

ÉNAP

ÉCOLE NATIONALE
D'ADMINISTRATION
PUBLIQUE

enap.ca

**Étude de cas sur le chantier d'intelligence numérique (CIN) :
Les enjeux et les défis de la valorisation des données
en éducation**

Par Christian Boudreau (Ph. D.)

Avril 2023

Table des matières

Introduction.....	1
L'origine du CIN	2
2. La GRICS comme partenaire incontournable	4
3. Un repositionnement stratégique de la GRICS.....	5
4. L'organisation du projet et les activités de soutien	7
5. Espace Données comme plateforme de valorisation du CIN	9
6. Une culture de données à instaurer et des pratiques de collaboration à pérenniser	12
7. Des données, des pratiques et des systèmes à harmoniser	14
8. Des mesures et règles d'accès aux données à moderniser	16
9. Des enjeux professionnels et éthiques à considérer	19
Conclusion.....	22
Annexes	24

Introduction

Depuis sa création, le ministère de l'Éducation (MEQ) a accès à des données permettant de produire des statistiques officielles (par exemple, taux de diplomation, taux de réussite aux épreuves ministérielles, taux de sortie sans diplôme et dépenses globales) pour assurer le suivi des activités éducatives à l'échelle nationale et pour élaborer ou ajuster ses programmes scolaires. Toutefois, la pandémie de la COVID-19 a montré que le MEQ ne disposait pas de données suffisantes pour bien mesurer les effets de cette pandémie sur le réseau scolaire (par exemple, la fermeture répétée de classes et l'isolement des élèves) et, plus particulièrement, pour « *avoir un portrait complet des retards d'apprentissage afin de mettre en place les mesures de rattrapage et de suivi appropriées* »¹. Quant aux démarches du MEQ pour accéder à de nouvelles données sur la situation qui prévalait dans le réseau en contexte pandémique, elles se sont révélées exigeantes et lentes : « *À chaque fois, c'était un parcours du combattant... et on devait collecter manuellement, à coups de sondages, de questionnaires, de reddition de comptes* »². De plus, ces démarches ont conduit à des collectes d'information le plus souvent incomplètes. Cela a été le cas lorsque le gouvernement a voulu documenter, à l'automne 2022, la pénurie d'enseignants liée à la pandémie. Le bilan hebdomadaire que le MEQ diffusait sur le Web était en effet incomplet du fait que certains centres de services et certaines commissions scolaires (ci-après CSS/CS) n'ont pas répondu au questionnaire du ministère, le plus souvent par manque de moyens ou de temps.

Cette étude de cas porte sur un projet mis en œuvre dans le réseau de l'éducation appelé le *Chantier d'intelligence numérique* (CIN), qui vise à offrir au MEQ et aux acteurs du milieu, en particulier aux centres de service scolaires, aux commissions scolaires et aux écoles, un accès simplifié à des données et à des outils d'aide à la décision (par exemple, tableaux de bord et modèles prédictifs) pour les appuyer dans la gestion de la réussite scolaire et autres activités dont ils ont la responsabilité. Tout en relatant l'histoire du projet, l'étude expose les retombées, les défis et les leçons qui accompagnent le déploiement de ce projet. Les résultats de l'étude s'appuient principalement sur des entretiens semi-dirigés menés auprès de quatre représentants du MEQ³ et de deux représentants de la GRICS⁴. Principal

¹ Vérificateur général du Québec (2022). *Rapport du Vérificateur général du Québec à l'Assemblée nationale pour l'année 2022-2023*, page 17.

² Pierre St-Arnaud (2022) « Le réseau de l'éducation entre dans le XXI^e siècle avec la numérisation de ses données », *La Presse canadienne*, article publié le 12 mars 2022.

³ Dont Frédéric Potok, directeur général en ressources informationnelles, Ferhat Badis, directeur de l'architecture en ressources informationnelles, Natalia Cruz, conseillère stratégique en cadre normatif et Ivan-Mauricio Parra, scientifique de données.

⁴ Dont Sébastien Gougeons, président-directeur général et Guillaume Trottier, vice-président, Valorisation des données.

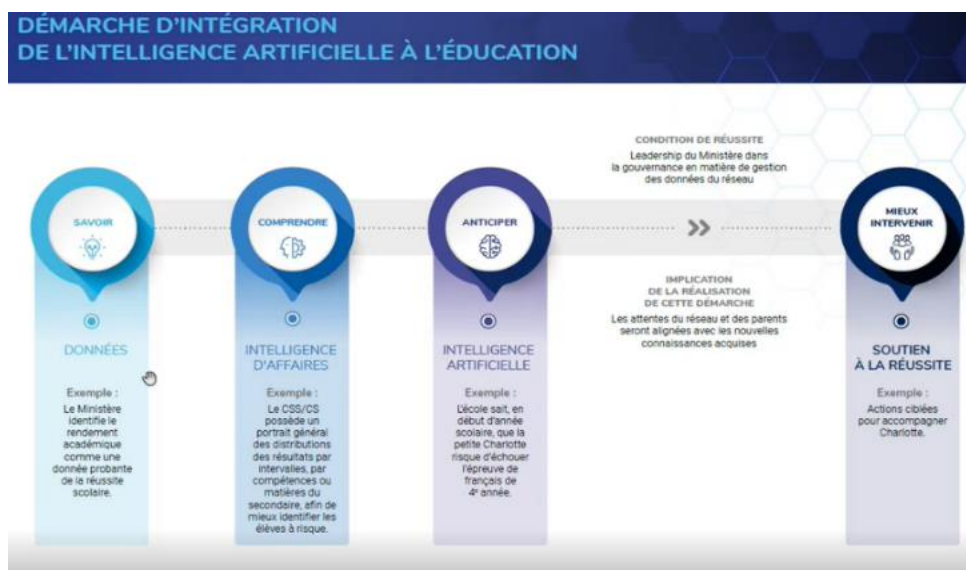
partenaire du MEQ dans le projet, la GRICS est un OBSL qui crée et soutient des systèmes de gestion administrative, scolaire et pédagogique pour le compte des centres de service scolaires et des commissions scolaires du Québec.⁵ Les propos des répondants sont en italiques dans le texte. Ils sont tantôt insérés entre des guillemets (« »), tantôt mis en retrait du texte. Nous tenons à remercier la disponibilité et la franchise des personnes rencontrées lors des entretiens. Leurs propos constituent la matière première de l'étude.

L'origine du CIN

Pour répondre en partie aux besoins d'information exacerbés par la pandémie de COVID-19, le MEQ a lancé le *Projet d'accès aux micro-données sur les élèves* à l'automne 2020. « *Le projet d'accès aux micro-données sur les élèves, c'est le volet précurseur du CIN.* » Ce projet visait à donner au MEQ l'accès à de nouvelles données détenues par les CSS/CS, en particulier des données sur le rendement académique (par exemple, notes des bulletins pour chaque matière et résultats aux épreuves obligatoires) et sur l'absentéisme et les retards (par exemple, nombre, motifs et durée). Ces données allaient permettre au MEQ de produire des indicateurs qui l'appuieraient dans sa gestion stratégique et opérationnelle du réseau.

Durant la mise à jour économique de l'automne 2020, le gouvernement du Québec a rendu disponibles aux ministères et organismes des budgets supplémentaires en fonction de la situation économique qui prévalait. Les autorités du MEQ ont demandé à leur *Direction générale des ressources informationnelles* de leur soumettre un projet pour financement.

Figure 1

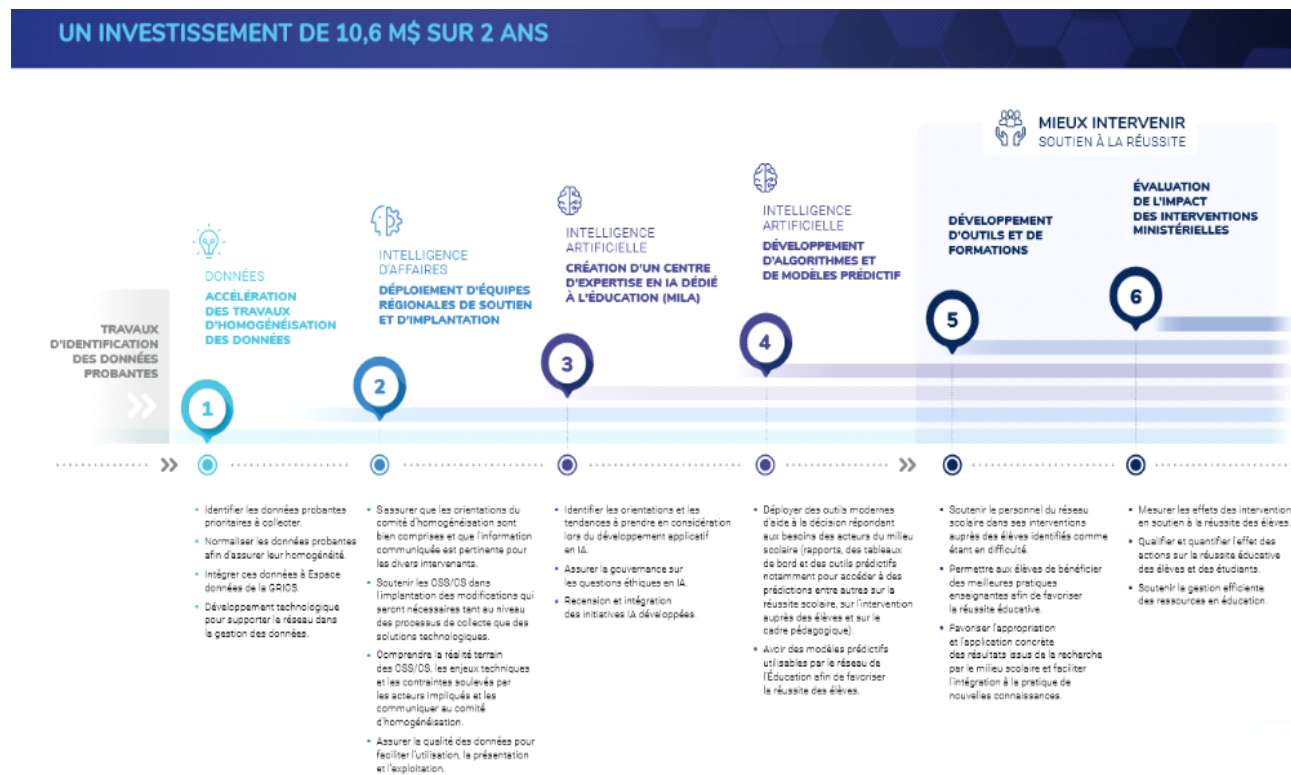


Le projet qui a été proposé, appelé *Démarche d'intégration de l'intelligence artificielle à l'Éducation*, s'est trouvé à élargir la portée du *Projet d'accès aux micro-données sur les élèves*.

⁵ Précisons que tous les membres du conseil d'administration de la GRICS sont des représentants de centres de service scolaires et des commissions scolaires.

Plus précisément, le nouveau projet s'est articulé autour de trois axes de valorisation des données en éducation, à savoir : 1) la priorisation des données probantes à collecter et leur harmonisation; 2) l'intelligence d'affaires, en particulier la création de tableaux de bord à partir de données et d'indicateurs de gestion pour assurer le suivi des activités scolaires et de l'utilisation des ressources; 3) l'intelligence artificielle, en particulier la conception de modèles de prédiction de la réussite ou de l'échec scolaire des élèves (voir figure 1). Contrairement au projet précédent, qui se limitait aux besoins informationnels du ministère, le nouveau projet s'adressait aussi aux autres acteurs du réseau, en particulier aux CSS/CS et aux directions d'école. Par cette valorisation des données, on souhaitait mieux outiller le MEQ et les autres acteurs du réseau dans leurs interventions de soutien à la réussite scolaire, notamment par le recours à des tableaux de bord, à des modèles de prédiction et autres outils d'aide à la décision. Préoccupées par les risques de décrochage plus importants pendant la pandémie, les autorités du ministère ont accueilli favorablement le projet proposé, si bien qu'un budget de 10,6 millions \$ sur 2 ans lui a été accordé. (Pour une ventilation du budget du projet et de ses postes de dépenses, voir la figure 2.) Ce projet de valorisation des données jetait les bases conceptuelles et financières de ce qu'il est désormais convenu d'appeler le *Chantier d'intelligence numérique (CIN)*.

Figure 2



2. La GRICS comme partenaire incontournable

Bien que le financement du CIN ait été officiellement annoncé le 14 mars 2022 par le ministre de l'Éducation de l'époque, M. Jean-François Roberge, les premiers travaux de ce projet ont commencé un an plus tôt, soit à l'hiver 2021. Une des premières actions posées par le ministère consistait à approcher des partenaires pour les appuyer dans la réalisation du CIN. Aux yeux du ministère, la GRICS s'est rapidement imposée comme un partenaire stratégique du CIN, compte tenu de son statut de fournisseur officiel de solutions technologiques de gestion administrative, scolaire et pédagogique pour le compte des CSS/CS. En effet, la « *presque la totalité des CSS/CS utilisent les principaux systèmes de mission de la GRICS* ». La GRICS est aussi l'hébergeur des données des CSS/CS collectées à partir de ces systèmes. La GRICS « *a déjà l'écosystèmes en place... constitué d'une trentaine de logiciels* », dont elle est l'éditeur.

[Bien que] la majorité de nos logiciels datent de plusieurs années [et] sont installés localement dans chaque centre de services scolaires... nos logiciels se parlent. Le logiciel de finance parle au logiciel de paye qui parle au logiciel de gestion scolaire.

Pour faciliter l'accès aux données générées par ses systèmes de mission implantés localement, la GRICS a créé une interface de programmation, qu'elle a installée sur les serveurs des CSS/CS. Appelé *le connecteur hybride*, cette interface permet à la GRICS d'aller chercher les données et de les ramener dans ses différents systèmes, et ce, de manière centralisée. « *On connaît les bases de données, comment elles sont faites* », affirme un représentant de la GRICS. Le fait d'avoir accès aux principales bases de données du milieu scolaire et de pouvoir les exploiter avec l'autorisation des CSS/CS conférait à la GRICS un rôle pivot dans le déploiement du CIN. La GRICS a joué (et joue encore aujourd'hui) le rôle de pont technologique entre le MEQ et les CSS/CS dans la réalisation du CIN. Le statut d'organisme public de la GRICS au sens de la *Loi sur les contrats des organismes publics* a aussi été facilitant pour le MEQ, puisqu'il a permis de resserrer les relations d'affaires entre les deux organisations.

Aux yeux de la loi sur les contrats des organismes publics, la GRICS est reconnue comme un organisme public... On est en train de faire un rapprochement très important avec la GRICS... On peut désormais la considérer comme un partenaire incontournable du ministère.

Enfin, la GRICS est apparue comme le fournisseur technologique tout indiqué pour les besoins du CIN, compte tenu de l'existence de la plateforme infonuagique, appelée *Espace Données*, qu'elle a récemment créée. « *Espace Données, qui est la plateforme de valorisation des données de la GRICS, a été choisi comme plateforme pour le chantier* ».

Grâce à sa participation au CIN, la GRICS a obtenu le financement nécessaire pour poursuivre le développement d'Espace Données et d'en faire une plateforme infonuagique plus ouverte et dynamique dans laquelle « *les CSS/CS pourront éventuellement travailler et y mettre leurs propres jeux de données* ». En tant que socle informationnel et technologique du CIN, Espace Données permettra à terme le passage « *d'une approche où [la GRICS] collecte et envoie des données au MEQ à une approche où le ministère peut lui-même aller chercher l'information dans l'Espace Données* », et ce, en respectant les règles de sécurité et de protection des renseignements personnels en vigueur. Cette plateforme permettra aux différents acteurs du réseau de l'éducation, en particulier au MEQ, aux CSS/CS et aux écoles, d'exploiter des données, dans le but de produire des tableaux de bord et de concevoir des modèles ou algorithmes de prédiction. (Pour une description plus détaillée d'Espace Données, voir la section 4.)

3. Un repositionnement stratégique de la GRICS

La mission de la GRICS consiste traditionnellement à « *fournir des logiciels de gestion scolaire ou administrative aux CSS... et autres services collectifs en TI* ». Or, la participation de la GRICS au CIN est venue bousculer sa façon habituelle de faire des affaires : de fournisseur de produits et services informatiques auprès des CSS/CS, la GRICS est désormais appelée à jouer le rôle d'un partenaire, certes stratégique, mais qui est tenu de collaborer avec d'autres acteurs du réseau.

Maintenant, on [la GRICS] n'est plus tout seul sur la glace. Ce n'est plus nous qui sommes la locomotive. On va partager le rôle de locomotive avec d'autres. On demeure un joueur important, mais on va contribuer dans un projet qui est plus large... avec des financements supplémentaires qui accélèrent la cadence de développement... d'outils de valorisation de la donnée.

Il faut préciser que les objectifs du CIN, essentiellement orientés vers la valorisation des données, ne sont pas étrangers aux récentes orientations stratégiques de la GRICS. Au contraire, ils font écho à la vision que la GRICS s'est donnée dans son dernier plan stratégique, adopté à l'automne 2021, puis déposé et présenté au réseau à l'hiver 2022 :

On s'est donné une nouvelle vision qui est de révolutionner l'éducation par l'intelligence numérique... On veut décloisonner... Permettre que l'information soit fluide, qu'elle s'échange... Donc, briser les silos, mettre les gens ensemble, et ça passe beaucoup par la valorisation des données... par un écosystème intégré ouvert pour le milieu de l'éducation et que la GRICS soit le principal maître d'œuvre de cet écosystème... Donc, on veut ouvrir notre plateforme (Espace Données) pour que les gens ne soient pas

prisonniers de la donnée... et qu'ils puissent prendre la donnée de l'extérieur puis la ramener dans notre écosystème pour la valoriser. Ça, c'est notre vision à long terme.

La GRICS a aussi créé, au même moment, une vice-présidence consacrée à la valorisation des données. Ce repositionnement stratégique de la GRICS résulte, entre autres, d'une prise de conscience à l'égard d'une perte de parts de marché, à la suite d'un éventuel déploiement de la plateforme gouvernementale intégrée SAGIR⁶, pour gérer les activités administratives du réseau.

Le gouvernement a pour projet gouvernemental de faire migrer le réseau vers SAGIR. Donc, nos solutions administratives vont quitter la GRICS dans un horizon de 5 à 10 ans. Ça a influencé notre plan stratégique. On perd une grosse partie de nos activités administratives... On doit recentrer notre mission... et consolider notre rôle de leader dans nos solutions de gestion scolaire.

Le MEQ et la GRICS ne sont pas les seuls organismes à vouloir valoriser les données en éducation. Avant que ne démarre le CIN, certains CSS/CS, comme le Centre de service scolaire du Val-des-Cerfs ou le Centre au Cœur-des-Vallées, ont créé et expérimenté des algorithmes pour prédire le décrochage scolaire sur leur territoire. « *Aux CSS Cœur-des-Vallées, en Outaouais, et Val-des-Cerfs, en Estrie, l'utilisation de l'intelligence numérique a permis de prévenir des échecs scolaires en dépistant, avec un taux supérieur à 90 %, les élèves les plus à risque de décrochage dès leur arrivée en secondaire 1⁷.* » Le CIN vise à fédérer ces diverses initiatives de valorisation des données en cours dans le réseau de l'éducation, en mutualisant les outils et les expertises qui y sont engagés.

La dynamique de collaboration entre le MEQ et la GRICS s'est grandement améliorée au fur et à mesure que leurs représentants joignaient leurs efforts, sur une base quasi quotidienne, et apprenaient à se connaître et à travailler ensemble à la réalisation du CIN. Le CIN s'apparente désormais à un projet fédérateur qui s'appuie sur la mise en commun de ressources financières, humaines, informationnelles et technologiques ainsi que sur des relations de collaboration soutenues entre le MEQ et certains acteurs du réseau, en particulier la GRICS.

On [GRICS] a passé beaucoup de temps à construire ce partenariat-là, c'est-à-dire de ne plus juste être un prestataire de services... mais de

⁶ Pour plus d'information sur SAGIR, voir lien suivant : <https://www.quebec.ca/gouvernement/faire-affaire-gouvernement/services-organisations-publiques/services-en-systemes-de-soutien-communs/solution-daffaires-en-gestion-integree-des-ressources-sagir/presentation-sagir>.

⁷ Pierre Saint-Arnaud, « Le réseau de l'éducation numérisera ses données », *La presse canadienne*, publié dans *La Presse*, 14 mars 2022.

travailler avec eux... Ça porte fruit présentement... la dynamique est devenue beaucoup plus collégiale aussi... Il y a des gens du ministère avec qui je travaille... et à qui je parle à tous les jours.

4. L'organisation du projet et les activités de soutien

Le MEQ ne disposait pas de ressources suffisantes pour accompagner l'ensemble des CSS/CS dans leur participation à la réalisation du CIN et dans la valorisation de leurs données, pas plus qu'il n'était « *en mesure d'aller chercher ces ressources externes lui-même directement* ». Il a donc fait appel à Collecto⁸ pour obtenir des ressources spécialisées pour accompagner les CSS/CS dans « *le développement des compétences en intelligence numérique* ». Ces ressources doivent être « *capables d'intervenir auprès des CSS/CS qu'ils soient débutants, intermédiaires ou avancés en matière de traitement de l'information* ». LEVIO est, quant à elle, la firme qui a fourni ces ressources humaines.

La GRICS est en train de former les consultants de LEVIO sur l'ensemble des produits qu'elle offre pour qu'ils les maîtrisent bien... pour qu'ils comprennent comment les données fonctionnent dans les systèmes opérationnels avant qu'on les amène dans l'entrepôt (Espace Données) et qu'on les valorise dans des tableaux de bord.

Un *centre de soutien en intelligence numérique* a été créé pour organiser le travail

Figure 3

Nom de la table régionale (numéros des régions administratives correspondantes)	Nb de CSS/CS
1 Laval/Lanaudière/Laurentides (13-14-15)	8
2 Montréal (06)	5
3 Bas St-Laurent/Gaspésie/Iles-de-la-Madeleine (01-11)	8
4 Saguenay/Lac-Saint-Jean (02)	4
5 Capitale-Nationale/Chaudière-Appalaches (03-12)	10
6 Mauricie/Centre-du-Québec (04-17)	5
7 Estrie/Montérégie (05-06)	15
8 Outaouais (07)	5
9 Abitibi-Témiscamingue/Nord-du-Québec (08-10)	8
10 Côte-Nord (09)	4

d'accompagnement de LEVIO et de la GRICS auprès des CSS/CS, afin que ceux-ci gagnent en compétence numérique. La principale stratégie de coordination adoptée par ce centre consiste à mobiliser les tables régionales actuelles et à les convertir en communautés de pratique dans le cadre du CIN (voir figure 3). « *Les CSS sont déjà habitués à travailler par région... dans des tables régionales* ». L'élément nouveau introduit par ce centre de soutien, « *ce n'est pas les communautés de pratique en tant que tel, mais l'accompagnement des*

communautés de pratique par l'ensemble des partenaires (MEQ, GRICS, prestataires de

⁸ Créé en 1997, Collecto est un organisme qui offre des services regroupés en éducation aux CSS/CS, aux cégeps, aux collèges privés et aux universités. Pour en savoir davantage sur cet organisme, consultez le lien suivant : <https://collecto.ca/#>

services, etc...) qui travaillent ensemble pour aider le réseau à transformer ses pratiques en intelligence numérique». Les communautés de pratique « visent à partager des ressources, des développements, des outils et des compétences. Donc, l'idée, c'est de faire monter en compétence numérique l'ensemble des CSS/CS ». Pour ce faire, chaque communauté doit se doter d'un plan de travail. « Ça peut être le développement de tableaux de bord ou d'un modèle de prédiction », en fonction des besoins et de la maturité de chacun. Le déploiement des communautés de pratique est présenté comme une façon d'amener les « CSS/CS les plus avancés à accompagner les moins avancés » dans la valorisation des données. Par ces communautés d'apprentissage, on souhaite aussi « mutualiser les efforts de l'ensemble du réseau pour le bénéfice du réseau ». Autrement dit, « une fois qu'il y a une bonne idée qui émane des communautés de pratique qui se met en place, il s'agit de reprendre cette solution... et de la rendre disponible à tous », notamment par le truchement d'Espace Données.

Espace Données deviendrait donc la plateforme de mutualisation des initiatives d'intelligence d'affaires ou d'intelligence artificielle du réseau. Ainsi, on espère éviter le travail en double, en s'assurant « qu'il n'y aura pas 2 ou 3 communautés de pratiques qui vont travailler sur le même sujet ». La mutualisation des efforts apparaît d'autant plus justifiée que certaines initiatives de valorisation des données requièrent la mobilisation d'importantes ressources, comme la conception et l'entraînement d'algorithmes prédictifs sur le décrochage scolaire. Enfin, les communautés de pratique sont perçues comme une façon d'instaurer un mode d'organisation pérenne de la valorisation des données et de la mutualisation des initiatives qui en découlent (par exemple, les tableaux de bord, les algorithmes et les expertises). En misant sur une structure bien établie (tables régionales), les responsables du CIN souhaitent, en effet, que les pratiques de collaboration et les initiatives issues des communautés de pratique se perpétuent au-delà de la durée du projet. « Parce que notre chantier a une date de fin, on veut que les CSS s'organisent par eux-mêmes pour travailler autour de l'intelligence numérique et de la valorisation des données ».

En complément des communautés de pratique, des escouades de données ont été créées avec pour mandat de prioriser et de normaliser les données à valoriser, afin de produire des indicateurs et des tableaux de bord qui répondent aux besoins des CSS/CS et du MEQ. Avec ces escouades, « on n'est pas dans des algorithmes d'intelligence artificielle, on est plutôt dans du business analytics ». Composées de représentants du MEQ, de la GRICS et des CSS/CS, ces escouades organisent leurs travaux de priorisation, de normalisation et d'intelligence d'affaires autour de thématiques qui revêtent un intérêt marqué pour le réseau. Deux premières escouades ont été créées à

l'été 2022 : une sur la réussite scolaire, l'autre sur la main-d'œuvre. Les responsables du CIN envisagent de créer trois autres escouades, d'ici la fin du projet, qui aborderaient respectivement les thèmes suivants : le financement, les infrastructures et la formation professionnelle aux adultes. S'inscrivant dans une démarche itérative, le travail des escouades vise à combler les besoins en matière de compétence, dans un contexte de pénurie de main-d'œuvre particulièrement criant en matière de valorisation des données.

Enfin, un centre d'expertise en intelligence artificielle a aussi été créé dans le cadre du CIN, qui regroupe des experts œuvrant au sein des organismes suivants : GRICS, Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique (OBVIA), Institut québécois d'intelligence artificielle (Mila) et Institut de valorisation des données (IVADO). Ces experts ont pour mandat d'accompagner et de conseiller les responsables du CIN dans le développement et le déploiement d'algorithmes en éducation, à la lumière des enjeux éthiques que l'utilisation de ces algorithmes peuvent soulever.

Le but, c'est d'utiliser la connaissance du milieu de la recherche et de l'enseignement universitaire... S'il y a des obstacles, on veut le savoir, s'il y a des réticences populaires, on veut le savoir, s'il y a des enjeux technologiques, on veut le savoir... Donc, eux (les experts) nous permettent d'identifier les orientations et les tendances à prendre en considération... L'idée est qu'on ne se retrouve pas dans les journaux.

5. Espace Données comme plateforme de valorisation du CIN

La plateforme technologique Espace Données a été officiellement lancée en juin 2021.

Figure 4



Elle est l'un des produits de la GRICS qui composent sa suite Mozaïk Data, laquelle se veut la porte d'entrée vers les données du réseau (voir figure 4)⁹.

Espace Données est une solution infonuagique centralisée qui permet aux acteurs du réseau, en particulier les CSS/CS,

⁹ LUMIX propose plusieurs dizaines de rapports en lien avec la réussite scolaire et Espace API offre des API permettant de faciliter l'intégration à des applications maison ou de tiers pour chaque organisme scolaire.

d'y verser des données et de les exploiter sous la forme de tableaux de bord, à partir des outils *Power BI* de Microsoft. L'accès à Espace Données est compartimenté, de manière à ce que chaque direction générale de CSS/CS et direction d'école n'accède qu'à ses propres données, et non à celles des autres. Pour exploiter Espace Données et ses tableaux de bord, la GRICS met à la disposition des directions générales des CSS/CSS et des directions d'école des interfaces de visualisation, dont Mozaïk-dVision et Mozaïk-dStudio.

Mozaïk-dVision permet la visualisation de tableaux de bord clés en main (préconçus) sur la réussite scolaire. Cette interface présente, entre autres, sous forme de rapports, les résultats scolaires (0 à 60 %, 61 à 74 % et 75 % à 100 %) en fonction des quatre compétences de base (écriture, lecture, résolution de problèmes et utilisation du raisonnement mathématique) et des niveaux scolaires au primaire, afin d'identifier les élèves à risque d'échec (voir annexe 1). Elle permet aussi de ventiler les résultats selon différents critères, en particulier le sexe (féminin ou masculin), la présence ou non d'un plan d'intervention, le groupe linguistique et l'immigration (voir annexe 2). Les directions générales (ou directions générales associées) de CSS/CS disposent d'une vue globale sur les données leur permettant de comparer les écoles entre elles (voir annexe 3). Les directions d'école ont un accès aux données sur la réussite scolaire qui, contrairement à celles des directions générales de CSS/CS, leur permettent de « *se rendre jusqu'au niveau de l'élève* ». Mentionnons que les directions générales de CSS/CS n'ont pas accès aux données des autres CSS/CS, pas plus que les directions d'école n'ont accès aux données des autres écoles. Au moment des entretiens, Espace Données ne contenait que des tableaux de bord sur les résultats et sur la réussite scolaire. D'autres tableaux de bord sont à venir. Par exemple, « *On a commencé à amener les données sur la main-d'œuvre dans Espace Données. Cependant, les tableaux de bord ne sont pas encore disponibles. On est en train de les développer* ».

Mozaïk-dVision entend aussi offrir des outils qui favorisent la réussite scolaire, en particulier des modèles de prédiction de la réussite/échec à l'épreuve ministérielle obligatoire d'écriture de la 4^e année du primaire, français langue d'enseignement, par école et par élèves (voir annexe 4). Alors que les directions d'école pourraient avoir accès à des données nominatives sur la prédiction de la réussite de leurs élèves à l'épreuve ministérielle, les directions de CSS/CS accéderaient à des données agrégées sur la probabilité de la réussite scolaire de leurs écoles, et non à celle des élèves (voir annexe 5). Ici aussi, il serait possible de ventiler les données (probabilités de réussite/échec) selon le sexe, la présence ou non d'un plan d'intervention, le groupe linguistique et l'immigration ainsi que d'ordonner les élèves selon leur probabilité

d'échec (voir annexe 6). Avec ces outils de prédiction, on souhaite mieux détecter les élèves à risque d'échec, afin « *de mettre des mesures en place pour éviter qu'ils tombent en échec. L'idée, c'est de prévenir l'échec en amont, de le voir d'avance et de l'éviter* ».

Quant à l'interface Mozaïk-dStudio, bien qu'elle utilise les mêmes données que Mozaïk-dVision, elle le fait en mode libre-service. Mozaïk-dStudio n'offre donc pas de tableaux préconçus, comme Mozaïk-dVision. En revanche, elle propose des fonctionnalités permettant à l'utilisateur de créer de nouveaux tableaux de bord sur mesure. Autrement dit, Mozaïk-dStudio permet « *de personnaliser des rapports en fonction des besoins de l'organisme scolaire à partir d'une sélection spécifique de jeux de données* ». Cet outil de valorisation des données s'adresse à « *des utilisateurs chevronnés en analytique de données* » qui maîtrisent bien les fonctionnalités de l'application *Power BI* de Microsoft.

Espace Données offre aussi au MEQ la possibilité d'accéder à des données provenant des CSS/CS et de créer des indicateurs nationaux et des tableaux de bord en utilisant l'interface libre-service Mozaïk-dStudio. Pour ce faire, le MEQ « *a demandé à la GRICS de créer un compartiment ministère* » dans Espace Données. Avec ce compartiment, le MEQ dispose d'un accès simplifié à des données provenant des CSS/CS qu'il peut exploiter. Cependant, le MEQ n'a accès à aucune donnée nominative. En effet, les données versées dans ce compartiment ont été dénominalisées par la GRICS, pour que le MEQ ne puisse identifier aucun individu; noms, prénoms, dates de naissance, codes postaux et autres éléments d'identification y ont été retirés. « *Au ministère, on s'intéresse à la province... L'individu comme unité n'est pas notre champ de préoccupations. C'est une vision globale des enjeux qui nous intéresse* ».

Au moment des entretiens, Espace Données se limitait aux données produites par les systèmes de mission développés par la GRICS. Autrement dit, « *la GRICS ne va pas fouiller dans des systèmes qui ne lui appartiennent pas pour aller chercher les données* », puis les verser dans Espace Données. La GRICS n'écarte pas pour autant l'idée d'inclure éventuellement d'autres sources de données dans Espace Données.

Quant aux enseignants, ils n'avaient pas accès à Espace Données au moment des entretiens. Dans l'éventualité où des accès leur seraient accordés, il faudra alors différencier ces accès « *pour que l'enseignant ne puisse pas voir les résultats d'un autre groupe (ou enseignant)* ». En somme, avec Espace Données, on souhaite « *avoir à un seul endroit l'ensemble des données du réseau à valoriser* ».

6. Une culture de données à instaurer et des pratiques de collaboration à pérenniser

Si l'intelligence artificielle et l'utilisation d'algorithmes constituent l'une des principales cibles du CIN dans la valorisation des données du réseau, les besoins plus immédiats de la majorité des CSS/CS semblent se situer surtout du côté de l'intelligence d'affaires et de la production de tableaux de bord. « *Avant tout, le premier besoin, c'est d'avoir des tableaux de bord... l'intelligence artificielle, c'est le summum... mais t'as pas toujours besoin de te rendre là* ». Lorsque le CIN a commencé, « *la plupart (des CSS/CS) n'était même pas rendu au BI (intelligence d'affaires)* ». Le degré de maturité numérique des CSS/SC serait à géométrie variable selon la disponibilité des expertises et autres ressources.

Il y a des CSS/CS qui sont bien équipés en termes de ressources humaines et informationnelles, d'autres qui ne le sont pas... Certains sont très à l'aise en BI, d'autres non. Certains sont des débutants en intelligence artificielle, d'autres sont assez avancés, et la plupart en savent très peu à ce sujet.

La capacité organisationnelle à valoriser les données tend aussi à varier selon la taille des CSS/CS.

Les grands CSS ont souvent de plus grandes équipes... Et plus tu as de grandes équipes, plus tu es en mesure de dédier des gens à monter des nouveaux tableaux de bord, à travailler la donnée, et cetera. À l'autre extrême, les très petits CSS... bien souvent ont une ou deux personnes en informatique qui fait un peu tout.

Bien que la quasi-totalité des CSS/CS aient accepté les conditions d'utilisation d'Espace Données et s'y soient connectés au moment des entretiens, « *ce n'est pas tout le monde qui l'utilise* ». Par exemple, pour utiliser l'outil libre-service Mozaïk-dStudio, « *encore faut-il avoir le personnel qui détient l'expertise... pour monter des tableaux de bord... Puis, c'est bien beau d'avoir des données... Mais une fois que t'as ça entre les mains, tu en fais quoi ?* » Quant aux CSS/CS qui ont développé leurs propres outils et plateformes de valorisation des données avant l'arrivée du CIN, « *ils ne voyaient pas nécessairement la pertinence de joindre Espace Données* ».

Il n'en demeure pas moins qu'une majorité des CSS/CS exploitent peu (ou pas du tout) les données qu'ils détiennent. La valorisation des données n'est ni innée, ni acquise, chez plusieurs acteurs du réseau, notamment en raison de compétences numériques

limitées. De l'avis de la majorité des répondants, un changement de culture et l'acquisition de nouvelles compétences et pratiques sont nécessaires pour amener les CSS/CS à valoriser leurs données et à mutualiser les initiatives qui en résultent. « *Si vous n'avez pas la culture de la donnée... les tableaux de bord vont dormir sur une tablette. Vous ne saurez pas quoi faire avec* ».

Différents pièges sont à éviter dans ce changement de culture et de pratiques en matière d'intelligence numérique. Un premier piège consisterait à inonder l'utilisateur d'indicateurs et de tableaux de bord qui lui seraient peu utiles pour suivre les activités de son organisation et pour éclairer ses décisions. Dans le déploiement d'outils de suivi et d'aide à la décision, la qualité doit prévaloir sur la quantité, afin de ne pas noyer ni décourager l'utilisateur. Un autre piège à éviter serait de traiter de façon uniforme l'ensemble des organisations du réseau et leurs intervenants au moment du déploiement des stratégies de valorisation des données. Ici, la maturité numérique des organisations et la volonté de se transformer doivent être prises en considération. Or,

Diffuser la culture de données en tenant compte des réalités du milieu, c'est tout un défi... Il y a certains centres de services scolaires qui sont en avant de la parade, on n'a pas besoin de les convaincre... Il y en a qui trouvent qu'on ne va pas assez vite : « Vous devriez être plus loin. Nous on a déjà fait ça. » Puis, j'en ai d'autres qui disent : « Pourquoi je ferais ça? On a bien d'autres chats à fouetter. »

Dans ce contexte, des mesures d'accompagnement personnalisées, comme les communautés de pratique, semblent plus appropriées pour faciliter le virage numérique des CSS/CS que des mesures de soutien gouvernemental mur-à-mur et homogènes. « *La beauté des communautés pratiques, c'est justement d'avoir accès à différents niveaux d'accompagnement.* » Dans ces communautés, les CSS/CS plus expérimentés dans la valorisation des données seraient appelés à accompagner les CSS moins expérimentés, afin qu'ils gagnent en intelligence numérique. Regroupant des acteurs habitués à collaborer entre eux, les communautés de pratique favoriseraient la mise en commun d'initiatives issues d'un même territoire ainsi que la pérennité de ces pratiques collaboratives. Fondée sur la mobilisation d'acteurs du milieu, une telle stratégie ne diminue pas pour autant l'importance du rôle du ministère dans le CIN. Au contraire, le fait que le MEQ ait été le principal promoteur du projet a facilité les démarches de la GRICS pour obtenir la participation des CSS/CS dans le déploiement d'Espace Données.

On (GRICS) est un organisme OSBL qui appartient à ses membres (CSS/CS). On n'est pas en autorité sur eux... On ne veut pas les brusquer... Avec l'aide du ministère, on a pu accélérer l'implantation [d'Espace Données].

Dans ce changement de culture et de pratiques, il est important que la GRICS donne l'exemple, ne serait-ce que par ses importantes responsabilités dans le déploiement technologique du CIN. Or, le virage culturel et managérial en cours à la GRICS, annoncé dans son plan stratégique et mis en branle dans le cadre du CIN, a rencontré de la résistance à l'interne. Cette résistance s'explique notamment par des changements profonds à la GRICS dans ses relations d'affaires avec les CSS/CS. Habitée à exercer un plein contrôle sur le développement et le déploiement de ses produits informatiques, la GRICS doit désormais gérer de façon plus collaborative le développement de ses produits, en particulier Espace Données qui, rappelons-le, dépend de données détenues par les CSS/CS.

C'est un gros changement qu'on a vécu et qui brasse les équipes à l'interne... parce qu'on est des propriétaires de produits habitués d'être en plein contrôle de leur feuille de route... C'est nouveau de se rapprocher autant des CSS et d'être autant à leur écoute.

Un des moyens mis de l'avant par la direction de la GRICS pour faciliter ce virage a été de créer dans l'organisation un secteur de valorisation de la donnée qui préconise une approche collaborative avec les CSS/CS. Quant à la création d'Espace Données, comme plateforme de valorisation et de mise en commun des ressources informationnelles du réseau, elle témoigne de cette volonté de la GRICS de collaborer davantage avec les CSS/CS et le MEQ en tant que partenaires, et non plus seulement comme fournisseurs de produits informatiques.

7. Des données, des pratiques et des systèmes à harmoniser

Rappelons que la création d'Espace Données vise à valoriser les données dont disposent les CSS/CS et à en permettre la réutilisation, afin d'assurer un meilleur suivi des activités scolaires et une meilleure coordination des interventions à l'échelle locale et nationale. Cependant, rendre disponibles les données de manière centralisée nécessite un important travail de collecte d'informations auprès des différents systèmes que la GRICS a installés dans les CSS/CS. « Il faut aller puiser la donnée dans 72 environnements. Les logiciels ne sont pas en fonuagique. Ils sont installés

localement, dans chacun des centres de services scolaires». La valorisation des données nécessite aussi un important travail d'homogénéisation des données, compte tenu de la diversité des pratiques, des définitions et des systèmes dans le réseau, notamment au moment de la saisie des données. C'est le cas des données sur l'absentéisme des enseignants.

Dans les grands centres de services scolaires, ça va être en partie automatisé. Ils vont rentrer les données dans le système assez rapidement. Par contre, dans un petit centre de service scolaire, ça peut être fait par l'adjointe administrative qui connaît l'ensemble des enseignants et enseignantes par cœur... Il y en a qui font ça dans un fichier Excel... On se retrouve donc avec des manières de rentrer les données, de les interpréter totalement différentes.

Pour ajouter à la complexité de ces enjeux de saisie et d'intégration des données, certains CSS/CS ont créé leurs propres applications qui, à l'instar des systèmes manuels (par exemple, Excel), ne s'interconnectent pas aux systèmes de la GRICS. « Comme l'information est ailleurs, il fallait qu'ils (CSS/CS) nous la poussent, c'est tout un défi ». Extraire les données de ces systèmes informatiques maison, pour les déposer dans Espace Données, nécessite le plus souvent des tâches de saisie supplémentaires, dont le remplissage de formulaires par les CSS/CS.

Des enjeux de normalisation des données se posent aussi dans le cas des systèmes de mission développés par la GRICS et déployés dans les CSS/CS. En effet, les CSS-SC peuvent personnaliser certains champs à saisir dans les applications de la GRICS qu'ils utilisent, que ce soit en ajoutant des options de réponses ou en permettant la saisie dans un champ libre.

Lorsqu'on veut ramener les données dans Espace Données, elles ne sont pas toutes normées de la même manière. Ce n'est (sic) pas des cases à cocher, nécessairement. Il peut même y avoir de l'information qui est rentrée en champ libre.

Si certaines données peuvent être homogènes et normées d'un CSS/CS à l'autre, comme les notes aux épreuves ou aux bulletins, d'autres le sont moins. C'est le cas des motifs d'absence, où les choix de réponses peuvent être plus ou moins précis et libres, selon les besoins des CSS/CS et de leurs utilisateurs. Les possibilités de réponses peuvent même être infinies quand la saisie se fait dans un champ libre, donc sans nomenclature ou menu déroulant.

Il en y a qui vont écrire absent pour maladie... absent en général... absent pour santé... absent pour personnel... Il y en a d'autres qui vont être très précis ... Il est allé chez le chiro, chez le médecin. Il a eu un accident.

L'enjeu de normalisation des données se révèle particulièrement problématique quand il est question de doter le MEQ d'indicateurs nationaux à partir de l'intégration de données saisies de manière disparate entre les CSS-CS.

Pour un centre de services scolaires, les données font du sens. Mais quand on met les données des 72 CSS/CS ensemble, des fois ça marche moins. On est en train de travailler avec le ministère de l'Éducation et les centres de services scolaires pour harmoniser les données.

Pour favoriser l'harmonisation des données, les responsables du CIN ont recours à deux stratégies : d'une part, une harmonisation dite *a posteriori*, dans laquelle « on collecte les données de façon différente au départ et après on les harmonise » et, d'autre part, une harmonisation dite *a priori*, dans laquelle « on essaie d'homogénéiser les données en amont » au moment de leur collecte. Des répondants reconnaissent qu'une harmonisation en amont des données simplifierait leur utilisation en aval, en particulier à l'échelle nationale, en diminuant les efforts d'intégration au moment de l'exploitation des données. Cependant, une harmonisation complète des données en amont est susceptible de rencontrer un enjeu de complexité, du fait que les acteurs du milieu se verraient dans l'obligation de changer leurs systèmes et leurs façons de faire, en particulier au moment de la saisie des données, soutiennent certains répondants. Cela n'empêche aucunement que la normalisation en amont de certaines données produites par les CSS/CS apparaît inévitable. C'est le cas des données administratives du réseau de l'éducation, qui devront éventuellement s'intégrer à la solution d'affaires en gestion intégrée des ressources (SAGIR), qui est déployée dans l'administration publique québécoise, depuis 2008, pour gérer les ressources humaines, financières et matérielles.

8. Des mesures et règles d'accès aux données à moderniser

Bien que la GRICS héberge la plateforme Espace Données, les données qui y sont versées ne lui appartiennent pas. Elles appartiennent aux CSS/CS, qui en sont les détenteurs ou les gardiens « en vertu de la loi sur l'accès à l'information et la protection des renseignements personnels ». Par conséquent, « la GRICS n'a pas le droit de les

transmettre au MEQ [ou autre tierce partie] sans l'autorisation des CSS ». Le MEQ n'a pas non plus le pouvoir légal d'accéder directement aux données détenues par les CSS/CS sur la réussite scolaire, sur l'absentéisme et sur beaucoup d'autres renseignements nominatifs, sans le consentement explicite de ces organismes. Aucune loi, ni décret ou règlement ne l'autorise explicitement à le faire.

Le MEQ a donc besoin du consentement des différents CSS/CS pour accéder à son compartiment de données dans Espace Données, et ce, même si les données y sont dénominalisées. Or, l'obtention du consentement des CSS/CS s'est avérée un exercice long et exigeant. Pour chacun des jeux de données demandés (par exemple, la réussite scolaire et l'absentéisme), le MEQ a dû justifier la nécessité d'y avoir recours, le traitement qu'il voulait en faire, la durée de conservation des données et bien d'autres choses encore.

Lorsqu'un organisme donne son consentement pour l'utilisation de la donnée, il faut lui dire précisément quelle donnée est visée ? Quelle est la finalité ? D'où vient cette donnée ? Qu'est-ce qu'on va en faire ? Combien de temps veut-on la conserver ? ... Comment les données vont être protégées... tout ça par écrit.

C'est la GRICS qui a eu à consacrer la majorité des efforts dans cette gestion du consentement, puisqu'elle a dû négocier la rédaction et la signature des ententes avec les CSS/CS, afin qu'ils l'autorisent à aller chercher des données dans leurs systèmes et à les verser dans Espace Données, notamment pour le compte du MEQ. « *C'est long de négocier une entente de consentement avec un grand nombre d'organismes qui ont toutes leurs réalités et leurs questionnements.* » Quant à l'entente de consentement :

... c'est un document légal, donc on a besoin d'avocats qui revoient ça, qui s'entendent entre eux, qu'on présente aux secrétaires généraux des centres de services scolaires... Il faut qu'on sache exactement quels jeux de données on va chercher en détail avant d'écrire le consentement... et qu'est-ce qu'on va faire avec.

La gestion du consentement s'est avérée d'autant plus exigeante qu'il a fallu refaire la démarche en entier pour ajouter dans la plateforme un deuxième jeu de données, portant cette fois-ci sur des données sur la main-d'œuvre et les effectifs scolaires (par exemple, nombre d'enseignants à temps plein/partiel ou psychoéducateurs à temps plein/partiel), après avoir obtenu un premier jeu des données portant sur la réussite scolaire. « *On a été obligé de refaire le consentement, donc... de redemander aux 72 DG*

(de CSS/CS), de consentir avant de pouvoir aller chercher la donnée». Qui plus est, l'accès par le MEQ à ce deuxième jeu de données sur la main-d'œuvre et les effectifs scolaires semble avoir suscité plus de questionnements que dans le cas du premier jeu, notamment en raison d'enjeux de relations de travail¹⁰.

En plus d'avoir à obtenir le consentement des CSS/CS, le MEQ a dû se plier à d'autres exigences pour accéder à Espace Données. D'une part, les données auxquelles le MEQ a accès dans la plateforme doivent être dénominalisées, comme nous l'avons indiqué précédemment. D'autre part, la direction des affaires juridiques du MEQ et sa direction d'accès à l'information n'autorisent pas le ministère à télécharger les données dénominalisées d'Espace Données pour les verser et les exploiter dans ses propres systèmes, compte tenu des possibilités de réidentification.

On (MEQ) est capable d'accéder aux données dénominalisées de l'entrepôt de la GRICS, mais on n'est pas capable de prendre ces données, de les mettre chez nous, de les croiser avec nos propres données... On n'est pas vraiment autorisé... Notre direction des affaires juridiques et notre service d'accès à l'information sont très sensibles au fait que, techniquement, nous pourrions, en croisant les données avec d'autres données, réidentifier l'élève.

Pour la majorité des répondants, le cadre juridique et réglementaire actuel paraît particulièrement contraignant en ce qui a trait à la gestion des données nominatives. De plus, ce cadre ne se semble pas tenir compte de la nouvelle réalité des projets d'intelligence artificielle, dans ce sens que les données à exploiter et les algorithmes à développer ne sont souvent connus qu'au moment de l'entraînement des modèles.

Avec l'intelligence artificielle, ce que tu veux, c'est de mettre des données ensembles pour laisser la machine déterminer quelles sont les données les plus probantes, quelles sont les données significatives. Donc, l'approche est complètement différente... On ne sait pas forcément à l'avance quelles sont les données les plus intéressantes d'un point de vue statistique...

Pour les répondants, une modernisation du cadre juridique et réglementaire s'impose, si l'on veut améliorer la fluidité de l'information nécessaire à la valorisation des données. Une telle modernisation semble nécessaire pour appuyer les dernières orientations

¹⁰ Au moment des entretiens, 14 CSS/CS n'avaient toujours pas consenti à ce que leurs données sur la main-d'œuvre soient versées dans Espace Données.

gouvernementales liées à la transformation numérique de l'État et des services, rappellent certains répondants. En réponse à ces contraintes juridiques et réglementaires, la GRICS a créé des CSS dits *virtuels*, constitués de données anonymisées qu'ils peuvent exploiter, afin de repérer les variables à insérer dans les modèles de prédiction et autres outils d'intelligence artificielle, au moment de leur entraînement.

On a accès à ce qu'on appelle des CSS virtuels. On a certains centres de services scolaires qui ont accepté qu'on utilise leurs données... anonymisées. Donc, on ne voit pas d'élèves. C'est une photo statique de leurs données... Une fois qu'on connaît la recette... les variables dont on a de besoin, il faut aller mettre ça dans le consentement et aller officiellement chercher ces données-là.

9. Des enjeux professionnels et éthiques à considérer

Bien que porteuse de réflexivité organisationnelle, la valorisation des données en éducation soulève des enjeux professionnels et éthiques, notamment quand il s'agit de recourir à l'intelligence artificielle. Le conception et l'utilisation d'algorithmes dans ce domaine suscitent, en effet, diverses préoccupations, soulignent les répondants. Une de ces préoccupations renvoie à une possible dérive de nature professionnelle, à savoir que les algorithmes sur la réussite scolaire ou sur l'absentéisme pourraient aussi servir à évaluer la performance des intervenants en éducation, en particulier le rendement des professeurs. À ce propos, des responsables du CIN se veulent rassurants, en soutenant que les algorithmes sur la réussite scolaire, développés par la GRICS ou par des CSS/CS, ne permettent pas une telle dérive.

Les variables qu'on utilise dans ces algorithmes, ce sont les variables sur l'élève. On n'a pas pris en considération les variables de l'enseignant... On évalue la réussite d'un élève basée sur ses expériences passées.

Une autre préoccupation soulevée par les répondants est liée à la possibilité que les algorithmes puissent se substituer au jugement des enseignants et autres intervenants du réseau. Cette préoccupation a aussi été évoquée par la Centrale des syndicats du Québec (CSQ)¹¹, à la suite d'un point de presse de l'ex-ministre de l'Éducation, Jean-François Roberge, sur le lancement du CIN, en mars 2022.

¹¹ Qui représente environ 120 000 membres du secteur de l'éducation.

La CSQ s'interroge notamment quant au respect du jugement professionnel du personnel du réseau. Un modèle (prédictif), aussi efficace soit-il, demeurera toujours simplement un modèle. Il faudra toujours quelqu'un, en bout de ligne, pour poser un jugement professionnel. Ces guides ne pourront rendre compte de l'ensemble du parcours d'un élève, de ses forces et de ses faiblesses, de son évolution comme seul peut le faire le personnel de l'éducation, souligne M. Gingras (président de la CSQ).

Pour plusieurs répondants, les algorithmes en éducation doivent être utilisés comme des outils d'aide à la décision, et non comme des outils qui remplacent le jugement de l'humain. « *L'intelligence artificielle peut nous mettre sur des pistes et on va confirmer après* ». Cela n'empêche pas que les algorithmes puissent aider à la détection de certains élèves à risque de décrochage, en raison de la « *variabilité de leurs résultats* », et qui ont tendance « *à passer sous le radar des professeurs... parce qu'ils n'ont jamais été en échec* ».

L'utilisation d'algorithmes pour suivre les parcours scolaires et pour intervenir auprès d'élèves à risque d'échec soulève aussi des préoccupations éthiques à l'égard des principaux concernés, soit les élèves eux-mêmes. Cette utilisation de l'intelligence artificielle peut, par exemple, accentuer le phénomène d'étiquetage et de discrimination vis-à-vis d'élèves déjà aux prises avec des difficultés d'ordre scolaire, y compris les élèves à risque d'échec modéré.

La petite Charlotte... si on sait qu'elle risque d'échouer, ne vient-on pas lui mettre une étiquette qui lui collera à la peau longtemps ? Surtout quand on va avoir toutes les données regroupées à un seul endroit.

Il faut également s'assurer que les algorithmes ne reproduisent pas (ou le moins possible) des biais cognitifs et sociaux associés au sexe, à l'ethnie, à l'âge, au salaire et autres caractéristiques qui pourraient marginaliser davantage des élèves appartenant à des groupes déjà discriminés. Il faut utiliser les algorithmes de prédiction en éducation avec d'autant plus de précautions que ces algorithmes sont le plus souvent constitués de données sensibles sur des mineurs, précisent des répondants.

Comme autre précaution, les personnes engagées dans la conception d'algorithmes doivent s'assurer que les données qu'ils utilisent dans les modèles de prédiction sont exactes et, dans la mesure du possible, exemptes de biais. Elles auront aussi la délicate tâche de sélectionner les données probantes à insérer dans le modèle de prédiction et

de leur accorder un poids selon leur pouvoir de prédiction respectif. Par exemple, « *est-ce qu'on donne plus de poids dans l'algorithme à la scolarité de la mère que du père* », sachant que les études tendent à démontrer que la scolarité de la mère a une influence plus déterminante sur la réussite scolaire de l'élève.

Par ailleurs, l'utilisation d'algorithmes sur la réussite scolaire, en particulier la détection d'élèves à risque d'échec, s'accompagne d'une certaine obligation d'agir pour corriger le problème ainsi que d'un besoin d'expliquer les motifs et les conséquences des actions à entreprendre. En effet, les parents pourraient être tentés et justifiés de demander des explications lorsque le modèle de prédiction catégorise leur enfant comme un élève à risque qui commande une intervention. « *Comment ça se fait que vous avez identifié mon jeune ? Qu'est-ce que vous allez faire maintenant ? Faut être capable d'expliquer.* »

À défaut d'offrir des explications compréhensibles sur les prédictions du modèle, l'utilisation d'algorithmes apparaîtra comme une boîte noire qui « *en vient à se programmer un peu tout seul* », à partir d'un traitement automatisé de données de plus en plus massives et diversifiées. Cette opacité est susceptible de contribuer au sentiment de méfiance face à l'intelligence artificielle déjà bien présent chez plusieurs intervenants du réseau et chez des parents. « *On a même des parents qui... se méfient des données tout court. On reçoit régulièrement des demandes de parents qui ne veulent pas que l'élève s'inscrive sur une base de données de la classe.* »

Il est raisonnable de croire que l'explicabilité en intelligence artificielle, c'est-à-dire la capacité de rendre compte de décisions algorithmiques, se complexifiera au fur et à mesure que les données et les variables exploitées dans les modèles de prédiction se multiplieront. Quoiqu'il en soit, l'utilisation d'algorithmes en éducation doit demeurer assujettie au jugement de l'humain, pour en faire un outil socialement acceptable, rappellent les répondants. Si l'algorithme peut aider à la détection de cas à risque, la décision ultime revient à l'humain et à son jugement, particulièrement en éducation. Il revient aussi au concepteur d'algorithmes de justifier le choix des données et des variables qu'il intégrera au modèle de prédiction, et de tenter de minimiser les biais associés à ce choix. Afin de guider de façon responsable le développement et l'utilisation d'algorithmes en éducation, les responsables du CIN travaillent à l'élaboration d'un cadre éthique destiné aux organismes et aux intervenants du réseau, avec la collaboration du centre d'expertise en intelligence artificielle.

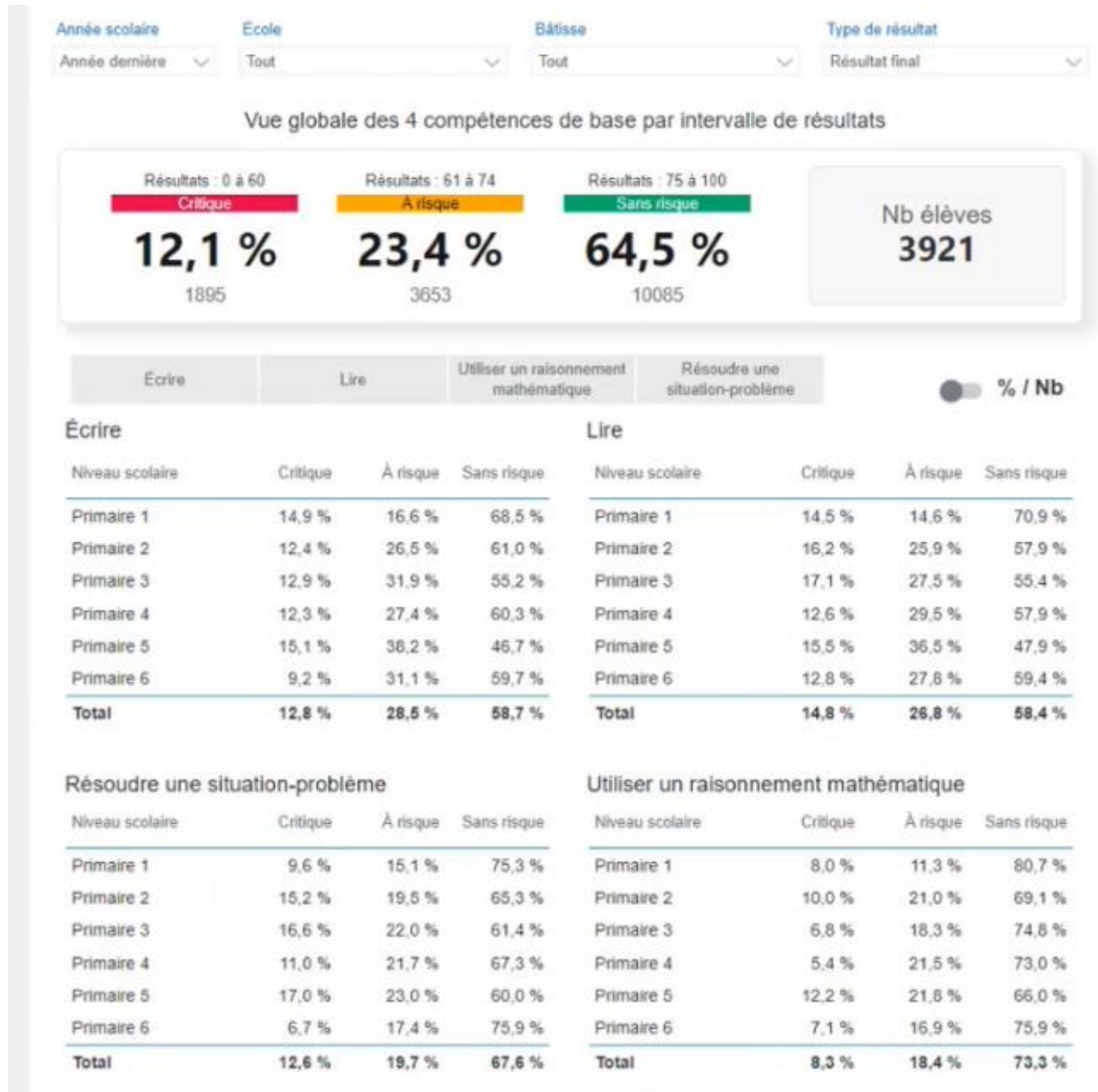
Conclusion

Avec la numérisation croissante des transactions et la montée en puissance des outils d'analyse et d'aide à la décision, la valorisation des données tiendra un rôle de plus en plus stratégique dans la mise en œuvre des politiques et dans la gestion des ressources et des services publics. Le CIN témoigne des potentialités stratégiques liées à la valorisation des données en éducation, par le développement d'outils d'aide à décision, qu'il s'agisse de tableaux de bord (intelligence d'affaires) ou de modèles de prédiction (intelligence artificielle). Ce projet montre, en effet, que les pratiques réflexives issues de la valorisation des données peuvent être utiles pour appuyer le MEQ dans l'élaboration des programmes scolaires et autres mesures d'encadrement nationales. La valorisation des données peut aussi être profitable aux acteurs du milieu, en particulier aux CSS/CS, aux écoles et aux intervenants, pour gérer la réussite scolaire, notamment à partir d'analyses, de suivis et d'interventions menées auprès de leurs élèves. Comme projet fédérateur sur la valorisation des données en éducation, le CIN a également permis de mettre en commun des ressources stratégiques provenant des principaux acteurs du milieu de l'éducation œuvrant au primaire et au secondaire. À notre connaissance, il s'agirait de la première expérience du genre au Québec, du moins en éducation.

Par ailleurs, l'étude du CIN montre que la valorisation des données pose d'importants enjeux dont les défis dépassent les considérations technologiques, elles-mêmes complexes. À ce propos, la gestion de l'information a certainement constitué l'un des enjeux les plus ardues à gérer par les responsables du CIN dans le cadre du projet. Si les données à valoriser sont nombreuses dans le réseau, l'accès à ces données et leur réutilisation ne vont pas de soi, en particulier quand il s'agit de doter le MEQ d'indicateurs nationaux. En effet, la réalisation du projet a nécessité la mise en marche d'activités parfois complexes en matière de gestion de l'information, dont l'obtention du consentement et l'harmonisation des données, des systèmes et des pratiques, sans oublier l'anonymisation de l'information et la différenciation des droits d'accès selon la compétence des intervenants.

La valorisation des données en éducation soulève aussi des enjeux de gouvernance. Ici, le défi consiste à faire migrer les partenaires d'une gouvernance en vase clos à une gouvernance en réseau qui favorise la mise en commun des données et des réalisations issues d'initiatives de valorisation (par exemple, tableaux de bord, modèles de prédiction et expertises).

Enfin, la valorisation des données en éducation soulève des préoccupations d'ordre professionnel et éthique qui nécessitent une importante réflexion sur les impacts potentiels de cette valorisation, notamment en ce qui a trait à l'évaluation de la performance des intervenants en éducation et à l'étiquetage ou à la discrimination de certaines clientèles étudiantes. Les responsables du CIN devront s'assurer que le déploiement des outils de valorisation seront à la fois utiles et acceptables pour les élèves ainsi que pour les intervenants qui les encadreront.



Annexe 2

Vue globale des 4 compétences de base par intervalle de résultats

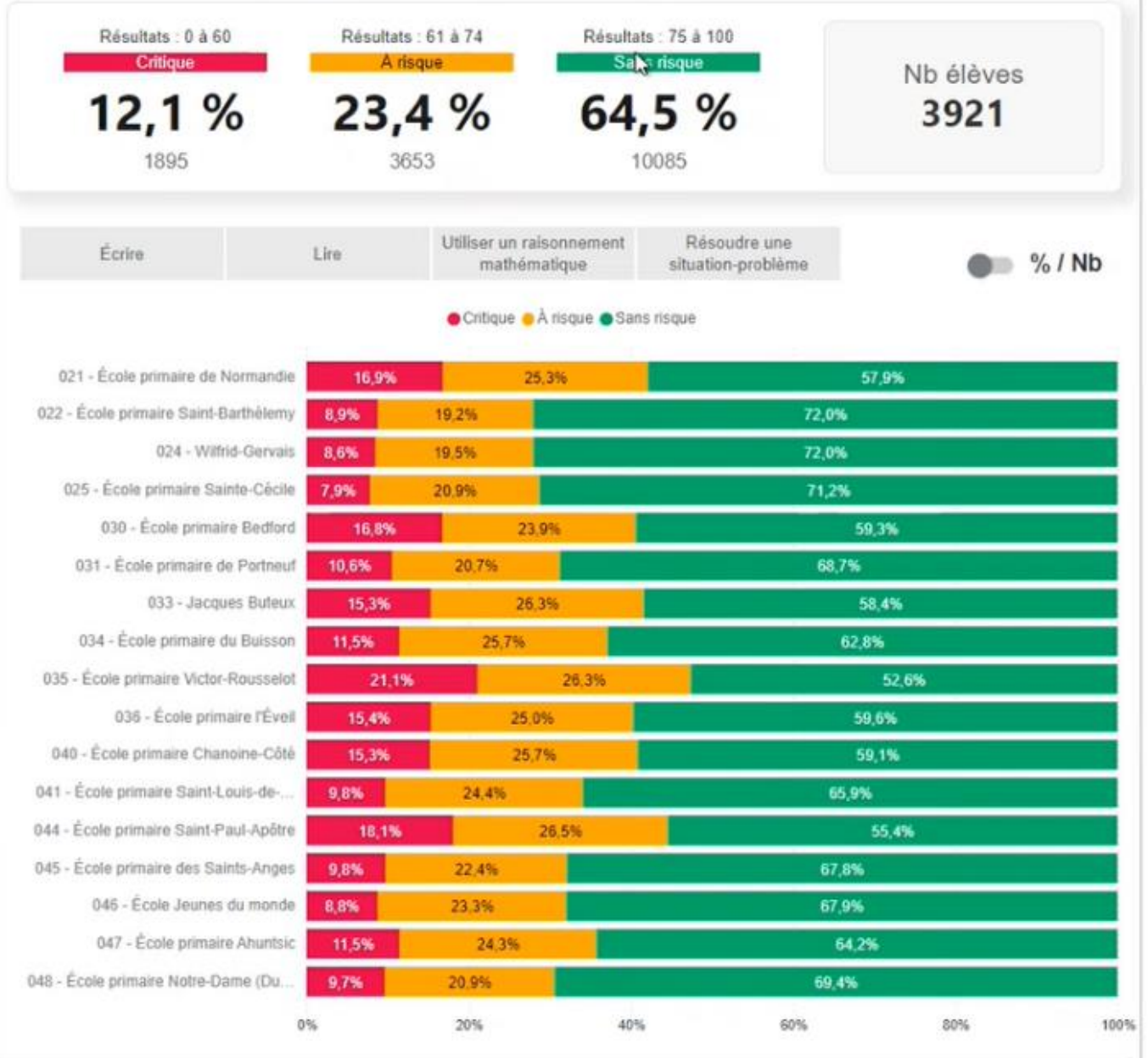


● Critique ● À risque ● Sans risque



Annexe 3

Vue globale des 4 compétences de base par intervalle de résultats



Propulsé par Espace Données © Tous droits réservés GRICS

Annexe 4¹²

Prédictions des réussites/échecs - Épreuve MEQ écriture 4e primaire, Français langue d'enseignement



2018-2019, en début de 3e étape

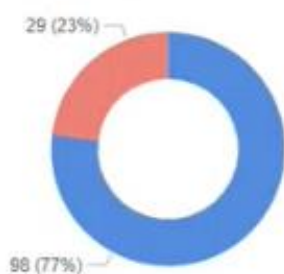
Moment de la prédiction

2018-2019, en début de 3e étape

École, groupe-repère

Tout

Prédictions



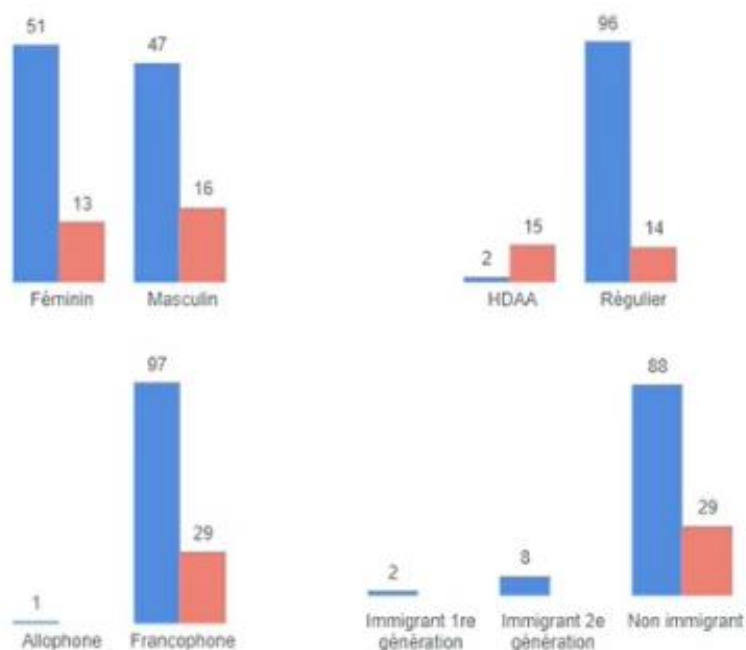
Prédit échec
Probabilité d'échec \geq 50 %

29

Prédit réussite
Probabilité d'échec < 50 %

98

Répartitions des prédictions de réussite et d'échec des élèves de la 4e année du primaire selon différentes ventilations



École	Groupe-repère	Fiche	Nom et prénom	Résultat	
				Échec	Réussite
097 - École primaire Bedford	401	750497	Massicotte Jason		●
097 - École primaire Bedford	401	762112	Chenette Brittany		●
097 - École primaire Bedford	401	763144	Lachance Johanne		●
097 - École primaire Bedford	401	763540	St-Roux Eugène		●
097 - École primaire Bedford	401	763797	Lehouillier Élodie		●
097 - École primaire Bedford	401	763839	Soucy-Turcotte Johanne	▲	
097 - École primaire Bedford	401	763953	Lévesque Christian-Tristan		●

¹² Les noms dans ce tableau sont fictifs.

Annexe 5

Prédictions des réussites/échecs - Épreuve MEQ écriture 4e primaire, Français langue d'enseignement



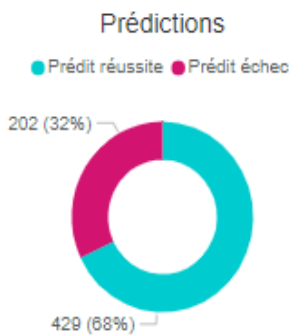
2021-2022, en début de 2e étape

Moment de la prédiction

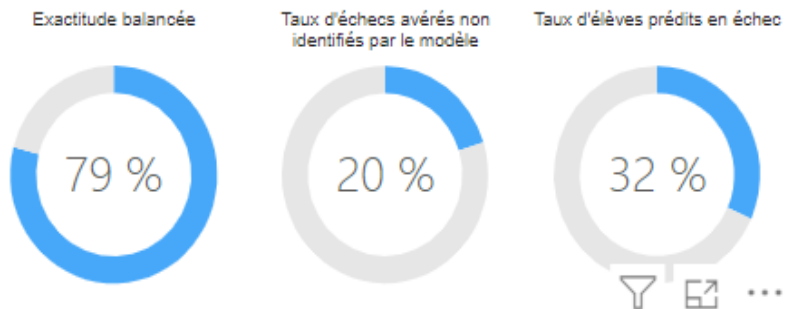
Dernière prédiction

École, groupe-repère

Tout



Performance du modèle prédictif basée sur l'historique



Prédictions par école



Annexe 6¹³

Prédictions des échecs - Épreuve MEQ écriture 4e primaire, Français langue d'enseignement



2018-2019, en début de 3e étape

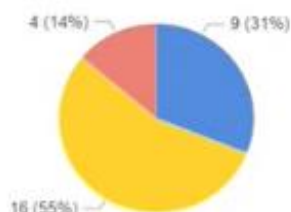
Moment de la prédiction

2018-2019, en début de 3e étape

École, groupe-repère

Tout

Probabilités d'échec



Prédit échec (élevé)
Probabilité d'échec > 84 %

4

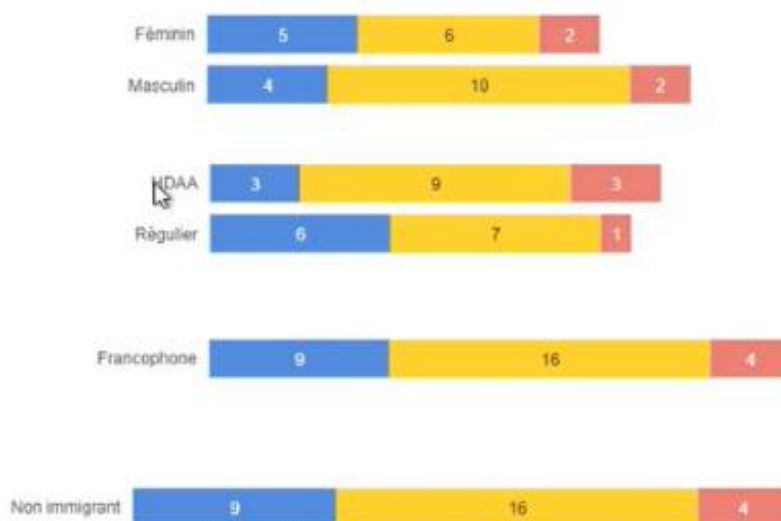
Prédit échec (moyen)
Probabilité d'échec > 64 % et <= 84 %

16

Prédit échec (faible)
Probabilité d'échec >= 50 % et <= 64 %

9

Répartitions des probabilités d'échecs des élèves de la 4e année du primaire selon différentes ventilations



École	Groupe-repère	Fiche	Nom et prénom	Probabilité d'échec
097 - École primaire Bedford	401	763839	Soucy-Turcotte Johanne	Prédit échec (élevé) ▲
097 - École primaire Bedford	403	780429	Cyr Arno	Prédit échec (élevé) ▲
097 - École primaire Bedford	403	783076	Falardeau Angélique	Prédit échec (élevé) ▲
097 - École primaire Bedford	403	838433	St-Laterreur Frédéric	Prédit échec (élevé) ▲
097 - École primaire Bedford	401	763995	Langlois Francis	Prédit échec (moyen) ▲
097 - École primaire Bedford	402	764142	Belleau Marcel	Prédit échec (moyen) ▲
097 - École primaire Bedford	402	767079	Leclerc Samuel-Nathan	Prédit échec (moyen) ▲

¹³ Les noms dans ce tableau sont fictifs.

enap.ca

QUÉBEC

MONTRÉAL

OUTAOUAIS

SAGUENAY

TROIS-RIVIÈRES

