord en considérant l'intervention comme « très pertinente ». Non loin derrière, deux novices et deux experts jugent le tout « pertinent ». Ainsi, la solution proposée semble répondre à leur besoin, soit de corriger la situation.

Verbatim du novice 2 en lien avec l'item 2

« Leur donner une nouvelle question me permet de les guider. Je peux revenir sur la matière et leur donner une chance de s'améliorer. Je peux aussi leur donner le choix de répondre ou non à la nouvelle question. »

Verbatim du praticien 1 en lien avec l'item 2

« Ça leur permet d'aller chercher des points supplémentaires. C'est vraiment pour les aider. »

À l'opposé du continuum, sous les catégories de réponses « peu pertinente » (n=1) et « non pertinente » (n=3), les participants se sentaient moins à l'aise avec l'effet surprise que proposait l'item 2. Plus précisément, les répondants n'étaient pas contre l'idée d'administrer une nouvelle question, mais bien de le faire dans un contexte où les élèves ne sont pas au courant de la situation, principe qui va à l'encontre d'un contexte évaluatif favorable. Il est à noter qu'aucun novice n'a eu cette réflexion.

Verbatim de l'expert 1 en lien avec l'item 2

« En administrant une question au prochain cours, les élèves ne sont pas prêts. C'est une surprise pour eux. C'est tout ce qui va à l'encontre des principes en lien avec la planification de l'épreuve. Il faut que cela soit connu à l'avance, que les élèves soient préparés et qu'ils sachent sur quoi ils vont être évalués. En arrivant avec une nouvelle question, on va encore les pénaliser. »

Verbatim du praticien 3 en lien avec l'item 2

« Je ne le ferais pas au cours suivant, car je ne voudrais pas que les élèves ne sachent pas sur quoi ils vont être évalués. Je ne fais pas de tests-surprises. »

Pour continuer, la Figure 35 présente la probabilité pour les répondants des différents profils à réaliser l'item 2.

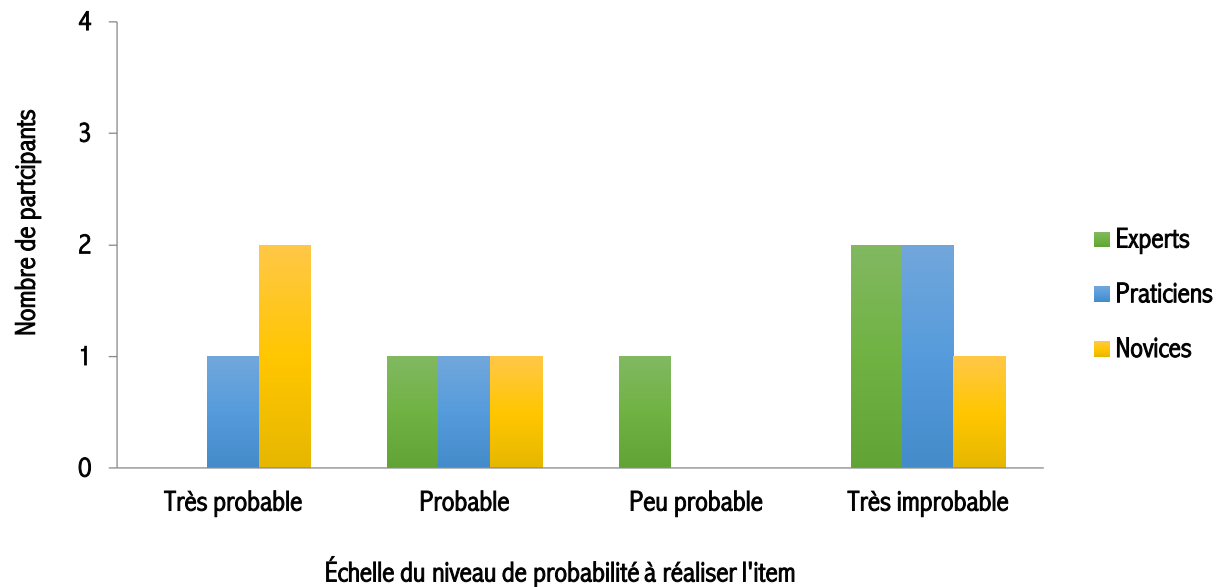


Figure 35. Répartition des réponses des répondants (n=12) à l'échelle B par profil

Tous profils confondus, la majorité des participants (n=5) ont considéré « très improbable » de réaliser l'intervention proposée par l'item 2, soit « d'administrer une nouvelle question au cours suivant ». Parmi les personnes ayant opté pour cette option, l'on peut dénombrer deux experts, deux praticiens contre un seul novice. Les trois autres novices, de leur côté, ont affirmé « très probable » (n=2) ou « probable » (n=1) qu'ils agissent ainsi. Ainsi, pour la majorité des individus issus de ce profil de novices, l'intervention présentée par l'item 2 semble être une bonne solution à la problématique présentée. À l'opposé, les experts ne semblent pas du même avis : trois d'entre eux jugent « très improbable », un autre « peu probable » et un dernier « probable » d'aller de l'avant avec l'intervention proposée. Enfin, les praticiens sont partagés : deux d'entre eux ont opté pour « très improbable », un « probable » et un autre « très probable ». Cette disparité dans les réponses peut, selon nous, être justifiée par le malaise qu'éprouvaient certains experts et certains praticiens à administrer une nouvelle question d'examen sans l'annoncer aux élèves préalablement.

4.3.6 Item 3 de la troisième vignette

Enfin, le troisième item en lien avec la troisième vignette souligne la possibilité d'annuler la question problématique présentée par la troisième vignette. Vous pouvez la lire comme suit :

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Vous décidez d'annuler cette question.		Les élèves révisent les autres questions de l'épreuve en espérant détecter d'autres questions lacunaires.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

La figure suivante présente les résultats obtenus à la première échelle.

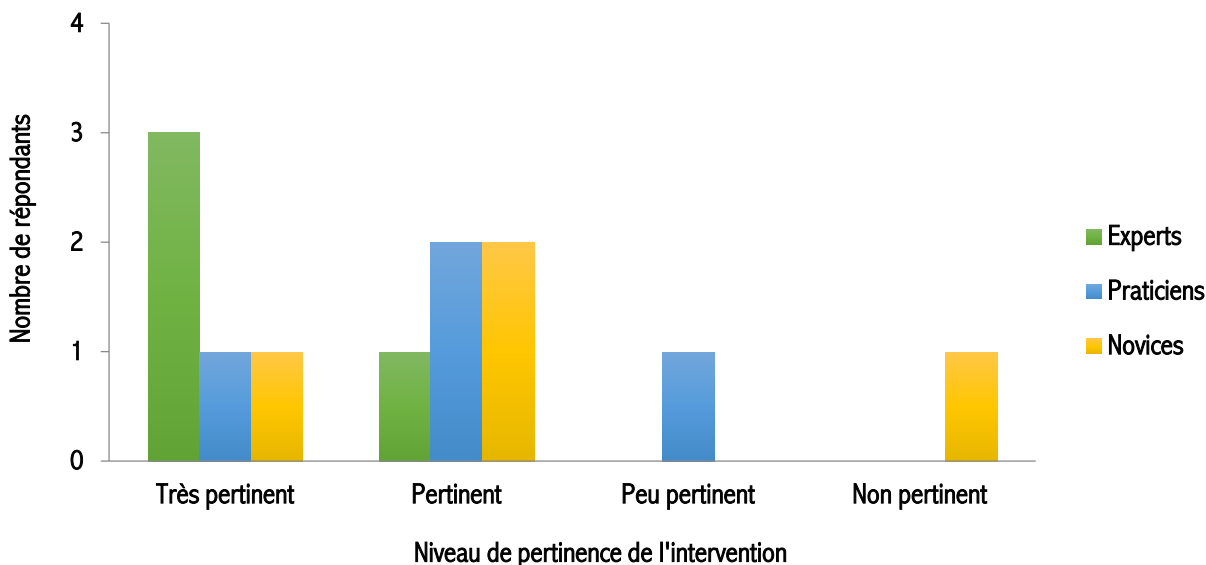


Figure 36. Répartition des réponses des répondants (n=12) à l'échelle A par profil

Au premier regard, la majorité des réponses des répondants (n=10) se trouvent à gauche du continuum, soit choisissant l'option « très pertinent » (trois experts, un praticien et un novice) ou « pertinent » (un expert, deux praticiens et deux novices). Seulement deux répondants divergent : un praticien a considéré « peu pertinent » d'agir ainsi et un novice « non pertinent ».

Verbatim du praticien 4 en lien avec l'item 3

« Je ne pense pas que j'annulerais la question. Je pense que j'adapterais ma correction en fonction de la réponse que les élèves m'ont donnée. Si j'ai mis cette question dans mon examen, c'est que je pense qu'elle avait lieu d'être. Je l'enlèverais pour l'année prochaine. »

Verbatim du novice 3 en lien avec l'item 3

« Je n'annulerais pas pour tout le monde. Je la garderais pour ceux qui ont bien compris. Je demanderais alors à ces élèves-là d'expliquer leur réponse à ceux qui n'ont pas compris. »

Ceci étant dit, l'item 3 permet de déceler une certaine tendance de similarité entre les trois profils. Par contre, les experts ont été les uniques répondants à se situer en majorité à l'extrême gauche du continuum de pertinence.

Verbatim de l'expert 3 en lien avec l'item 3

« Dans une épreuve à choix de réponses où tu as plusieurs questions mesurant la même chose, c'est ce qui faut faire. Quand c'est une question à réponse construite, là c'est plus délicat. »

Verbatim de l'expert 1 en lien avec l'item 3

« La question ne donne pas l'information souhaitée. Je trouve donc très pertinent de la retirer. »

Pour continuer, la Figure 37 présente la distribution des réponses des répondants selon le niveau de probabilité à réaliser l'item 3, soit « d'annuler la question ». Une majorité d'experts (n=3/4 experts) ont déterminé comme étant « très probable » qu'ils décident d'agir ainsi contre un seul ayant résolu comme ceci étant « probable ». Chez les praticiens, les réponses sont plus variées : un praticien a opté pour l'option « très probable », deux autres pour « probable » et un dernier pour « peu probable ». Un scénario autant varié se dessine chez les novices : deux d'entre eux ont répondu « très probable », un « peu probable » et un autre « très improbable ». Il semblerait que le fait que les novices soient moins portés à annuler la question problématique soit justifié par l'inquiétude de confronter les élèves pour qui celle-ci a été bien réussie.

Verbatim du novice 3 en lien avec l'item 3

« Je ne sais pas trop ce que je dirais aux élèves qui ont bien réussi. Ils ne seraient pas contents et je les comprends. »

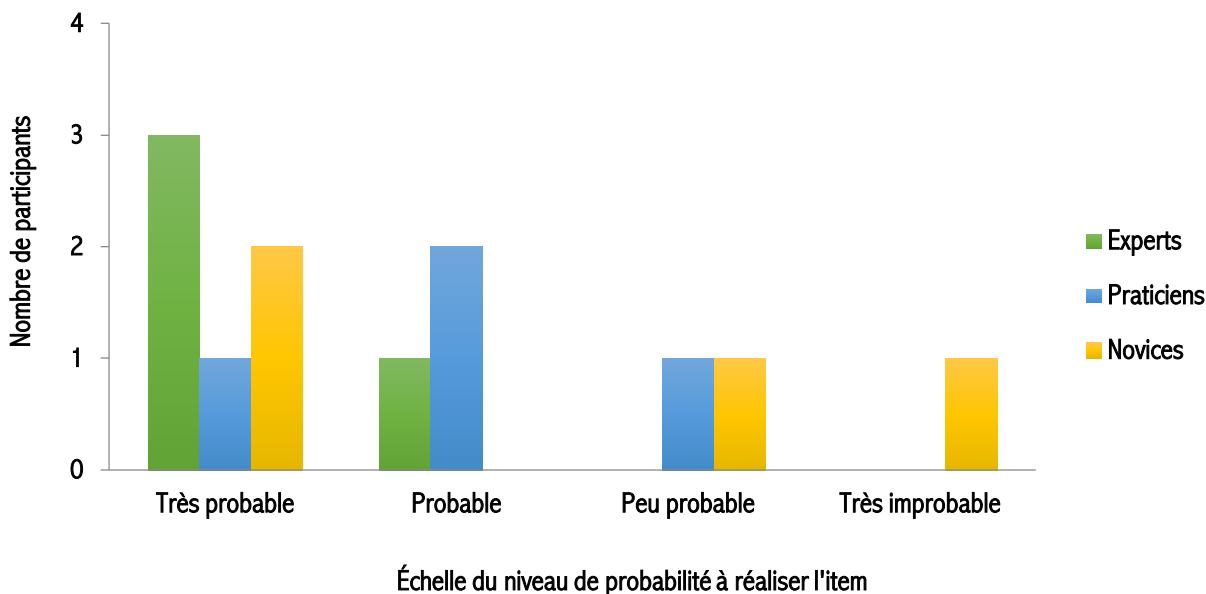


Figure 37. Répartition des réponses des répondants (n=12) à l'échelle B par profil

4.3.7 Synthèse de la vignette 3

En conclusion, la troisième vignette permet, encore une fois, de déceler une meilleure aptitude chez les experts que chez les novices et les praticiens à identifier les concepts abordés par le scénario ainsi que les données manquantes susceptibles d'influencer l'exercice du jugement en évaluation. À la lecture des résultats obtenus au premier item, les praticiens et les experts démontrent un raisonnement similaire en soulignant l'importance de se questionner quand des résultats inattendus se produisent lors d'une évaluation. À l'opposé, les novices se sentent moins alarmés par ce genre de situation si cela avantage les élèves. L'item 2, quant à lui, ne permet pas de constater autant de consensus. En effet, les réponses obtenues à l'échelle A sont variées au sein des trois profils. L'analyse des justifications données par les sujets permettent de mieux comprendre ce qui cause une telle variance. Il semblerait que l'intervention d'administrer une nouvelle question soit bien reçue par l'ensemble des trois profils de sujets, mais que certains praticiens et experts critiquent l'effet surprise qui la contextualise. Enfin, le troisième item démontre un accord entre les experts et les praticiens. Par contre, certains novices se sentent plus fragilisés par l'intervention proposée si cela n'avantage pas les élèves. La peur de décevoir l'élève est donc au cœur de leur décision. En somme, la troisième vignette permet de rendre compte du

jugement professionnel en évaluation des sujets et d'en déceler certaines similitudes ou différences entre les profils de répondants.

4.4 Vignette 4

La Figure 38 présente la quatrième vignette composant notre test de jugement professionnel en évaluation des apprentissages. Nous rappelons que celle-ci a été lue à voix haute en plus d'être accessible à partir du support papier remis à chaque participant.

Les questions se rapportent à la mise en situation qui suit :

Un parent mécontent vous demande une rencontre afin que vous lui expliquiez le résultat particulièrement décevant de son enfant lors du dernier examen.

Figure 38. Cas présenté dans le cadre de la vignette 4 du TJPEA

4.4.1 Authenticité de la vignette 4

À la question « *Trouvez-vous que la vignette présentée représente une situation possible de se dérouler en salle de classe ou dans une école ?* », tous les participants (n=12) ont répondu par l'affirmative.

4.4.2 Identification des concepts

Dans un souci de validité, nous avons posé la question suivante à chaque sujet : « *Quels concepts sont abordés par le cas présenté par la vignette ?* ». Le prochain tableau dresse les concepts disciplinaires relevés par les participants et ceux anticipés par le concepteur de l'instrument, soit un expert en évaluation des apprentissages.

Tableau 15

Tableau comparatif entre les concepts anticipés par le concepteur de la vignette 4 et ceux relevés par les trois groupes de participants (experts, praticiens et novices)

Profils	Concepteur N=1	Experts N=4	Praticiens N=4	Novices N=4
Concepts				
Démarche en évaluation	Oui 1/1	Oui 1/4	Non	Non
Communication avec le parent	Oui 1/1	Oui 4/4	Oui 4/4	Oui 4/4
Analyse d'items	Oui 1/1	Oui 1/4	Oui 2/4	Non
Transparence	Non	Oui 2/4	Non	Non

Selon le point de vue du concepteur, trois principaux concepts liés à l'évaluation des apprentissages émergent de la quatrième vignette du TJPEA : la démarche en évaluation, la communication avec le parent et l'analyse d'items. Du côté des experts, ces trois concepts sont ressortis : la démarche en évaluation (1/4 experts), la communication avec le parent (4/4 experts) et l'analyse d'items (1/4 experts). À ceux-là s'est ajoutée la transparence des résultats (2/4 experts). Chez les praticiens, tous (n=4) ont identifié le concept de la communication avec le parent et la moitié (n=2) celui de l'analyse d'items. Enfin, chez les novices, un seul concept a été relevé, soit celui de la communication avec le parent (n=4).

De manière plus globale, les experts restent les répondants ayant identifié un plus grand nombre de concepts (4). De leur côté, les praticiens relèvent deux concepts et un seul pour les novices. Peu importe le profil, la communication avec le parent reste la seule notion relevée par tous les répondants (n=12).

4.4.3 Données manquantes à la vignette 4

Les figures suivantes présentent les réponses obtenues à la question « *Quelles informations non présentées par la vignette pourraient vous être utiles ?* », par groupe de répondants. La démarche utilisée dans le cadre des trois précédentes a été préconisée.

Tableau 16

Liste des codes et des sous-codes utilisés en lien avec les données manquantes de la vignette 4 du TJPEA

CODE 1 Profil de l'élève	CODE 2 Tâche	CODE 3 Contexte
•Sous-codes 1 <ul style="list-style-type: none">• Niveau d'effort• Notes antérieures	•Sous-codes 2 <ul style="list-style-type: none">• Qualité de l'épreuve• Qualité de la clé de correction• Pondération de l'épreuve• Note obtenue par l'élève• Matière visée par l'épreuve	• Sous-codes 3 <ul style="list-style-type: none">• Moment de l'année• Historique comportemental du parent• Attentes formulées par le parent face à la rencontre

À partir des réponses obtenues à cette question par tous les répondants, trois grandes catégories de données manquantes (codes principaux) se sont créées : celles liées au profil de l'élève, celles liées à la tâche et celles liées au contexte. Pour chacune de ces catégories, des sous-codes ont été identifiés. Ici, la majorité des questionnements émis par nos participants étaient en lien avec la tâche, code pour lequel se sont associés cinq sous-codes, et ce, contrairement aux deux autres catégories : deux sous-codes pour le « profil de l'élève » et trois profils pour le « contexte ».

4.4.3.1 Données manquantes par les novices

Parmi les données manquantes souhaitées par les participants novices, l'historique comportemental du parent, la matière visée par la tâche, la qualité de l'épreuve et le niveau d'effort de l'élève semblent les plus populaires, car elles ont été demandées par la moitié des novices (2/4 novices). Un seul novice a fait la demande d'en savoir davantage sur le moment de l'année et la qualité de la correction effectuée. Bref, au total, sept données manquantes ont été soulevées par ce profil.

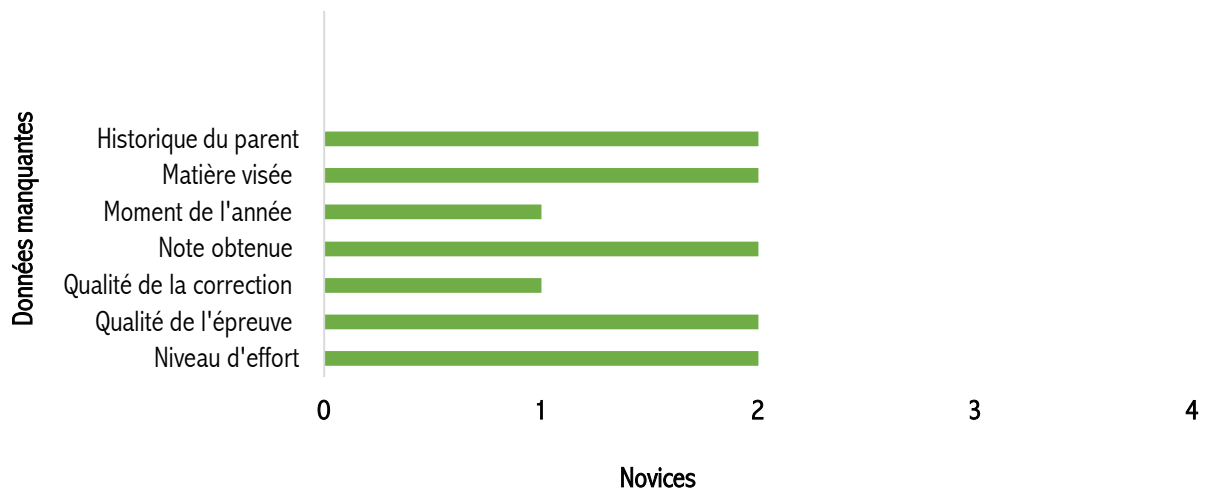


Figure 39. Données manquantes identifiées par les novices (n=4)

4.4.3.2 Données manquantes par les praticiens

Parmi les réponses obtenues chez les praticiens concernant les données manquantes à la vignette 4, les plus fréquentes sont les suivantes : la note obtenue par l'élève (4/4 praticiens) et l'historique comportemental du parent (3/4 praticiens). La moitié (2/4 praticiens) aurait souhaité connaître le moment de l'année où se déroule la situation présentée ainsi que le niveau d'effort de l'élève en salle de classe. Un seul praticien a demandé à en connaître davantage sur la pondération de l'épreuve. Enfin, au total, cinq données manquantes ont été relevées par le groupe de praticiens, dont la majorité d'entre elles l'ont été par plus de la moitié du groupe.

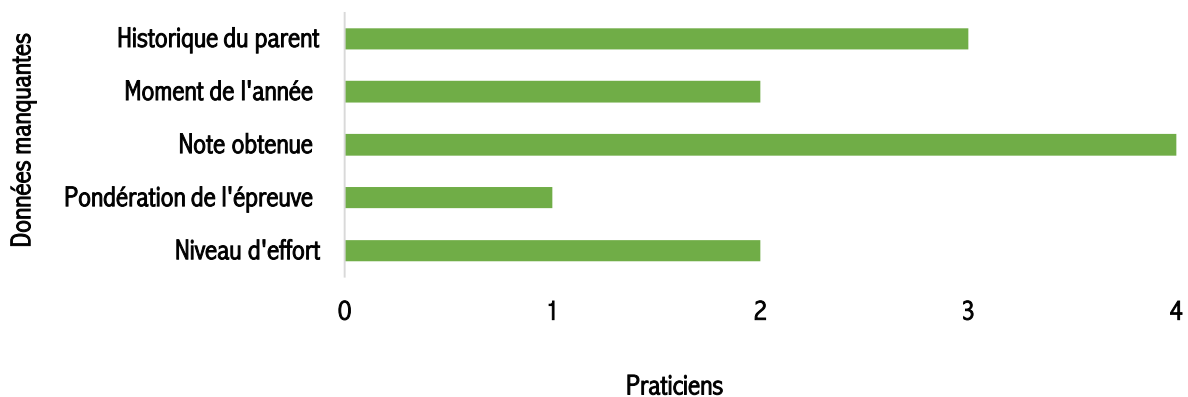


Figure 40. Données manquantes identifiées par les praticiens (n=4)

4.4.3.3 Données manquantes par les experts

Parmi les données manquantes identifiées par les experts, les plus fréquentes touchent les attentes du parent face à la rencontre, la note obtenue par l'étudiant et ses notes antérieures (3/4 experts). Pour continuer, la moitié des répondants de ce groupe (n=2) a mentionné l'influence possible de la matière visée, du moment de l'année, de la qualité de la correction et de la qualité de l'épreuve. Enfin, un seul expert a fait la mention du niveau d'effort comme étant une donnée manquante. Au total, huit informations manquantes ont été relevées par ce groupe dont la majorité d'entre elles ont été mentionnées par plus de la moitié des répondants experts.

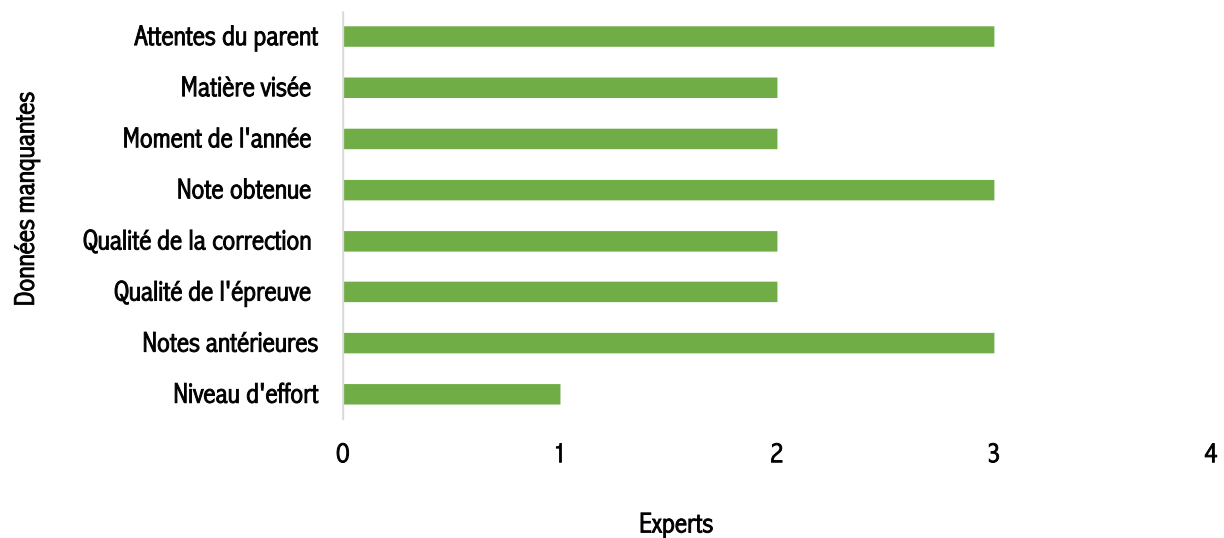


Figure 41. Données manquantes identifiées par les experts (n=4)

4.4.3.4 Synthèse comparative des données manquantes par profil de participants

Afin de répondre à notre question de recherche « *À partir des réponses des participants aux vignettes du TJPEA, pourrions-nous démontrer des différences entre le jugement en évaluation des novices, des praticiens et celui des experts ?* », nous présentons la synthèse des tableaux précédents, à la lumière des différences et des similitudes des données manquantes identifiées par les trois groupes de sujets.

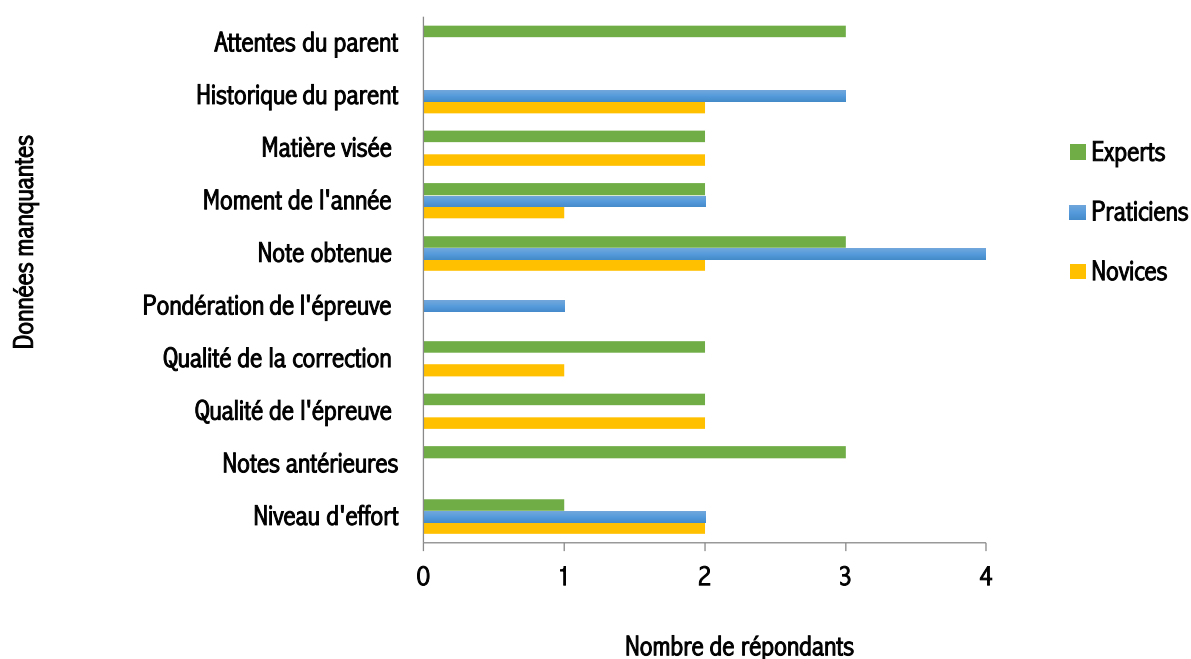


Figure 42. Synthèse des données manquantes identifiées par tous les participants (n=12)
Il faudrait souligner que les praticiens ne remettent pas en question

Comme nous le témoigne la Figure 42, seulement trois données manquantes ont été relevées par les trois profils de répondants : le moment de l'année où se déroule la situation, la note obtenue par l'élève ainsi que son niveau d'effort. Pour ces trois cas, ce sont les praticiens qui ont été les plus nombreux à identifier les données manquantes (huit praticiens, contre six experts et cinq novices). Certaines données, telles que les attentes du parent face à la rencontre et les notes antérieures des élèves n'ont été demandées que par les experts. Sinon, les novices et les experts ont pareillement demandé à en savoir davantage sur la matière visée par la tâche, la qualité de la correction ainsi que la qualité de l'épreuve. Nous tenons à souligner, par ailleurs, que les praticiens ne remettent pas en question la qualité de l'épreuve ni la correction, c'est-à-dire leur propre travail.

Malgré le fait que les trois profils de participants aient identifié un nombre total similaire de données manquantes (sept pour les novices, cinq pour les praticiens et huit pour les experts), la fréquence par donnée par participant n'est pas la même. Sous cet angle, les novices relèvent l'absence d'information 12 fois, les praticiens 10 et les experts 18. Ainsi, les experts sont de ceux qui nomment un plus grand nombre de facteurs susceptibles d'influencer leur jugement face à la

situation présentée. Les praticiens ont, quant à eux, huit fréquences derrières; les novices six fréquences derrières.

La quatrième vignette du TJPEA présente des résultats différents des trois antérieures. Au regard des données manquantes, les experts ont été plus performants en termes d'identification. Par contre, le groupe de novices a aussi obtenu de bons résultats en relevant plus de données manquantes que les praticiens. Toutefois, il faut souligner que cet écart est relativement mince, soit deux identifications de plus pour les novices que les praticiens.

4.4.4 Item 1 de la quatrième vignette

Toujours en lien avec la quatrième vignette du TJPEA, voici le premier item qui a été présenté aux participants.

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Vous rencontrez le parent et lui résumez tout ce que vous avez dit et fait en classe avec vos élèves afin qu'ils se préparent à l'examen.		Le parent juge que vos résultats sont arbitraires et non appuyés.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Dans un premier temps, les sujets devaient se situer sur l'échelle de type Likert présentée en lien avec la pertinence de l'intervention indiquée. La figure suivante illustre bien la distribution des réponses obtenues.

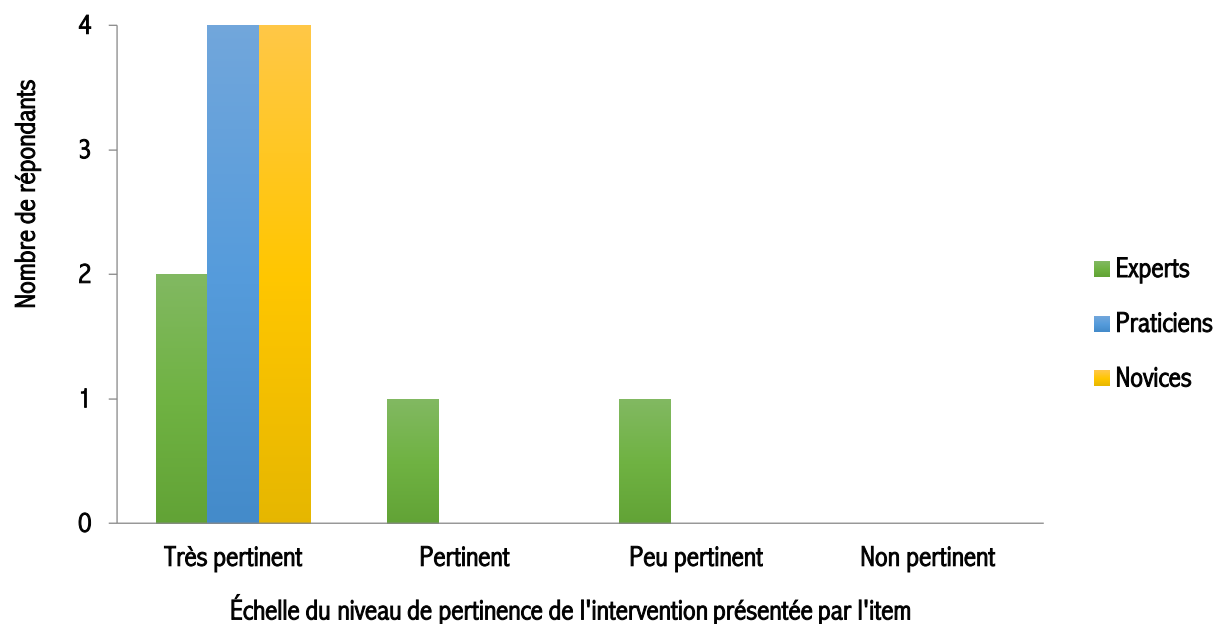


Figure 43. Répartition des réponses des profils de répondants (n=12) à l'échelle A

Au premier regard, nous pouvons remarquer que tous les praticiens (n=4) et les novices (n=4) ont déterminé « très pertinente » l'intervention proposée par l'item 1, soit celle de « résumer tout ce que vous avez dit et fait en classe avec vos élèves afin qu'ils se préparent à l'examen ». Chez les experts, cette option a été choisie que par la moitié des répondants (n=2). Les deux autres ont jugé l'action proposée par l'item 1 comme étant « pertinente » (n=1) et « peu pertinente » (n=1). D'un point de vue comparatif, nous pouvons remarquer une plus grande variance dans les réponses obtenues par les experts. À l'opposé, les praticiens et les novices sont entièrement unanimes.

Verbatim de l'expert 4 en lien avec l'item 1

« Je ne pense pas que l'enseignant doit tout donner ça. Quand le parent va voir l'enseignant, il ne veut pas savoir ce qui a été fait avec tous les élèves... Il veut son enfant à lui. Comment ça se fait que Johnny n'a pas réussi, lui? Si l'enseignant se met à expliquer en long et en large, ce n'est pas ça qu'il veut. Il veut savoir comment était son enfant en classe. L'intervention 1 est beaucoup trop globale. »

Verbatim du praticien 1 en lien avec l'item 1

« J'ai toujours des éléments que j'ai fait en classe pour justifier la pertinence de l'évaluation. Les parents sont mis au courant par le portail... tout ce qu'il faut faire pour se préparer à l'examen... lire ça à ça... J'en ai des traces. »

Verbatim du novice 1 en lien avec l'item 1

« C'est le droit du parent de savoir pourquoi son enfant a eu tel résultat. Moi, je ne trouverais pas ça choquant, mais normal... Je ne pense pas que j'aurais de la difficulté à expliquer à un parent pourquoi son enfant a eu un résultat si décevant. Je parlerais de son comportement, son intérêt face à la matière... Je ferais en sorte d'accumuler les preuves. »

Pour continuer, le graphique suivant démontre la répartition des réponses des répondants en lien avec la deuxième échelle de l'item1, soit celle de la probabilité à réaliser l'action proposée.

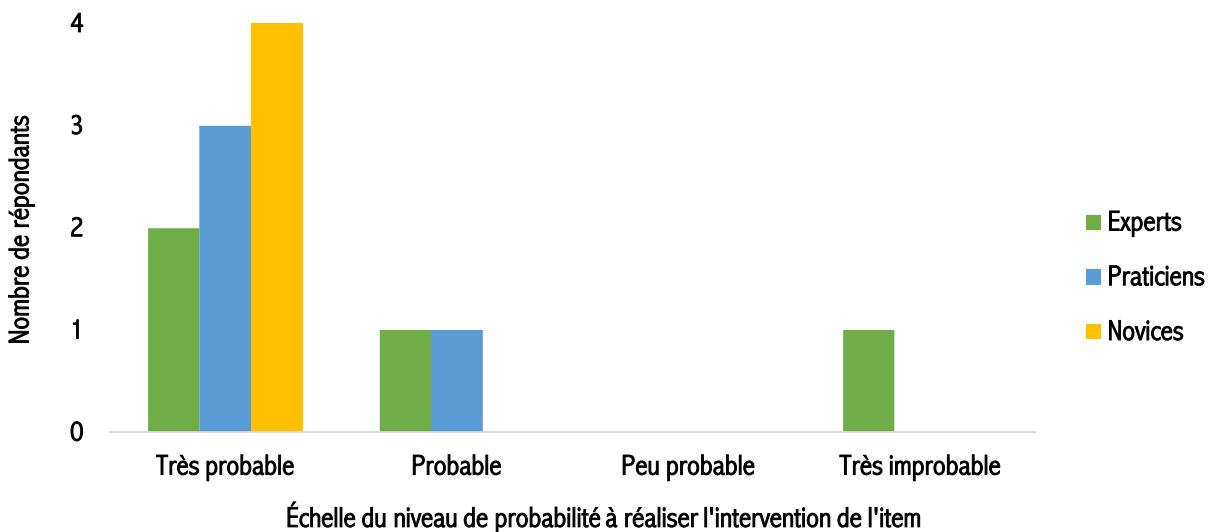


Figure 44. Répartition des réponses des répondants (n=12) à l'échelle B par profil

À première vue, la majorité des répondants (n=9) ont dit « très probable » qu'ils réalisent l'intervention présentée par l'item 1. Parmi eux, un plus grand nombre de novices (n=4) et de

praticiens (n=3) contre un plus petit nombre d'experts (n=2). En effet, d'autres experts ont aussi jugé « probable » (n=1) ou même « très improbable » (n=1) d'agir de la sorte. Enfin, un seul praticien n'a pas opté pour la même réponse que les membres de son profil en répondant « probable ». Malgré cette réponse différente, les extraits de *verbatim* suivants illustrent tout de même une convergence entre les répondants experts.

Verbatim de l'expert 4 en lien avec l'item 1

« Si on est en début de carrière, c'est probablement ce que l'on va décider de faire. Plus on avance en carrière, plus on se rend compte que ce qui intéresse le parent c'est son enfant, pas tout ce qui a été fait en classe avec l'ensemble des élèves. »

Verbatim de l'expert 3 en lien avec l'item 1 ayant répondu « très improbable »

« Le parent veut savoir si son enfant a été bien préparé pour faire l'examen... Ce n'est pas avec les élèves... C'est avec son enfant. L'enseignant pourrait situer l'enfant dans les élèves. Est-ce que son enfant était présent? Attentif? Est-ce qu'il a fait tous les exercices? Quand les parents arrivent c'est toujours à propos de leur enfant. »

4.4.5 Item 2 de la quatrième vignette

Voici le deuxième item présenté aux participants, et ce, en lien avec la quatrième vignette.

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Vous rencontrez le parent et évitez de lui expliquer vos décisions pédagogiques, car vous jugez qu'il n'est pas en mesure de bien comprendre toutes vos décisions.		Le parent demeure mécontent.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

La figure suivante présente la répartition des réponses des répondants à l'échelle A de l'item 2 de la quatrième vignette.

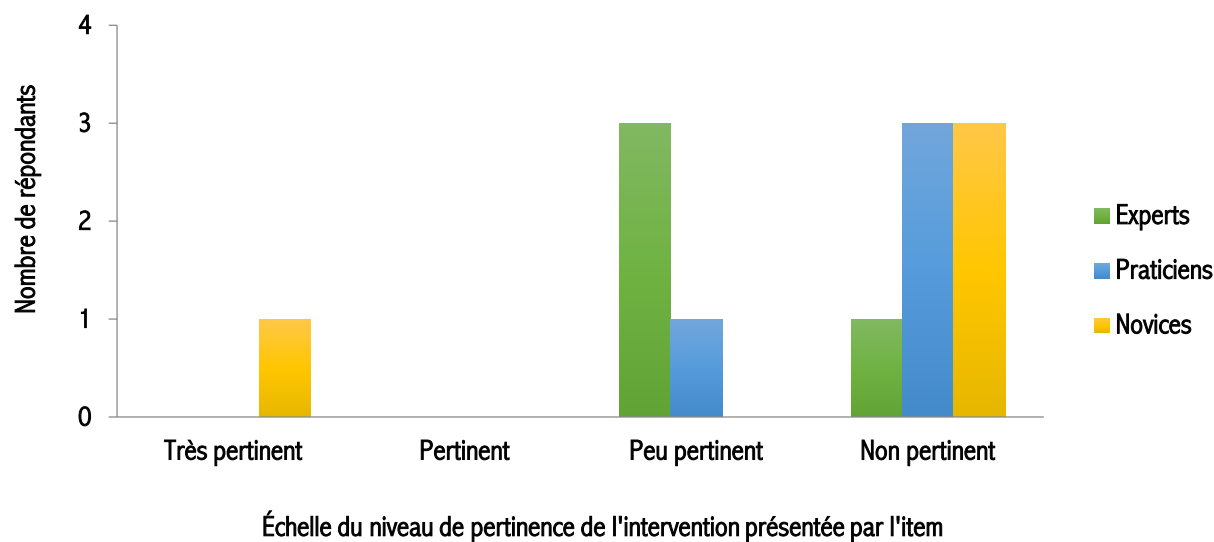


Figure 45. Répartition des réponses des répondants (n=12) à l'échelle A par profil

La répartition des réponses obtenues à l'item 2 est variée, et ce, selon les profils de répondants. Tout d'abord, une majorité de novices (n=3) et de praticiens (n=3) a jugé l'intervention présentée par l'item 2 comme étant « non pertinente ». Un seul expert est allé de l'avant avec cette même option. Plusieurs d'entre eux ont plutôt jugé l'intervention comme étant « peu pertinente », tout comme un praticien du même avis. Enfin, un seul novice a déterminé « très pertinent » d'intervenir comme le proposait l'item 2.

Dans une perspective comparative, l'item 2 de la quatrième vignette nous démontre un accord majoritaire entre le profil des novices et celui des praticiens. Par contre, chez les experts, les réponses ont été moins unanimes. Les *verbatim* suivants illustrent bien cette nuance dans l'exercice de leur jugement.

Verbatim de l'expert 3 en lien avec l'item 2

« C'est une question de transparence. On ne peut pas prendre le parent pour un demeuré non plus. Il faut que tu lui expliques. Je ne pense pas que les enseignants en général aient un langage très conceptuel. Ils devraient être en mesure d'expliquer dans des mots pour que le parent comprenne. »

Verbatim de l'expert 4 en lien avec l'item 2

« Jouer à la cachette n'est jamais une bonne chose. »

Verbatim de l'expert 2 en lien avec l'item 2

« Ce n'est pas toujours agréable de faire face aux parents. C'est une source d'anxiété pour bien des enseignants, mais les parents ont ce droit-là. C'est probablement dans l'intérêt de l'élève de rendre le parent partenaire de cette aventure scolaire. Il faut éviter de fermer la porte aux parents. »

Pour continuer, la Figure 46 présente la probabilité pour les répondants des différents profils à réaliser l'item 2. Tous profils confondus, la majorité des participants (n=10) a considéré « très improbable » de réaliser l'intervention proposée par l'item 2, soit « rencontrer le parent sans lui expliquer vos décisions pédagogiques ». Parmi les personnes ayant opté pour cette option, l'on peut dénombrer quatre experts, trois praticiens et trois novices. Un seul novice et un seul praticien ne sont pas allés dans le même sens en affirmant respectivement « très probable » et « peu probable » d'agir ainsi. Par conséquent, les résultats obtenus démontrent clairement une similitude quant au désir d'impliquer les parents dans le processus de réussite de leur enfant et de dialoguer à propos du sujet.

Verbatim du praticien 1 ayant répondu « peu probable » en lien avec l'item 2

« Ça peut dépendre. Parfois, il y a des décisions pédagogiques qui sont plus difficiles à expliquer comme les contraintes ministérielles. Des fois, je ne dis pas tout, mais je vais quand même mentionner l'essentiel. Par contre, je veux que le parent comprenne et je vais vulgariser. Je veux qu'il soit capable d'aider son enfant. »

Verbatim du novice 1 ayant répondu « très probable » en lien avec l'item 2

« Ce n'est pas donné à tout le monde de comprendre ce qu'un enseignant peut faire. Parfois, en partant, on voit que les parents ne sont pas aptes à comprendre les décisions pédagogiques. »

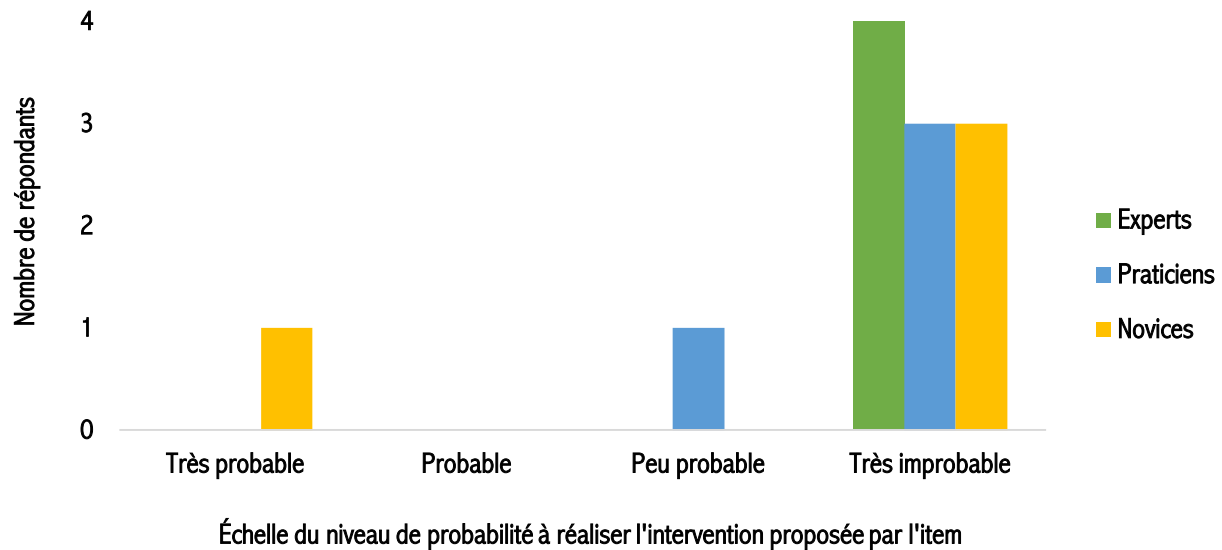


Figure 46. Répartition des réponses des répondants (n=12) à l'échelle B par profil

4.4.6 Item 3 de la quatrième vignette

Enfin, le troisième item en lien avec la quatrième vignette souligne la possibilité de rencontrer le parent avec un collègue. Vous pouvez la lire comme suit :

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Vous rencontrez le parent avec un collègue qui appuie sans contredit votre décision, mais sans en préciser les raisons.		Le parent accepte votre décision qui est appuyée par l'un de vos collègues.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

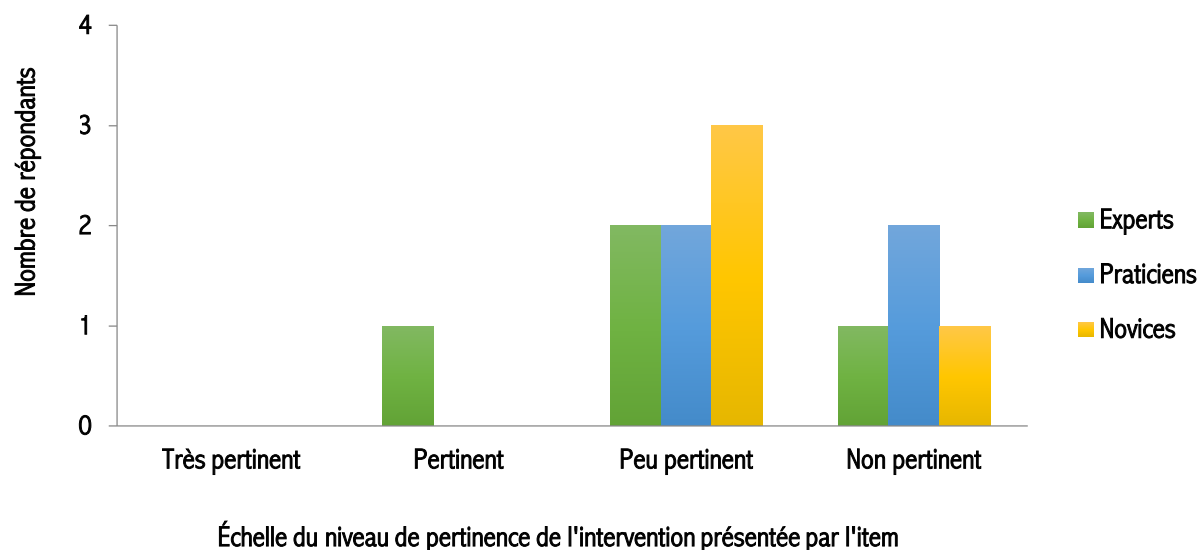


Figure 47. Répartition des réponses des répondants (n=12) à l'échelle A par profil

Au premier regard, la majorité des réponses des répondants (n=11) se trouvent à droite du continuum, soit en optant pour l'option « peu pertinent » (deux experts, deux praticiens et trois novices) ou « non pertinent » (un expert, deux praticiens et un novice).

Verbatim de l'expert 1 en lien avec l'item 3

« Ce n'est pas pertinent. Tu amènes un collègue pour dire ce que tu dis. C'est juste une manière de calmer le jeu... de dire au parent que je ne suis pas tout seul. Ça pourrait être fait par quelqu'un qui n'est pas en confiance. »

Verbatim du novice 2 en lien avec l'item 3

« La partie que je n'aime pas c'est de ne pas en justifier les raisons. C'est comme si j'amenais ma gang pour m'appuyer. C'est le deux contre un. Je n'aime pas ça. Il faut être capable de se justifier. »

Verbatim du praticien 3 en lien avec l'item 3

« C'est quasi un geste d'intimidation à l'encontre du parent... »

Seul un expert diverge en considérant « pertinent » d'agir ainsi. Le *verbatim* suivant explique pourquoi.

Verbatim de l'expert 4 en lien avec l'item 3

« Ça ajoute du poids au jugement de l'enseignant. Par contre, cela dépend qui est le collègue. Il faudrait que cela soit un collègue qui enseigne aussi à l'élève. Par exemple, un prof de maths et un prof de sciences où l'élève a aussi certaines difficultés ça ajoute du poids. »

Ceci étant dit, l'item 3 permet tout de même de déceler une similarité entre les trois profils.

Pour continuer, la Figure 48 présente la distribution des réponses des répondants selon le niveau de probabilité à réaliser l'item 3. Une majorité de répondants (deux individus/profil) ont jugé « très improbable » de rencontrer le parent accompagné d'un collègue. Cinq autres répondants ont également jugé ceci « peu probable » (deux novices, deux praticiens et un expert). Enfin, un seul expert a jugé « probable » de réaliser l'intervention proposée.

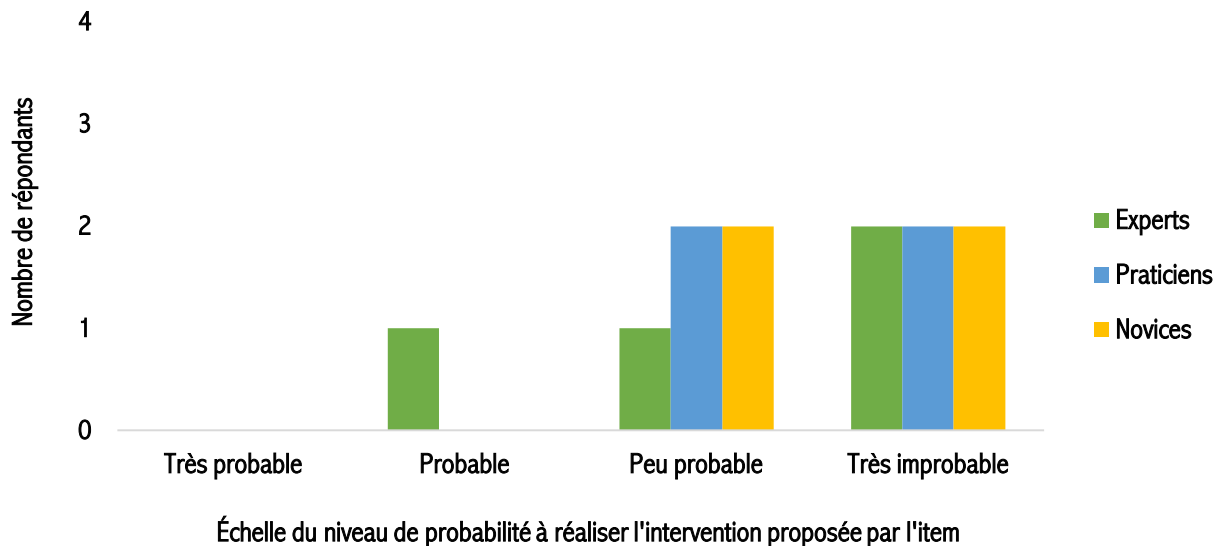


Figure 48. Répartition des réponses des répondants (n=12) à l'échelle B par profil

Les justifications données par nos sujets démontrent clairement que peu de personnes ressentent le besoin de se faire accompagner dans le cadre d'une situation similaire. Au contraire, plusieurs n'en comprenaient pas la pertinence et y percevaient plutôt un certain « acte de pouvoir » envers le parent.

Enfin, l'expert pour lequel la réponse est divergente du reste des participants se justifie de la manière suivante :

Verbatim de l'expert 1 en lien avec l'item 3

« Cela peut être favorable de se faire accompagner quand cela est pertinent pour la situation de l'élève. »

Ainsi, bien que la réponse à l'échelle B de cet expert soit différente du reste des sujets constituant son profil, nous pouvons comprendre que son choix est justifié par le souci qu'il porte à la réussite de l'élève. Subséquemment, ce sujet perçoit l'intervention proposée comme une ressource supplémentaire fournie au parent afin de mieux saisir ce qui se passe avec son enfant. Bref, l'analyse du *verbatim* semble illustrer un accord entre son avis et ceux des autres répondants (tous profils confondus).

4.4.6 Synthèse de la vignette 4

En conclusion, la quatrième vignette permet de déceler une meilleure aptitude chez les experts que chez les novices et les praticiens à identifier les concepts abordés par le scénario ainsi que les données manquantes susceptibles d'influencer l'exercice du jugement en évaluation. Contrairement aux trois vignettes précédentes de ce TJPEA, celle-ci permet d'illustrer une similitude de raisonnement en lien avec les trois items présentés. En effet, malgré certaines variances au sein des réponses obtenues aux échelles A et B, les justifications données étaient ressemblantes : le parent se doit d'être informé du cheminement scolaire de son enfant, la transparence est de vigueur dans la communication des résultats et tout doit être mis en place pour assurer la réussite de l'élève. En somme, la quatrième vignette permet de rendre compte du jugement professionnel en évaluation des sujets et d'en déceler des similitudes entre les profils de répondants.

4.5 Synthèse des résultats

La dernière section de ce chapitre fera ressortir les principales conclusions qui se dégagent de cette étude en lien avec la question générale de recherche, soit : *Au final, les vignettes présentées par le TJPEA permettent-elles de rendre compte du jugement en évaluation des enseignants ?* ainsi que la question spécifique : *À partir des réponses des participants aux vignettes du TJPEA, pourrions-nous démontrer des différences entre le jugement en évaluation des novices, des praticiens et celui des experts ?* Les résultats obtenus aux quatre vignettes composant le TJPEA nous mènent à tirer les conclusions suivantes.

- 1) On constate que les experts et les praticiens sont davantage en mesure de repérer les données manquantes que les novices; ce qui est cohérent avec la littérature. Plus précisément, les experts ont obtenu une plus grande fréquence d'identification des données manquantes à la première et à la quatrième vignette. Les praticiens en ont relevé plus à la deuxième et à la troisième. Toutes vignettes confondues, les novices ont observé un moins grand nombre de données.

Sur le plan qualitatif, les données manquantes relevées par un profil de répondants le sont parfois par les deux autres, parfois pas. La présente collecte de données n'a donc pas permis d'identifier clairement de tendances quant aux raisons expliquant pourquoi certaines données sont rapportées par un profil en lien avec un item et pas par d'autres.

- 2) On relève que les experts et les praticiens sont davantage en mesure de nommer les concepts en évaluation abordés par les vignettes que les novices. Pour l'ensemble des scénarios du TJPEA, les experts ont soutiré des situations présentées un plus grand nombre de concepts liés à la discipline ciblée que les praticiens. Ces derniers ont, de leur côté, reconnu un plus grand nombre que les novices. Enfin, contrairement aux experts, les praticiens et les novices ont éprouvé de la difficulté à nommer les concepts en utilisant des termes spécifiques. Cette situation a demandé à l'étudiante-chercheuse de soutirer des participants de nombreuses reformulations et d'inviter les sujets praticiens et novices à « dire dans leurs mots » ce qu'ils croyaient être les thèmes liés à l'évaluation.

- 3) On remarque que, selon les vignettes et les items, la variabilité ou le consensus des réponses aux échelles A et B sont présents. Le plus souvent, les experts et les praticiens ont démontré des similitudes dans leur raisonnement en optant pour des réponses similaires (vignette 1, items 1 et 2; vignette 3, items 1 et 2). Dans d'autres cas, le consensus a été remarqué au sein des trois profils (vignette 2, items 1, 2 et 3; vignette 4, items 1, 2 et 3).
- 4) L'analyse des *verbatim* des justifications données aux items nous a permis de constater une plus grande profondeur réflexive chez les experts. Plus précisément, la qualité de l'argumentation des experts s'est démarquée par (1) une meilleure connaissance des concepts en évaluation, (2) une argumentation plus étayée en termes de nuances et de recadrages des scénarios présentés et (3) une plus grande propension à faire des liens et à arrimer la théorie à la pratique. Les praticiens, de leur côté, ont partagé des propositions, mais en moins grand nombre et de manière moins spécifique. Enfin, chez les novices, les réponses étaient plus souvent de l'ordre de « l'intuitif », car très peu d'entre eux étaient aptes à argumenter et à appuyer leur démarche par des concepts liés à la discipline de l'évaluation des apprentissages.
- 5) Les réponses obtenues aux échelles A et B (pertinence de l'intervention et probabilité à réaliser l'item) nous indiquent que les participants sont cohérents entre les deux. De ce fait, quand le répondant se prononce relativement au niveau de pertinence de l'intervention fournie par un item, sa réponse quant à sa probabilité à réaliser celle-ci va dans le même sens. Cette cohérence entre les deux échelles peut être expliquée par la perception de faisabilité de l'intervention proposée au répondant, ou encore par un effet limité de la « désirabilité sociale ». Comme on reconnaît au construit du jugement en évaluation une « sensibilité », nous considérons prometteur de constater une telle cohérence. En contexte pédagogique à faibles enjeux comme celui de l'évaluation formative, l'outil du TJPEA pourrait ainsi faciliter l'exercice du jugement en permettant aux individus de se prononcer sans nécessairement se sentir « jugés » par d'autres.
- 6) Nous jugeons que l'instrument du TJPEA, pour rendre compte des nuances de jugement entre les profils de répondants (novices, praticiens et novices), se doit de combiner : (1) la

réponse du sujet aux échelles de Likert A et B et (2) une justification. C'est cette dernière, dans certains cas, qui nous a permis de mieux comprendre les facteurs susceptibles d'influencer la réponse donnée par le participant. Par exemple, au troisième item de la deuxième vignette, un expert et un novice ont tous deux répondu « très pertinent ». Par contre, la justification nous a permis de constater que les deux sujets le faisaient pour des raisons différentes. Le novice a trouvé l'intervention « très pertinente », car il « n'avait pas envie de décourager ses élèves », tandis que l'expert a opté pour la même réponse, mais pour pouvoir « analyser l'outil ».

En somme, nous soulignons la limite de l'utilisation unique des scores aux échelles de Likert comme le propose le gabarit original du test de concordance de script dont cet instrument est inspiré.

Conclusion

*« La science progresse en indiquant l'immensité de l'ignoré. »
Pauwels*

Pour conclure cet essai, le cinquième chapitre présente un bilan du travail de recherche mené, et ce, au regard des questions générale et spécifique ciblées par l'étude. S'enchainent ensuite les limites du projet accompli ainsi que les pistes d'études éventuelles.

5.1 Bilan du travail de recherche

Cette étude exploratoire qualitative interprétative a été menée auprès de douze participants catégorisés en trois profils différents : quatre étudiants de première année à la formation des maîtres au secondaire (novices), quatre enseignants d'expérience oeuvrant dans une école secondaire québécoise (praticiens) et quatre professeurs universitaires spécialisés en évaluation des apprentissages (experts). L'objectif premier du présent projet était de mieux comprendre le processus de développement du jugement en évaluation des apprentissages des futurs enseignants par la passation d'un test de jugement professionnel en évaluation des apprentissages (TJPEA). Plus précisément, l'on cherchait à déterminer si les vignettes constituant le TJPEA permettraient de démontrer des différences de jugement entre les trois profils de répondants.

L'originalité de cette étude émane entre autres du fait que l'instrument en question soit inspiré des travaux issus de la pédagogie des sciences de la santé, plus précisément de ceux en lien avec le potentiel pédagogique du test de concordance de script (TCS). Ce dispositif est utilisé dans le cadre de la formation des futurs médecins, plus précisément à des visées de développement du raisonnement clinique. Nous avons établi que plusieurs similitudes existaient entre ce dernier et le jugement professionnel en évaluation des apprentissages qu'exercent les enseignants. Dans une perspective où aucun outil n'est disponible à son développement et que la compétence à évaluer est une tâche prépondérante de la profession enseignante, il devient novateur d'aborder ce défi. En effet, nombre de chercheurs discutent du jugement en évaluation, mais très peu se sont risqués à emprunter le sinueux chemin de l'instrumentation. En nous intéressant aux propriétés de l'instrument du TJPEA développé par Dionne, nous devenons la première étude à documenter

cette avenue. En ce sens, nous espérons que d'autres empruntent aussi cette route, car il est impératif d'avoir une collection d'instruments, afin de fournir un matériel crédible et scientifiquement fondé aux formateurs d'enseignants.

Pour continuer, la collecte de données s'est réalisée par le biais d'entrevues semi-dirigées guidées par la passation de quatre vignettes distinctes composant un TJPEA. Les données de cette recherche ont contribué à identifier les différents facteurs susceptibles d'influencer le jugement professionnel en évaluation des répondants. Cela a également permis de comprendre davantage la dynamique de progression du développement de la compétence à évaluer, notamment en identifiant les différents schèmes de pensée entre les profils de répondants. Les résultats ont également fait ressortir les différents obstacles à l'exercice du jugement évaluatif des sujets. En ce sens, l'appropriation de la compétence 5 « évaluer la progression des apprentissages et le degré d'acquisition des compétences des élèves pour les contenus à apprendre » du *Référentiel des compétences professionnelles des enseignants* (MEQ, 2001) n'est pas si simple.

Malgré l'encadrement que propose la formation des maîtres dans les universités québécoises, l'unique cours destiné à l'évaluation des apprentissages ne semble pas suffisant pour préparer la relève à la réalité évaluative des enseignants. Les résultats obtenus par la passation du TJPEA paraissent donner espoir quant au potentiel pédagogique possible en contexte de formation initiale et continue que peut jouer l'instrument à l'égard de ce défi. Plus précisément, la situation authentique que propose chacune des vignettes permet au répondant de se questionner et de s'affirmer quant à un choix à faire. Cette réflexion favorise un apprentissage actif et un ancrage de concepts vus en contexte théorique universitaire, et ce, sans même devoir attendre d'être confronté au terrain des écoles. Nous pensons que cela constitue un terreau favorable au développement d'une compétence telle que celle ciblée par notre étude.

5.2 Limites à l'étude

Cette recherche présente des limites que nous voudrions aborder au cours des prochaines lignes. Tout d'abord, malgré l'importante quantité de données recueillies, seulement quatre individus constituent chaque catégorie de profils (novices, praticiens et experts). Les données obtenues ne peuvent donc pas être généralisées à l'ensemble des groupes ciblés. Toutefois, comme

la visée de cette étude était essentiellement exploratoire, nous tenons à mentionner que cela n'était pas le but recherché. Il aurait tout de même été intéressant de raffiner les critères de sélection des individus, surtout en lien avec la sélection des experts. Des études ont été menées dans ce sens avec le TCS, mais jamais dans le cadre du présent instrument.

En outre, l'étude a recueilli de l'information quant aux différences et aux similitudes du jugement des répondants confrontés aux quatre vignettes, et ce, par le biais d'entrevues semi-dirigées. Ces dernières comportent des limites qui ont pu influencer les données récoltées. À cet effet, Savoie-Zajc (2009, p.357) souligne que « l'interviewé peut être mû par un désir de rendre service ou d'être bien vu par le chercheur, limitant ainsi la crédibilité des messages communiqués ». Conjointement, les participants ont pu vivre certaines réticences ou blocages faisant en sorte que « l'entrevue n'ait pas permis d'engager une véritable discussion sur un sujet » (idem, 2009). Enfin, une validation de chaque retranscription du *verbatim* auprès des sujets aurait confirmé une compréhension juste des justifications fournies.

5.3 Pistes de recherches éventuelles

Considérant l'objectif poursuivi, cette étude a permis de reconnaître un potentiel pédagogique à l'instrument du TJPEA en rendant compte du jugement en évaluation des répondants. Ce projet a contribué, au plan théorique, à la production de connaissances sur les facteurs possibles d'influencer l'exercice du jugement évaluatif des enseignants. Au plan instrumental, cette recherche possède un caractère novateur; ce qui a permis de reconnaître les adaptations possibles du TCS à d'autres construits complexes comme le jugement professionnel en évaluation. Au plan praxéologique, cette initiative devrait fournir aux futurs étudiants à la formation des maîtres un outil de formation susceptible de guider leurs pratiques évaluatives et, éventuellement, de développer plus rapidement leur jugement professionnel en évaluation.

Nous croyons fermement que cette étude permet des pistes nouvelles pour de futurs travaux en lien avec les pratiques évaluatives des enseignants. Il convient de préciser que nos résultats ne permettent pas d'identifier clairement tous les facteurs possibles permettant d'influencer le

jugement des répondants, mais que les réponses obtenues par la passation du TJPEA indiquent tout de même la présence de différences de jugement en évaluation des novices, des praticiens et des experts. Quiconque s'intéresse au jugement en évaluation des apprentissages pourrait être tenté de préciser cette observation par le biais d'une expérimentation à plus grande échelle d'un TJPEA composé de plus de scénarios et où un plus grand nombre de sujets sont invités à justifier leurs réponses. Un exercice intéressant à mener dans ce sens serait de catégoriser les vignettes présentées en fonction des construits qu'elles caractérisent. Ainsi, il serait possible de modéliser les processus cognitifs impliqués et de tester à nouveau l'instrument. En conclusion, d'éventuelles retombées positives permettraient alors à cet instrument d'agir à titre de levier favorable à l'émergence de pratiques évaluatives exemplaires et cohérentes aux aspects théoriques et éthiques auxquels est confrontée la population enseignante.

Références

- Allal, L. (2007). Régulations des apprentissages : orientations conceptuelles pour la recherche et la pratique en éducation. Dans L. Allal et L. Mottier Lopez (dir.), *Régulation des apprentissages en situation scolaire et en formation* (p. 7-23). Bruxelles, Belgique : De Boeck.
- Allal, L. et L. Lafortune, L. (2008). À la recherche du jugement professionnel. Dans Lafortune, L. et Allal, A. (dir.), *Jugement professionnel en évaluation. Pratiques enseignantes au Québec et à Genève* (p. 1-10). Québec, Canada : Presses de l'Université du Québec.
- Baker, E. L. (1997). Model-based performance assessment. *Theory into Practice*, 36(4), 247-54.
- Baribeau, A. (2009). *Analyse des pratiques d'évaluation d'enseignants de français, langue d'enseignement, pour établir un bilan des apprentissages au premier cycle du secondaire* (mémoire de maîtrise, Université du Québec à Trois-Rivières, Canada). Récupéré des archives électroniques de l'UQTR : <http://depot-e.uqtr.ca/1187/1/030125025.pdf>
- Bélair, L. M. (1999). *L'évaluation dans l'école. Nouvelles pratiques*. Paris, France : ESF Éditeur.
- Bélair, L. M. et Dionne, E. (2009). Évaluation des apprentissages dans le contexte québécois. *Mesure et évaluation en éducation*, 32(3), 77-100.
- Black, P. et Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education*. 5(1), 7-74.
- Blais, J.-G., Charlin, B., Grondin, J., Lambert, C., Loye, N. et Gagnon, R. (2011). L'estimation de l'accord entre les experts lors du calibrage d'un test de concordance de script avec le modèle à facettes de Rasch. Dans Raïche, G., Paquette-Côté, K. et Magis, D. (dir) : *Des mécanismes pour assurer la validité de l'interprétation de la mesure en éducation* (p. 139-162). Ste-Foy, Québec : Presses de l'Université du Québec.
- Bressoux, P. et Pansu, P. (2003). *Quand les enseignants jugent leurs élèves*. Paris, France : Presses universitaires de France.
- Brousselle, A., Champagne, F., Contandriopoulos, A-P. et Hartz, Z. (2009). *L'évaluation : concepts et méthodes*. Québec, Canada : Les Presses de l'Université de Montréal.
- Burnett, N. (2008). The Delors Report: A guide towards Education for All. *European Journal of Education*, 43(2), 181-187.
- Bursztejn, A. C., Cuny, J. F., Adam, J. L., Sido, L., Schmutz, J. L., De Korwin, J. D., Latarche, C., Braun, M. et Barbaud, A. (2011). Usefulness of the script concordance test in dermatology. *Journal of the European Academy of Dermatology and Venereology*, 25(12), 1471-1475.

- Buysse, A. (2011). Les mécanismes inductifs et déductifs dans la genèse des savoirs des enseignants en formation. Dans Maubant, P., Clénet, J. et Poisson, D. (dir.), *Débats sur la professionnalisation des enseignants. Les apports de la formation des adultes* (p. 267-307). Sainte-Foy, Québec : Presses de l'Université du Québec.
- Caire, F., Sol J. C., Charlin B., Isidori, P. et Moreau, J. J. (2004). Le test de concordance de script (TCS) comme outil d'évaluation formative des internes en neurochirurgie : implantation du test sur Internet à l'échelle nationale. *Pédagogie médicale*, 5, 87-94.
- Carrière, B., Gagnon, R., Charlin, R., Downing, S., et Bordage, G. (2009). Assessing clinical reasoning in pediatric emergency medicine : validity evidence for a script concordance test. *Annals of Emergency Medicine*, 53 (5), 647-652.
- Cattonar, B. (2008). L'entrée dans le métier d'enseignant : un moment important d'élaboration identitaire. Dans Portelance, L., Mukamurera, J., Martineau, S. et Gervais, C. (dir.), *L'insertion dans le milieu scolaire. Une phase cruciale du développement professionnel de l'enseignant* (p. 87-103). Québec, Québec : Les Presses de l'Université Laval.
- Champagne, F., Contandriopoulos, A. P., Brousselle, A., Hartz, Z. et Denis, J. L. (2009). L'évaluation dans le domaine de la santé : concepts et méthodes. Dans Brousselle, A., Champagne, F., Contandriopoulos, A. P. et Hartz, Z. (dir.), *L'évaluation : concepts et méthodes* (p. 35-56). Québec, Canada : Les presses de l'Université de Montréal.
- Charlin, B., Gagnon, R., Pelletier, J., Coletti, M., Abi-Rizk, G., Nasr, C., Sauvé, E. et Vleuten, C. (2006). Assessment of clinical reasoning in the context of uncertainty : the effect of variability within the reference panel. *Medical Education*, 18, 22-27.
- Charlin, B., Brailovsky, C. A., Brazeau-Lamontagne, L., Samson, L. et Leduc, C. (1998). Script questionnaires : their use for assessment of diagnostic knowledge in radiology. *Med Teacher*, 20, 567-571.
- Charlin, B., Tardif, J. et Boshuizen, H. P. (2000). Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Academic Medicine*, 75(2), 182-190.
- Charlin, B., Gagnon, R., Sibert, L. et Van der Vleuten, C. (2002). Le test de concordance de script : un instrument d'évaluation du raisonnement clinique. *Pédagogie Médicale*, 3, 135-144.
- Charlin, B., Gagnon, R., Kazi-Tani, D. et Thivierge, R. (2005). Le test de concordance comme outil d'évaluation en ligne du raisonnement des professionnels en situation d'incertitude. *Revue internationale des technologies en pédagogie universitaire*, 2(2), 22-27.
- Charlin, B., Gagnon, R., Sauvé, É. et Coletti, M. (2007). Composition of the panel of reference for concordance tests: Do teaching functions have an impact on examinees ranks and absolute scores? *Medical Teacher*, 29(1), 49-53.

- Claire, M. (2014). Étude exploratoire du raisonnement clinique chez les étudiants en soins infirmiers à l'aide de la carte conceptuelle. *Recherche en soins infirmiers*, (2), 85.
- Crooks, T. J. (1988). The Impact of classroom evaluation practices on students. *Review of educational research*. 58(4), 438-481.
- De Ketele, J. M. (2000). En guise de synthèse : Convergences autour des compétences. Dans Bosman, C., Gerard, F. M. et Roegiers, X. (dir.), *Quel avenir pour les compétences ?* (p. 187-191). Bruxelles, Belgique : De Boeck.
- De Ketele, J. M. et Gerard, F. M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences. *Mesure et évaluation en éducation*, 28(3), 1-26.
- De Ketele, J. M. (2009). L'évaluation de l'école par les standards internationaux. *Revue Internationale d'éducation*, numéro 52, 69-76.
- Delors, J. (dir.) (1996). *L'éducation : un trésor est caché dedans, rapport à l'UNESCO de la Commission internationale sur l'éducation pour le vingt et unième siècle*. Paris, France : Éditions UNESCO / Éditions Odile Jacob.
- Deslauriers, J. P. (1991). *Recherche qualitative : guide pratique*. Montréal, Canada : McGraw-Hill.
- 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. (2014). Un canevas d'item pour évaluer la compétence d'investigation scientifique en laboratoire. *McGill Journal of Education*, 49(1), 89-109.
- Dionne, E. et Simoes Forte, L. A. (2012). *Le Test de Jugement Professionnel en Évaluation des Apprentissages (TJPEA) : un outil visant à mesurer des compétences avec des télévotants*. Communication présentée au Colloque de la Maison des technologies de formation et d'apprentissage Roland-Giguère (MATI Montréal), Montréal, Canada.
- Dory, V., Gagnon, R., Vanpee, D. et Charlin, B. (2012). How to construct and implement script concordance tests : insights from a systematic review. *Medical Education*, 46, 552-563.
- Downing, S. M. et Haladyna, T. M. (2006). *Handbook of test development*. Mahwah, N. J.: L. Erlbaum.
- Draxler, A. (2010). Un monde de rapports ? *La lettre du Norrag*, 43, 27-30.

- Drapeau, M. (2004). Les critères de scientificité en recherche qualitative. *Pratiques Psychologiques*, 10(1), 79-86.
- Dumont, K. (2013). *Étude exploratoire du potentiel diagnostique des questions d'un test de concordance de scripts (TCS) pour évaluer le raisonnement clinique infirmier* (mémoire de maîtrise, Université de Montréal, Canada). Récupéré des archives de l'Université de Montréal : <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/9941>
- Durand, M. J. et Chouinard, R. (2012). *L'évaluation des apprentissages. De la planification de la démarche à la communication des résultats* (2e éd.). Montréal, Canada : Éditions Marcel Didier.
- Durand, M. J. (2010). Documenter le jugement des enseignants en langue seconde au 3e cycle du primaire au Québec dans le cadre du bilan des apprentissages de fin de cycle. *Revue des sciences de l'éducation de McGill*, 45(1), 129-150.
- Ericsson, K.A. et Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215-251.
- Eva, K. (2005). Ce que tout enseignant devrait savoir concernant le raisonnement clinique. *Pédagogie Médicale*, 6(4), 225-234.
- Fontaine, S., Savoie-Zajc, L., Cadieux, A., Lainey, C., Leclerc, M., et Labelle, J. (2013). L'impact des CAP sur le développement de la compétence des enseignants en évaluation des apprentissages. *Éducation et Francophonie*, 41(2), 10-34.
- Fournier, J. P., Thiercelin, D., Pulcini, C. et al. (2006). Clinical reasoning assessment in emergency medicine : script concordance tests are more efficient to detect clinical experience than rich context multiple choice questions. *Pédagogie Médicale*, 7, 20-30.
- Fournier, J. P., Demeester, A. et Charlin, B. (2008). Script Concordance Tests: Guidelines for Construction. *BMC Medical informations and Decision Making*, 8(18).
- Figari, G. (2001). *L'activité évaluative réinterrogée : Regards scolaires et socioprofessionnels*. Paris, France : De Boeck.
- Gagnon, R., Charlin, B., Coletti, M., Sauvé, E. et Van der Vleuten, C. (2005). Assessment in context of uncertainty: How many members are needed on the panel of reference of a script concordance test? *Medical Education*, 39, 204-291.
- Gagnon, R., Charlin, B., Lambert, C., Carrière, B. et Van der Vleuten, C. (2008). Script Concordance Testing: More Cases or More Questions? *Adv in Health Sci Educ*, 14(13), 1382-4996.
- Gérard, F. M. (2002). L'indispensable subjectivité de l'évaluation, *Antipodes*, numéro 156, 26-34.

- Gérard, F. M. (2008). *Évaluer les compétences. Guide pratique*. Bruxelles, Belgique : De Boeck.
- Gohier, C., Bednarz, N., Gaudreau, L., Pallascio, R. et Parent, G. (1999). *L'enseignant, un professionnel*. Québec, Canada : Presses de l'Université de Québec.
- Groves, M., Scott, I. et Alexander, H. (2002). Assessing clinical reasoning: A method to monitor its development in a PBL curriculum. *Medical Teacher*, 24(5), 507-515.
- Guba, E. G. et Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, États-Unis : Sage.
- Hattie, J. et Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Hébert, M. H. (2013). *Validation d'une épreuve pour rendre compte du niveau de développement des compétences du programme de mathématique pour l'enseignement primaire* (thèse de doctorat, Université Laval, Québec, Canada). Récupérée des archives de l'Université Laval : www.theses.ulaval.ca/2013/29810/29810.pdf
- Hobbs, N. (1975). *Issues in the classification of children* (vol. 1). San Francisco, États-Unis : Jossey-Bass.
- 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.
- Huberman, M. et Miles, M. B. (1991). *Analyse des données qualitatives : recueil de nouvelles méthodes*. Bruxelles, Belgique : De Boeck.
- Huberman, A. M. et Miles, M. B. (2002). *The qualitative researcher's companion*. Thousand Oaks, CA : Sage Publications.
- Hudson, S. M., Beutel, D. A. et Hudson, P. B. (2008, janvier). *Beginning teachers' perceptions of their induction into teaching*. Communication présentée à la 8^e Practical Experiences in Professional Education International Conférence. Édimbourg, Écosse.
- Jeffrey, D. (2013). *L'éthique dans l'évaluation scolaire*. Québec, Canada : Presses de l'Université Laval.
- Jonnaert, P. (2002). *Compétences et socioconstructivisme. Un cadre théorique*. Bruxelles, Belgique : De Boeck.
- Jonnaert, P. (2011). Curriculum, entre modèle rationnel et irrationalité des sociétés. *Revue internationale d'Éducation*. Sèvres, numéro 56, 135-145.
- Jorro, A. (2000). *L'enseignant et l'évaluation*. Paris, France : De Boeck.

- Jorro, A. (2006). L'éthos de l'évaluateur : entre imaginaire et posture. Dans Figari, G. et Mottier Lopez, L. (dir.), *Recherche sur l'évaluation en éducation* (p. 67-75). Paris : L'Harmattan.
- Jouquan, J., Parent, F. et Audétat, M. C. (2013). Des analogies entre le raisonnement médical et l'évaluation formative. *Revue Française de Linguistique Appliquée*, volume 18, 93-106.
- Karsenti, T. et Savoie-Zajc, L. (2004). *La recherche en éducation : Étapes et approches*. Sherbrooke, Québec : Éditions du Centre de ressources pédagogiques.
- King, L. et Macleod Clark, J. M. (2002). The role of intuition and the development of expertise in surgical ward and intensive care nurses. *Journal of Advanced Nursing*, 37(4), 322- 329.
- Lafortune, L. (2006). *Exercice et développement du jugement professionnel, Accompagner l'évaluation des apprentissages dans l'école québécoise. Aide à l'apprentissage et reconnaissance des compétences*. Récupéré du site de l'Université du Québec à Trois-Rivières : https://oraprdnt.uqtr.quebec.ca/pls/public/gscw030?owa_no_site=1
- Lafortune, L. (2007). Le développement du jugement professionnel : perspective socioconstructiviste. Dans Bélair, L. M., Laveault, D. et Lebel, C. (dir.). *Les compétences professionnelles en enseignement et en évaluation* (p. 7-24). Ottawa, Canada : Presses de l'Université d'Ottawa.
- Lambert C. (2005). *Le test de concordance de scripts : un outil pour évaluer le raisonnement clinique des résidents en radio-oncologie* (mémoire de maîtrise, Université de Montréal, Canada). Récupéré des archives de l'Université de Montréal : <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/17828>
- Latreille, M. E. (2012). *Évaluation du raisonnement clinique d'étudiantes et d'infirmières dans le domaine de la pédiatrie* (mémoire de maîtrise, Université d'Ottawa, Ottawa). Récupéré des archives de l'Université d'Ottawa : <http://hdl.handle.net/10393/22698>
- Laurier, M., Tousignant, R. et Morissette, D. (2005). *Les principes de la mesure et de l'évaluation des apprentissages* (3^e éd.). Montréal, QC : Gaëtan Morin.
- Laveault, D. et Grégoire, J. (2002). Introduction aux théories des tests en psychologie et en sciences de l'éducation (2e éd.). *Revue des sciences de l'éducation*, 30(1), 197–198.
- Laveault, D. (2005). Le jugement professionnel de l'enseignant: quel en est impact sur l'acte d'évaluer? *Mesure et évaluation en éducation*, 28(2), 93-114.
- Laveault, D. (2008). Le jugement professionnel d'évaluation scolaire : enjeux, tensions et synergies nouvelles. *Revue suisse des sciences de l'éducation*, 3, 483-500.
- Le Boterf, G. (1994). *De la compétence. Essai sur un attracteur étrange*. Paris, France : Les éditions d'organisation.

- Leclercq, D., Nicaise, J. et Demeuse, M. (2004). Docimologie critique : des difficultés de noter des copies et d'attribuer des notes aux élèves. Dans Demeuse, M. (dir.), *Introduction aux théories et aux méthodes de la mesure en science psychologiques et en sciences de l'éducation* (p. 273-292). Liège, Belgique : Les Éditions de l'Université de Liège.
- Lecointre, M. (1997). *Les enjeux de l'évaluation*. Paris, France : L'Harmattan.
- Legendre, M. F. (2001). Sens et portée de la notion de compétence dans le nouveau programme de formation. *Revue de l'association québécoise des enseignants de français langue seconde*, 23(1), 12-30.
- Legendre, M. F. (2004). Approches constructivistes et nouvelles orientations curriculaires. D'un curriculum fondé sur l'approche par objectifs à un curriculum axé sur le développement de compétences. Dans Jonnaert, P. et Masciotra, D. (dir.), *Constructivisme et choix contemporains. Hommage à Ernst von Glasersfeld* (p. 51-91). Montréal, Canada : PUQ.
- Legendre, R. (2005). *Dictionnaire actuel de l'éducation*. Montréal, Canada : Guérin.
- Lemay, V. (2000). *Évaluation scolaire et justice sociale*. Montréal, Canada : ERPI.
- Leroux, J. L. (2009). *Analyse des pratiques évaluatives d'enseignantes et d'enseignants dans une approche par compétences au collégial* (thèse de doctorat, Université de Sherbrooke, Sherbrooke). Récupérée en ligne des archives de l'Université de Sherbrooke : <http://savoirs.usherbrooke.ca/bitstream/handle/11143/947/NR62810.pdf?sequence=1&isAllowed=y>
- Leroux, J. L. (2014). Évaluer pour faire apprendre, Dans Ménard, L. et St-Pierre, L. (dir.), *Enseigner pour faire apprendre au collégial et à l'université* (p. 330-353). Montréal, Canada : Éditions Chenelière.
- Leroux, J. L. (2016). Le jugement évaluatif d'enseignants du supérieur dans un contexte d'évaluation des compétences. Dans Mottier Lopez, L. et Tessaro, W. (dir.), *Processus de jugement dans les pratiques d'évaluation et de régulation des apprentissages* (p. 169-196). Berne, Suisse : Peter Lang.
- Lessard-Hébert, M., Goyette, G. et Boutin, G. (1995). *Recherche qualitative : fondements et pratiques*. Montréal, Canada : Éditions Nouvelles.
- Lincoln, Y. et Guba, E. (1989). Ethics: The failure of positivist science. *The Review of Higher Education*, 12(3), 221.
- Lineberry, M., Kreiter, C. D. et Bordage, G. (2013). Threats to validity in the use and interpretation of script concordance test scores. *Medical Education*, 47, 1175-1183.
- Lubarsky, S., Dory, V., Duggan, P., Gagnon, R. et Charlin, B. (2013). Script concordance testing: from theory to practice: AMEE guide no. 75. *Medical Teacher*, 35(3), 184-193.

- Lubrasky, S., Gagnon, R. et Charlin, B. (2013). Scoring the Script Concordance Test: not a black and white issue. *Medical Education*, 47, 1152–1161.
- Maray, C. (2002). *L'enseignement secondaire et ses enseignants*. Bruxelles, Belgique : Éditions De Boeck.
- Marie, I., Sibert, L., Roussel, F., Hellot, M-F., Lechevallier, J. et Weber, J. (2005). Le test de concordance de scripts : un nouvel outil d'évaluation du raisonnement et de la compétence clinique en médecine interne? *La revue de médecine interne*, 26, 501-507.
- Mayer, R. (1991). *Méthodologie de recherche pour les intervenants sociaux*. Boucherville, Québec : G. Morin.
- Mayer, R., Deslauriers, J., Groulx, L., Poupart, J., Pires, A. et Savoie-Zajc, L. (2000). *La recherche qualitative : Enjeux épistémologiques et méthodologiques*. Ottawa, Canada : Presses de l'Université de Toronto.
- Ménard, L. et L. St-Pierre (2014). *Pour un enseignement cohérent au cégep et à l'université*. Montréal, Canada : Édition Chenelière.
- Merle, P. (1996). *L'évaluation des élèves. Enquête sur le jugement professoral*. Paris, France : Presses universitaires de France.
- Merle, P. (1998). *Sociologie de l'évaluation scolaire*. Paris, France : Presses universitaires de France.
- Ministère de l'Éducation du Québec (1983). *Guide d'évaluation en classe : primaire, introduction générale*. Québec, Québec : Gouvernement du Québec.
- Ministère de l'Éducation du Québec (1988). *Éléments de docimologie, fascicule 3. L'évaluation formative*. Québec, Québec : Gouvernement du Québec.
- Ministère de l'Éducation du Québec (2001). *La formation à l'enseignement : les orientations, les compétences professionnelles*. Québec : Gouvernement du Québec.
- Ministère de l'Éducation du Québec (2002). *L'évaluation des apprentissages au préscolaire et au primaire : Cadre de référence*. Québec : Gouvernement du Québec.
- Ministère de l'Éducation du Québec (2003a). *Plan de mise en œuvre de la politique d'évaluation des apprentissages*. Québec : Gouvernement du Québec.
- Ministère de l'Éducation du Québec (2003b). *Politique d'évaluation des apprentissages : formation générale des jeunes, formation générale des adultes, formation professionnelle*. Québec : Gouvernement du Québec.

- Ministère de l'Éducation, du Loisir et du Sport (2006). *Programme de formation de l'école québécoise. Enseignement secondaire, premier cycle*. Québec, Canada : Gouvernement du Québec.
- Ministère de l'Éducation, du Loisir et du Sport (2008). *Échelles des niveaux de compétences*. Québec, Canada : Gouvernement du Québec.
- Ministère de l'Éducation, du Loisir et du Sport (2011). *Cadre d'évaluation des apprentissages*. Modèle révisé. Québec, Canada : Gouvernement du Québec.
- Morrisette, J. et Legendre, M. F. (2011). L'évaluation des compétences en contexte scolaire : des pratiques négociées. *Education Sciences and Society*, 2(2), 120-132.
- Mottier Lopez, L. et Allal, L. (2008). Le jugement professionnel en évaluation : un acte cognitif et une pratique sociale située. *Revue suisse des sciences de l'éducation*, 30, 465-482.
- Mottier Lopez, L. et Allal, L. (2010). Le jugement professionnel en évaluation : quelles triangulations méthodologiques et théoriques ? Dans Paquay, L., Van Nieuwenhoven, C. et Wouters, P., *L'évaluation, levier du développement professionnel ? Tensions, dispositifs, perspectives* (p. 237-250). Bruxelles, Belgique : De Boeck.
- Mottier Lopez, L. et Allal, L. (2008). Le jugement professionnel en évaluation : un acte cognitif et une pratique sociale située. *Revue Suisse Des Sciences De l'Éducation*, volume 3, 465-482.
- Mukamurera, J. (2011). Les conditions d'insertion et la persévérance dans la profession enseignante. Dans Lacourse, F., Martineau, S. et Nault, T. (dir.), *Profession enseignante. Démarches et soutien à l'insertion professionnelle* (p. 37-58). Anjou, Canada : CEC.
- Nendaz, M., Charlin, B., Leblanc, V. et Bordage, G. (2005). Le raisonnement clinique : données issues de la recherche et implications pour l'enseignement. *Revue internationale francophone d'éducation médicale*, 65(4), 235-254.
- Nizet, I. (2014). Formation des enseignants du secteur de l'éducation des adultes à l'évaluation des compétences : défis et perspectives. *Revue des sciences de l'éducation*, 40(2), 373-396.
- Norman, G. (2005). Research in clinical reasoning. Past history and current trends. *Medical Education*, 39, 418-427.
- Paquay, L., Altet, M., Charlier, E. et Perrenoud, P. (1996). *Former des enseignants professionnels : quelles stratégies? quelles compétences ?* Bruxelles, Belgique : De Boeck.
- Paquay, L. (2002). *L'évaluation des compétences chez l'apprenant : pratiques, méthodes et fondements*. Louvain, Belgique : Presses de l'Université de Louvain.

- Paucard-Dupont, S. et Marchand, C. (2014). Étude exploratoire du raisonnement clinique chez les étudiants en soins infirmiers à l'aide de la carte conceptuelle. *Recherche en soins infirmiers*, numéro 177, 85-112.
- Perrenoud, P. (1998). *L'évaluation des élèves. De la fabrication de l'excellence à la régulation des apprentissages*. Bruxelles, Belgique : De Boeck.
- Poupart, J. (1997). L'entretien de type qualitatif : considérations épistémologiques, théoriques et méthodologiques. Dans Poupart, J., Deslauriers, J. P., Groulx, L. H. et al., *La recherche qualitative: enjeux épistémologiques et méthodologiques* (p. 173-209). Montréal, Canada : Gaëtan Morin éditeur.
- Proulx, J. P. (2005). La réforme au secondaire et les défis des enseignants ou le point de vue du Conseil supérieur de l'éducation. Dans Biron, D., Cividini, M. et Desbiens, J. F., *La profession enseignante au temps des réformes* (p. 159-170). Sherbrooke, Canada : Éditions du Centre de ressources pédagogiques.
- Rey, B., Carette, V., Defrance, A. et Kahn, S. (2003). *Les compétences à l'école, apprentissage et évaluation*. Bruxelles, Belgique : De Boeck.
- Rey, O. (2012). Le défi de l'évaluation des compétences. *Veille et Analyses*, 76. Récupéré en ligne : <http://ife.ens-lyon.fr/vst/DA-Veille/76-juin-2012.pdf>.
- Roegiers, X. (2000). *Une pédagogie de l'intégration. Compétences et intégration des acquis dans l'enseignement*. Bruxelles, Belgique : De Boeck.
- Roegiers, X. (2001). *Une pédagogie de l'intégration*. Bruxelles, Belgique : De Boeck.
- Romainville, M. (1996). L'irrésistible ascension du terme « compétence » en éducation... *Enjeux*, numéro 37-38, 132-142.
- Savoie-Zajc, L. (2009). L'entretien semi-dirigé. Dans Gauthier, B. (dir.) *Recherche sociale de la problématique à la collecte des données* (p. 337-360). Québec, Canada : Presses de l'Université du Québec.
- Savoie-Zajc, L. (2013). Le jugement professionnel. Dans Fontaine, S., Savoie-Zajc, L. et Cadieux, A. (Ed.), *Évaluer les apprentissages. Démarche et outils d'évaluation pour le primaire et le secondaire* (p. 104-116). Anjou, Québec : Éditions CEC.
- Savoie-Zjac, L. (2004). La recherche qualitative/interprétative en éducation. Dans Karsenti, T. et Savoie-Zajc, L., *La recherche en éducation : étapes et approches* (p.142-150). Sherbrooke, Canada : Éditions du Centre de ressources pédagogiques.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Saint-Laurent, Québec : ERPI

- Schön, D. (1983). *The reflective practitioner: How professionals think in action*. New York, États-Unis : Basic Books.
- Sibert, L., Charlin, B., Gagnon, R., Corcos, J., Khalaf, A., et Grise, P. (2001). Evaluation du raisonnement clinique en urologie : l'apport du Test de Concordance de Script. *Progrès en urologie*, 11, 1213-1219.
- Solano-Flores, G., Shavelson, R. J. et 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.
- Tanner, C. A. (2006). Thinking like a nurse: a research-based model of clinical judgment in nursing. *Journal of Nursing Education*, 45(6), 204-211.
- Tardif, J. (2006). *L'évaluation des compétences. Documenter le parcours de développement*. Montréal, Canada : Chenelière Éducation.
- Tardif, M. et Lessard, C. (2004). *La profession d'enseignant aujourd'hui : Évolutions, perspectives et enjeux internationaux*. Louvain-la-Neuve, Belgique : De Boeck.
- Tochon, F. V. (1996). Rappel stimulé, objectivation clinique, réflexion partagée. Fondements méthodologiques et applications pratiques de la rétroaction vidéo en recherche et en formation. *Revue des sciences de l'éducation*, 22(3), 467-502.
- Tousignant, R. et Morissette, D. (1990). *Les principes de la mesure et de l'évaluation des apprentissages*, 2^e édition. Boucherville, Canada : Gaëtan Morin.
- Van Der Maren, J. M. (2003). *La recherche appliquée en pédagogie*. Bruxelles, Belgique : De Boeck.
- Van der Maren, J. M. et Loye, N. (2011). À propos de quelques difficultés de l'évaluation des compétences. Réflexions critiques. *Education Sciences et Society*, 2(2), 40-53.
- Van der Maren, J.-M. (1996, 26 octobre). *La recherche qualitative peut-elle être rigoureuse?* Communication présentée au Colloque international "Recherche Qualitative et Gestion", HEC, Montréal.
- Vanhulle, S. (2009). Évaluer la professionnalité émergente des futurs enseignants. Dans Mottier Lopez, L. et M. Crahay (dir.), *Évaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*. Bruxelles, Belgique : De Boeck.
- Vermersch, P. (2003). *L'entretien d'explicitation*. Paris, France : Édition Sociale Française.
- Vermersch, P. (2006). *L'entretien d'explicitation* (5^e éd.). Issy-les-Moulineaux, France : Édition Sociale Française

Vial, M. (2006). Les relations entre formation et évaluation : perspectives de recherches. *Mesure et évaluation en éducation*, 29(1), 81-98.

Wiggins, G. P. (1993). *Assessing student performance : Exploring the purpose and limits of testing*. San Francisco, États-Unis : Jossey-Bass.

Wiggins, G. (1998). *Educative Assessment : Designing Assessment to Inform and Improve Student Performance*. San Francisco, États-Unis : Jossey-Bass.

Annexe A

Courriel d'invitation

Présentation du projet

Titre du projet : Application du test de concordance de scripts : étude exploratoire du développement de la compétence à évaluer les élèves chez les enseignants

But de l'étude

L'objectif de cette étude exploratoire qualitative interprétative est de mieux comprendre le processus de développement du jugement en évaluation des apprentissages des futurs enseignants par la passation d'un test de jugement professionnel en évaluation des apprentissages. Cette recherche vise à étudier les pratiques des évaluateurs et, plus particulièrement, les choix méthodologiques qu'ils mettent en œuvre. Elle vise également à rendre compte de l'influence des parties prenantes sur lesdits choix. Enfin, la finalité de cette étude permettra éventuellement de proposer des solutions à la relève enseignante et de favoriser le développement d'un jugement en évaluation juste et équitable pour tous les élèves fréquentant les écoles québécoises de langue française.

Recrutement

Étudiants

Le recrutement des 4 participants ainsi que la sélection des sujets de la catégorie étudiante se fera selon les critères suivants :

- Être inscrit à un programme de formation des maîtres (volet secondaire) offert en français au Canada
- Avoir le français comme langue de scolarisation depuis la première secondaire

Experts

Les 4 participants composant le groupe d'experts seront sélectionnés. Les candidats devront également répondre aux critères prédéterminés suivants :

- Avoir le français comme langue de scolarisation depuis la première secondaire
- Enseigner un cours d'évaluation des apprentissages au niveau universitaire
- Détenir une expérience d'enseignement d'au moins 10 ans

Praticiens

- Les 4 participants composant le groupe de praticiens devront répondre aux critères prédéterminés suivants:
 - Être enseignant dans une école secondaire depuis plus de 10 ans
 - Détenir un brevet d'enseignement valide dans sa province de pratique
 - Avoir le français comme langue de scolarisation depuis la première secondaire

Bienfaits de la participation à étude

Participer à cette recherche présente plusieurs bénéfices :

Une participation à cette recherche donne l'occasion de discuter et d'échanger avec une chercheuse-enseignante qui s'intéresse aux problèmes que le participant peut rencontrer sur le terrain en contexte d'évaluation de ses élèves. De plus, cela offre une chance de développer ou de renforcer sa pratique réflexive dans son rôle d'évaluateur. Enfin, il participe à l'enrichissement des connaissances et contribue au développement des solutions que la chercheuse a pour objectifs de proposer afin de réduire l'écart entre la théorie et la pratique de terrain. Ainsi, cela entraînera de meilleures évaluations, de meilleurs praticiens ainsi qu'une meilleure image de notre discipline.

La participation à cette recherche

Participer à un entretien semi-dirigé (enregistré sur bande audio) consistant à répondre à un test de jugement en évaluation des apprentissages (TJPEA). Dans le cadre de cette passation, on présentera des vignettes suivies de 4 questions. Pour chaque item, on invite le participant à partager oralement son opinion face aux contenus des situations présentées et à justifier ses actions. L'entretien durera entre 60 et 90 minutes.

Annexe B

Formulaire de consentement

Titre du projet

Application du test de concordance de scripts :
étude exploratoire du développement de la compétence à évaluer les élèves chez les
enseignants

Invitation à participer

Je suis invité(e) à participer à la recherche nommée ci-haut et qui est menée par Jennifer Smith.

But de l'étude

L'objectif de cette étude exploratoire qualitative interprétative est de mieux comprendre le processus de développement du jugement en évaluation des apprentissages des futurs enseignants par la passation d'un test de jugement professionnel en évaluation des apprentissages (TJPEA). La présente étude s'intéresse aux caractéristiques du jugement en évaluation des enseignants à partir de l'instrument du TJPEA, mais aussi à explorer un domaine d'intérêt en évaluation pour lequel il n'existe peu ou pas de données disponibles. Enfin, la finalité de ce projet de recherche permettra éventuellement de proposer des pistes didactiques à la relève enseignante et de favoriser le développement d'un jugement en évaluation juste et équitable pour tous les élèves.

Participation

Je participerai à un entretien semi-dirigé (enregistré sur bande audio) consistant à répondre à un test de jugement en évaluation des apprentissages (TJPEA). Dans le cadre de cette passation, on me présentera des vignettes suivies de 4 questions. Pour chaque item, je serai invité(e) à partager oralement mon opinion face aux contenus des situations présentées et à justifier mes actions. L'entretien durera entre 60 et 90 minutes.

Risque

Les données ne serviront en aucun cas à porter un jugement sur les pratiques du participant comme enseignant ou encore sur son statut d'étudiant. La chercheuse a précisé qu'il n'y a pas de risques prévisibles.

Bienfaits

Ma participation à cette recherche me donne l'occasion de discuter et d'échanger avec une chercheuse-enseignante qui s'intéresse aux problèmes que je peux rencontrer sur le terrain en

contexte d'évaluation de mes élèves. De plus, cela m'offre une chance de développer ou de renforcer ma pratique réflexive dans mon rôle d'évaluateur. Enfin, je participe à l'enrichissement des connaissances et je contribue au développement des solutions que la chercheuse a pour objectifs de proposer afin de réduire l'écart entre la théorie et la pratique de terrain. Ainsi, cela entrainera de meilleures évaluations, de meilleurs praticiens ainsi qu'une meilleure image de notre discipline.

Confidentialité et anonymat

J'ai l'assurance de la chercheuse que l'information que je partagerai avec elle restera strictement confidentielle. Je m'attends à ce que le contenu ne soit utilisé que pour la rédaction de sa thèse et des publications futures.

L'anonymat et la confidentialité sont garantis de la façon suivante : une fois le recrutement réalisé, la chercheuse me demandera de m'auto-attribuer un code qui sera composé de trois lettres en s'assurant que cela ne donne pas d'indice sur leurs initiales, et un chiffre soit, par exemple, FAD7. Ainsi, elle sera la seule détentrice de cette information et personne ne sera en mesure de m'identifier. Elle anonymisera également le rapport d'évaluation. Enfin, elle ne fera jamais référence à mon identité, que soit ce soit dans sa thèse ou tout autre type de publications ou communications orales.

Conservation des données

En ce qui concerne la protection physique des données, celles-ci seront conservées de manière sécuritaire, "sous clés", dans un casier dans le bureau du directeur de thèse de la chercheuse principale qui ferme également à clé, et ce, pendant toute la durée de l'étude. Quant à la protection technique des données, celles-ci seront anonymisées/codées dans tous les documents relatifs à l'étude (enregistrements audio, fichiers de transcriptions, etc.), chaque participant devenant un "code" sur tous les documents qui sont relatifs à leur participation. Enfin, les données se trouveront sur le disque dur d'un seul ordinateur dans le bureau du directeur de thèse qui ferme à clé et elles seront conservées pendant 5 ans à partir de la fin de la maîtrise de la chercheuse principale à des fins de publications, de communications scientifiques, mais aussi pour une potentielle nouvelle recherche inscrite dans la lignée des résultats de la thèse. Elles seront détruites à la fin du mois de mai 2021.

Participation volontaire

Ma participation à la recherche est volontaire et je suis libre de me retirer en tout temps, et/ou refuser de répondre à certaines questions, sans subir de conséquences négatives. Si je choisis de me retirer de l'étude, la chercheuse me demandera mon consentement afin d'utiliser mes données.

Acceptation:

Je, _____, accepte de participer à cette recherche menée par Jennifer Smith de la Faculté d'éducation de l'Université d'Ottawa, laquelle est supervisée par le directeur de la chercheuse, le professeur Eric Dionne.

Pour tout renseignement additionnel concernant cette étude, je peux communiquer avec la chercheuse ou son directeur.

Pour tout renseignement sur les aspects éthiques de cette recherche, je peux m'adresser au Responsable de l'éthique en recherche de l'Université d'Ottawa.

Il y a deux copies du formulaire de consentement, dont une copie que je peux garder.

Signature du participant :

Date :

Signature de la chercheuse :

Date:

Annexe C

Liste des thèmes

Liste des contenus théoriques tirés de la synthèse des ouvrages ciblés (Dionne et Forte, 2013)

Thème	Vignette	Item
Interprétation des résultats	1	4
Affectivité et évaluation	2	4
Cohérence enseignement-apprentissage-évaluation	3	4
Fonctions de l'évaluation	4	4
Modèles d'enseignement et d'apprentissage	5	4
Complexité des apprentissages	6	4
Démarche d'évaluation	7	4
Validité	8	4
Fidélité	9	4
Interprétation des scores	10	3
Qualité de l'évaluation, biais, aspects culturels	11	3
Rédaction des items, erreur de mesure	12	3
Instruments d'évaluation	13	3
Auto-évaluation-éthique-équité	14	4
Équité	15	2
Communication des résultats	16	3
Alignement	17	3
Jugement professionnel	18	3
Progression et structure de notation	19	3
Équité	20	3
Effets indésirables (<i>teach to the test</i>)	21	2
Processus de notation	22	3
Planification de l'évaluation	23	3
Éthique	24	3
Erreur de mesure	25	3
Total	25	83

Annexe D

Liste des ouvrages synthétisés

- Airasian, P. W., Engemann, J. et Gallagher, T. L. (2007). *Classroom assessment: Concepts and applications*. Toronto, Canada : McGraw Hill.
- Auger, R., Séguin, S. P. et Nézet-Séguin, C. (2000). *Formation de base en évaluation des apprentissages*. Montréal, Canada : Éditions Logiques.
- Durand, M. J. et Chouinard, R. (2006). *L'évaluation des apprentissages : de la planification de la démarche à la communication des résultats*. Montréal, Canada : Hurtubise.
- Laurier, M., Tousignant, Robert et Morissette, Dominique (2005). *Les principes de la mesure et de l'évaluation des apprentissages*. Montréal, Canada : G. Morin.
- Liu, X. (2010). *Essentials of science classroom assessment*. Los Angeles, États-Unis : SAGE.
- Louis, R. et Bernard, H. (2004). *L'évaluation des apprentissages : théorie et pratique*. Montréal, Canada : Beauchemin.
- Marzano, R. (2006). *Classroom Assessment & Grading that Work*. Alexandria, États-Unis : Association for Supervision & Curriculum Development.
- McMillan, J. (2014). *Classroom assessment : Principles and practice for effective standards-based instruction*. New York, États-Unis : Pearson.
- Oosterhof, A. (2009). *Developing and using classroom assessments*. Upper Saddle River, États-Unis : Merrill-Pearson.
- Popham, W. (2007). *Classroom assessment : What teachers need to know*. Boston, États-Unis : Pearson.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Saint-Laurent, Canada : ERPI.
- Shermis, M., et Di Vesta, Francis J. (2011). *Classroom assessment in action*. Lanham, États-Unis : Rowman & Littlefield.
- Stiggins, R. J., Arter, J., Chappuis, J. et Chappuis, S. (2004). *Classroom assessment for student learning : Doing it right-using it well*. Portland, États-Unis : Assessment Training Institute.
- Tardif, J. (2006). *L'évaluation des compétences : Documenter le parcours de développement*. Montréal, Canada : Chenelière-éducation.
- Taylor, C. et Nolen, S. B. (2008). *Classroom assessment: Supporting teaching and learning in real classrooms*. Upper Saddle River, États-Unis : Pearson, Merrill Prentice Hall.

Annexe E

Vignettes du TJPEA

Vignette 1

Les questions se rapportent à la mise en situation qui suit :
 Lors de la comptabilisation des notes pour le bulletin, vous remarquez qu'un de vos élèves est en échec de peu en raison des notes extrêmement basses qu'il a obtenues au début de l'étape. Par contre, si vous prenez en compte ses notes plus récentes, vous jugez qu'il a fait un progrès exceptionnel.

Option 1

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Vous décidez de modifier la pondération des évaluations de cet élève afin d'augmenter le poids des plus récentes.		L'élève termine son étape avec une bien meilleure note et sa confiance augmente drastiquement.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Option 2

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Vous décidez de modifier la pondération des évaluations de cet élève afin d'augmenter le poids des plus récentes.		La situation se fait apprendre par des élèves et des parents de la classe réclament des traitements similaires.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Option 3

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
Votre direction vous indique que ses directives sont les suivantes : ce genre de modifications ne sont pas acceptables et que vous ne pouvez rien changer aux résultats.		L'élève termine son étape avec un échec et semble avoir un déclin d'intérêt et de motivation face à l'école.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Vignette 2

Les questions se rapportent à la mise en situation qui suit :
 À la suite de la correction d'une évaluation sommative, la distribution des notes d'un groupe d'élèves s'éloigne substantiellement de ce à quoi vous êtes habitué.

Option 1

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
1. Vous ne faites rien de particulier puisqu'il s'agit probablement de l'effet du hasard.		Les notes reviennent à la normale lors de l'évaluation suivante.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Option 2

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
2. Vous modifiez les résultats afin d'attribuer à ces élèves les notes qu'ils atteindraient en temps normal.		La distribution des notes ressemble aux précédentes.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Option 3

Vous décidez de prendre la décision suivante :		À la suite de votre décision, vous remarquez que :		
3. Vous réévaluez votre enseignement afin de déterminer la source de la variation.		La matière vue était beaucoup moins dynamique et stimulante que celle vue précédemment.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Vignette 3

Les questions se rapportent à la mise en situation qui suit :
Lors de la correction d'une épreuve écrite une question a été comprise d'une toute autre façon que ce que vous aviez prévu initialement lors de la confection de cette épreuve.

Option 1

Vous décidez de réaliser l'intervention suivante :		À la suite de votre intervention, vous remarquez que :		
1. Vous conservez les résultats des élèves sans rien changer.		Les notes des étudiants réputés très performants sont plus faibles pour cette question.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Option 2

Vous décidez de réaliser l'intervention suivante :		À la suite de votre intervention, vous remarquez que :		
2. Vous administrez une nouvelle question au cours suivant afin remplacer la question problématique.		Les élèves sont mécontents de devoir répondre à la nouvelle question.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Option 3

Vous décidez de réaliser l'intervention suivante :		À la suite de votre intervention, vous remarquez que :		
3. Vous décidez d'annuler cette question.		Les élèves révisent les autres questions de l'épreuve en espérant détecter d'autres questions lacunaires.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Vignette 4

Les questions se rapportent à la mise en situation qui suit :
Un parent mécontent vous demande une rencontre afin que vous lui expliquiez le résultat particulièrement décevant de son enfant lors du dernier examen.

Option 1

Vous décidez de réaliser l'intervention suivante :		À la suite de votre intervention, vous remarquez que :		
1. Vous rencontrez le parent et lui résumez tout ce que vous avez dit et fait en classe avec vos élèves afin qu'ils se préparent à l'examen.		Le parent juge que vos résultats sont arbitraires et non appuyés.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Option 2

Vous décidez de réaliser l'intervention suivante :		À la suite de votre intervention, vous remarquez que :		
2. Vous rencontrez le parent et évitez de lui expliquer vos décisions pédagogiques, car vous jugez qu'il n'est pas en mesure de bien comprendre toutes vos décisions.		Le parent demeure mécontent.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Option 3

Vous décidez de réaliser l'intervention suivante :		À la suite de votre intervention, vous remarquez que :		
3. Vous rencontrez le parent avec un collègue qui appuie sans contredire votre décision, mais sans en préciser les raisons.		Le parent accepte votre décision qui est appuyée par l'un de vos collègues.		
A. Vous considérez que votre intervention était :	Très pertinente	Pertinente	Peu pertinente	Non pertinente
B. La probabilité que vous ayez réalisé cette intervention est :	Très probable	Probable	Peu probable	Très improbable

Annexe F

Questionnaire d'entrevue

1. Recueil de données d'ordre général

Pourriez-vous me préciser quelques informations d'ordre général à votre sujet :

- Votre âge ?
- Votre sexe?
- Votre durée d'exercice professionnel ou étudiant ?
- Votre milieu d'enseignement ou d'études ?
- Le titre des cours enseignés ou suivis en évaluation ?

2. Phase d'explicitation

1. Lecture de la vignette.			
1. Pourrait-on qualifier la situation présentée comme étant possible d'être vécue par un enseignant? Pourquoi?	2. Quels concepts/notions liés à l'évaluation sont évoqués?	3. À cette étape, y a-t-il de l'information dont vous ne disposez pas à laquelle vous auriez aimé avoir accès? Pourquoi?	4. Y a-t-il de l'information que vous jugez non pertinente? Pourquoi?
2. Lecture de l'intervention			
1. Est-ce que vous jugez cette intervention authentique? Pourquoi?	2. Quels concepts/notions liés à l'évaluation sont évoqués?	3. À cette étape, y a-t-il de l'information dont vous ne disposez pas à laquelle vous auriez aimé avoir accès? Pourquoi?	4. Y a-t-il de l'information que vous jugez non pertinente? Pourquoi?
3. Lecture de la suite...			
1. Est-ce que vous jugez cette réaction possible? Pourquoi?		3. À cette étape, y a-t-il de l'information dont vous ne disposez pas à laquelle vous auriez aimé avoir accès? Pourquoi?	4. Y a-t-il de l'information que vous jugez non pertinente? Pourquoi?
4. Lire la pertinence...			
1. Sur le plan pédagogique, vous considérez que votre intervention était... Pourquoi?			
5. Lire la probabilité...			

3. Phase de conclusion

L'entretien arrive à son terme. Auriez-vous d'autres commentaires ou précisions à ajouter ?

Je vous remercie du temps et de la confiance que vous m'avez accordés. Je vous tiendrai au courant de l'avancement du présent projet de recherche.