

**ÉLABORATION ET ÉVALUATION  
DES RAPPORTS DIAGNOSTIQUES  
DES DONNÉES DU PIRLS 2011:  
PERCEPTIONS DES ENSEIGNANTS  
AU PRIMAIRE, DES CONSEILLERS  
PÉDAGOGIQUES ET DES  
ORTHOPÉDAGOGUES\***



Dan Thanh Duong Thi\*\*

Nathalie Loye\*\*\*

**Résumé:** *La demande croissante pour des évaluations à grande échelle s'accompagne d'une forte pression pour rendre ces évaluations plus informatives sur l'apprentissage des élèves. Or, ces évaluations fournissent les résultats des élèves sous forme de scores globaux et de sous-scores, ce qui renseigne peu sur leurs forces et faiblesses. Par ailleurs, les recherches en approche diagnostique cognitive suggèrent qu'il est possible de décomposer la lecture en connaissances et habiletés possibles à diagnostiquer grâce à des modélisations psychométriques. Des épreuves à grande échelle ont donc le potentiel de fournir aux enseignants des rapports diagnostiques contenant des rétroactions détaillées sur les forces et les faiblesses des élèves. Cet article décrit l'élaboration de rapports diagnostiques à partir des données du PIRLS 2011 avec un panel d'experts et rapporte les résultats de l'évaluation de ces rapports auprès d'enseignants au primaire, de conseillers pédagogiques et d'orthopédagogues.*

**Mots-clés:** *Approche diagnostique cognitive. Élaboration et évaluation des rapports diagnostiques. Lecture. Évaluation à grande échelle.*

## Contexte

Une bonne maîtrise des habiletés en lecture en début des parcours scolaires exerce une influence positive sur la réussite future des

\* Recebido em: 08.02.2020. Aprovado em: 06.10.2020.

\*\* Professeure en méthodologies d'évaluation et de recherche en éducation au Département d'éducation et de pédagogie à l'Université du Québec à Montréal. *E-mail:* duong\_thi.dan\_thanh@uqam.ca.

\*\*\* Professeure en mesure et évaluation au Département d'administration et fondements de l'éducation, Université de Montréal. *E-mail:* nathalie.loye@umontreal.ca.

élèves (GOUVERNEMENT DU QUÉBEC, 2012; ROMANO *et al.*, 2010). En effet, un élève ayant des difficultés en lecture en 1<sup>re</sup> année a 9 chances sur 10 d'être en difficulté à la fin de sa 4<sup>e</sup> année et ce taux reste de 75% au secondaire (GOUVERNEMENT DU QUÉBEC, 2012). De plus, les résultats pendant les premières années du primaire peuvent prédire de façon significative le niveau d'études complété vers l'âge de 22 ans (ENTWISLE, ALEXANDER; OLSON, 2005). Au Québec, environ 70% des décrocheurs âgés de 15 à 26 ans ne maîtrisent pas suffisamment les habiletés en lecture pour répondre aux exigences d'une société du savoir (GOUVERNEMENT DU QUÉBEC, 2006). Il est donc important d'intervenir dès que possible afin d'assurer la réussite des élèves en lecture au primaire (GOUVERNEMENT DU QUÉBEC, 2012).

Les dernières années, la demande croissante pour des évaluations à grande échelle au primaire s'accompagne d'une forte pression pour rendre ces évaluations plus informatives sur l'état de l'apprentissage des élèves (CHIU, SEO; DOUGLAS, 2009). Or, ces évaluations ne fournissent que les résultats des élèves sous forme de scores globaux ou de sous-scores, ce qui nous renseigne peu sur leurs forces et leurs faiblesses et ne guide pas les enseignants dans la proposition de pistes d'interventions appropriées (LEIGHTON, GIERL, 2007).

Par ailleurs, l'approche diagnostique cognitive (ADC) offre la possibilité d'obtenir des informations détaillées sur les forces et faiblesses des élèves (LEIGHTON; GIERL, 2007). Combinant la psychologie cognitive et la psychométrie, cette approche s'appuie sur deux composantes: 1) l'analyse du contenu des items pour identifier les habiletés sous-jacentes et 2) les modèles psychométriques représentant les liens entre ces items et ces habiletés (LEE, SAWAKI, 2009; YANG, EMBRETSON, 2007). Les habiletés sont en général identifiées par les experts du domaine à partir des modèles théoriques sous-jacents aux contenus visés par le test et dans des protocoles verbaux des élèves. Par exemple, des études réalisées avec le *Test of English as a foreign language* (TOEFL) et le *Michigan English Language Assessment Battery* (METLAB) confirment que la lecture peut être décomposée en un ensemble d'habiletés possibles à diagnostiquer grâce à des modélisations psychométriques (JANG, 2005; LI, 2011).

Or, les recherches de Jang (2005) et de Li (2011) sont basées sur des tests administrés aux adultes. Peu d'études ont été réalisées avec un test en lecture du primaire, une tranche d'âge qui nécessite pourtant

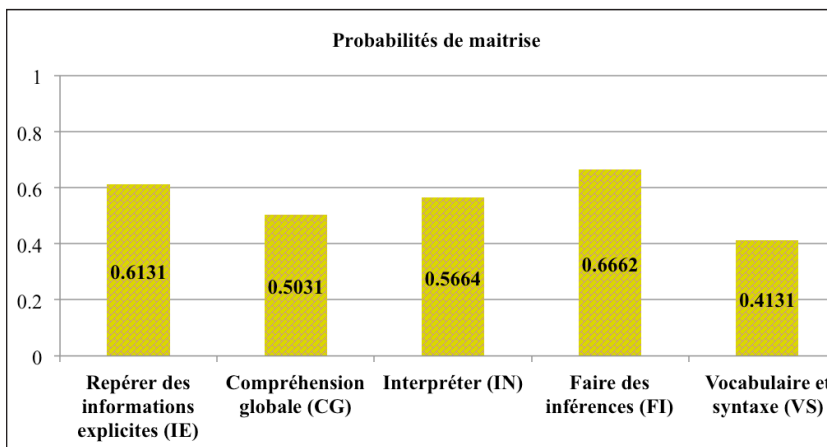
tout particulièrement des pistes d'intervention, ce qui a justifié notre choix du Programme international de recherche en lecture scolaire (PIRLS) 2011. Ainsi, dans une étude réalisée en 2018 (DUONG THI, 2018), nous avons démontré la faisabilité d'obtenir un diagnostic détaillé, fiable et nuancé à partir des modélisations des résultats de 4762 élèves canadiens ayant pris le livret 13 du PIRLS 2011. Dans cette étude, cinq habiletés, telles que présentées dans le tableau 1, ont été identifiées par un panel d'experts.

**Tableau 1** - Description des habiletés identifiées par le panel d'experts

Habiletés	Définition
(A1) <i>Repérer des informations explicites</i>	Localiser et reconnaître des informations explicites exprimées dans le texte pour répondre aux questions
(A2) <i>Compréhension globale</i>	Former une compréhension globale d'un paragraphe ou de l'ensemble du texte
(A3) <i>Interprétation</i>	Clarifier le sens des idées ou des configurations complexes et interpréter des relations
(A4) <i>Faire des inférences</i>	Comprendre les informations qui ne sont pas explicitement exprimées en faisant des inférences ou des prédictions
(A5) <i>Vocabulaire et syntaxe</i>	Exprimer des idées dans une grammaire correcte et compréhensible de l'anglais écrit

Source: Duong Thi (2018, p. 136).

Ainsi, notre étude, à l'instar de nombreuses autres études (GIERL, CUI, HUNKA, 2008; JANG, 2005; TEMPLIN, HENSON, 2006), confirme que l'ADC donne la possibilité de fournir des rétroactions précises et fiables à chaque élève relativement aux habiletés identifiées, ce qui permet une influence positive sur l'enseignement et sur l'apprentissage en lecture. La figure 1 présente le portrait global de maîtrise des habiletés pour l'ensemble des 4762 élèves. Parmi les cinq habiletés identifiées, *Faire des inférences* est celle la plus maîtrisée par 66,62 % des élèves, viennent ensuite *Repérer des informations explicites* (61,31 %) et *Interpréter* (56,64 %). Les habiletés *Compréhension globale* (50,31 %) et *Vocabulaire et syntaxe* (41,31 %) semblent celles les plus difficiles à maîtriser.



**Figure 1** - Le portrait global de maîtrise des habiletés de 4762 élèves du livret 13 du PIRLS 2011

Source: Duong Thi (2018, p. 151).

Dans un autre ordre d'idée, les concepteurs d'épreuves à grande échelle portent plus d'attention au fait de répondre aux besoins des décideurs en politiques éducatives plutôt qu'à l'organisation et à la communication des résultats aux enseignants (GOODMAN, HAMBLETON, 2004). Les recherches sur l'élaboration et l'évaluation de rapports mettent en évidence les difficultés des enseignants dans l'interprétation de tels rapports en raison du jargon statistique (VEZZU; VANWINKLE; ZAPATA-RIVERA, 2012). En outre, même si ces rapports leur semblent très importants, les enseignants ne sont généralement pas impliqués activement dans leur processus d'élaboration (VEZZU; VANWINKLE; ZAPATA-RIVERA, 2012). Ils sont en effet rarement invités à fournir des commentaires sur les contenus des rapports, leur compréhension de ces contenus ou sur les formats des rapports élaborés.

Pour contribuer à documenter le développement et l'évaluation par des enseignants de tels rapports, la présente étude est basée sur les profils diagnostiques obtenus des modélisations des données de 4762 élèves canadiens ayant fait le livret 13 du PIRLS 2011 (DUONG THI, 2018). Les profils correspondent aux probabilités de maîtrise des cinq habiletés présentées dans le tableau 1. Cet article a pour objectif de présenter les résultats du processus de l'élaboration puis de l'évaluation de ces rapports destinés aux enseignants.

## CADRES DE RÉFÉRENCES POUR L'ÉLABORATION ET L'ÉVALUATION DES RAPPORTS

De nombreux cadres de référence ont été développés pour guider le processus de l'élaboration et de l'évaluation des rapports, à savoir celui de Hambleton et Zenisky (2010), de Vezzu, Vanwinkle et Zapata-Rivera (2012) et de Zapata-Rivera, Vanwinkle et Zwick (2012). Malgré les différences du public visé et des objectifs des rapports, ces cadres de référence se composent des étapes principales suivantes: 1) identifier le public visé et analyser ses besoins; 2) documenter les recherches et les formats de rapports qui existent; 3) développer et réviser des rapports et 4) évaluer et valider des rapports auprès du public visé. La section suivante vise donc à décrire plus en détail chacune de ces étapes.

### Identifier le Public Visé et Analyser ses Besoins

L'identification du public visé et l'analyse de ses besoins sont une étape indispensable dans l'élaboration des rapports, car les utilisateurs des rapports sont des acteurs centraux qui font des inférences et prennent des décisions à partir des résultats communiqués (ZAPATA-RIVERA; KATZ, 2014). L'analyse du public visé doit porter sur leurs besoins, leurs connaissances et leurs attitudes (FLOWER, 1985). Les besoins renvoient aux objectifs et aux raisons pour lesquelles ce public visé lit les rapports. Les connaissances portent non seulement sur ce que le public sait déjà, mais aussi sur ce qu'il faut lui fournir pour mieux comprendre les rapports. Par exemple, les enseignants connaissent bien leurs élèves et leurs niveaux de performance. Ils ont aussi de bonnes connaissances sur les objectifs et le contenu des évaluations ainsi que sur les échelles de mesure (ZAPATA-RIVERA; KATZ, 2014). Toutefois, leurs connaissances sont limitées au niveau des erreurs de mesure ou de l'interprétation des résultats et du langage utilisé dans les rapports (HAMBLETON; SLATER, 1997). Finalement, les *attitudes* réfèrent aux sentiments ou biais qui peuvent influencer l'interprétation des résultats (FLOWER, 1985; VEZZU *et al.*, 2012). L'identification du public visé et l'analyse de ses besoins permettent donc de prendre les décisions sur le design des rapports (ZAPATA-RIVERA; KATZ, 2014).

## Documenter les Recherches et les Formats de Rapports qui Existent

La recherche de Goodman et Hambleton (2004) attire l'attention des chercheurs du domaine, car elle fournit un guide très précis et complet sur l'élaboration des rapports pour le *National Assessment of Educational Progress* (NAEP) (RYAN, 2006). Les auteurs ont identifié des faiblesses, tant sur la forme que sur le fond, qui nuisent à l'interprétation des résultats. Au niveau de la forme, les rapports développés montrent un manque de variété des formats de présentation des résultats et une surcharge d'informations. Quant au fond, le niveau élevé de connaissances statistiques et le manque de guide d'interprétation avec des informations sur l'importance des résultats d'évaluation peuvent dérouter les utilisateur (GOODMAN, HAMBLETON, 2004; ZENISKY, HAMBLETON, SIRECI, 2009).

Les recherches de Zapata-Rivera, Vanwinkle et Zwick (2012) et Zapata-Rivera et Katz (2014) s'intéressent à l'interprétation des rapports interactifs en ligne du programme *Cognitively Based Assessment of, for, and as Learning* (CBAL). Trois types de rapports sont élaborés et validés avec ou sans guide d'interprétation des rapports. Ils contiennent des informations sur les échelles de mesure, le niveau de performance, le guide d'interprétation, les liens d'accès aux sources supplémentaires et les recommandations pour les enseignants. Les résultats de l'évaluation indiquent que les informations doivent être présentées d'une manière concise en évitant de longs paragraphes. Les auteurs soulignent l'importance des guides d'interprétation et du glossaire permettant aux enseignants de comprendre des concepts sophistiqués.

Le premier rapport diagnostique, *Score Plus Report*, a été élaboré pour communiquer les résultats en mathématique, en littérature critique et en écrit de *Preliminary Standard Assessment Test/National Merit Scholarship Qualifying Test*. Les informations diagnostiques sur les trois habiletés les moins maîtrisées sont présentées avec des stratégies de remédiation (ROBERTS et GIERL, 2010). Un autre format de rapport diagnostique conçu à partir des résultats du TOEFLIBT, contient des probabilités de maîtrise des habiletés, des indices de discrimination des items et la description des habiletés (JANG, 2005).

De leur côté, Roberts et Gierl (2010) s'intéressent au design des rapports à partir des résultats d'un test en mathématique avec une re-ension très riche sur des principes de présentation des informations textuelles et non textuelles. Le profil de maîtrise des habiletés de l'élève est présenté en trois parties: les directives pour lire le rapport; l'examen

des réponses avec la performance de chaque habileté et un résumé de la performance de l'élève. Le rapport est accompagné d'un guide avec la description détaillée des habiletés et des questions souvent posées lors de l'interprétation des résultats.

## Développer et Réviser des Rapports

Plusieurs guides d'élaboration des rapports efficaces ont émergé dans la littérature et varient selon les programmes d'évaluation et les provinces. À travers des recherches empiriques, Hambleton et Slater (1997), Goodman et Hambleton (2004), Roberts et Gierl (2010), Ryan (2006) et Zenisky, Hambleton et Sireci (2009) ont proposé un ensemble de principes afin d'assurer une bonne qualité des rapports que nous avons regroupé en quatre critères principaux: accessibilité, utilité, lisibilité et fiabilité.

Selon ces auteurs, l'accessibilité renvoie à accorder une importance à la personnalisation des rapports pour produire une variété de rapports selon les types d'informations disponibles, l'objectif et le public visé. L'utilité fait référence à la pertinence et à la quantité raisonnable des informations présentées pour éviter la surcharge. La lisibilité vise à assurer la clarté, la simplicité et la compréhensibilité d'une variété de formats de présentations en évitant le jargon statistique, en utilisant différentes couleurs selon les niveaux de performances et des légendes pour les expliquer. La fiabilité consiste à assurer la justesse des résultats présentés. Par exemple, les sous-scores présentés doivent être comparables entre les individus et le groupe, et fournir des informations plus détaillées que les scores totaux.

Roberts et Gierl (2010) ont d'ailleurs proposé des principes de présentations des informations textuelles et graphiques dans lesquels l'attention devrait être accordée à la structure interne et externe du texte. La structure interne du texte comprend des techniques d'organisation et de séquences permettant de fournir des balises pour comprendre son contenu. Par exemple, un paragraphe doit contenir au moins une phrase qui exprime son idée-clé et une autre pour conclure le problème annoncé.

La structure externe comprend des techniques comme l'accessibilité de la structure, la typographie et la disposition des textes. *L'accessibilité de la structure* du texte sera plus facile avec une table des matières, un glossaire, des objectifs et un sommaire (ROBERTS; GIERL, 2010). Pour la *typographie*, l'utilisation de différents types de caractères et de couleurs peut attirer l'attention des utilisateurs sur des idées importantes et créer

le contraste (HORTON, 1991). Le recours à la disposition horizontale ou verticale du texte permet de renforcer la structure hiérarchique du document au niveau visuel (ROBERTS; GIERL, 2010).

Pour la présentation des résultats sous forme de tableaux et de graphiques, il est recommandé d'utiliser : a) le contraste pour exprimer les informations importantes; b) les redondances des indices visuels pour accentuer les informations présentées (couleurs différentes); c) la proximité pour regrouper les éléments similaires et d) un alignement commun pour marquer la structure visuelle des informations (CLEVELAND et MCGILL, 1985). L'utilisation d'un tableau est recommandée pour une petite quantité de données où il faut fournir des chiffres exacts pour les comparaisons (TUFTE, 1996) tandis que les graphiques peuvent mieux communiquer les changements et les tendances des données. Le graphique et le texte doivent être intégrés dans une même page pour faciliter l'interprétation (ROBERTS; GIERL, 2010). Ces principes de la présentation des résultats peuvent être transférés au contexte d'un rapport diagnostique. Toutefois, le processus de l'élaboration des rapports est complété seulement si ces rapports sont validés auprès du public visé.

### Évaluer et Valider des Rapports Auprès du Public Visé

Dans les études empiriques, les méthodes pour valider les rapports reposent sur des entrevues de groupe ou des questionnaires. Les questions portent principalement sur l'interprétation des représentations graphiques; les préférences des formats de représentation; l'efficacité des représentations; le contenu des rapports et les suggestions pour améliorer des rapports. Le tableau 2 propose une synthèse des dimensions à retenir lors de la validation des rapports avec des exemples de questions à poser.

Les résultats de ces études suggèrent que les difficultés des enseignants sont liées à l'interprétation des notions statistiques comme la fiabilité, les percentiles, les vrais scores et les échelles de mesure (HAMBLETON, SLATER, 1997; ZAPATA-RIVERA, VANWINKLE, ZWICK, 2012). De plus, la compréhension des rapports dépend des formats de représentation (ZAPATA-RIVERA, VANWINKLE, ZWICK, 2012), de la présence ou non des guides d'interprétation (IMPARA, ZAPATA-RIVERA, KATZ, 1991; ZENISKY, HAMBLETON, SIRECI, 2009) et de la familiarité avec les cours de statistiques (HAMBLETON; SLATER, 1997).



En effet, les enseignants ont plus de difficultés à interpréter les boîtes à moustaches que les graphiques avec des icônes empilées ou la courbe de distribution des scores (ZAPATA-RIVERA, VANWINKLE, ZWICK, 2012) et les rapports sans guide d'interprétation (IMPARA *et al.*, 1991; ZENISKY, HAMBLETON, SIRECI, 2009). Ces recherches montrent clairement que les enseignants ont besoin d'informations additionnelles sur les concepts en statistique pour comprendre le contenu présenté (ZWICK, ZAPATA-RIVERA, HEGARTY, 2014).

Jusqu'à présent, il n'existe qu'une seule recherche portant sur l'évaluation de rapports diagnostiques. Elle vise à vérifier l'utilité des informations diagnostiques dans l'apprentissage en lecture, mais pas à examiner la présentation visuelle des résultats. Or, la nature des résultats des modélisations de l'ADC nécessite le recours à des éléments textuels et à des représentations graphiques. Le fait que les résultats se présentent sous forme de niveaux de maîtrise des habiletés de l'élève et non de scores suscite de nombreuses questions chez des enseignants, peu familiers avec une telle approche.

À partir des cadres de référence d'élaboration et d'évaluation des rapports, nous retenons qu'il est important de documenter les recherches et les rapports qui existent, de concevoir le protocole du design des rapports, d'élaborer les rapports avec un panel d'experts, de développer un questionnaire pour l'évaluation des rapports, de prévalider le questionnaire et de procéder à l'évaluation des rapports auprès du public visé. Ainsi, cette étude vise à élaborer des rapports avec un panel d'experts à partir des profils diagnostiques des élèves issus de l'étude de Duong Thi (2018) et à évaluer ces rapports auprès des enseignants pour les valider.

**Tableau 2** - Synthèse des dimensions de l'évaluation de rapports

Dimensions	Sous-dimensions	Types des questions	Exemples de questions	Références
Préférence	Format de présentations préférées des enseignants	- QCM sur le format de représentation préféré - Questions ouvertes pour justifier les choix	- Parmi les formats des rapports présentés, lequel préférez-vous?	Zapata-Rivera <i>et al.</i> (2013)

à suivre

Dimensions	Sous-dimensions	Types des questions	Exemples de questions	Références
Évaluation de la qualité des rapports	Perceptions des enseignants sur la présentation textuelle et graphiques selon les quatre critères d'un bon rapport:  - Accessibilité - Lisibilité - Utilité - Fiabilité	- Énoncés de type Likert à quatre points (degré d'accord ou de désaccord) sur les différentes parties du rapport - Questions ouvertes pour améliorer les rapports	-Les chiffres présentés dans les barres facilitent le positionnement de l'élève par rapport au groupe. -Les niveaux de maîtrise des habiletés de l'élève sont faciles à comparer avec ceux du groupe	Vezzu et al. (2012)
	Compréhension et interprétation des résultats	- QCM portent sur l'interprétation du profil de l'élève	Dans le graphique sur le profil de maîtrise des habiletés du groupe, pour l'habileté Interpréter, quel groupe a le pourcentage le plus élevé?	
Sociodémographie	Informations contextuelles qui influencent l'évaluation de la qualité des rapports, la compréhension et la préférence	-Genre, tranche d'âge, domaines de formation, diplôme obtenu, années d'expérience, connaissances en statistiques, etc.	-Quel est votre genre? -Quelle est votre tranche d'âge? -Quel est votre nombre d'années d'expérience?	

Source: Duong Thi (2018, p. 166).

## Méthodologie et résultats

### Phase 1: Élaboration Des Rapports Diagnostiques Avec Un Panel D'experts

#### Participants

Un panel de cinq experts provenant de différents milieux éducatifs a été constitué pour élaborer les rapports. Les experts devaient répondre au moins à l'un des critères suivants: a) avoir une très bonne compréhension en ADC; b) avoir de l'expertise dans le domaine de langues et c) avoir de l'expérience en enseignement au primaire. Notre panel d'experts comprenait une enseignante au primaire, une conseillère pédagogique, la responsable du service de mesure et d'évaluation d'une firme de *testing*, une experte en mesure et évaluation et la chercheuse.

#### Déroutement

L'élaboration des rapports s'est déroulée pendant quatre rencontres en groupe. Lors de la première rencontre, la chercheuse a présenté les sept profils les plus fréquents issus de l'étude réalisée avec les données du PIRLS 2011 (DUONG THI, 2018). Les experts ont discuté pour identifier les profils qu'ils jugeaient les plus représentatifs d'élèves du primaire au Québec afin, dans un 2<sup>e</sup> temps, de choisir celui pour lequel serait développé le rapport. Le profil jugé typique renvoie à un élève qui maîtrise bien les habiletés *Repérer des informations explicites* et *Faire des inférences*, mais moins *La compréhension globale*, *Interpréter* et *Vocabulaire et Syntaxe*.

La deuxième rencontre visait à collecter des propositions sur le contenu et sur la forme des rapports à élaborer selon le protocole du design issu de la revue de littérature et en examinant des exemples de rapports. Les experts ont finalement décidé d'élaborer trois formats de rapports pour le profil type retenu, afin de pouvoir expérimenter plusieurs propositions.

La chercheuse s'est ensuite basée sur ces propositions pour développer les premières versions de rapports selon trois formats différents. La troisième rencontre a permis de discuter et commenter ces premières versions. Une fiche d'évaluation portant sur les différentes parties du rapport a permis de guider les discussions et de recueillir les commentaires. Cette fiche contient une liste de vérifications qui touchent tant la

forme que le fond, élaborée à partir des critères d'un bon rapport, tirés de la revue de littérature. La chercheuse a ainsi pu faire évoluer les trois formats de rapport qui ont été discutés et finalisés lors de la quatrième rencontre. Le questionnaire d'évaluation à l'attention des enseignants utilisé dans la phase 2 a été ébauché durant la quatrième rencontre sur la base des éléments mis en évidence par la revue de littérature (voir le tableau 2).

### Résultats: trois formats de rapports diagnostiques

Les experts ont établi que les rapports doivent être simples, clairs et interprétables pour faciliter la compréhension des enseignants. Le nombre de pages doit se limiter à deux afin de favoriser l'impression recto verso ou la consultation en ligne. Ainsi, la première page contient: 1) l'identification de l'élève; 2) les directives pour lire le rapport; 3) le profil de maîtrise des habiletés et 4) une description du profil de l'élève et des pistes d'intervention. La deuxième page présente la partie 5), soit le guide d'interprétation. Le contenu des rapports est semblable dans les trois formats (les parties 1, 2, 4). Seules les parties 3) le profil de maîtrise des habiletés (graphiques) et 5) le guide d'interprétation diffèrent.

#### *L'identification de l'élève*

Pour l'identification de l'élève, les experts ont considéré qu'il est peu important d'indiquer le nom de l'enseignant. Il suffit d'inscrire le code du groupe et de l'école. Par contre, il est utile de donner la date de naissance de l'élève, qui peut nous renseigner sur son niveau de retard. Avec ces propositions des experts, la première partie contient ces informations permettant de personnaliser le rapport de l'élève: nom et prénom, date de naissance, genre, niveau, groupe et école.

#### Les directives pour lire le rapport

Les directives pour lire les rapports doivent contenir des informations sur l'objectif du rapport, le nombre d'habiletés évaluées et le lien avec la deuxième page du rapport. La description des objectifs du rapport est déplacée dans le guide d'interprétation afin d'éviter la surcharge des informations sur la première page.

## Le profil de maîtrise des habiletés

Les experts ont insisté sur l'importance d'intégrer le profil de l'élève dans celui du groupe pour fournir un portrait global du groupe. Afin d'assurer la simplicité et l'interprétabilité des formats de présentation, nous avons choisi le diagramme en barres pour présenter les profils de l'élève et du groupe. Les habiletés sont classées en ordre de difficulté et présentées en différentes couleurs selon le niveau de maîtrise. L'utilisation des étoiles ou des flèches est recommandée pour mieux montrer le positionnement de l'élève par rapport au groupe.

Dans le format A (figure 2), le profil de l'élève et celui du groupe sont présentés dans deux graphiques séparés. Le graphique à gauche présente le profil de l'élève sur les cinq habiletés. La ligne verticale en bleu représente le point de coupure de 0,5 qui permet de juger si l'élève maîtrise ou non une habileté. Ce seuil a été suggéré dans de nombreuses recherches sur le sujet, dont celle de Li (2011), de Lee et Sawaki (2009) et de De la Torre (2009). Le graphique à droite indique les pourcentages de maîtrise des habiletés du groupe. Les barres sont distinguées en trois couleurs selon le niveau de maîtrise des habiletés: rouge pour les niveaux de 0 à 0,30, jaune pour les niveaux de 0,31 à 0,70 et vert pour les niveaux de 0,71 et 1,0. Les experts ont jugé important d'ajouter un groupe intermédiaire (0,31 et 0,70) au lieu d'avoir seulement deux groupes symétriques, afin de mieux classifier les élèves.

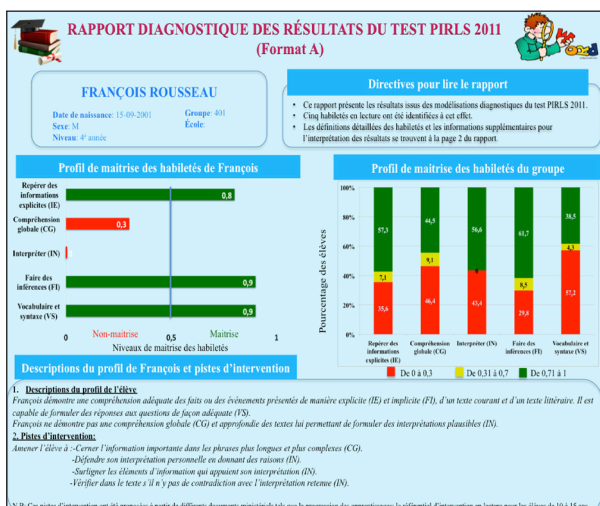
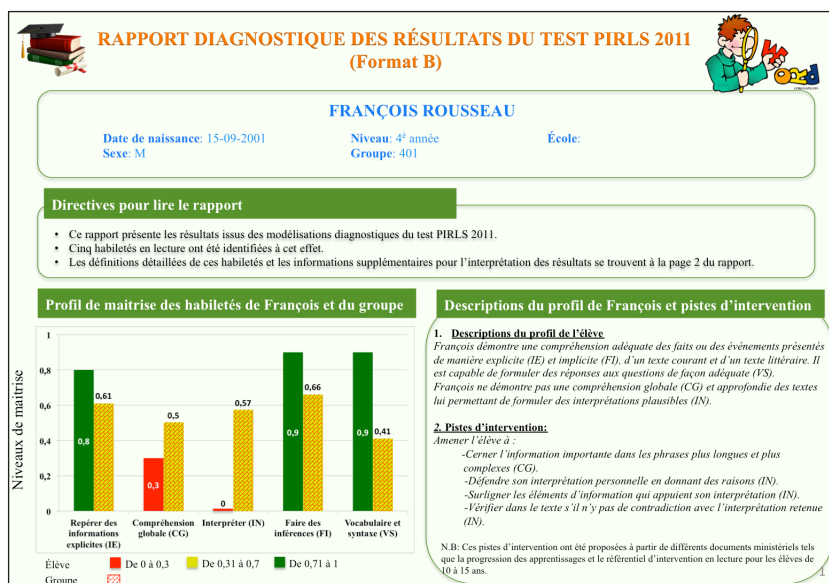


Figure 2 - Format A du rapport  
 Source: Duong Thi (2018, p. 161).

Dans le format B (figure 3), le profil de l'élève et celui du groupe sont présentés dans le même graphique avec trois couleurs selon le niveau de maîtrise. Les chiffres figurant sur les barres correspondent aux niveaux de maîtrise de l'élève et du groupe. Par exemple, pour «*Compréhension globale*», le niveau de maîtrise de l'élève est 0,3, donc 30% tandis que celui du groupe est 0,5, donc à 50%. L'élève a donc un niveau de maîtrise plus faible que celui du groupe.

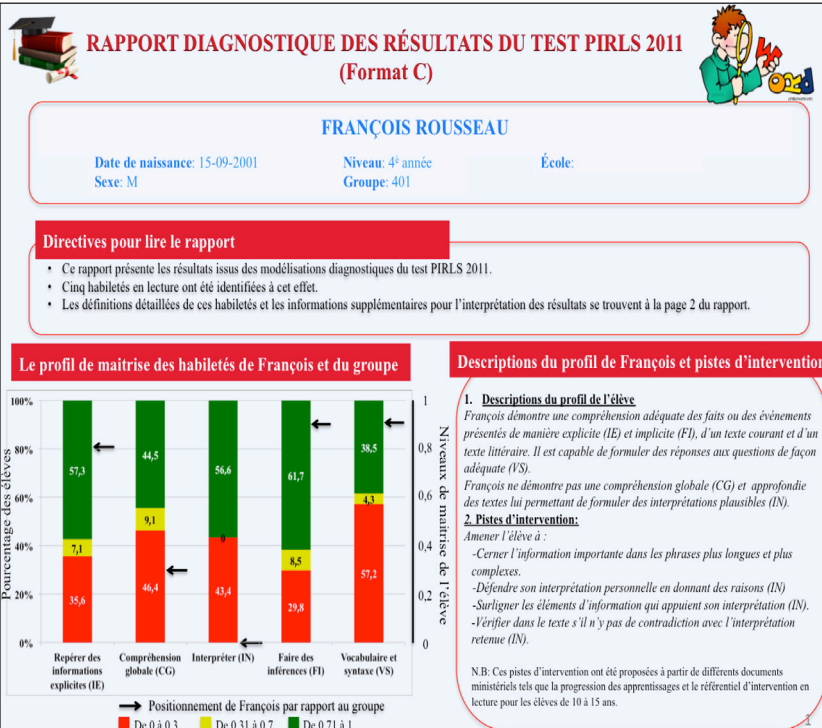
Figure 3 - Format B du rapport



Source: Duong Thi (2018, p. 162).

Finalement, dans le format C (figure 4), le profil de l'élève et celui du groupe sont présentés dans le même graphique et les niveaux de maîtrise se distinguent par trois couleurs. Le diagramme en barres représente les pourcentages des élèves dans trois groupes tandis que les flèches indiquent le positionnement de l'élève par rapport au groupe. Par exemple, pour «*Repérer des informations explicites*», l'élève a un niveau de maîtrise de 0,8 et fait donc partie du groupe vert qui regroupe 57,3% d'élèves.

Figure 4 - Format C du rapport



Source: Duong Thi (2018, p. 163).

## Une Description du Profil de L'élève et des Pistes D'intervention

Cette partie doit tracer le portrait individuel de l'élève, avec les habiletés maîtrisées et non maîtrisées, ses forces, des aspects à améliorer et des pistes d'intervention. Celles-ci doivent être détaillées et rédigées dans un langage accessible aux enseignants. Avec leurs expériences en salle de classe et en recourant aux documents ministériels comme *La Progression des apprentissages* et le *Référentiel d'intervention en lecture des élèves de 10 à 15 ans*, l'enseignante au primaire et la conseillère pédagogique ont reformulé quatre pistes d'intervention qui proposent des activités concrètes que l'élève doit réaliser pour améliorer les habiletés non maîtrisées.

## Le guide d'interprétation

Chaque format de rapport est accompagné d'un guide d'interprétation (figure 5). Selon les experts, il est important de préciser les objectifs du rapport et de donner des descriptions détaillées des habiletés identifiées. Cependant, il est peu utile de fournir des exemples de questions, car les enseignants n'ont pas accès aux épreuves. Les informations sur l'ADC sont limitées pour éviter la surcharge d'informations. Ainsi, le guide d'interprétation se compose de trois parties: a) les objectifs du rapport diagnostique; b) la description des habiletés en lecture qui fournit des définitions détaillées sur les cinq habiletés reformulées par le panel d'experts (voir le tableau 1) et c) le mode d'emploi pour lire le graphique qui contient des informations brèves sur les modélisations du PIRLS 2011, l'explication du terme *niveau de maîtrise* et les directives pour l'interprétation du graphique avec un exemple concret.

**GUIDE D'INTERPRÉTATION DES RÉSULTATS**

**Objectifs du rapport diagnostique**

Ce rapport vise à :

- dresser un portrait des forces et des points à améliorer de l'élève sur les cinq habiletés en lecture identifiées pour le test PIRLS 2011 et
- proposer des pistes d'intervention.

**Descriptions des habiletés en lecture**

**Repérer des informations explicites (IE):** Extraire et reconnaître des informations explicites exprimées dans le texte pour répondre aux questions.

**Compréhension globale (CG):** Former une compréhension globale d'un paragraphe ou de l'ensemble du texte.

**Interpréter (IN):** Acquérir une compréhension plus approfondie du texte en combinant les connaissances antérieures et les informations présentées dans le texte. Les liens à établir ne sont pas seulement implicites; ils peuvent également être ouverts à l'interprétation de la lectrice ou du lecteur (CMEC, 2012).

**Faire des inférences simples (FI):** Déduire des informations justes à partir des indices du texte.

**Vocabulaire et syntaxe (VS):** Exprimer des idées dans une grammaire correcte et compréhensible de l'anglais écrit.

**Mode d'emploi pour lire le graphique**

- Les résultats présentés dans ce rapport ont été obtenus à partir des analyses avec des modèles de classification diagnostique.
- Le résultat de chaque habileté est sous-forme des niveaux de maîtrise qui varie entre 0 et 1. À titre d'exemple, un élève qui a obtenu le résultat de 0,7 pour l'habileté « **Faire des inférences** » a 70% de chance de maîtriser cette habileté.

**3. Pour le profil de maîtrise des habiletés:**

- Les niveaux de maîtrise des habiletés se distinguent en trois couleurs: ■ pour les élèves ayant les niveaux de 0 à 0,3; ■ pour les élèves ayant les niveaux de 0,31 à 0,7; ■ pour les élèves ayant les niveaux de 0,71 à 1.
- Les nombres qui figurent sur les barres représentent les pourcentages des élèves. Par exemple, pour l'habileté « **Compréhension globale** »:
  - 46,4 % des élèves ont un niveau de maîtrise de 0 à 0,3;
  - 9,1 % des élèves ont un niveau de 0,31 à 0,7;
  - 44,5 % des élèves ont un niveau de maîtrise de 0,71 à 1.
- Les flèches correspondent au positionnement de l'élève par rapport au groupe. À titre d'exemple, pour l'habileté « **Repérer des informations explicites** », l'élève obtient un niveau de maîtrise de 0,8 et fait donc partie du groupe ■.

2

Figure 5 - Guide d'interprétation du format C

Source - DUONG THI (2018, p. 164).



## Phase 2 - Évaluation des Rapports Diagnostiques Auprès des Enseignants Au Primaire, des Conseillers Pédagogiques et des Orthopédagogues

### **Instrument**

L'évaluation des rapports a été faite à l'aide d'un questionnaire conçu en s'inspirant des questions de la recherche de Zapata-Rivera, Vezzu et Vanwinkle (2013) et de Vezzu, Vanwinkle et Zapata-Rivera (2012) et de quatre critères d'un bon rapport dont les exemples se trouvent dans le tableau 2. Le questionnaire contient 55 questions portant sur: 1) la préférence des rapports (quatre QCM avec trois choix de réponse); 2) l'évaluation de la qualité des rapports (32 questions de type Likert de quatre points selon le degré d'accord ou désaccord); 3) la compréhension des rapports (neuf QCM de trois choix de réponse avec une seule réponse correcte) et 4) les informations contextuelles (neuf questions) portant sur le poste occupé, l'ancienneté, la formation, etc. Ce questionnaire a été prévalidé par deux conseillers pédagogiques et un expert en mesure et évaluation pour vérifier la compréhension des items avant d'être traité sur *Survey Monkey*. Pour les questions portant sur la préférence, trois formats de rapports ont été affichés pour que les participants choisissent un format préféré. Quant à l'évaluation de la qualité et de la compréhension des rapports, chaque participant évalue un seul format de rapport attribué aléatoirement par *Survey Monkey*. Cela permet d'assurer une répartition équitable des participants pour les trois formats de rapports.

### **Participants**

L'échantillon se compose de 98 participants, dont 67 enseignants au primaire (68,4%), 15 orthopédagogues (15,3%) et 16 conseillers pédagogiques (16,3%). La répartition est équitable entre la tranche d'âge de 20 à 39 ans (49%) et celle des 40 ans et plus (51%). Quant à l'ancienneté, 31,6% des participants avaient moins de 10 ans d'expérience; 52% avaient de 11 à 20 ans d'expérience et 16,3% des participants avaient plus de 20 ans d'expérience. Près de la moitié des participants ont obtenu un baccalauréat (48%) tandis que 23,5% ont fait un microprogramme ou un Diplôme d'études supérieures spécialisées (DESS) et 28,6% ont obtenu une maîtrise ou un doctorat. Quant aux domaines de formation, 57,1% des participants sont formés en enseignement au primaire ou préscolaire,

24,5% en adaptation scolaire et 18,4% en psychopédagogie et d'autres domaines.

## **Analyse des données**

Les données ont été analysées avec SPSS. Des analyses descriptives ont permis de déterminer le format de préférence. Des tests Chi 2 ont été utilisés pour étudier les liens entre la préférence des participants et les variables contextuelles. Des analyses factorielles exploratoires pour les 32 items sur l'évaluation de la qualité ont permis d'identifier les dimensions du questionnaire. L'ANOVA à un facteur de classification et les test T de Student ont permis d'analyser les liens entre les perceptions des participants sur la qualité des rapports et certaines variables contextuelles.

## **Résultats**

### La préférence des formats de rapports

La tableau 3 présente la préférence des participants sur les rapports élaborés. Pour la question 1, la moitié des participants a choisi le format B, ensuite le format C (30,6%) et finalement le format A (19,4%). Dans la question 2 sur le format qui permet de comprendre le mieux les niveaux de maîtrise des habiletés de l'élève, le format B est le plus choisi (52%), le format C arrive au 2<sup>e</sup> rang (29,6%), suivi par le format A (18,4%).

La question 3 porte sur le format qui favorise la compréhension du positionnement de l'élève par rapport au groupe. Ici encore, le format B est le plus choisi (46,9%), le format C vient au deuxième rang (38,8%) et le format A est choisi par 14,3% des participants. Selon nous, les participants préfèrent les formats B et C, car le profil de l'élève et celui du groupe sont présentés sur le même graphique, ce qui facilite le positionnement de l'élève par rapport au groupe.

La question 4 porte sur le format de rapport que, selon les répondants, les enseignants vont préférer. C'est, là encore, le format B qui est le plus choisi, par 51% des participants, suivi du format C (29,6%) et finalement du format A (19,4%). Les participants ont tendance à choisir pour les enseignants le même format que celui qu'ils préfèrent à titre individuel, ce qui explique la répartition semblable des participants dans les questions 1 et 4. Nous concluons que le format B est celui que préfèrent les participants, suivi du format C et finalement du format A.

**Tableau 3** - Préférence des participants sur les trois formats de rapports

Question	Format A	Format B	Format C
1-Parmi les trois formats de rapports présentés, lequel préférez-vous?	19,4 %	50,0 %	30,6 %
2-Lequel, parmi ces trois formats, vous permet de comprendre le mieux les niveaux de maîtrise des habiletés de l'élève?	18,4 %	52,0%	29,6 %
3-Lequel, parmi ces trois formats, vous permet de comprendre le positionnement de l'élève par rapport au groupe?	14,3 %	46,9 %	38,8 %
4-Lequel, parmi ces trois formats, pensez-vous que les enseignants vont préférer?	19,4 %	51,0 %	29,6 %

Source: Duong Thi (2018, p. 171).

#### Lien entre la préférence et des variables contextuelles

Les résultats descriptifs montrent que la préférence pour un format est la même chez les enseignants, les orthopédagogues et les conseillers pédagogiques. Autrement dit, les participants choisissent majoritairement le format B, puis le format C et puis le format A. Les résultats des tests de Chi 2 confirment que la préférence des rapports des participants est constante peu importe les postes occupés.

Ces mêmes résultats ont été constatés pour le nombre d'années d'expérience, le diplôme obtenu et le suivi ou non des cours en méthodes quantitatives au cours de cinq dernières années. Pour des tranches d'âges, le test de Chi2 n'est significatif que pour la question 2 ( $\text{Chi}2 = 6,024$ ;  $p = 0,049$ ), ce qui confirme un lien de dépendance entre les tranches d'âge des participants et le format de rapport préféré. Les participants de 20 à 39 ans sont plus nombreux à choisir le format B tandis que ceux dans la tranche d'âge de 40 ans et plus choisissent davantage les formats A et C. Cependant, la probabilité associée au test (0,049) est très proche du seuil de rejet de 0,05.

Quant aux domaines de formation, les tendances sont les mêmes pour les questions 1 et 4: les participants formés en enseignement au primaire, au préscolaire et en adaptation ont les mêmes préférences. Ils sont les plus nombreux à choisir le format B, puis le format C et enfin

le format A. Les participants formés en psychopédagogie, choisissent le format A en premier lieu, puis le format B et enfin le format C. Les tests de Chi 2 qui sont significatifs pour les questions 1 ( $\text{Chi}2 = 13,663$ ;  $p = 0,008$ ) et 4 ( $\text{Chi}2 = 14,284$ ;  $p = 0,006$ ) confirment que le domaine de formation joue un rôle sur la préférence des rapports. Les participants en psychopédagogie et dans d'autres domaines préfèrent que le profil de l'élève et celui du groupe soient présentés dans deux graphiques séparés permettant de mieux voir le portrait individuel de l'élève.

**Tableau 4** - Lien entre la préférence et des variables contextuelles

Informations contextuelles	Résultats du Chi 2	Questions			
		Q1	Q2	Q3	Q4
Poste occupé	Chi 2	0,604	0,312	0,466	1,121
	p	0,739	0,856	0,792	0,571
Tranche d'âge	Chi 2	3,050	6,024	1,163	4,13
	p	0,218	0,049* <sup>1</sup>	0,559	0,127
Expériences	Chi 2	2,019	4,118	1,591	2,557
	p	0,732	0,390	0,810	0,634
Diplôme obtenu	Chi 2	9,289	5,005	5,303	8,506
	p	0,054	0,287	0,258	0,075
Domaine de formation	Chi 2	13,663	7,980	7,606	14,284
	p	0,008*	0,092	0,107	0,006*
Suivi des cours méthodes quantitatives	Chi 2	2,064	0,723	1,677	1,773
	p	0,356	0,697	0,432	0,412

Source: Conception des auteures.

Étude de la fiabilité et de la dimensionnalité des questions sur l'évaluation de la qualité des rapports

Nous avons fait des analyses factorielles exploratoires pour les 32 questions de l'évaluation de la qualité des rapports. Bien qu'une taille de l'échantillon de 100 est recommandée comme seuil minimal pour les analyses factorielles (FIELD, 2009), les recherches empiriques montrent

qu'il est possible de faire des analyses avec des échantillons plus petits lorsque le nombre de facteurs est limité à une ou deux dimensions et que ces facteurs sont bien définis (LOYE et BARROSO DA COSTA, 2013). Nous avons effectué des analyses avec la méthode d'extraction par factorisation en axes principaux, la rotation orthogonale de Varimax et une saturation plus élevée que 0,30.

Les résultats suggèrent la suppression de deux items dont l'indice de saturation est plus petit que 0,30. La solution à un facteur avec 30 items explique 38,060 % de la variance totale. Le test de sphéricité de Barlett ( $\chi^2(435) = 1511,181$ ,  $p = 0,000$ ) ainsi que la mesure Kaiser-Meyer-Olkin (KMO = 0,760) ont indiqué que la solution factorielle est acceptable. L'analyse de la fiabilité pour ces 30 items suggère un Alpha de Cronbach de 0,943, ce qui est considéré comme excellent. Ainsi, nous avons décidé de retenir ces 30 items pour former le facteur *Évaluation de la qualité globale des rapports*.

Lien entre l'évaluation de la qualité des rapports et des variables contextuelles

Les analyses préliminaires ont été faites pour vérifier les postulats du Test T de Student et de l'ANOVA à un facteur. Les résultats des tests Kolmogorov-Smirnov suggèrent que le postulat sur la normalité est respecté.

Les analyses descriptives montrent que les participants ont des perceptions très positives des rapports évalués. Ainsi, les participants donnent l'évaluation la plus positive au format B ( $m = 3,13$ ;  $s = 0,57$ ), puis au format C ( $m = 3,11$ ;  $s = 0,36$ ) et enfin au format A ( $m = 3,07$ ;  $s = 0,37$ ). Les conseillers pédagogiques sont les plus positifs ( $m = 3,24$ ;  $s = 0,43$ ), suivis des orthopédagogues ( $m = 3,08$ ;  $s = 0,29$ ) et des enseignants au primaire ( $m = 3,07$ ;  $s = 0,46$ ). Cependant, les résultats des ANOVA à un facteur confirment que les perceptions sur la qualité des rapports sont les mêmes, peu importe le format évalué ( $F_{(2)} = 0,021$ ;  $p = 0,921$ ) et les postes occupés ( $F_{(2)} = 0,716$ ;  $p = 0,493$ ).

Les participants ayant moins de 10 ans d'expérience ont les perceptions les plus positives ( $m = 3,15$ ;  $s = 0,35$ ) comparativement aux participants les plus expérimentés ( $m = 3,11$  et  $s = 0,48$ ) et ceux ayant de 11 à 20 ans d'expérience ( $m = 3,08$ ;  $s = 0,46$ ). Toutefois, les résultats de l'ANOVA à un facteur montrent que l'ancienneté n'influence pas leurs perceptions sur la qualité des rapports ( $F_{(2)} = 0,319$ ;  $p = 0,870$ ).

Les participants titulaires d'une maîtrise ou d'un doctorat ont les perceptions les plus positives des rapports évalués ( $m = 3,13$ ;  $s = 0,49$ ). Ils sont suivis par ceux qui ont un baccalauréat ( $m = 3,11$ ;  $s = 0,44$ ) et enfin par ceux qui ont suivi un microprogramme ou un diplôme d'études supérieures spécialisées (DESS) ( $m = 3,07$ ;  $s = 0,34$ ). Quant aux domaines de formation, les participants formés en psychopédagogie et dans d'autres domaines ( $m = 3,23$ ;  $s = 0,45$ ) semblent plus positifs que leurs collègues formés en enseignement au primaire ( $m = 3,11$ ;  $s = 0,49$ ) et en adaptation scolaire ( $m = 3,00$ ;  $s = 0,27$ ). Malgré les différentes tendances observées, rien de significatif ne ressort. Les perceptions des participants sur la qualité des rapports ne varient donc pas selon le diplôme obtenu ( $F_{(2)} = 0,913$ ;  $p = 0,913$ ) ni le domaine de formation ( $F_{(2)} = 1,150$ ;  $p = 0,323$ ).

Quant aux tranches d'âge, les participants ayant de 20 à 39 ans sont plus positifs ( $m = 3,14$ ;  $s = 0,34$ ) que leurs collègues plus âgés ( $m = 3,08$ ;  $s = 0,50$ ). Les participants ayant suivi des cours en méthodes quantitatives ont des perceptions plus positives ( $m = 3,31$ ;  $s = 0,38$ ) sur la qualité des rapports que leurs collègues qui n'en ont pas suivis ( $m = 3,06$ ;  $s = 0,43$ ). Toutefois, les résultats des tests T indiquent que les tranches d'âges et le fait d'avoir suivi ou non des cours en méthodes quantitatives n'influencent pas leurs perceptions sur la qualité des rapports.

En résumé, les participants ont des perceptions très positives sur la qualité des rapports. Malgré les différentes tendances observées, leurs perceptions sont constantes peu importe le format de rapport évalué, les postes occupés, la tranche d'âge, l'ancienneté, le diplôme obtenu, les domaines de formation et le fait d'avoir suivi ou non des cours en méthodes quantitatives. Nous donnerons plus d'explications sur ces résultats dans la discussion. Les résultats de l'évaluation de la compréhension des rapports n'ont pas été présentés dans cet article à cause de la mauvaise qualité psychométrique des questions, ce qui ne nous permet pas d'aller plus loin avec des analyses inférentielles.

**Tableau 5** - Lien entre l'évaluation de la qualité des rapports et les variables contextuelles

	Évaluation de la qualité des rapports	N	Moyenne	Écart-type	Résultats de l'ANOVA ou du test T
Format de rapports	A	19	3,07	0,37	F(2) = 0,021 p = 0,921
	B	21	3,13	0,57	
	C	28	3,11	0,36	
Poste occupé	Enseignants au primaire	46	3,07	0,46	F(2) = 0,716 p = 0,493
	Orthopédagogues	10	3,08	0,29	
	Conseillers pédagogiques	12	3,24	0,43	
Expérience	Moins de 10 ans	20	3,15	0,35	F(2) = 0,319 p = 0,870
	11-20 ans	35	3,08	0,46	
	20 ans et plus	13	3,11	0,48	
Diplôme	Baccalauréat	30	3,11	0,44	F(2) = 0,913 p = 0,913
	Microprogramme ou DESS	16	3,07	0,34	
	Maitrise ou doctorat	22	3,13	0,49	
Domaine de formation	Enseignement au primaire	34	3,11	0,49	F(2) = 1,150 p = 0,323
	Enseignement en adaptation	20	3,00	0,27	
	Psychopédagogie et autre	14	3,23	0,45	
Tranche d'âge	20-39 ans	30	3,14	0,34	t(66)=0,541
	40 ans et plus	38	3,08	0,50	P = 0,591
Suivi des cours méthodes quantitatives	Oui	13	3,31	0,38	t(66) = 1,924
	Non	55	3,06	0,43	P = 0,059

Source: Conception des auteures.

## Discussion

Cet article décrit les démarches de l'élaboration des rapports diagnostiques avec un panel d'experts et les résultats de l'évaluation des rapports auprès d'enseignants du primaire, de conseillers pédagogiques et d'orthopédagogues en vue de les valider. En comparaison avec des rapports conçus dans d'autres études, nos rapports présentent une amélioration importante sur l'aspect visuel et sur le contenu. Nous avons utilisé trois types de graphiques différents pour présenter le profil de l'élève avec trois couleurs selon les niveaux de maîtrise des habiletés. En effet, le recours aux redondances des indices visuels avec les différentes

couleurs et le regroupement des élèves selon leurs niveaux de maîtrise accentuent les informations présentées, ce qui a été appuyé par Cleveland et McGill (1985) et Kosslyn (1994). Nous présentons les résultats dans un graphique plutôt que dans un tableau, ce qui permet de mieux communiquer les changements et les tendances des données (SHAH, MAYER et HEGARTY, 1999).

Au niveau du contenu, nous n'incluons que des informations personnelles de l'élève, ce qui renforce le critère de l'accessibilité et la personnalisation des rapports. Les directives pour lire les rapports ont été présentées brièvement pour éviter la surcharge d'informations et laisser de l'espace à la présentation des résultats. Dans la recherche de Jang (2005), ces informations sont jumelées avec le mode d'emploi pour l'interprétation des graphiques. La présentation des directives dans la première partie constitue une amélioration de notre recherche, ce qui oriente mieux les enseignants lors de l'interaction avec ces rapports.

La description du profil de l'élève et des pistes d'intervention sont rédigées avec soin dans un langage accessible aux enseignants, en se basant sur différents documents ministériels. Cela permet d'assurer la cohérence entre les pistes d'intervention proposées, le contenu de formation et la compréhension des rapports des enseignants, ce qui diffère des rapports élaborés par Roberts et Gierl (2010) et Jang (2005). En effet, les pistes d'intervention proposées par ces auteurs sont très sommaires. Roberts et Gierl (2010) ont seulement résumé les forces et les faiblesses de l'élève sans pistes d'amélioration tandis que ces éléments sont absents dans les rapports de Jang (2005).

Dans le guide d'interprétation, nous avons décidé de limiter les informations sur les modélisations diagnostiques pour faciliter la compréhension des enseignants. Nous avons ajouté la description détaillée des objectifs du rapport et le mode d'emploi pour lire les rapports avec des exemples concrets d'interprétation des résultats. À notre avis, ces exemples rendent le guide d'interprétation plus utile pour les enseignants et assurent un équilibre entre les éléments textuels et non textuels.

L'évaluation des rapports est réalisée à l'aide d'un questionnaire administré en ligne. Une valeur ajoutée de notre étude est que nous avons fait des analyses factorielles pour définir la structure de 32 questions sur l'évaluation de la qualité des rapports, ce que les recherches précédentes n'ont pas fait. Les résultats suggèrent un regroupement de 30 items pour l'évaluation de la qualité globale des rapports. La suppression de deux items ne change pas la qualité du questionnaire ni la quantité d'informa-



tions obtenues. De plus, le coefficient Alpha de Cronbach qui dépasse le seuil minimal requis (0,7) assure la fidélité de l'instrument.

Pour la préférence, les résultats descriptifs révèlent que les participants préfèrent le format B dans les quatre questions, suivi du format C et finalement, le format A. La préférence la plus grande pour le format B s'explique par la familiarité du type de diagramme présenté, ce qui a été souligné dans la recherche de Zapata-Rivera, Vezzu et Vanwinkle (2013). La préférence plus grande pour les formats B et C se justifie par le fait que le profil de l'élève et celui du groupe sont présentés sur le même graphique plutôt que sur deux graphiques (format A). Cette idée s'appuie sur le fait que les enseignants ont des hésitations dans l'interprétation des graphiques (ZAPATA-RIVERA, VEZZU, VANWINKLE, 2013), ce qui les pousse à choisir un format de rapport avec un type de graphique plus simple (format B) ou qui contient peu de graphiques (B et C).

Par ailleurs, l'utilisation de deux axes différents pour les niveaux de maîtrise des habiletés et les pourcentages de maîtrise du groupe (format C) peut dérouter les participants, ce qui explique qu'ils choisissent moins ce format que le format B. Le format A contient deux graphiques qui ne fournissent pas nécessairement les mêmes types d'information, car l'un présente le profil de maîtrise des habiletés de l'élève tandis que l'autre fournit les pourcentages des groupes selon leur niveau de maîtrise des habiletés. La présentation des résultats de l'élève et de ceux du groupe dans deux graphiques séparés rend plus difficile le positionnement de l'élève par rapport au groupe, ce qui explique le pourcentage le plus faible dans la question 3 pour le format A.

La préférence des participants ne varie pas selon les postes occupés, l'ancienneté, le diplôme obtenu et le suivi ou non des cours en méthodes quantitatives. Ces résultats corroborent en grande partie ceux de la recherche de Zapata-Rivera, Vezzu et Vanwinkle, (2013) qui montrent qu'il n'y a pas de lien entre la préférence des rapports et l'ancienneté, le nombre de cours en statistiques suivis, le degré d'aisance avec les ordinateurs et la formation professionnelle en mesure. Cependant, la préférence des participants varie selon la familiarité avec les termes statistiques. Ainsi, les participants les plus familiers avec ces concepts préfèrent les formats de présentation plus abstraits, comme les courbes, tandis que ceux dont la familiarité avec les termes statistiques est plus faible choisissent des graphiques concrets, comme le diagramme avec des icônes empilées.

Par ailleurs, notre recherche montre que la préférence des participants varie selon leur domaine de formation pour les questions 1

et 4. Ces résultats se justifient par le fait que la question 1 porte sur la préférence personnelle des rapports tandis que la question 4 porte sur le format que le participant croit que les enseignants préféreront. Donc, les participants ont généralement tendance à choisir le même format dans ces deux questions. C'est une manière de renforcer la croyance sur leur propre préférence.

Autrement dit, si le participant préfère le format A, il croit que les enseignants auront le même choix. Une des explications possibles peut être liée aux caractéristiques du domaine de formation. Par exemple, les participants formés en psychopédagogie s'intéressent surtout au processus d'apprentissage de chaque individu. Ils préfèrent donc que le profil de l'élève soit séparé de celui du groupe pour mieux l'interpréter. Toutefois, nous n'avons pas trouvé de recherche empirique qui appuie cette idée. Une autre explication est que ces deux questions pourraient être redondantes. Il serait donc souhaitable d'ajouter une question ouverte pour mieux comprendre la justification des participants.

En général, les participants ont des perceptions très positives sur la qualité des rapports. Les résultats suggèrent que l'évaluation de la qualité la plus positive est donnée au format B, puis au format C et enfin au format A. Bien qu'il existe des tendances différentes dans l'évaluation de la qualité des rapports, les résultats des tests inférentiels indiquent que les perceptions des participants sur la qualité des rapports sont constantes, peu importe les variables contextuelles. Ces résultats corroborent en partie ceux des recherches de Zwick, Zapata-Rivera et Hegarty (2014) et de Zapata-Rivera, Vanwinkle et Zwick (2012). À la différence de notre recherche, celle de Zapata-Rivera, Vezzu et Vanwinkle (2013) montre que les perceptions des participants varient selon les trois formats de rapports développés. Cependant, les formats de présentations de graphiques sont si différents (courbe, boîte à moustache et icônes empilées) qu'ils peuvent générer différents niveaux d'interprétation. Dans notre cas, trois formats de présentation représentent des similitudes, ce qui explique que les participants ont des perceptions constantes sur les trois formats.

## **Conclusion**

Notre étude vise à répondre aux grands besoins des enseignants de recevoir des informations détaillées à partir des évaluations à grande

échelle à travers des rapports diagnostiques. Nous avons conçu trois formats de rapports fiables, interprétables et accessibles avec un panel d'experts provenant de différents milieux éducatifs, ce qui a assuré la variété des commentaires lors de l'élaboration et de l'évaluation des rapports. Ces rapports répondent bien aux différents critères d'un bon rapport, suggérés dans la revue de la littérature. De plus, nous avons réussi à développer et à valider un questionnaire sur l'évaluation des rapports, ce qui n'existait pas dans les recherches précédentes. Des analyses factorielles exploratoires montrent que la fiabilité et la fidélité de notre instrument sont très bien assurées. De plus, les résultats sur l'évaluation des rapports confirment qu'ils sont grandement appréciés par le public visé provenant de différents milieux éducatifs avec une préférence constante et une évaluation très positive des rapports élaborés. Ces qualités de l'instrument conçu et des rapports élaborés suggèrent donc l'excellent potentiel de leur applicabilité dans plusieurs milieux éducatifs différents. Les résultats de l'évaluation des rapports ont fait ressortir le format de rapport préféré des participants. Par ailleurs, le fait que la préférence pour le format B résiste aux différences entre les divers groupes sociodémographiques nous semble un résultat vraiment intéressant, car il fournit une balise importante pour développer des rapports en grand nombre dans un contexte semblable.

De nombreuses recherches devront encore se pencher sur l'élaboration des tests diagnostiques en lecture au primaire. Les données obtenues permettront de générer des rapports en grand nombre avec des logiciels professionnels et une importance devra être accordée à la compréhension du contenu des rapports. Finalement, les stratégies de recrutement des participants doivent prendre en considération la répartition équitable selon des groupes sociodémographiques.

Notre étude établit le pont entre les résultats du PIRLS 2011 et l'ADC grâce à des rapports diagnostiques compréhensibles et interprétables destinés aux enseignants. En effet, le lien entre les résultats de la recherche et l'inclusion scolaire a été tissé, car en identifiant des difficultés en lecture pour chaque élève et en proposant des pistes d'interventions appropriées, elle permet donc de mieux accompagner des élèves en difficulté et de faciliter leur processus d'intégration dans l'apprentissage en lecture au primaire, la compétence la plus importante à développer pour assurer leur réussite scolaire.

# THE DEVELOPMENT AND EVALUATION OF DIAGNOSTIC REPORTS ON PIRLS 2011 DATA: PERCEPTIONS OF PRIMARY TEACHERS, PEDAGOGICAL COUNSELORS AND REMEDIAL TEACHERS

**Abstract:** *The growing demand for large-scale assessments is concurrent with strong pressure to make these assessments more informative about students' learning. Indeed, current assessments represent students' grades as global scores and subscores, which say little about their strengths and weaknesses. However, research in cognitive diagnostic approaches suggests that it is possible to break down reading into knowledge and skills that can be diagnosed with the help of psychometric modelling. Large-scale trials could provide teachers with diagnostic reports including detailed feedback on student strengths and weaknesses. This article describes the development of diagnostic reports based on PIRLS 2011 data with a panel of experts, and relates the results of evaluations of these reports by primary teachers, pedagogical counsellors and remedial teachers.*

**Keywords:** *Cognitive diagnostic approach. Diagnostic report development and evaluation. Reading. Large-scale assessment.*

## Nota

1 Le test est significatif au seuil de 0.05.

## Références

CHIU, C.; SEO, M.; DOUGLAS, J. Cluster analysis for cognitive diagnosis: An application to the 2001 PIRLS reading assessment. *IERI Monograph Series: Issues and Methodologies in Large-scale Assessment*, Princeton, v. 2, p. 137-159, 2009.

CLEVELAND, W. S.; MCGILL, R. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of American Statistical Association*, Virginia, n. 79, p. 531-554, 1985.

DE LA TORRE, J. A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement*, Chicago, v. 33, n. 3, p. 163-183, 2009.

DUONG THI, D.T. *Modélisation, élaboration et évaluation des rapports à visée diagnostique des données du PIRLS 2011*. 2018. Thèse de doctorat, Faculté des sciences de l'éducation, Université de Montréal. Montréal, 2018.

ENTWISLE, D. R.; ALEXANDER, K. L.; OLSON, L. S. First grade and educational attainment by age 22: a new story. *American journal of sociology*, Chicago, v. 110, n. 5, p. 1458-1502, 2005.

FIELD, A. *Discovering Statistics Using SPSS*. London, UK: Sage publication, 2009.

FLOWER, L. *Problem-solving strategies for writing (2nd ed.)*. New York, NY: Harcourt Brace Jovanovich, 1985.

GIERL, M. J.; CUI, Y.; HUNKA, S. Using connectionist models to evaluate examinees' response patterns to achievement tests. *Journal of Modern Applied Statistical Methods*, Detroit, v. 7, n. 1, p. 234-245, 2008.

GOODMAN, D. P.; HAMBLETON, R. K. Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education*, Lincoln, v. 17, n. 2, p. 145-220, 2004.

GOUVERNEMENT DU QUÉBEC. La scolarité: un élément clé des compétences en littératie. *Institut de la statistique du Québec, Développer nos compétences en littératie: un défi porteur d'avenir. Rapport québécois de l'Enquête internationale sur l'alphabétisation et les compétences des adultes (EIACA), 2003*. Québec: Institut de la statistique, 2006.

GOUVERNEMENT DU QUÉBEC. *Les facteurs liés à la réussite aux épreuves obligatoires de français en sixième année du primaire: un tour d'horizon*. Québec: Institut de la statistique, 2012.

HAMBLETON, R. K.; SLATER, S. *Are NAEP executive summary reports understandable to policy makers and educators?* (CSE Technical Report 430). 1997. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, 1997.

HAMBLETON, R. K.; ZENISKY, A. *Improvements to student score reporting: Steps and more use of suitable methodologies*. Invited presentation at the ETS Conference on Score Reporting, Princeton, NJ, 2010.

HORTON, W. Overcoming chromophobia: A guide to the confident and appropriate use of color. *IEEE Transactions on Professional Communication*, Macon, v. 34, n. 3, p. 160-173, 1991.

IMPARA, J. C.; DIVINE, K. P.; BRUCE, F. A.; LIVERMAN, M. R.; GAY, A. Does interpretive test score information help teachers?. *Educational Measurement: Issues and Practice*, Iowa, v. 10, n. 4, p. 16-18, 1991.

JANG, E. E. *A validity narrative: effects of reading skills diagnosis on teaching and learning in the context of NG TOEFL*. 2005. Thèse (Doctorat). Faculté de psychologie, University d'Illinois, Champaign, 2005.

- KOSSLYN, S. M. *Elements of graph design*. New York, NY: Freeman, 1994.
- LEE, Y.-W.; SAWAKI, Y. Cognitive diagnosis approaches to language assessment: An overview. *Language Assessment Quarterly*, New York, v. 6, n. 3, p. 172-189, 2009.
- LEIGHTON, J.P.; GIERL, M. J. (org.), *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge, États-Unis: Presses universitaires de Cambridge, p. 119-145, 2007.
- LEIGHTON, J. P.; GIERL, M. J. Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice*, Iowa, v. 26, n. 2, p. 3-16, 2007.
- LI, H. A cognitive diagnostic analysis of the METLAB reading test. *Spaan Fellow Working Papers in Second or Foreign Language Assessment*, Michigan, n. 9, p. 17-46, 2011.
- LOYE, N.; BARROSO DA COSTA, C. Hiérarchiser les besoins de diagnostic en mathématique en FP à l'aide d'un modèle de Rasch. *Mesure et évaluation en éducation, Montréal*, v. 36, n. 2, p. 59-85, 2013.
- ROBERTS, M. R.; GIERL, M. J. Developing score reports for cognitive diagnostic assessments. *Educational Measurement: Issues and Practice*, Iowa, v. 29, n. 3, p. 25-38, 2010.
- ROMANO, E.; BABCHISHIN, L.; PAGANI, L. S.; KOHEN, D. School readiness and later achievement: Replication and extension using a nationwide Canadian survey. *Developmental Psychology*, Michigan, v. 46, n. 5, p. 995-1007, 2010.
- RYAN, J. M. Practices, issues, and trends in student test score reporting. In: DOWNING, S.M.; HALADYNA, T. M. (org.). *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates, p. 677-710, 2006.
- SHAH, P.; MAYER, R. E.; HEGARTY, M. Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph construction. *Journal of Educational Psychology*, Washington, v. 91, n. 4, p. 690-702, 1999.
- TEMPLIN, J.; HENSON, R. A. Measurement of psychological disorders using cognitive diagnosis models. *Psychological methods*, Washington, v. 11, n. 3, p. 287-305, 2006.
- TUFTE, E. R. *Visual explanations*. Cheshire, CT: Graphics Press, 1996.
- VEZZU, M.; VANWINKLE, W.; ZAPATA-RIVERA, D. *Designing and evaluating an interactive score report for students*. Princeton, NJ: ETS, 2012.
- YANG, X.; EMBRETSON, S. E. Construct validity and cognitive diagnostic assessment. In: LEIGHTON, J. P. GIERL, M. J. (org.), *Cognitive diagnostic*

*assessment for education: Theory and applications*. Cambridge, États-Unis: Presses universitaires de Cambridge, p. 119-145, 2007.

ZAPATA-RIVERA, D.; KATZ, I. R. Keeping your audience in mind: applying audience analysis to the design of interactive score reports. *Assessment in Education: Principles, Policy & Practice*, Oxford, v. 21, n. 4, p. 442-463, 2014.

ZAPATA-RIVERA, D.; VANWINKLE, W.; ZWICK, R. Applying Score Design Principles in the Design of Score Reports for CBAL™ Teachers. *In: Research Memorandum No. RM-12-20*. Princeton: Educational Testing Service, p. 1-18, 2012.

ZAPATA-RIVERA, D.; VANWINKLE, W.; ZWICK, R. Applying Score Design Principles in the Design of Score Reports for CBAL™ Teachers. *In: Research Memorandum No. RM-12-20*. Princeton: Educational Testing Service, p. 1 18, 2012.

ZAPATA-RIVERA, D.; VEZZU, M.; VANWINKLE, W. Exploring Teachers' Understanding of Graphical Representations of Group Performance. *In: Research Memorandum No. RM-12-20*. Princeton: Educational Testing Service, p. 1-29, 2013.

ZENISKY, A. L.; HAMBLETON, R. K.; SIRECI, S. G. Getting the message out: An evaluation of NAEP score reporting practices with implications for disseminating test results. *Applied Measurement in Education*, Lincoln, 2009, v. 22, n. 4, p. 359-375.

ZWICK, R.; ZAPATA-RIVERA, D.; HEGARTY, M. Comparing graphical and verbal representations of measurement error in test score reports. *Educational Assessment*, California, CHIU, C.; SEO, M.; DOUGLAS, J. Cluster analysis for cognitive diagnosis: An application to the 2001 PIRLS reading assessment. *IERI Monograph Series: Issues and Methodologies in Large-scale Assessment*, Princeton, v. 2, p. 137 159, 2009.