

PROVISOIRE : Qu'est-ce qui se cache derrière les points et les coches ?

Les résultats de l'élève à l'évaluation sont indiqués selon une échelle à quatre niveaux (de 1 à 4) en lecture, en écriture et en mathématiques.

Nous incluons également une représentation visuelle de la performance de l'élève en mathématiques et en lecture sous forme d'un point indiquant le niveau auquel se situe l'élève. Nouveau : le niveau en écriture sera aussi représenté par un point pour l'année scolaire 2022–2023.

Si, par exemple, on a un point complètement à la gauche du niveau 3, cela signifie que l'élève se situe à un bas niveau 3, s'approchant d'un niveau 2, en lecture.

Rendement en lecture	Manque d'information	Niveau 1 au-dessous des attentes	Niveau 2 se rapproche des attentes	Niveau 3 répond aux attentes	Niveau 4 surpasse les attentes
					

Si on a un point complètement à la droite du niveau 3, cela signifie que l'élève se situe à un haut niveau 3, s'approchant d'un niveau 4, en lecture.

Rendement en lecture	Manque d'information	Niveau 1 au-dessous des attentes	Niveau 2 se rapproche des attentes	Niveau 3 répond aux attentes	Niveau 4 surpasse les attentes
					

Ce document fournit une brève description du processus de correction des évaluations, puis explique la façon dont on a obtenu les scores en lecture et en mathématiques et enfin les scores en écriture.

Correction des évaluations et des examens

Après le déroulement des évaluations ou des examens, les écoles mettent les cahiers d'évaluation dans des boîtes, qu'elles renvoient au ministère de l'Éducation et du Développement de la petite enfance. Au ministère, nous comptons et organisons des milliers de livrets. Dans le cas de sessions de correction en ligne, nous découpons et numérisons les livrets (ce processus sera décrit plus tard). Dans le cas de sessions de correction en personne, nous mélangeons et comptons les livrets pour ensuite les envoyer à une session centrale de correction.

Lors des sessions de correction, les enseignants corrigent les questions à réponses construites (deux tâches de rédaction pour les évaluations et examens en littérature et plusieurs tâches pour ceux de Mathématiques 10). Après une formation approfondie portant sur la grille de correction, qui s'appuie sur des exemples présélectionnés de copies d'élèves, les enseignants se chargent d'attribuer des scores aux évaluations et aux examens. Aux moins deux enseignants attribuent un score à chaque tâche de rédaction. Si les scores ne sont pas adjacents (p. ex., si un livret reçoit un 2 et un 4 pour les Idées de la tâche 1), un troisième enseignant ou une chef de table tranche la question.

Lors du processus de numérisation, nous coupons la reliure de tous les livrets et nous les numérisons à l'aide de numériseurs à haut débit.

Le logiciel de numérisation examine chaque question pour laquelle l'élève a noirci une bulle et détermine s'il a choisi A, B, C ou D. Le logiciel examine également les scores que les enseignants ont indiqués pour l'écriture lors des sessions de correction en personne.

Lorsqu'il n'y a pas de réponse à la question ou qu'il semble y avoir plusieurs réponses, le logiciel envoie une image de la question concernée à un coordonnateur d'évaluation, qui se charge de faire une vérification. Cette personne peut vérifier si l'élève a répondu à la question en ne noircissant que légèrement la réponse, en cochant la bulle au lieu de la noircir ou en marquant une croix. Elle s'occupe ensuite de mettre à jour la réponse afin d'indiquer ce qu'elle voit sur l'image.

Pour chaque question, il n'y a qu'une réponse correcte. Si l'élève n'a vraiment pas répondu ou a noirci plusieurs réponses sans indiquer celle qu'il préférerait, on considère que la réponse de l'élève est incorrecte.

Si le nombre de questions sans réponse est trop élevé, on considère qu'il est impossible de faire une évaluation équitable du niveau de l'élève et on indique alors la mention de « pas de données suffisantes ».

Pourquoi y a-t-il ui

☒ C'est pour dire

☐ C'est pour indi

☒ C'est pour mel

☐ C'est pour sigr

pour dire

Cet élève a noirci deux réponses, mais a également écrit l'une des réponses en dessous.

La copie a été examinée par un coordonnateur lors du processus de numérisation et celui-ci a indiqué que la réponse fournie par écrit était celle que l'élève voulait choisir.

Vérification de la qualité : examen des items de l'évaluation

Une fois que les cahiers d'évaluation et d'examen ont été numérisés et corrigés, nous examinons les résultats produits par les questions en procédant à une analyse des items. On considère que les questions ont produit de bons résultats quand elles sont claires, que la bonne réponse est claire et que les élèves qui ont obtenu globalement de bons résultats pour l'ensemble du test ont obtenu de bons résultats pour cette question en particulier. Les questions qui n'ont pas produit de bons résultats (par exemple, les questions pour lesquelles il semble y avoir deux réponses correctes ou qui sont énoncées de façon confuse) sont éliminées de l'évaluation.

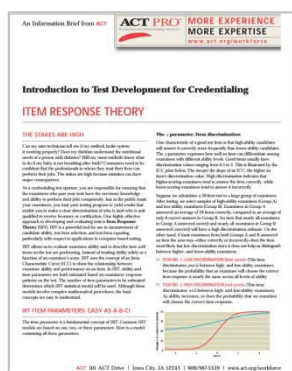
Les questions qui sont très faciles et celles qui sont très difficiles peuvent fournir des informations très utiles sur le niveau des élèves. Les questions très faciles permettent d'obtenir des informations sur les élèves qui se situent à un niveau d'aptitude plus faible et les questions très difficiles permettent d'obtenir des informations sur les élèves qui se situent à un niveau d'aptitude plus élevé. Nous essayons d'inclure un vaste éventail de questions, afin de prendre en compte tous les niveaux d'aptitude chez les élèves.

Nous calculons le score de l'élève en nous fondant exclusivement sur les questions qui répondent à nos lignes directrices pour les résultats.

Le score de l'élève en lecture et en mathématique est calculé à partir des trois paramètres de la théorie de la réponse d'item (TRI). Le score représente l'aptitude de l'élève telle qu'elle est mesurée par l'évaluation. Le calcul est très différent du calcul effectué pour obtenir le « pourcentage de réponses correctes » qu'on utilise traditionnellement pour la correction des tests.

Dans la TRI, le score dépend non seulement du nombre de questions pour lesquelles l'élève a donné la réponse correcte, mais aussi des *caractéristiques de chaque question* (par exemple, du niveau de difficulté de la question).

La méthode de calcul du score est très différente avec la TRI, mais le score attribué à l'élève au bout du compte est généralement très proche du score que l'élève recevrait si on comptabilisait le nombre de réponses correctes. C'est parce que les deux méthodes sont bonnes pour mesurer le niveau d'aptitude de l'élève.



L'article ci-contre fournit une bonne introduction sur la TRI :

[Item Response Theory \(IRT\): An Introduction
Nathan Thompson, PhD \(2016\)
https://assess.com/item-response-theory/](https://assess.com/item-response-theory/)

Lecture et mathématiques : fixer des points de coupure pour chaque niveau

Dans la TRI, on attribue un score non seulement à chaque ÉLÈVE, mais également à chaque ITEM de l'évaluation. Ce score représente la difficulté de l'item, mais aussi l'aptitude de l'élève¹.

Une fois qu'on a vérifié la qualité des items et fait l'analyse de la TRI, on produit un cahier avec un classement des items. Ce cahier fournit un classement de tous les items de l'évaluation, en allant du moins difficile au plus difficile.

Nous invitons ensuite des enseignants et d'autres spécialistes de l'éducation à une session lors de laquelle on définit les attentes. Les participants examinent attentivement chaque item dans le cahier, en prenant en compte la difficulté de l'item et les processus de réflexion sur lesquels l'élève doit s'appuyer pour produire la réponse correcte (*exigence cognitive*).

La difficulté et l'exigence cognitive sont liées, mais il ne faut pas les confondre. Quand on a, par exemple, une question qui demande à l'élève de trouver un mot dans un texte à lire, le processus cognitif concerné est simple (extraction). Si le mot se trouve dans un texte long et compliqué, cependant, il est possible que la question soit malgré tout assez difficile.

Nous fournissons également aux participants des descriptions détaillées des différents niveaux. Ces descriptions indiquent les types de connaissances, de compétences et d'aptitudes sur lesquelles les élèves doivent s'appuyer pour répondre aux questions à chaque niveau. Au niveau 1, par exemple, l'élève doit être capable d'extraire des informations du texte directement. Au niveau 4, il doit être capable d'établir des liens compliqués entre différents éléments d'informations dans le texte.

Avec ces descriptions et le cahier contenant le classement des items, chaque participant insère un signet à l'endroit où il pense qu'on passe du niveau 1 au niveau 2, à l'endroit où il pense qu'on passe du niveau 2 au niveau 3 et à l'endroit où il pense qu'on passe du niveau 3 au niveau 4. Au bout du compte, ce sont ces signets fournis par les enseignants et les spécialistes de l'éducation qui constituent la base sur laquelle on s'appuie pour définir le point de coupure pour chaque niveau.

Comme les points de coupure se fondent sur les scores concernant les aptitudes et non sur le nombre d'items auxquels les élèves ont répondu correctement, il est possible de faire facilement des comparaisons entre items, évaluations/examens et résultats des élèves d'une année à l'autre.

Attribution des niveaux

Une fois qu'on a les points de coupure indiquant les scores auxquels on passe d'un niveau au suivant, il est relativement facile d'attribuer un niveau à chaque élève. Si, par exemple, le point de coupure entre le niveau 3 et le niveau 4 correspond au score 1,3 et si l'élève a un score de 1,5, alors l'élève se situe au niveau 4.

Il est cependant clair, par ailleurs, qu'il y a tout un éventail d'aptitudes à chaque niveau. Certains élèves se situant au niveau 3 sont plus près du niveau 2 et certains sont plus près du niveau 4.

¹. On définit généralement la difficulté d'un item, dans la TRI, comme étant le score auquel l'élève a 50 p. 100 de chances de fournir la réponse correcte.

Lecture et mathématiques : placer le point pour chaque élève

L'éventail des aptitudes varie d'un niveau à l'autre. Il y a, par exemple, relativement peu d'élèves au niveau 4, mais ces élèves ont un vaste éventail d'aptitudes. Il y a beaucoup plus d'élèves au niveau 3, mais l'éventail des aptitudes qu'ils représentent est nettement moins vaste.

La situation est semblable à ce qui arrive pour la taille des êtres humains : la plupart des gens se situent à quelques centimètres de la moyenne, mais l'éventail des tailles est très vaste et il y a quelques personnes qui sont extrêmement grandes et d'autres, extrêmement petites.

Pour chaque élève, nous examinons, pour placer le point, l'éventail des aptitudes à chaque niveau et nous situons le score de l'élève dans cet éventail. Si le score de l'élève tombe au milieu, alors le point se trouvera au milieu du niveau. Si l'élève se situe vers le maximum, alors le point se trouvera à proximité du niveau suivant.

Il y a plus d'une centaine d'endroits auxquels on peut placer le point à chaque niveau. (Il y a des milliers de scores différents dans la province.) Si le point touche le bord supérieur du niveau 4 ou le bord inférieur du niveau 1, alors c'est qu'on a atteint les limites de la représentation graphique. Il est possible que l'élève ait un niveau d'aptitude qui va au-delà de l'éventail représenté par la figure, mais nous pensons que l'intervalle choisi pour cette figure est une bonne représentation de la variété des niveaux d'aptitude des élèves².



Sun Mingming, joueur de basketball le plus grand au monde, qui mesure 2,36 m

Lecture et mathématiques : scores pour les volets

Pour les volets (p. ex. genres, domaines d'études), la coche ne peut se situer qu'à deux endroits : à gauche (« pourrait exiger une attention supplémentaire ») ou à droite (« répond aux attentes »).

peut nécessiter plus d'attention	répond ou surpasse les attentes
✓	
	✓

C'est parce qu'il y a beaucoup moins de points dans les volets individuels que dans le score global à l'évaluation.

Nous fixons les attentes du programme d'études pour chaque volet, en nous appuyant sur de multiples sources : jugements portés sur l'aptitude dans le score global à partir de la session de placement des signets, jugement professionnel des spécialistes, examen attentif des programmes d'études, consultations, etc.

² Nous avons choisi un intervalle de trois écarts-types par rapport à la moyenne. Tous les scores au-delà de +3 ou de -3 sont arrondis à +3 ou -3.

Nous calculons le score de l'élève dans chaque sous-domaine et le comparons au point de coupure. S'il se situe au-dessus du point de coupure, nous cochons la colonne « répond aux attentes ». S'il se situe en dessous, nous cochons la colonne « pourrait exiger une attention supplémentaire ».

Écriture : score pour chaque volet

Dans les évaluations en lecture et en écriture, il y a deux tâches de rédaction. Pour obtenir un score en écriture, il faut que l'élève ait tenté de faire les deux tâches et ait écrit suffisamment de choses pour qu'on puisse corriger chaque tâche. Sinon, on attribue à l'élève la mention « pas de données suffisantes » en écriture.

Les deux tâches de rédaction sont corrigées par les enseignants à l'aide de la grille analytique (disponible sur notre site Web à <http://plans.ednet.ns.ca>). Cette grille se fonde sur quatre catégories : idées, organisation, utilisation de la langue et convention. Dans chacune de ces catégories, on attribue à l'élève un score de 1 à 4.

Chacune des deux tâches de rédaction est corrigée séparément par deux enseignants différents. Le deuxième enseignant ne voit pas les scores attribués par le premier, ce qui garantit l'indépendance des deux corrections. Pour garantir l'équité et la fiabilité des résultats, si deux scores d'écriture sont très différents pour une composante (séparés par plus d'un niveau), un troisième enseignant ou un chef de table attribue un score et les deux scores les plus près sont conservés.

Comme au moins deux enseignants examinent chacune des deux tâches, il y a quatre jugements distincts portés sur les aptitudes de l'élève en écriture dans chaque catégorie. Pour produire un score pour chaque catégorie, nous calculons la médiane des scores. Si la médiane tombe entre deux niveaux, nous arrondissons au niveau supérieur, parce que nous savons que l'élève a au moins deux scores supérieurs à la médiane.

Voici un exemple fictif de résultats d'un élève en écriture et la façon dont le score est calculé pour chaque composante.

Catégorie	Tâche 1		Tâche 2		Médiane	Score final
	Premier correcteur	Deuxième correcteur	Premier correcteur	Deuxième correcteur		
Idées	2	3	3	4	3	Niveau 3
Structure du texte	3	4	4	4	4	Niveau 4
Style	1	2	3	3	2,5	Niveau 3
Conventions de l'écrit	3	4	4	4	4	Niveau 4

Ci-dessous se trouve un exemple de la façon dont ces scores étaient rapportés jusqu'à l'année scolaire 2021–2022. Chaque critère individuel était affiché séparément avec un niveau.

Rendement en écriture	Manque d'information	Niveau 1 au-dessous des attentes	Niveau 2 se rapproche des attentes	Niveau 3 répond aux attentes	Niveau 4 surpasse les attentes
Idées				✓	
Structure du texte					✓
Style				✓	
Conventions de l'écrit					✓

Nouveau en 2022–2023 : Le rendement en écriture des élèves recevra un score global (avec un point semblable à ceux en lecture et en mathématiques) et les scores des composantes porteront la mention « pourrait exiger une attention supplémentaire » (niveau 1 et 2) ou « répond aux attentes ou surpasse les attentes » (niveau 3 et 4).

L'attribution de scores en écriture suivra le même processus qu'auparavant, mais après la session d'attribution de scores, nous utiliserons la TRI pour produire un score rapporté à une échelle pour chaque élève en écriture. Nous organiserons aussi des sessions pour fixer les attentes afin de fixer les notes de passage en écriture. Les avantages de ce nouveau système sont notamment :

- plus simple interprétation des scores en écriture au fil du temps,
- accent sur l'écriture comme processus global tout en fournissant de l'information détaillée sur les quatre composantes de l'écriture aux régions et aux écoles,
- résultats plus fiables vu que le score global est composé d'un plus grand nombre de scores,
- capacité à évaluer l'écriture au fil du temps.

Nouveau rapport des élèves proposé (comparable à ceux en lecture et en mathématiques)

Rendement en écriture

L'élève a fait preuve d'un rendement de niveau 3 dans cette évaluation.

Les élèves qui écrivent au niveau 3 peuvent généralement communiquer clairement par écrit. Ils présentent une idée principale claire avec des détails pertinents et structurent leur texte en utilisant correctement les transitions. Leur langage contribue à une écriture claire et fluide et ils utilisent généralement correctement les conventions.

Rendement en écriture	Manque d'information	Niveau 1 au-dessous des attentes	Niveau 2 se rapproche des attentes	Niveau 3 répond aux attentes	Niveau 4 surpasse les attentes

Le tableau ci-dessous montre le rendement de l'élève pour les composantes d'écriture de cette évaluation. Cette information peut aider les enseignants à examiner de plus près le rendement de l'élève dans les différentes composantes de la lecture.

Volets de l'écriture	Manque d'information	pourrait exiger une attention supplémentaire	répond ou surpasse les attentes
Idées			✓
Structure du texte		✓	
Style			✓
Conventions de l'écrit		✓	

Pour de plus amples renseignements

Veuillez consulter notre site Web (<http://plans.ednet.ns.ca>) ou communiquer avec Sarah Curry (sarah.curry@novascotia.ca).