

9. ÉVALUATION

Les questionnaires de l'élève, du personnel enseignant et de l'école comportaient des questions sur l'évaluation, présentées sous deux points de convergence principaux. Le premier consistait à étudier les méthodes d'évaluation utilisées par l'école et le personnel enseignant ainsi que leur incidence sur le rendement en lecture des élèves. Le second concernait la connaissance, l'utilisation et l'incidence des évaluations à grande échelle, y compris les évaluations provinciales/territoriales, le PIRS, le PPCE et le PISA.

Méthodes d'évaluation en classe

Les élèves étaient priés de dire à quelle fréquence on leur attribuait des notes ou cotes à l'aide d'une gamme de méthodes d'évaluation, y compris les tests à réponses choisies et à réponses construites, les compositions ou présentations et autres. Les réponses relatives aux différents tests ont fait l'objet d'une analyse factorielle qui a produit les trois facteurs montrés au tableau 9.1.

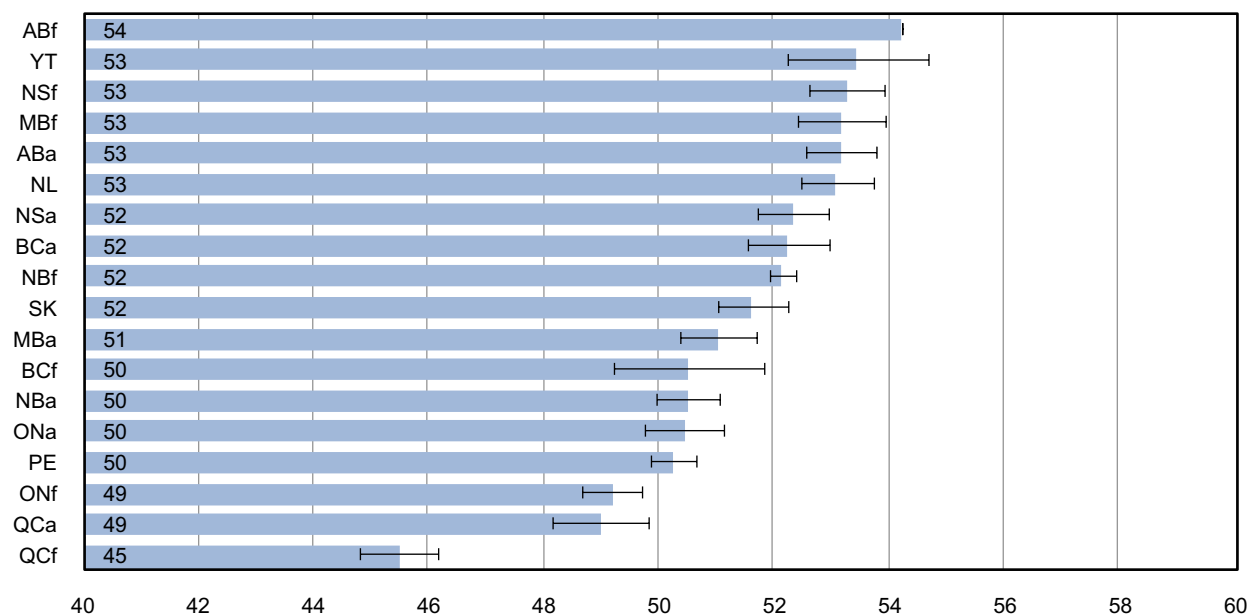
Tableau 9.1 Items du questionnaire et facteurs portant sur les réponses des élèves aux questions d'évaluation

Facteurs	Items
Items à réponse courte	Questions vrai/faux ou questions de correspondance Questions à choix multiple Phrases à compléter
Items à réponse longue	Questions à réponse courte (une phrase ou deux) Questions à réponse longue (un paragraphe ou plus)
Compositions/présentations	Compositions (une page ou plus) Présentations, discours et autres performances

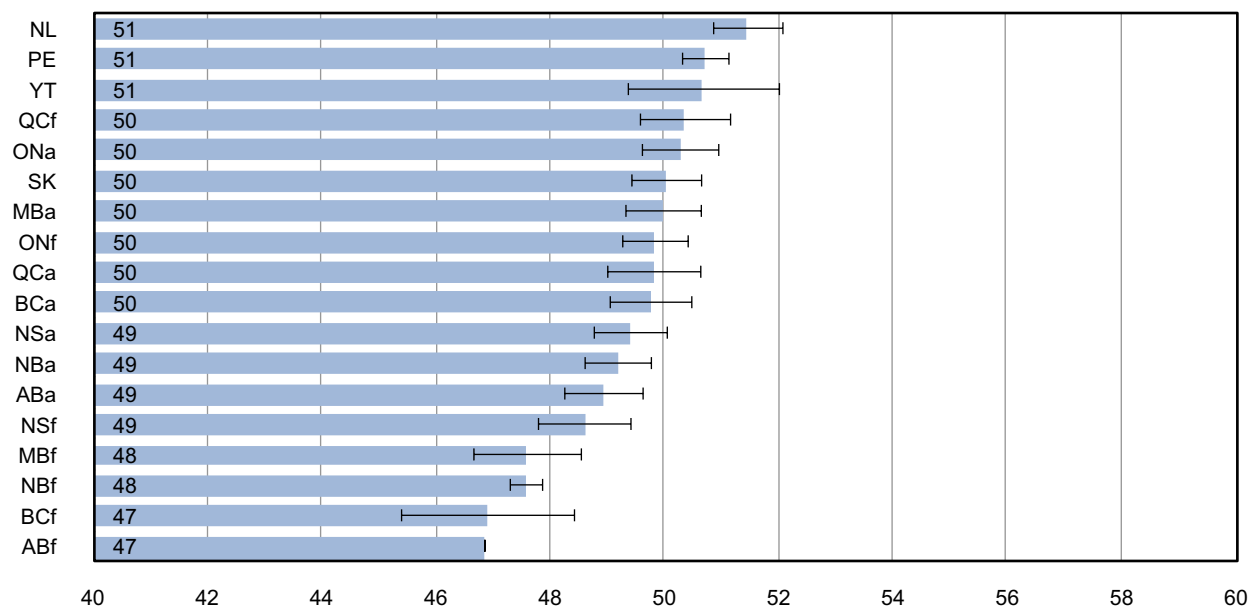
Les graphiques 9.1 à 9.3 montrent les scores factoriels moyens pour chacun de ces facteurs, par instance et selon la langue. Le graphique 9.1 montre que les items à réponse courte sont utilisés à une fréquence plus grande que la moyenne canadienne dans la plupart des instances. Toutefois, le Québec (francophone) et dans une moindre mesure, le Québec (anglophone) et l'Ontario (francophone) les utilisent de façon significativement inférieure à la plupart des autres instances. De manière générale, les différences entre les populations pour ce qui est

des items à réponse longue sont moins marquées que pour les items à réponse courte. Les populations francophones sont celles qui utilisent le moins les items à réponse longue. Cependant, dans ce cas, l'Ontario (francophone) et le Québec (francophone) se rapprochent de la moyenne canadienne sur le plan de l'utilisation. Les différences observées quant à l'utilisation des compositions/présentations sont également relativement petites et aucune tendance marquée n'est constatée dans la distribution pour les différentes populations.

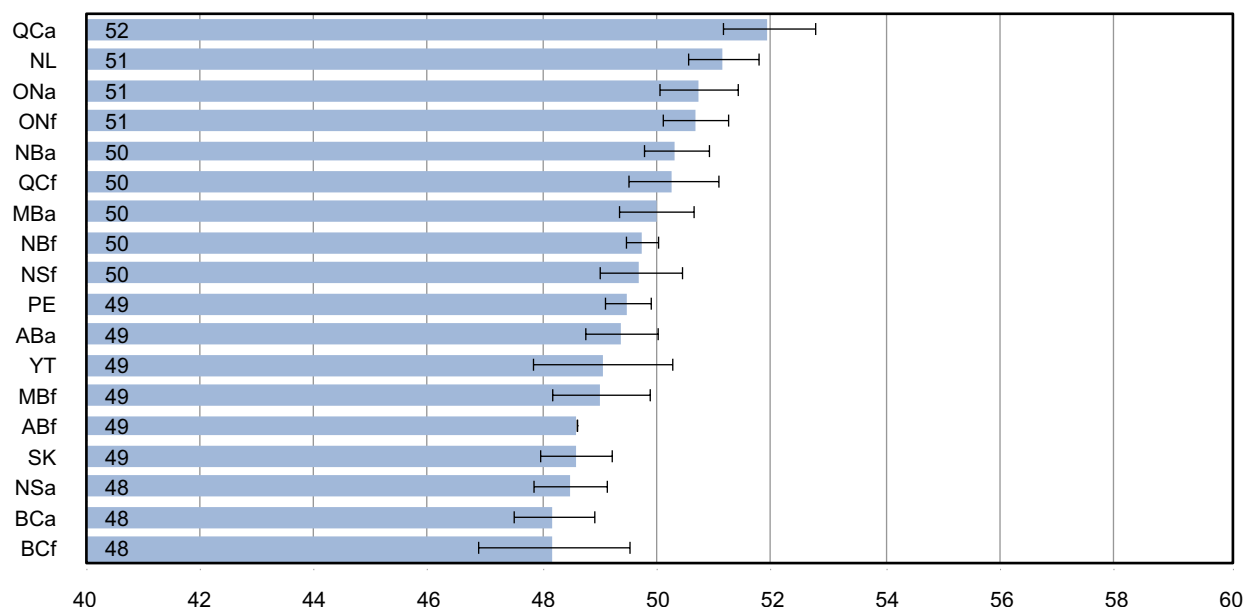
Graphique 9.1 Scores factoriels moyens d'après les réponses des élèves au sujet des items à réponse courte de l'évaluation, par instance et selon la langue



Graphique 9.2 Scores factoriels moyens d'après les réponses des élèves au sujet des items à réponse longue de l'évaluation, par instance et selon la langue



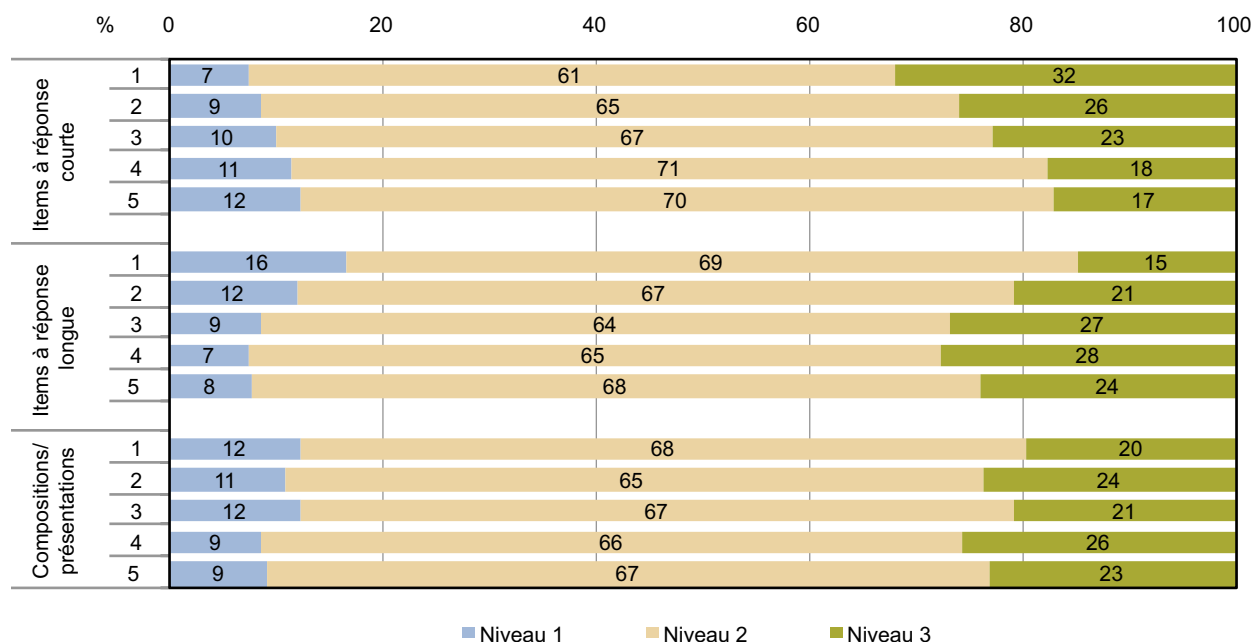
Graphique 9.3 Scores factoriels moyens d'après les réponses des élèves au sujet des compositions/présentations de l'évaluation, par instance et selon la langue



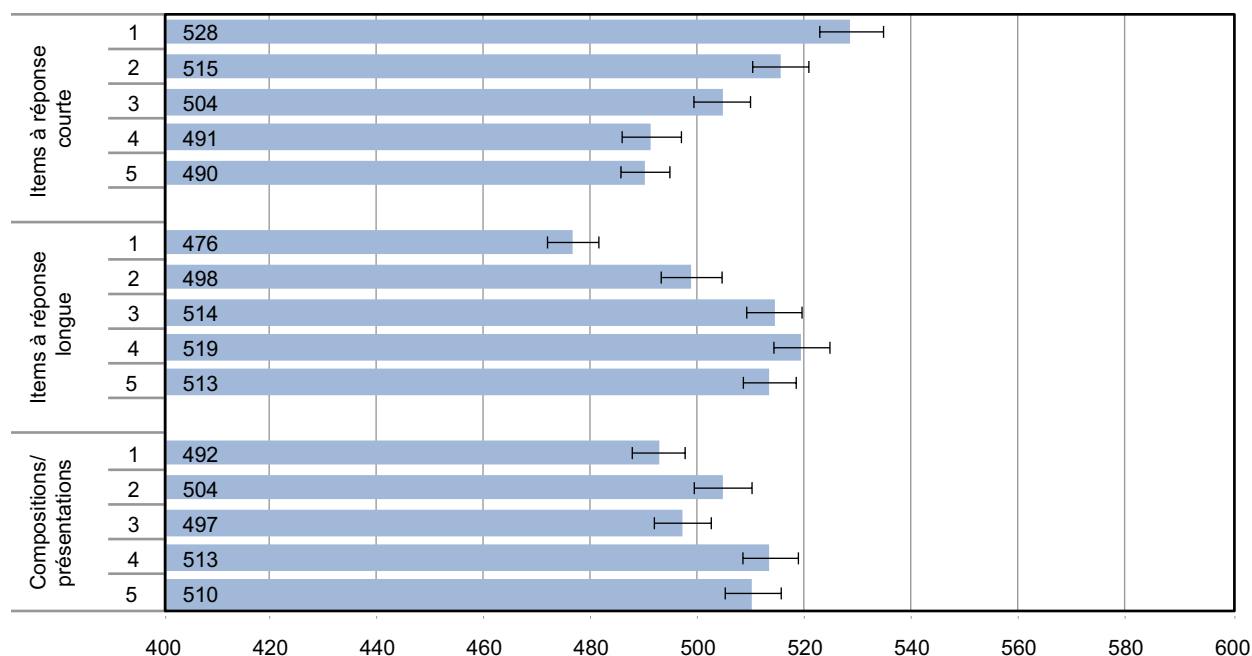
Les graphiques 9.4 et 9.5 montrent les niveaux de compétence en lecture et les scores moyens en lecture pour ces méthodes d'évaluation. Une association claire entre la fréquence d'utilisation des items à réponse courte et un rendement en lecture inférieur est constatée. À l'inverse pour ce qui est de

l'utilisation fréquente des items à réponse longue, le rendement en lecture est plus élevé bien que la courbe s'aplanisse au-delà du troisième quintile. Il y a également une légère tendance vers un rendement plus élevé pour ceux qui utilisent le plus les compositions/présentations.

Graphique 9.4 Niveaux de compétence en lecture d'après les réponses des élèves au sujet des méthodes d'évaluation utilisées par quintile



Graphique 9.5 Scores moyens en lecture d'après les réponses des élèves au sujet des méthodes d'évaluation utilisées par quintile



Le personnel enseignant a répondu à un ensemble de questions parallèles sur les méthodes d'évaluation. Dans ce cas, l'analyse factorielle n'a produit que deux facteurs, l'utilisation des compositions/présentations étant combinée à l'utilisation des items à réponse longue et à l'utilisation des items à réponse courte (une phrase ou deux) qui se répartissent également sur les deux facteurs, tel que montré dans le tableau

9.2. Cette structure est légèrement moins raffinée mais ne contredit pas la précédente. Il est très plausible que le personnel enseignant ne fasse pas de distinction nette entre les compositions/présentations et d'autres formes d'évaluation à réponse longue. Une solution à deux facteurs a donc été utilisée pour présenter les résultats obtenus d'après l'utilisation de ces résultats.

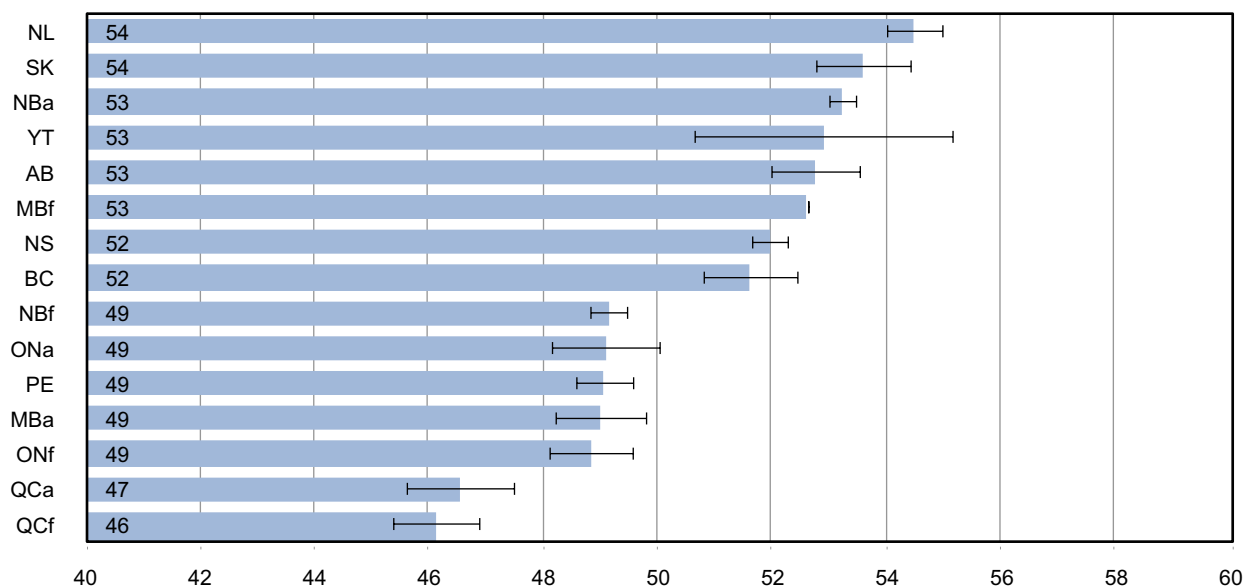
Tableau 9.2 Items du questionnaire et facteurs portant sur les réponses du personnel enseignant aux questions de l'évaluation

Facteurs	Items
Items à réponse courte	Questions vrai/faux ou questions de correspondance Questions à choix multiple Phrases à compléter Questions à réponse courte (une phrase ou deux)
Items à réponse longue	Questions à réponse courte (une phrase ou deux) Compositions (une page ou plus) Présentations, discours et autres performances Questions à réponse longue (un paragraphe ou plus)

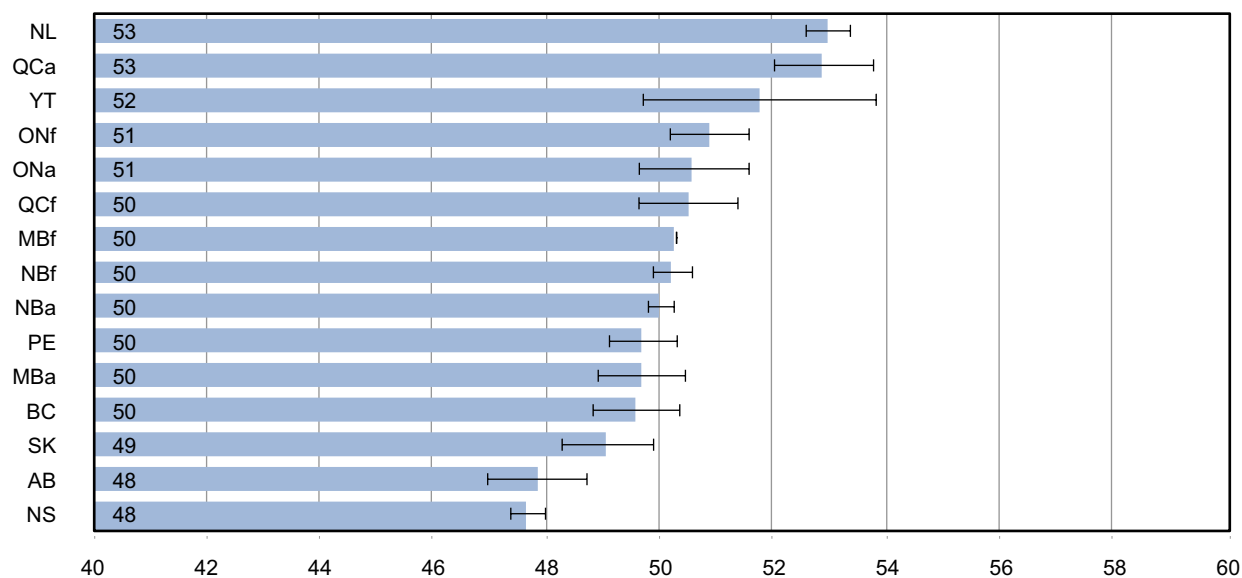
Les graphiques 9.6 et 9.7 montrent les scores factoriels moyens pour ces deux facteurs, par instance et selon la langue. Dans les deux cas, la tendance est relativement semblable à celle que donnent les facteurs obtenus pour les élèves. Le facteur des items à réponse courte produit trois grappes distinctes, mais aucune tendance géographique ou linguistique claire n'est constatée parmi les populations. Ici encore, le Québec (francophone et

anglophone) se distingue par une fréquence moindre de l'utilisation de cette méthode d'évaluation. Les différences entre les populations à l'égard du facteur des items à réponse longue sont plus petites. Terre-Neuve-et-Labrador et le Québec (anglophone) les utilisent significativement plus souvent et la Nouvelle-Écosse utilise moins souvent que la plupart des autres instances les items à réponse longue.

Graphique 9.6 Scores factoriels moyens d'après les réponses du personnel enseignant au sujet des items à réponse courte de l'évaluation, par instance et selon la langue



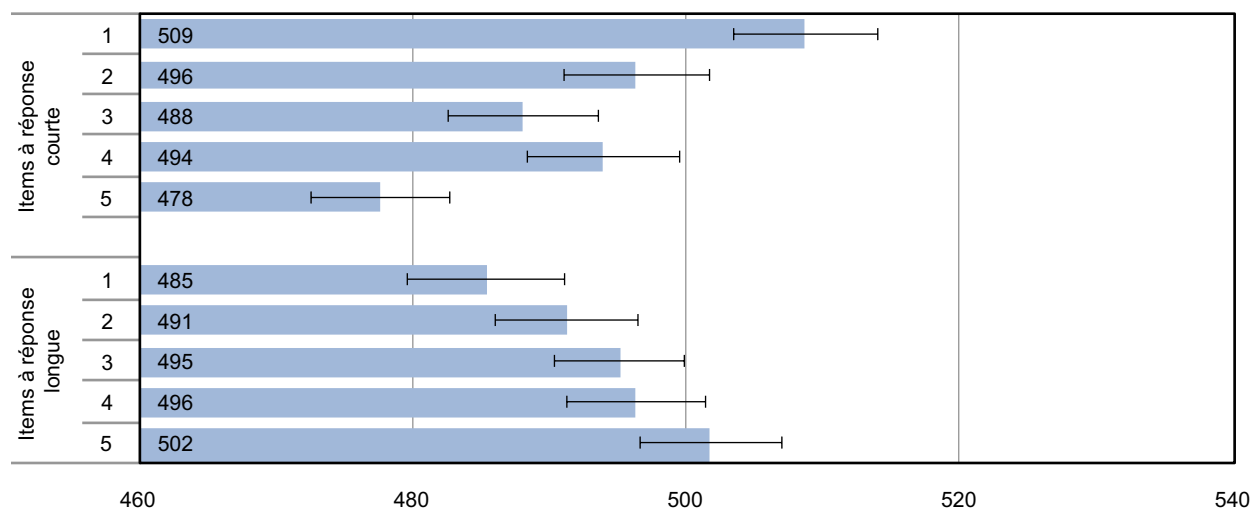
Graphique 9.7 Scores factoriels moyens d'après les réponses du personnel enseignant au sujet des items à réponse longue de l'évaluation, par instance et selon la langue



Le graphique 9.8 montre les scores moyens en lecture des enseignantes et enseignants qui utilisent ces deux méthodes d'évaluation. La tendance semble montrer une diminution du rendement à mesure qu'augmente la fréquence d'utilisation des items à réponse courte (le quintile 4 étant une anomalie par rapport à cette

tendance). L'inverse se produit en ce qui concerne l'utilisation des items à réponse longue, bien que les différences entre catégories adjacentes soient plus petites et généralement statistiquement non significatives.

Graphique 9.8 Scores moyens en lecture du personnel enseignant d'après leurs réponses au sujet des méthodes d'évaluation par quintile



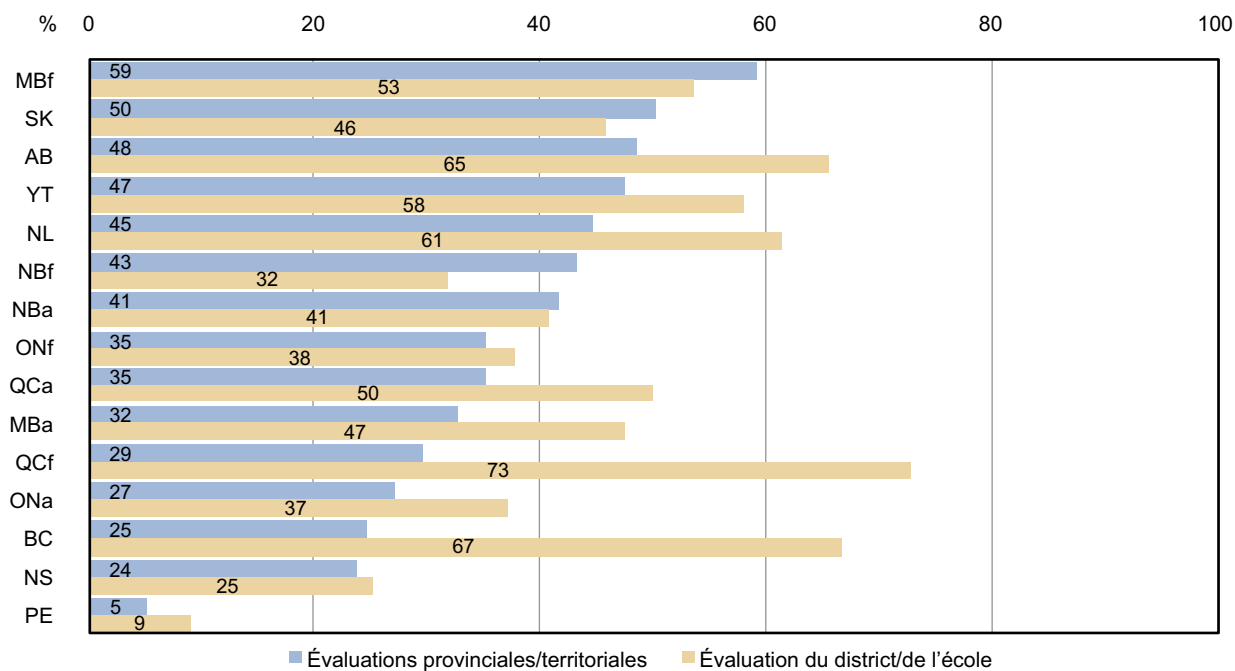
Éléments de l'évaluation contribuant à la note ou cote finale des élèves

Les questions à ce sujet portaient sur l'utilisation des évaluations ou tests employés à l'échelle de la province/du territoire et du district scolaire/de l'école, d'évaluations conçues par le personnel enseignant, de travaux, projets ou devoirs faits par les élèves et d'un portfolio dont le résultat était intégré dans la note ou cote finale des élèves. Les enseignantes et enseignants étaient priés d'indiquer à quelle fréquence ils utilisaient ces formes d'évaluation à cette fin, sur une échelle de fréquence à cinq points (jamais, 1 ou 2 fois par année, 3 à 5 fois par année, environ chaque mois et plus d'une fois par mois). Ils devaient dire en outre s'ils utilisaient d'autres critères de comportement dans le calcul de la note ou cote finale, et leurs réponses ont été rapportées à l'aide d'une échelle dichotomique (oui ou non).

Les évaluations à l'échelle de la province/du territoire, du district scolaire/de l'école servent rarement à l'attribution des notes ou cotes aux élèves, de sorte que les catégories représentant une

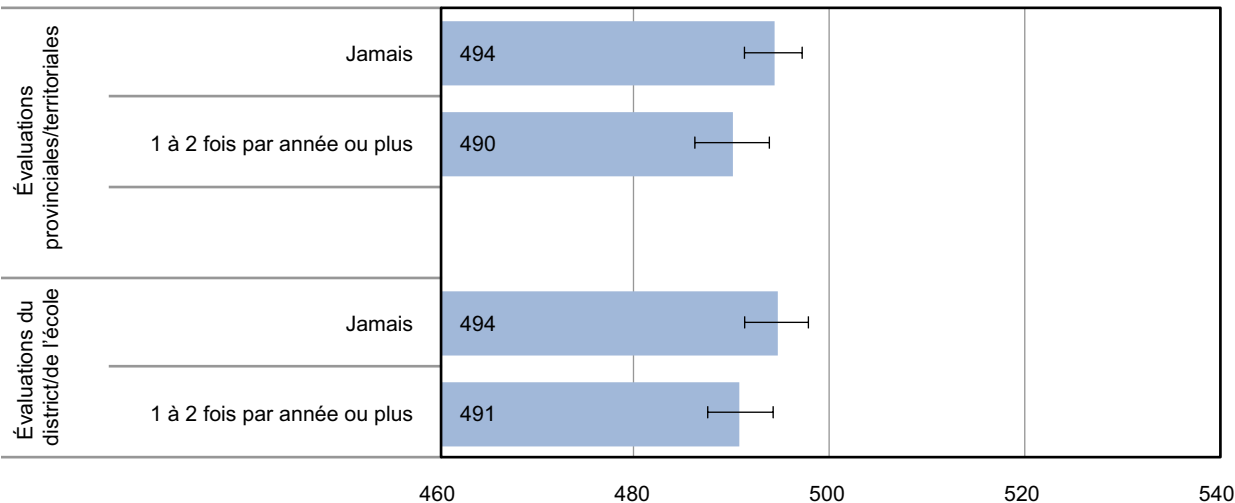
ou deux fois par année ou plus ont été combinées. Le graphique 9.9 montre le pourcentage d'enseignantes et enseignants qui disent utiliser ce genre de tests pour le calcul de la note ou cote une ou deux fois par année ou plus. Il y a beaucoup de variations entre les instances pour ces deux mesures. Ainsi, environ 60 p. 100 du personnel enseignant du Manitoba (francophone) mais seulement 5 p. 100 de celui de l'Île-du-Prince-Édouard intègrent les évaluations à l'échelle provinciale/territoriale dans le calcul de la note ou cote. Les enseignantes et enseignants du Québec (francophone) et ceux de la Colombie-Britannique se distinguent par un pourcentage plus élevé qui disent utiliser les tests à l'échelle du district/de l'école et le faible pourcentage qui emploient les tests à l'échelle de la province/du territoire. Par contre, une grande proportion d'enseignantes et enseignants de l'Alberta, du Yukon et de Terre-Neuve-et-Labrador utilisent les deux formes.

Graphique 9.9 Pourcentage d’enseignantes et enseignants qui utilisent une ou deux fois par année ou plus les évaluations provinciales/territoriales ou du district/de l’école dans le calcul de la note ou cote, par instance et selon la langue



Le graphique 9.10 montre l’incidence de l’utilisation de ces différentes formes d’évaluation sur le rendement en lecture. Dans un cas comme dans l’autre, la différence entre l’utilisation et la non-utilisation n’est pas statistiquement significative.

Graphique 9.10 Scores moyens en lecture d’après l’utilisation des évaluations provinciales/territoriales ou du district/de l’école dans le calcul de la note ou cote

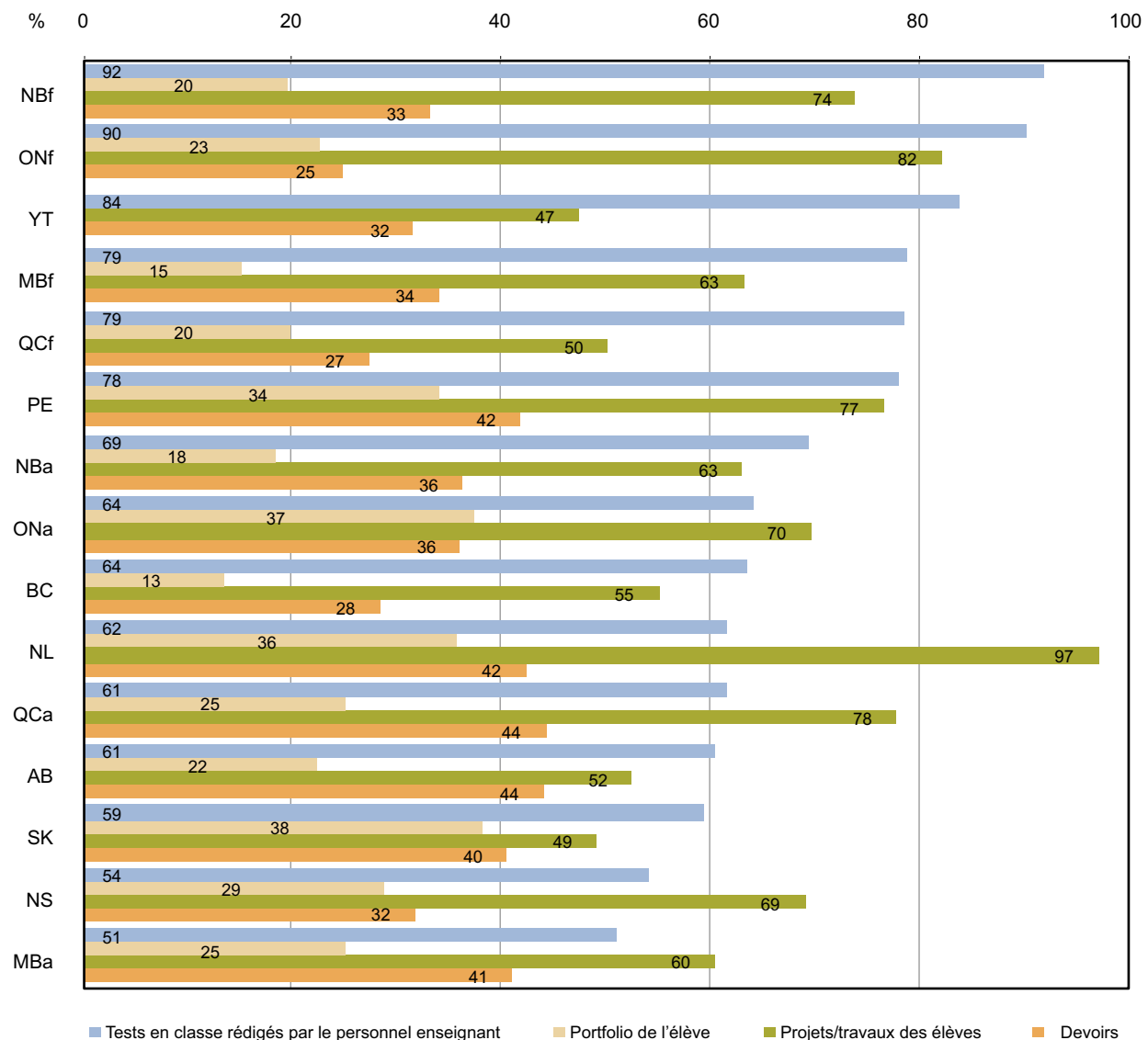


Comme on pouvait s’y attendre, l’évaluation à l’échelle de la classe est utilisée beaucoup plus souvent que l’évaluation externe. Le graphique 9.11 montre le pourcentage d’enseignantes et enseignants qui disent utiliser une fois par mois ou plus des tests qu’ils ont eux-mêmes conçus, les

portfolios des élèves, les projets/travaux des élèves et les devoirs. En général, les tests rédigés par le personnel enseignant et les projets/travaux des élèves sont utilisés plus souvent que les portfolios des élèves ou les devoirs. Ce sont les portfolios des élèves qui sont le moins utilisés. Les populations francophones sont généralement celles où le

personnel enseignant utilise le plus des tests qu'il conçoit lui-même. Les enseignantes et enseignants de Terre-Neuve-et-Labrador sont ceux qui utilisent le plus souvent les projets/travaux faits par les élèves. Cette méthode montre des variations considérables entre les populations.

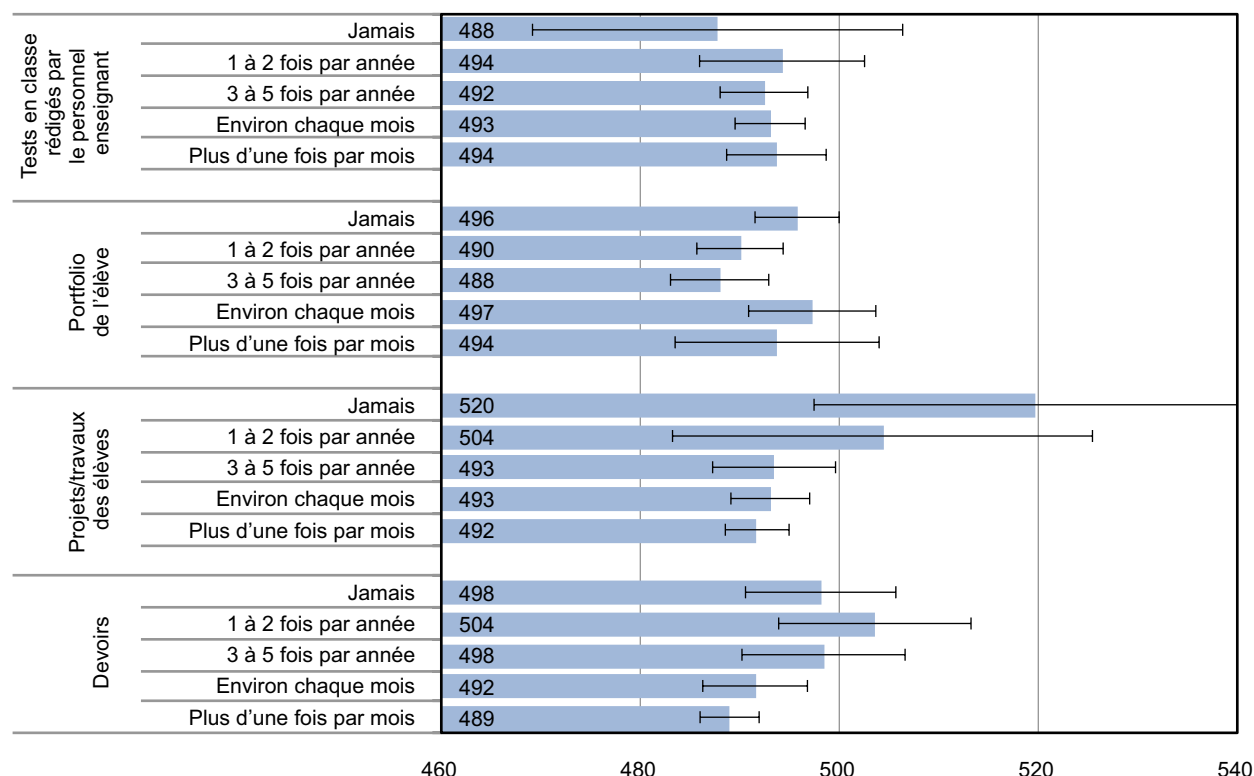
Graphique 9.11 Pourcentage d'enseignantes et enseignants qui utilisent une ou deux fois par année ou plus différentes méthodes d'évaluation en classe pour attribuer les notes ou cotes, par instance et selon la langue



Le graphique 9.12 montre les effets de la fréquence de l'utilisation de ces méthodes d'évaluation sur le rendement en lecture. Ici encore, aucune de ces différences n'est statistiquement significative. Il n'y a pas non plus de différence significative pour ce qui est de l'utilisation de tests rédigés par le personnel

enseignant ou de portfolios des élèves. Toutefois, on constate une légère tendance à la baisse du rendement en lecture à mesure qu'augmente l'utilisation des projets/travaux faits par les élèves ainsi qu'aux devoirs.

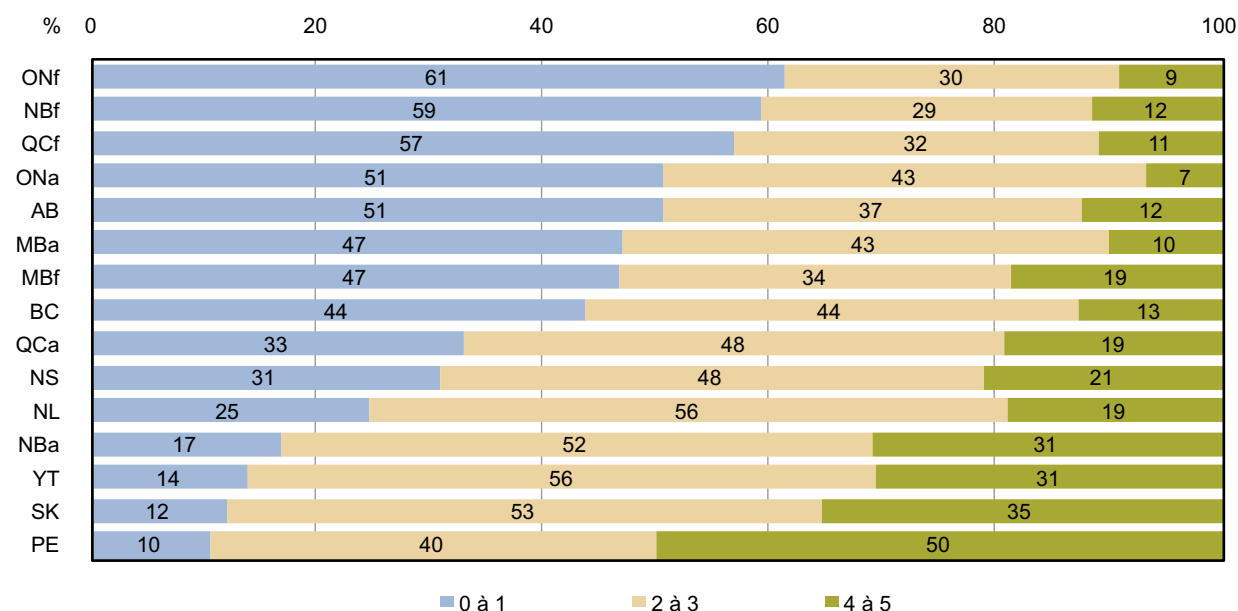
Graphique 9.12 Scores moyens en lecture du personnel enseignant d'après l'utilisation des évaluations en classe pour l'attribution de la note ou cote



Parmi les critères qui ne font pas appel aux connaissances scolaires mais qui peuvent servir à l'évaluation figurent la présence, la participation en classe, l'amélioration, l'effort et le comportement. Dans chaque cas, le personnel enseignant était prié de dire simplement s'il tenait compte de ces éléments dans l'attribution de la note ou cote. Les réponses par « oui » ont été additionnées pour produire une échelle de 0 à 5 selon le nombre de ces éléments utilisés. Ces derniers ont ensuite été répartis en trois catégories (0-1, 2-3 et 4-5) pour simplifier la présentation.

Le graphique 9.13 montre le pourcentage d'enseignantes et enseignants qui utilisent 0 à 1, 2 à 3 ou 4 à 5 pour attribuer les notes ou cotes. On constate de grandes variations entre les populations pour ce qui est du pourcentage d'enseignantes et enseignants qui utilisent les critères de 0 à 1 et de 4 à 5. Les variations sont moins marquées pour ceux qui disent utiliser les critères de 2 à 3. Trois des quatre populations francophones sont celles qui utilisent le moins ces critères.

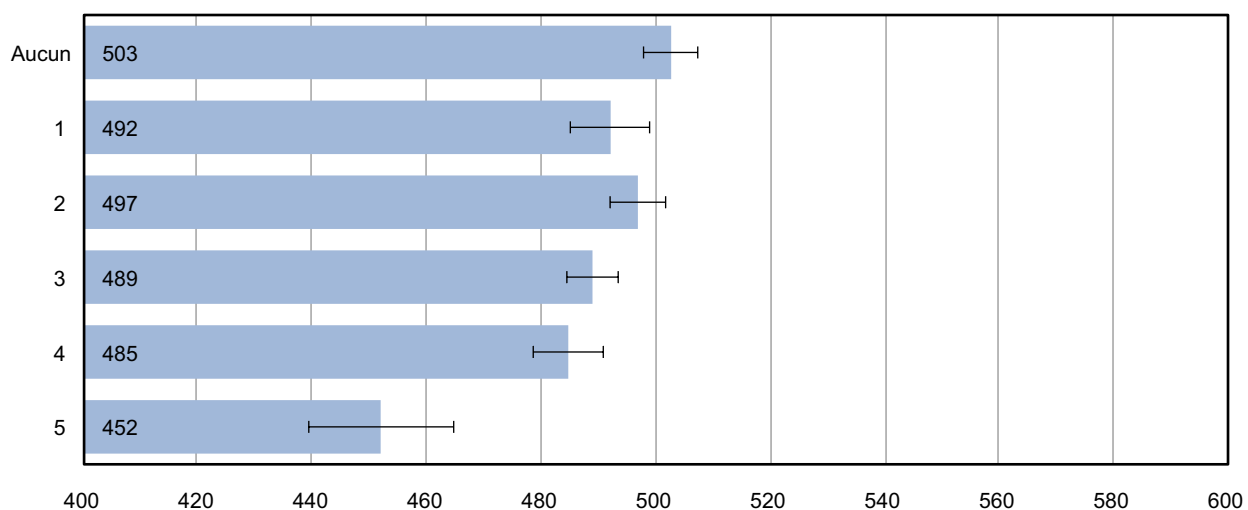
Graphique 9.13 Pourcentage d'enseignantes et enseignants qui utilisent des critères non liés aux connaissances scolaires pour attribuer les notes ou cotes, par instance et selon la langue



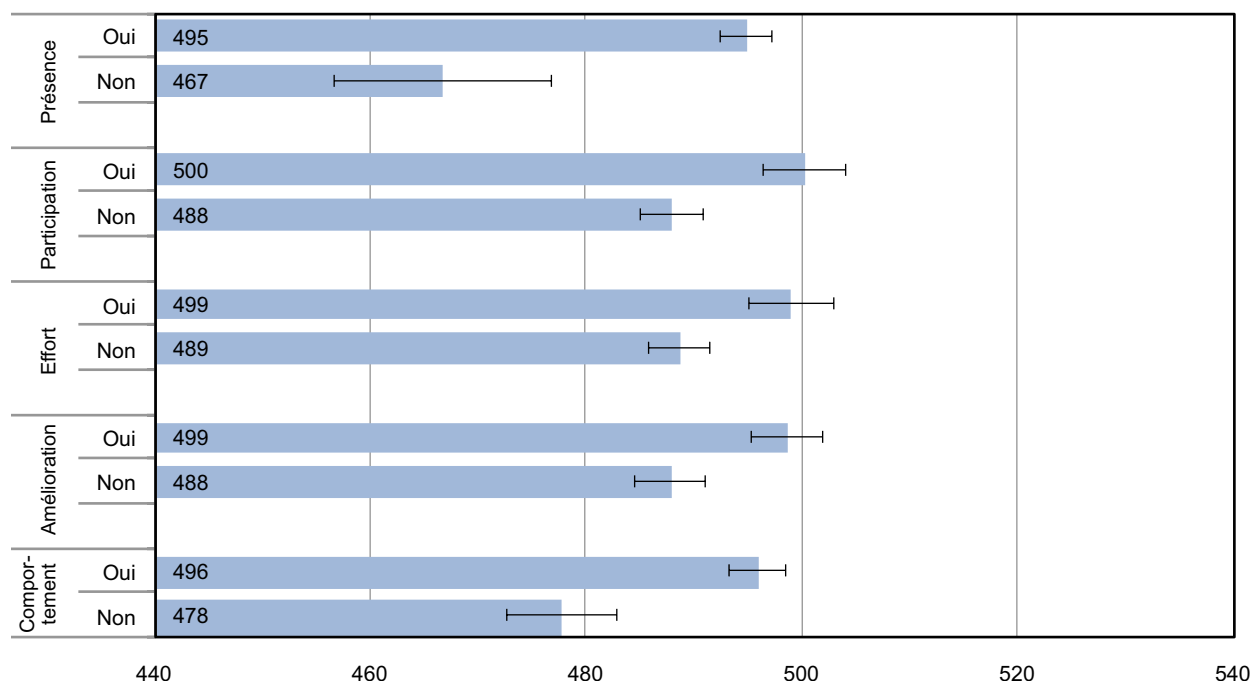
Le graphique 9.14 montre les scores moyens en lecture selon le nombre de critères non liés aux connaissances scolaires utilisés pour attribuer les notes ou cotes. La tendance ici révèle une diminution du rendement en lecture en fonction de l'augmentation du nombre de fois où ces critères sont

utilisés. Le graphique 9.15 montre les résultats de l'analyse des critères individuels. On voit beaucoup plus clairement que le rendement des élèves dont l'enseignante ou enseignant utilise ces critères est plus faible.

Graphique 9.14 Scores moyens en lecture d'après le nombre de critères non liés aux connaissances scolaires utilisés pour attribuer les notes ou cotes



Graphique 9.15 Scores moyens en lecture du personnel enseignant d'après l'utilisation de critères spécifiques non liés aux connaissances scolaires pour attribuer les notes ou cotes



Grilles de notation

Les grilles de notation sont des énoncés qui décrivent les résultats désirés et le niveau de rendement attendu par rapport à ces résultats au terme d'une tâche d'apprentissage donnée. Elles servent principalement à attribuer une note quand les critères de notation sont d'ordre qualitatif. Toutefois, elles peuvent servir aussi à informer les élèves des attentes et à organiser les tâches d'apprentissage. Les grilles de notation sont maintenant largement utilisées lors de l'enseignement des arts du langage pour expliciter les résultats désirés et les attentes.²⁶

Le personnel enseignant et les élèves devaient répondre à des questions sur l'utilisation des grilles de notation dans les cours d'arts du langage. Presque toutes les enseignantes et tous les enseignants disent utiliser des grilles de notation. Aucune différence n'est constatée dans le rendement en lecture entre les élèves des enseignantes et enseignants qui les utilisent et ceux, peu nombreux, qui n'en utilisent pas. Les réponses du

personnel enseignant ne sont donc pas reproduites en détail ici.

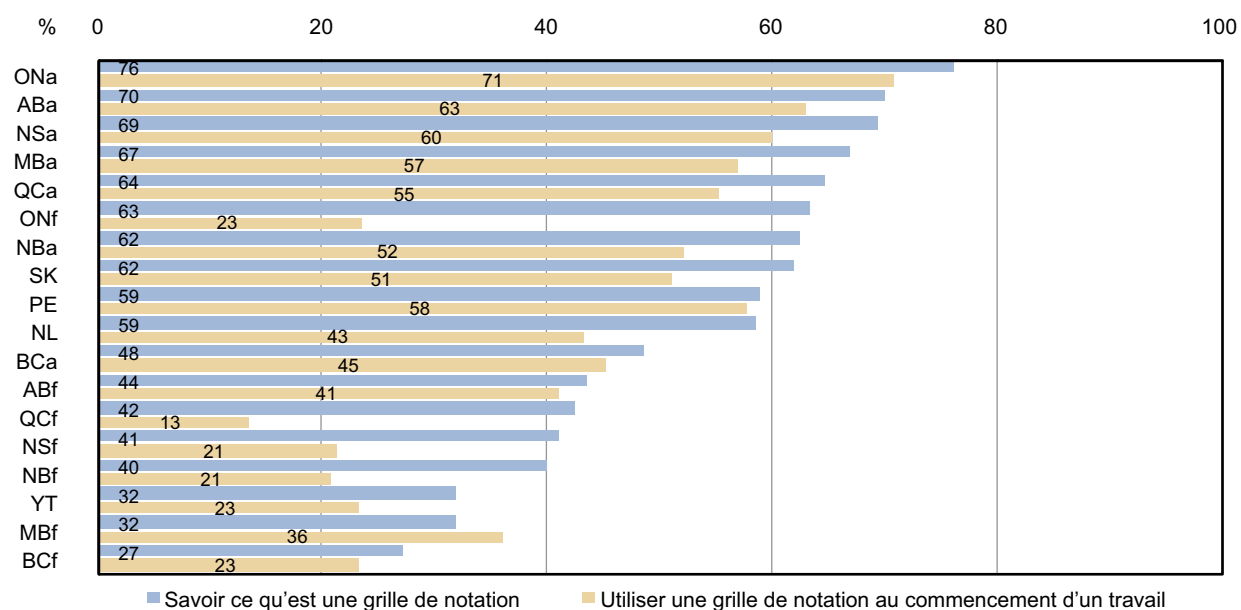
Les réponses des élèves brossent un portrait un peu différent. En général, deux tiers des élèves, environ, ont dit savoir ce qu'est une grille de notation et 50 p. 100 de ce groupe ont dit que leurs enseignantes et enseignants en utilisent souvent. Les graphiques 9.16 et 9.17 montrent les réponses détaillées par instance et selon la langue (les pourcentages manquants dans le graphique 9.17 résultent du nombre substantiel d'élèves qui ont répondu « je ne sais pas » à cette question). Ces graphiques montrent clairement que les élèves de la plupart des populations francophones connaissent moins et utilisent moins les grilles de notation que leurs homologues anglophones. L'Ontario (francophone) fait exception. L'écart entre le nombre d'élèves qui savent ce qu'est une grille de notation et le nombre de ceux qui l'utilisent est aussi plus grand dans quatre des sept populations

²⁶ Un exemple de grille de notation se trouve dans le *Rapport de l'évaluation des élèves de 13 ans en lecture, mathématiques et sciences* du PPCE-13 de 2007 [Toronto : Conseil des ministres de l'Éducation (Canada)], tableau 2-3 à la page 12 et les exemples d'accompagnement sont aux pages 16 et 17. Ils décrivent ce que les élèves devraient être capables de faire à chacun des trois niveaux de l'échelle de rendement en lecture du PPCE.

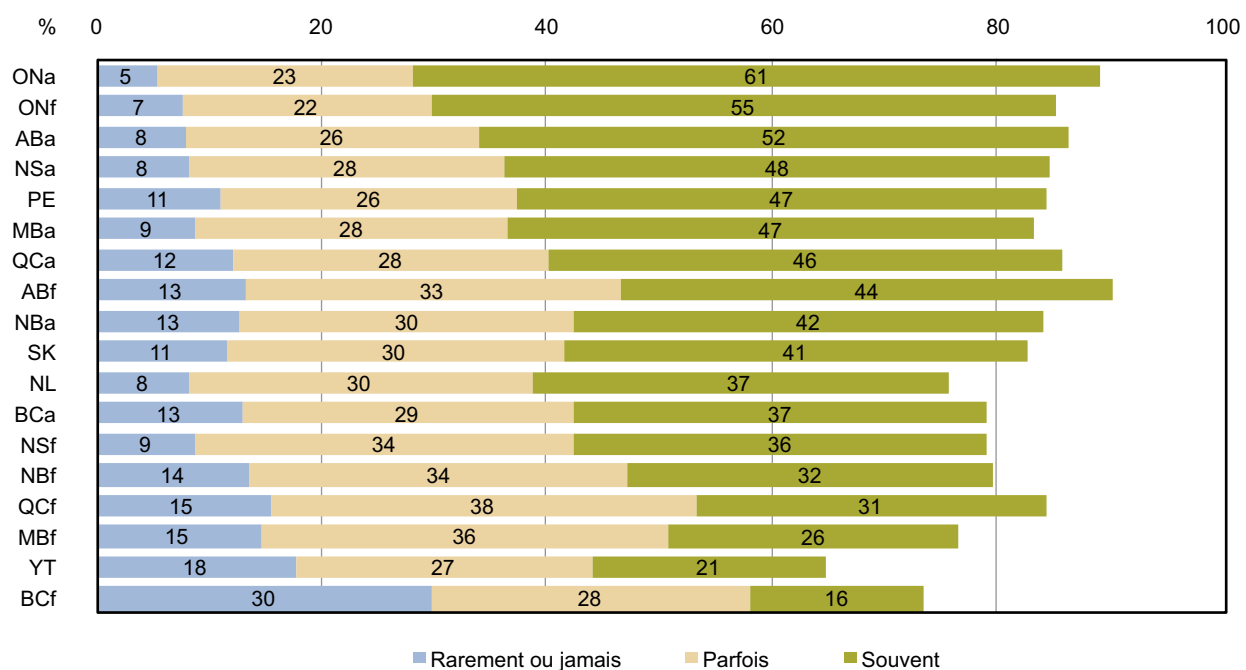
francophones. Il est particulièrement marqué en Ontario, où le nombre d'élèves qui savent ce qu'est une grille de notation est particulièrement grand dans les populations anglophone et francophone, mais où l'utilisation au début des travaux est beaucoup moins

grande chez les francophones que chez les anglophones. Les deux populations de l'Ontario rapportent une utilisation fréquente des grilles de notation pour l'attribution de la note.

Graphique 9.16 Pourcentage d'élèves qui savent ce qu'est une grille de notation et qui en reçoivent une au commencement d'un travail, par instance et selon la langue



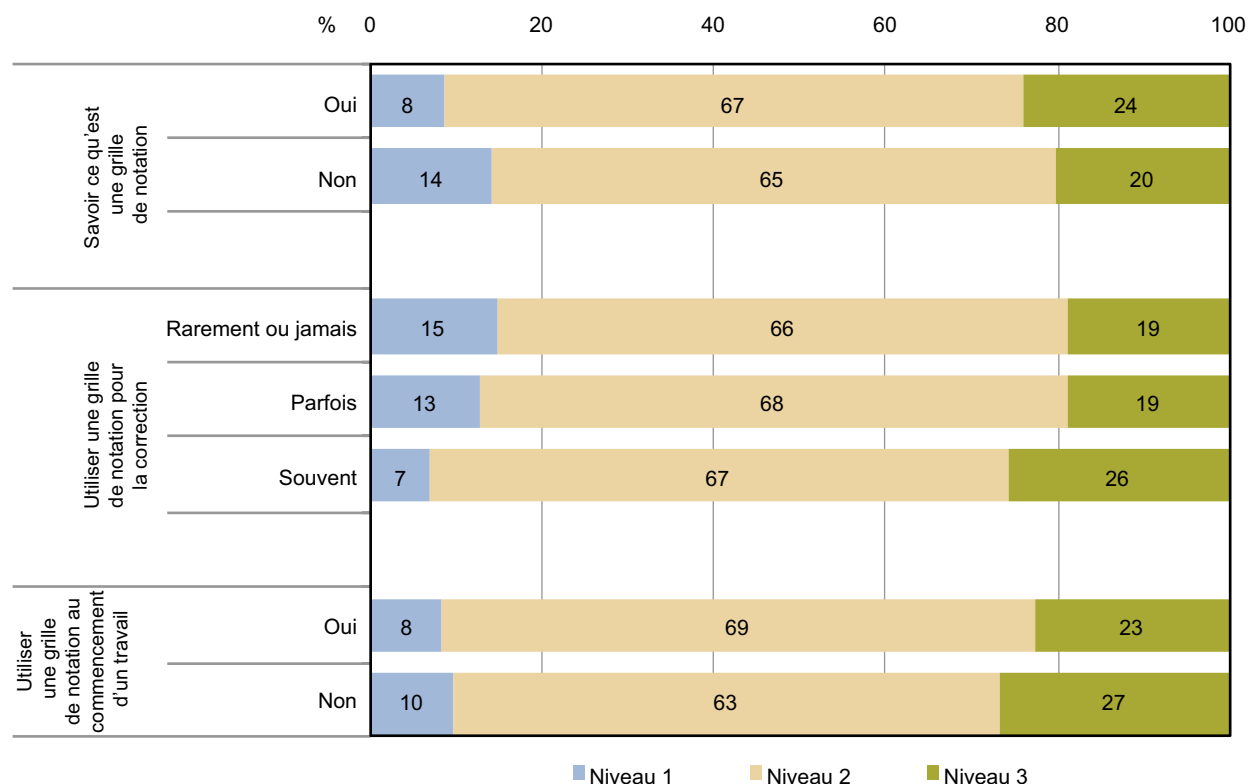
Graphique 9.17 Fréquence d'utilisation des grilles de notation pour la correction d'après les réponses des élèves, par instance et selon la langue



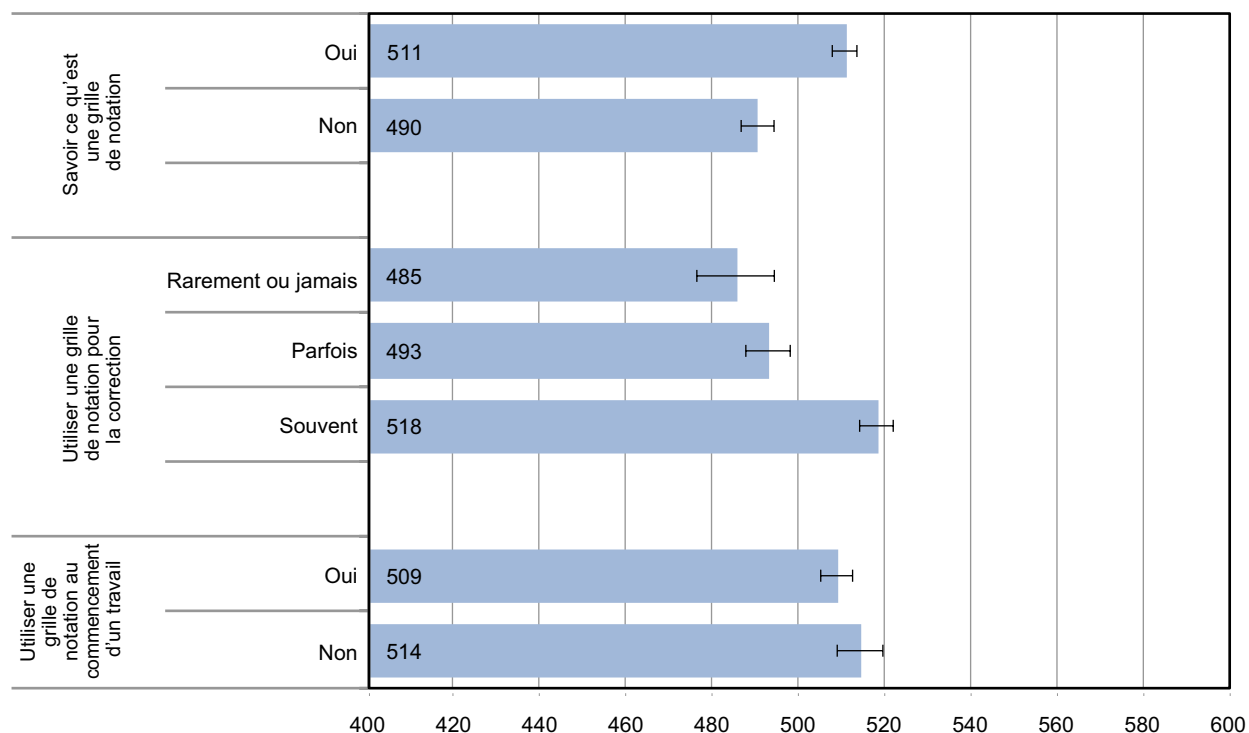
Le graphique 9.18 montre les niveaux de compétence et le graphique 9.19, les scores moyens en lecture des élèves qui disent connaître et utiliser les grilles de notation. Savoir ce qu'est une grille de notation et l'utiliser pour la correction sont deux variables

associées positivement au rendement en lecture. Toutefois, l'utilisation des grilles de notation au moment de commencer un travail ne paraît pas avoir d'effet significatif.

Graphique 9.18 Niveaux de compétence en lecture d'après la connaissance et l'utilisation des grilles de notation



Graphique 9.19 Scores moyens en lecture d'après la connaissance et l'utilisation des grilles de notation



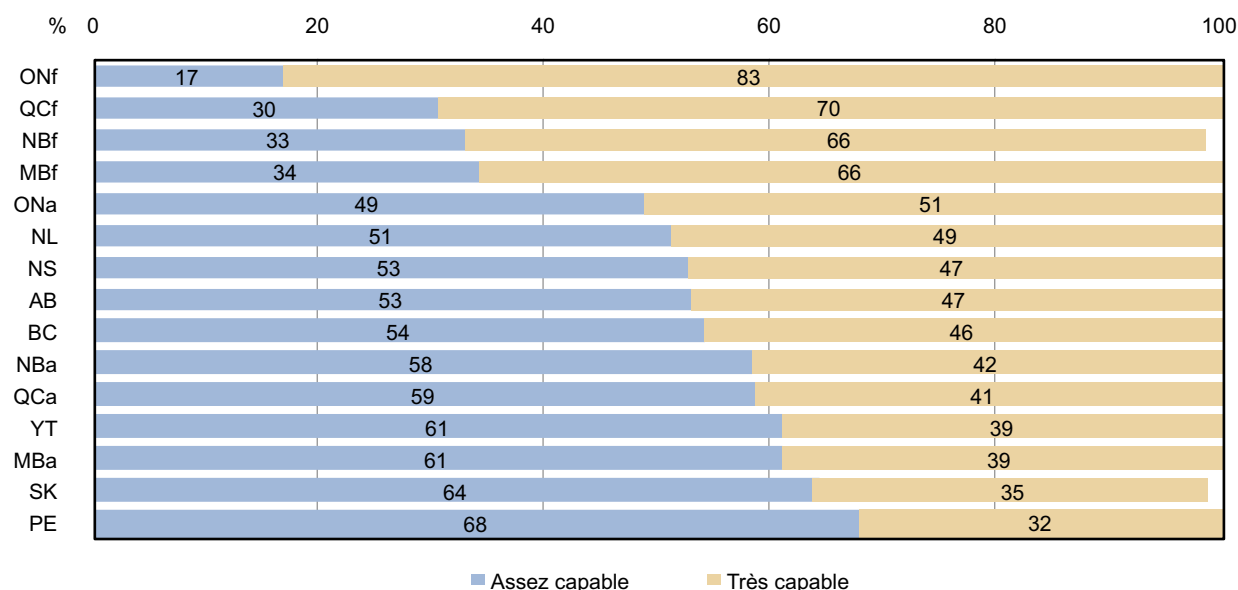
Habileté en évaluation du personnel enseignant

Les enseignantes et enseignants étaient priés de décrire leur habileté à l'égard de neuf types de tâches d'évaluation, y compris la conception de diverses formes de tests, la création et l'utilisation d'une grille de notation, l'interprétation des résultats et l'attribution des notes ou cotes finales. Presque tout le personnel enseignant s'est dit « assez capable » ou « très capable » d'exécuter la plupart de ces tâches. Une note composite a été conçue en calculant la moyenne des réponses aux neuf items et en arrondissant le résultat au nombre entier le plus proche. Le résultat était un indice composite à

deux valeurs, soit « assez capable » et « très capable ». (Moins de 1 p. 100 ayant la note composite « pas du tout capable », il n'a pas été tenu compte de ce niveau.)

Le graphique 9.20 montre les notes composites relatives à l'habileté à l'égard de diverses tâches d'évaluation, par instance et selon la langue. Il en ressort clairement que le personnel enseignant des populations francophones s'estime plus habile que celui des populations anglophones.

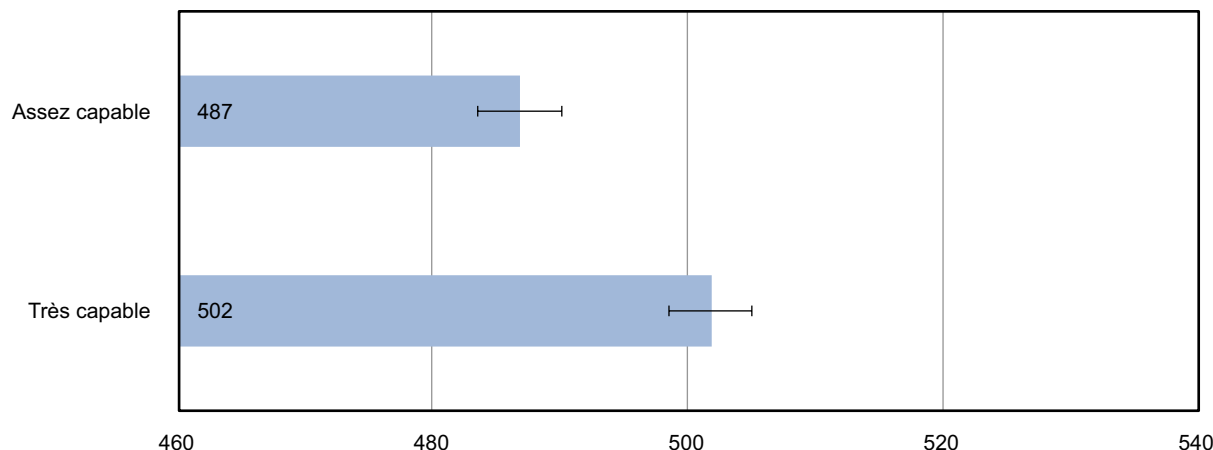
Graphique 9.20 Note composite du personnel enseignant d'après l'habileté en évaluation, par instance et selon la langue



Le graphique 9.21 montre que les enseignantes et enseignants qui se disent « très capable » d'évaluer

ont de meilleurs scores moyens en lecture que ceux qui se disent seulement « assez capable ».

Graphique 9.21 Scores moyens en lecture du personnel enseignant d'après la note composite de l'habileté en évaluation



Effets multivariés

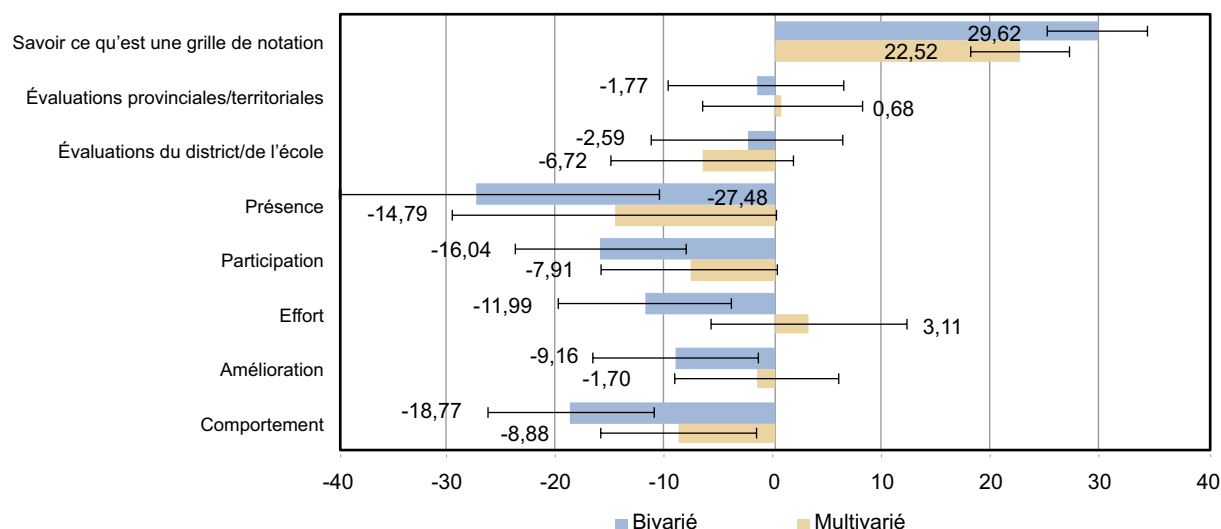
Étant donné le grand nombre de variables d'évaluation et les différences marquées entre les échelles utilisées, les variables ont été regroupées en grappes et les effets ont été reportés sur trois échelles.

Le graphique 9.22 montre les coefficients de régression des variables dichotomiques oui/non. L'effet, au niveau des élèves, de la connaissance des grilles de notation est relativement marqué et positif dans les deux modèles.²⁷ L'utilisation par le personnel enseignant d'évaluations

employées à l'échelle provinciale/territoriale ou du district/de l'école pour attribuer une note ou cote aux élèves n'a pas d'effet significatif. Presque tous les critères d'attribution des notes ou cotes non liés aux connaissances scolaires ont un effet significativement négatif dans le modèle bivarié. Cet effet est atténué dans une certaine mesure dans le modèle multivarié, mais trois de ces effets restent significativement négatif.

²⁷ Les deux autres variables relatives aux grilles de notation au niveau des élèves, soit l'utilisation des grilles pour faire les travaux et l'utilisation pour la correction, sont omises du modèle parce que ces questions n'ont pas été posées aux élèves qui ont indiqué ne pas savoir ce qu'est une grille de notation. Leur inclusion dans le modèle aurait résulté par un grand nombre de données manquantes pour toutes les variables du modèle, et non seulement pour ces variables spécifiques.

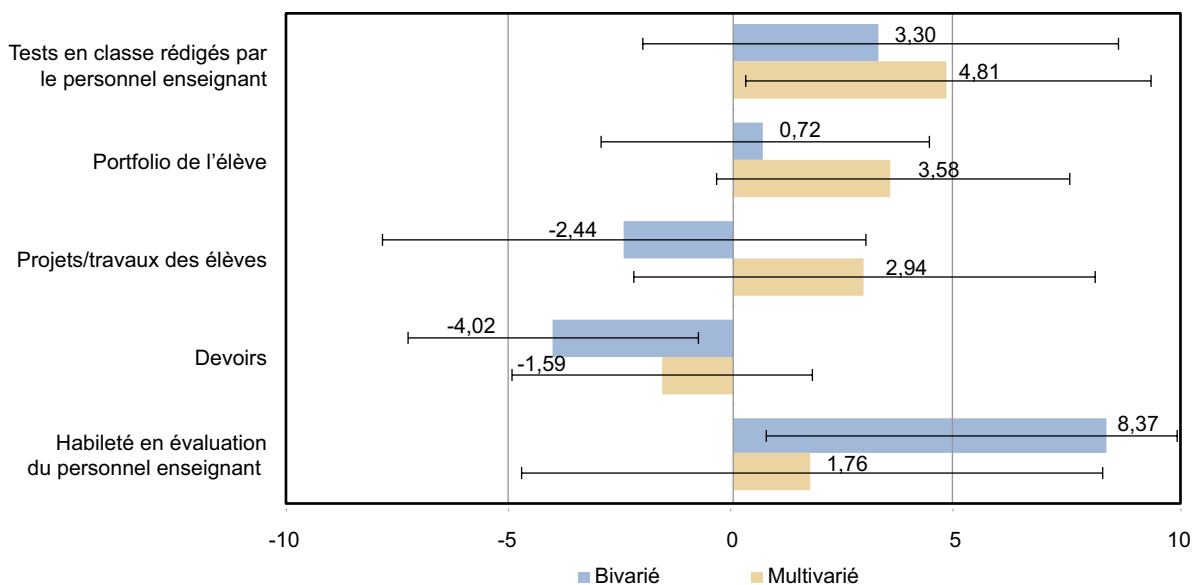
Graphique 9.22 Coefficients de régression portant sur les variables d'évaluation dichotomiques



Le graphique 9.23 montre que la plupart des variables d'évaluation au niveau du personnel enseignant n'ont pas d'effet significatif. Une exception notable, toutefois, est celle des tests rédigés par le personnel enseignant, dont l'effet devient significatif une fois toutes les autres variables neutralisées. L'effet de

l'habileté en évaluation du personnel enseignant est atténué dans le modèle multivarié par rapport au modèle bivarié, vraisemblablement à cause d'autres effets, tels celui de la connaissance des grilles de notation.

Graphique 9.23 Coefficients de régression portant sur les variables d'évaluation assorties d'une échelle



Le graphique 9.24 montre les effets des trois facteurs concernant les élèves et des deux facteurs concernant le personnel enseignant en rapport avec les types d'items utilisés. L'effet de l'utilisation des items à réponse courte (p. ex., vrai/faux, choix multiple) est négatif à la fois par rapport aux réponses du personnel enseignant et aux réponses des élèves, malgré le fait que l'effet multivarié sur le personnel enseignant ne soit pas statistiquement significatif. L'effet des items à réponse longue (p. ex.,

une phrase ou plus, un paragraphe ou plus) est significativement positif pour les élèves. L'effet du recours à la composition est notablement positif pour les élèves dans le modèle bivarié mais non dans le modèle multivarié. Ces deux derniers effets sont absorbés dans le facteur des items à réponse longue pour le personnel enseignant. Ce facteur n'est pas statistiquement significatif ni dans un modèle ni dans l'autre.

Graphique 9.24 Coefficients de régression portant sur les facteurs d'évaluation

