



## Explications relatives aux rapports classe et école au niveau de l'enseignement fondamental et secondaire

<b>1</b>	<b>Explications sur les compétences</b> .....	<b>2</b>
1.1	Que mesurent les ÉpStan? Qu'est-ce qu'elles ne mesurent pas ? .....	2
1.2	Comment sont évaluées les compétences? .....	7
1.3	Quelle est la différence entre les ÉpStan et les devoirs en classe .....	7
<b>2</b>	<b>Quelle est la fiabilité des ÉpStan? Résultats d'étude issus de l'enseignement fondamental</b> .....	<b>8</b>
<b>3</b>	<b>Évaluation des résultats</b> .....	<b>9</b>
3.1	La métrique ÉpStan.....	9
3.2	Qu'est-ce que l'intervalle de résultats attendus ? .....	9
<b>4</b>	<b>Explication sur la motivation à apprendre</b> .....	<b>11</b>
4.1	Motivation à apprendre par matière .....	11
4.2	Motivation générale à apprendre .....	12
4.3	Climat au niveau de la classe et au niveau de l'école .....	12
<b>5</b>	<b>Références</b> .....	<b>12</b>

# 1 Explications sur les compétences

## 1.1 Que mesurent les ÉpStan? Qu'est-ce qu'elles ne mesurent pas ?

Les ÉpStan consistent à évaluer les performances des élèves dans certains domaines de compétence à l'aide de tests standardisés. Les épreuves utilisées dans le cadre de ces tests de compétence se basent sur les standards de formation élaborés par le MENJE pour les différents cycles de l'enseignement fondamental et secondaire. Dans ces standards éducatifs, les compétences sont définies comme suit : « *la capacité de mettre en œuvre un ensemble organisé de connaissances, d'habiletés et d'attitudes qu'un élève doit mobiliser dans un contexte donné lui permettant de fournir une réponse adéquate à une problématique complexe* » (Jonnaert, cité dans MENFP, 2007, p. 3).

Suivant cette définition, la meilleure façon d'évaluer les compétences des élèves est de leur présenter des exercices comportant des problèmes complexes. Or, que signifie exactement « complexe » dans ce contexte ? La résolution de « problèmes complexes » requiert l'utilisation intégrée de connaissances et de compétences, du coup, elle est donc généralement longue, exigeante et nécessitant une grande motivation de la part des élèves.

Dans le cadre des ÉpStan, nous distinguons différents degrés de complexité, en fonction des exigences du problème à résoudre. Ainsi, l'objectif d'une tâche de lecture peut consister à localiser et à nommer des informations explicitées dans le texte (complexité moindre) ou à tirer une conclusion à partir de ce qui a été lu (complexité plus élevée).

Il convient également de souligner que les tests de compétence ÉpStan n'évaluent pas toutes les compétences définies par le ministère de l'éducation nationale dans le plan d'études (2011). En effet, d'une part, les formats de tâches standardisés ne permettent pas d'évaluer tous les domaines de compétence. D'autre part, le temps disponible pour réaliser les ÉpStan est très limité. Pour ces raisons, les ÉpStan se composent d'une sélection représentative de compétences clés pouvant être évaluées de la manière la plus économique et standardisée possible.

Évidemment, les exercices ÉpStan qui visent la compréhension de l'écrit et, le cas échéant, la compréhension de l'oral, ne permettent d'évaluer que deux domaines partiels de la compétence linguistique des élèves. Ces deux domaines de compétence jouent toutefois un rôle central dans le quotidien scolaire et d'apprentissage des élèves, dans la mesure où seule la capacité de réception de la langue écrite permet d'accéder à un enseignement supérieur. En outre, la compréhension de l'écrit et, le cas échéant, la compréhension de l'oral peuvent être mesurées de manière relativement économique, hautement standardisée et objective.

### Compétences évaluées au cycle 2.1:

- **Mathématiques au cycle 2.1.** Les épreuves de mathématiques portent sur le contenu des domaines (a) « espace et formes » (MENFP, 2011, p. 26, 116), (b) « nombres et opérations » (MENFP, 2011, p. 28, 117), et (c) « grandeurs et mesures » (MENFP, 2011, p. 30, 118).

- **Précurseurs de la compréhension de l'écrit au cycle 2.1.** Les épreuves de test visant à évaluer ces compétences préliminaires (MENFP, 2011, p. 8) portent principalement sur les compétences suivantes : conscience phonologique, attention auditive, discrimination visuelle, et première lecture et écriture (p. ex. lettres et mots simples d'une ou de deux syllabes).
- **Compréhension de l'oral en luxembourgeois au cycle 2.1.** Les épreuves de test visant à évaluer la compréhension orale en luxembourgeois (MENFP, 2011, p. 6) se réfèrent à différents types de textes oraux (p. ex. dialogue, récit, reportage). Les exercices présentés couvrent différentes compétences (cf. MENFP, 2011, p. 6) telle que comprendre son interlocuteur, localiser et comprendre des informations ou encore, utiliser des stratégies d'écoute (p. ex. reconnaître des bruits de fond).
- **Compréhension de l'oral en allemand au cycle 2.1.** Dans l'épreuve de compréhension de l'oral en allemand du cycle 2 (MENFP, 2011, p. 6), les élèves sont confrontés à différents formats de texte audio (p.ex. conversations, histoires courtes ou émissions radio pour enfants). On leur pose des questions sur ces textes d'écoute qui visent différentes compétences (cf. MENFP, 2011, p. 12) : p.ex. suivre une conversation et comprendre un interlocuteur (qui donne des instructions), comprendre globalement des textes d'écoute (comme dans cette question : Que font les enfants dans l'histoire ?), localiser et comprendre des informations plus ciblées (comme dans la question : Où est-ce que les enfants se rencontrent ?) et employer des stratégies d'écoute (comme p.ex. reconnaître des bruits de fond).

### Compétences évaluées au cycle 3.1:

- **Mathématiques au cycle 3.1.** Les exercices de mathématiques portent sur le contenu (a) du domaine « nombres et opérations » (MENFP, 2011, p. 28-29, 121) et (b) d'un domaine combiné « espace et formes » (MENFP, 2011, p. 26-27, 120) et « grandeurs et mesures » (MENFP, 2011, p. 30-31, 122). On distingue également deux compétences mathématiques générales qui visent (a) la « résolution de problèmes et la modélisation » (cf. MENFP, 2011, p. 115) et (b) les « compétences de base spécifiques » (cf. MENFP, 2008, p. 11). Ces dernières sont définies comme des connaissances et des compétences mathématiques qui s'appliquent de manière isolée, hors contexte et sans aucun travail de transfert. La distinction entre les exercices contextualisés (résolution de problèmes et modélisation) et les exercices décontextualisés (compétences de base spécifiques) permet de couvrir implicitement le domaine de contenu « résolution de problèmes d'arithmétique » (MENFP, 2011, p. 32-33, 123).
- **Compréhension de l'écrit en allemand au cycle 3.1.** Les épreuves de test visant à évaluer les compétences en lecture de l'allemand (MENFP, 2011, p. 16-17) portent sur deux types de textes différents : (a) les textes continus (p. ex. récit, conte, texte factuel) et (b) les textes discontinus (p. ex. recette, instructions). Elles sont représentées (comme pour PISA) dans un rapport de deux à un. L'élaboration du test distingue en outre deux sous-compétences qui visent à (a)

« localiser et comprendre des informations » (cf. MENFP, 2011, p. 82-83) et à (b) « interpréter des informations et connaître et utiliser des stratégies de lecture » (cf. MENFP, 2011, p. 82-83).

- **Compréhension de l'oral en allemand au cycle 3.1.** Les épreuves de test destinées à évaluer les compétences d'écoute en allemand (MENFP, 2011, p. 12-13) portent sur différents types de textes audio (p. ex. dialogue, reportage, pièce radiophonique). L'élaboration du test distingue les sous-compétences qui visent (a) à « localiser et comprendre des informations » (cf. MENFP, 2011, p. 79-80) et (b) à « interpréter des informations et connaître et utiliser des stratégies d'écoute » (cf. MENFP, 2011 p. 79-80).

#### **Compétences évaluées au cycle 4.1:**

- **Mathématiques au cycle 4.1.** Les exercices de mathématiques portent sur le contenu (a) du domaine « nombres et opérations » (MENFP, 2011, p. 28-29, 125) et (b) d'un domaine combiné « espace et formes » (MENFP, 2011, p. 26-27, 124) et « grandeurs et mesures » (MENFP, 2011, p. 30-31, 126). On distingue également deux compétences mathématiques générales qui visent (a) la « résolution de problèmes et la modélisation » (cf. MENFP, 2011, p. 115) et (b) les « compétences de base spécifiques » (cf. MENFP, 2008, p. 11). Ces dernières sont définies comme des connaissances et des compétences mathématiques qui s'appliquent de manière isolée, hors contexte et sans aucun travail de transfert. La distinction entre les exercices contextualisés (résolution de problèmes et modélisation) et les exercices décontextualisés (compétences de base spécifiques) permet de couvrir implicitement le domaine de contenu « résolution de problèmes d'arithmétique » (MENFP, 2011, p. 32-33, 127).
- **Compréhension de l'écrit en allemand au cycle 4.1.** Les épreuves de test visant à évaluer les compétences en lecture de l'allemand (MENFP, 2011, p. 16-17) portent sur deux types de textes différents : (a) les textes continus (p. ex. récit, poème, texte factuel) et (b) les textes discontinus (p. ex. reportage, commentaire). Elles sont représentées (comme pour PISA) dans un rapport de deux à un. L'élaboration du test distingue en outre deux sous-compétences qui visent à (a) « localiser et comprendre des informations » (cf. MENFP, 2011, p. 88-89) et à (b) « interpréter des informations et connaître et utiliser des stratégies de lecture » (cf. MENFP, 2011, p. 88-89).
- **Compréhension de l'écrit en français au cycle 4.1.** Les épreuves de test visant à évaluer les compétences en lecture de français (MENFP, 2011, p. 24-25) portent sur deux types de textes différents : (a) les textes continus (p. ex. récit, conte) et (b) les textes discontinus (p. ex. recette, annonce). Elles sont représentées (comme pour PISA) dans un rapport de deux à un. L'élaboration du test distingue en outre deux sous-compétences qui visent à (a) « localiser et comprendre des informations » (cf. MENFP, 2011, p. 104-105) et à (b) « interpréter des informations et connaître et utiliser des stratégies de lecture » (cf. MENFP, 2011, p. 104-105).

## Compétences évaluées en 7<sup>e</sup>:

- Mathématiques en 7<sup>e</sup>.** Le test de mathématiques en 7<sup>e</sup> couvre les compétences à atteindre à la fin du quatrième cycle d'apprentissage de l'enseignement fondamental dans les quatre sous-domaines suivants : « nombres et opérations » (MENFP, 2011, p. 28-29, 129), « espace et forme » (MENFP, 2011, p. 26-27, 128), « grandeurs et mesures » (MENFP, 2011, p. 30-31, 130) et « résolution de problèmes d'arithmétique » (MENFP, 2011, p. 32-33, 131). Le test se compose de 23 à 28 problèmes à texte réduit, présentés en français ou en allemand (les élèves peuvent changer de langue à tout moment et ainsi choisir la langue qui leur semble la plus appropriée pour résoudre le problème en question). La plupart des problèmes présentent des formats semi-ouverts ou à choix multiples. Ils visent à explorer dans quelle mesure les élèves sont capables d'appliquer correctement les contenus enseignés dans des situations mathématiques connues ainsi que nouvelles.
- Compréhension de l'écrit en allemand en 7<sup>e</sup>.** Le test de compréhension de l'écrit en allemand couvre les deux sous-compétences : « repérer les informations d'un texte et les exploiter » (compétence 1 - C1) et « interpréter des informations et mobiliser des stratégies / techniques de lecture » (compétence 2 - C2) (MENFP, 2011, p. 16-17). Le test se compose de quatre à six textes continus (p. ex. des récits ou des textes factuels) ou discontinus (p. ex. des tableaux, des illustrations ou des cartes), d'une longueur de 100 à 550 mots environ chacun, qui abordent des thématiques liées à la vie quotidienne des élèves, tels que les loisirs, la société, l'amitié ou la famille. La difficulté du test se base sur les consignes du cycle d'apprentissage précédent, à savoir le cycle 4 de l'enseignement fondamental.
- Compréhension de l'écrit en français en 7<sup>e</sup>.** Le test de compréhension de l'écrit en français couvre les deux sous-compétences : « repérer les informations d'un texte et les exploiter » (compétence 1 - C1) et « interpréter/analyser des informations et mobiliser des stratégies / techniques de lecture » (compétence 2 - C2) (MENFP, 2011, p. 22-23). Le test se compose de cinq à sept textes continus (p. ex. des histoires simples, des récits ou des textes factuels) ou discontinus (p. ex. des illustrations ou des annonces), d'une longueur de 50 à 500 mots environ chacun. Les textes abordent des thématiques du quotidien des élèves, telles que la famille, l'école ou l'amitié. La difficulté du test se base sur les consignes du cycle d'apprentissage précédent, à savoir le cycle 4 de l'enseignement fondamental.

La durée de test est de 45 minutes par matière pour tous les élèves, tandis que l'ordre de présentation varie.

## Compétences évaluées en 5<sup>e</sup>:

- Mathématiques en 5<sup>e</sup>.** Le test de compétence en mathématiques met l'accent sur des exercices qui traitent des modèles et des problèmes mathématiques, dont le degré de complexité peut varier considérablement. Tous les exercices utilisés peuvent être classés dans l'un des trois domaines de contenu mathématique : (1) nombres et opérations (MENFP 2008b ; 2008c, p. 21-23), (2)

plan et figures spatiales (MENFP 2008b ; 2008c, p. 17-19) ou (3) dépendance et variation (MENFP 2008b ; 2008c, p. 25-27). Les élèves peuvent traiter les exercices de mathématiques en français ou en allemand.

**Compréhension de l'écrit en allemand en 5<sup>e</sup>.** Le test de compréhension de l'écrit en allemand (MENFP 2008a, S. 34-35, 39-41, 50-52, 68) couvre les deux sous-compétences : « repérer les informations d'un texte et les comprendre/exploiter » (compétence 1 - C1) et « analyser et interpréter des informations et en tirer des conclusions (basées sur les connaissances) » (compétence 2 - C2). Ces sous-compétences peuvent être déduites de la description des quatre niveaux d'apprentissage compris dans le plan d'études (cf. MENFP 2008a, p. 68). Chaque test de lecture en allemand se compose de cinq à huit textes continus (p. ex. des récits ou des textes factuels) ou discontinus (p. ex. des tableaux, des illustrations ou des instructions de construction), d'une longueur de 100 à 750 mots environ chacun, qui abordent des thématiques telles que les loisirs, la société, l'amitié ou la formation professionnelle.

- **Compréhension de l'écrit en français en 5<sup>e</sup>.** Le test de compréhension de l'écrit en français se base également sur les deux sous-compétences : « repérer les informations d'un texte et les comprendre/exploiter » (compétence 1 - C1) et « analyser et interpréter des informations et en tirer des conclusions (basées sur les connaissances) » (compétence 2 - C2). En effet, pour le français également, ces sous-compétences peuvent être déduites de la description des quatre niveaux d'apprentissage compris dans le plan d'études (MENFP 2008a, p. 68). Chaque test de lecture en français se compose de quatre à sept textes continus (p. ex. des récits ou des textes factuels) ou discontinus (p. ex. des tableaux ou des illustrations), d'une longueur de 100 à 750 mots environ chacun, qui abordent des thématiques telles que l'amitié, les loisirs ou la formation professionnelle.

Pour les trois types d'enseignement « Enseignement secondaire classique », « Enseignement secondaire général - voie d'orientation » et « Enseignement secondaire général - voie de préparation », des versions de test différentes sont créées. Or, elles contiennent toujours un nombre minimum de textes et d'exercices se recoupant afin de garantir leur comparabilité.

En conséquence, trois versions de test sont créées pour chacun des trois domaines de compétence. Les versions de test sont nommées en fonction du type d'enseignement :

- *Enseignement secondaire classique (ESC)*
- *Enseignement secondaire général – voie d'orientation (ESG)*
- *Enseignement secondaire général – voie de préparation (ESG-VP)*

Chaque version de test se compose d'une part d'exercices spécialement conçus pour le niveau de performance moyen du type d'enseignement en question. D'autre part, chaque version de test contient toujours un nombre minimum d'exercices qui sont traités par tous les élèves. De cette manière, il est possible d'effectuer des comparaisons de compétences entre les différents types d'enseignement à l'aide de méthodes psychométriques modernes.

Quelle que soit la version du test, la durée du test pour tous les élèves est de 45 minutes par matière. Cependant, l'ordre de présentation des tests varie de manière aléatoire.

Les ÉpStan fournissent donc un instantané très standardisé mais incomplet, alors que les bilans ou les notes de bulletin reposent sur une évaluation complète et à long terme des performances scolaires. Par conséquent, les résultats des ÉpStan sont à considérer comme étant complémentaires et non pas en concurrence avec les bilans ou les notes de bulletin.

## 1.2 Comment sont évaluées les compétences ?

Dans une première phase, chaque exercice est attribué à un niveau de difficulté théorique (p. ex. « inférieur au niveau *Socle* », « niveau *Socle* », « niveau *Avancé* », « supérieur au niveau *Avancé* »). Cette première classification des exercices est effectuée dans les groupes de développement de tests en tenant strictement compte des descripteurs respectifs (MENFP, 2011 & MENFP 2008a ; 2008b ; 2008c) et avant même que des indicateurs empiriques de difficulté ne soient disponibles pour les exercices en question. Dans une deuxième phase, après la collecte des données, les difficultés théorique et empirique des exercices sont comparées entre elles. Si elles ne concordent pas pour un exercice, soit une reclassification théorique est effectuée - si cela semble justifiable sur la base des descripteurs théoriques relatifs aux différents domaines et niveaux de compétence - soit l'exercice en question n'est pas pris en compte dans le calcul final des socles ou des niveaux d'apprentissage.

Cette procédure, présentée ici de manière très simplifiée, permet aux groupes de développement de tests de s'assurer que les exercices considérés comme difficiles ou simples de manière théorique sont également difficiles ou simples de manière empirique. De plus, cela permet de s'assurer que les difficultés empiriques des niveaux d'apprentissage respectifs sont relativement similaires, de sorte que, par exemple, tous les exercices permettant de calculer le niveau *Socle* ou le niveau d'apprentissage 2 en mathématiques se situent à peu près dans la même plage de difficulté empirique.

## 1.3 Quelle est la différence entre les ÉpStan et les devoirs en classe

Les devoirs en classe traditionnels portent en premier lieu sur la matière enseignée en classe, que les élèves sont censés maîtriser. Dans ce format de test, on part du principe que les élèves qui ont suivi les cours sont capables de résoudre tous les exercices. Si tous les exercices ont été résolus de manière correcte, les élèves satisfont à l'attente de base, à l'état théorique. En règle générale, les devoirs en classe ne vont toutefois pas au-delà de cet état théorique à vérifier, de sorte qu'ils n'offrent guère aux très bons élèves la possibilité de démontrer leurs connaissances et leurs compétences. Ce format de test ne permet que d'identifier les performances des élèves qui n'atteignent pas l'état théorique visé. Cette approche axée sur les déficits est très fortement ancrée dans notre culture des examens. Ainsi, la correction des devoirs en classe ne met guère d'accent sur ce que les élèves font correctement. Elle met surtout en évidence ce que les élèves ne réussissent pas (encore). En général, les devoirs en classe traditionnels ne considèrent moins ce que les élèves savent faire que leurs faiblesses ou de leur écart par rapport à l'état théorique visé.

En revanche, les ÉpStan couvrent l'éventail complet des exercices ayant des niveaux de difficulté variables : le niveau atteint par (presque) tout le monde, le niveau que tout

le monde devrait (au moins) atteindre et le niveau qui va (bien) au-delà. Par conséquent, il ne faut pas s'attendre à ce que tous les élèves résolvent correctement tous les exercices des ÉpStan. Au contraire, les élèves sont censés exploiter pleinement leurs compétences en traitant le plus grand nombre possible d'exercices (plus ou moins difficiles). L'objectif principal des ÉpStan n'est donc pas d'identifier les faiblesses des élèves, mais de déterminer leurs compétences (partielles).

Dans ce sens, l'approche d'évaluation adoptée par les ÉpStan repose clairement sur les principes de l'enseignement par compétences. Cela ne signifie pas pour autant que les exercices utilisés peuvent être considérés comme représentatifs de cette forme d'enseignement. La situation est un peu plus complexe.

Les ÉpStan visent à comparer les performances des élèves entre elles et avec un standard. Dans ce contexte, les domaines de compétence testés ont été choisis a) parce qu'ils sont d'une importance essentielle pour le quotidien à l'école et l'apprentissage et b) parce qu'ils peuvent être évalués de manière standardisée, objective et fiable avec relativement peu d'efforts. Il en va de même pour les formats des exercices choisis. Ainsi, les ÉpStan utilisent principalement des formats d'exercice fermés et semi-ouverts. En revanche, la situation d'enseignement met le focus sur le processus d'apprentissage des élèves qu'il convient d'encourager de préférence par des formats d'exercice ouverts.

## **2 Quelle est la fiabilité des ÉpStan? Résultats d'étude issus de l'enseignement fondamental.**

L'un des objectifs centraux des ÉpStan est d'évaluer les compétences des élèves de la manière la plus standardisée possible, c'est-à-dire de manière objective et fiable - en tenant compte du contexte donné. Des mesures telles que l'utilisation exclusive de formats d'exercice fermés et semi-ouverts, la mise à disposition d'instructions de correction détaillées (pour éviter une correction trop douce ou trop sévère) et d'un masque d'encodage numérique sont mises en place afin d'atteindre cet objectif. Il reste la question de savoir si des correcteurs différents parviennent aux mêmes évaluations de compétences pour chaque élève.

Pour y répondre, nous avons mené une étude complémentaire sur les ÉpStan 2010/11 (pour le cycle 3.1) et sur les ÉpStan 2015/16 (pour les cycles 2.1 et 3.1). Pour chaque année académique, un échantillon aléatoire de 30 cahiers d'exercices remplis a été sélectionné pour chacun des quatre types d'épreuve. Dans une première phase, toutes les corrections notées dans ces cahiers par le titulaire de la classe ont été rendues illisibles. Dans une deuxième phase, les 120 cahiers d'exercices ont été corrigés par un(e) deuxième correcteur•trice formé•e à cet effet, en suivant strictement les instructions de correction données dans le manuel. Au total, pour 3400 (ÉpStan 2010/11) respectivement 3000 (ÉpStan 2015/16, cycle 2.1) et 4400 (ÉpStan 2015/16, cycle 3.1) réponses d'élèves, des corrections faites par deux correcteurs différents sont disponibles. Le résultat le plus important montrait que les corrections étaient hautement concordantes et que cette concordance restait stable au cours des deux années considérées. En valeur absolue, la concordance était de 97,4 % (ÉpStan 2010/11) respectivement de 97,2 % (ÉpStan 2015/16, pour le cycle 2.1) et de 98,0 % (ÉpStan 2015/16, pour le cycle 3.1). En tenant compte du fait que la concordance entre les correcteurs pourrait également être due au hasard, la concordance corrigée était encore

de 95,3 % (ÉpStan 2010/11) respectivement de 96,1 % (ÉpStan 2015/16, pour le cycle 2.1) et de 94,9 % (ÉpStan 2015/16, pour le cycle 3.1). En résumé, il apparaît donc que les consignes de correction précises et surtout leur application minutieuse par les enseignant•e•s permettent d'évaluer les compétences des élèves de manière extrêmement fiable et objective. Etant donné que les procédures des ÉpStan au cycle 4.1 sont quasiment identiques à celles du cycle 3.1, nous pouvons très probablement considérer que le constat positif du cycle 3.1 s'applique également au cycle 4.1. Au niveau de l'enseignement secondaire, les réponses des élèves sont directement enregistrées et traitées dans la plate-forme de test. Aucun encodage manuel n'est donc nécessaire de la part des enseignant(e)s.

### 3 Évaluation des résultats

#### 3.1 La métrique ÉpStan

L'objectif général de l'ÉpStan est de comparer les performances des élèves sur une base longitudinale, c'est-à-dire au fil du temps, afin de mettre en évidence les effets possibles des réformes de l'éducation. Ainsi, il n'est pas possible d'avoir recours aux mêmes exercices pour chaque année scolaire, il faut éviter que les élèves soient préparés de manière ciblée au test et que les résultats soient biaisés. C'est pourquoi, les performances des élèves aux ÉpStan ne sont pas simplement compilées en sommes de points, mais elles sont converties en valeurs de compétence déterminées à l'aide de modèles statistiques issus de la *Item Response Theory* (IRT). Ces valeurs de compétences peuvent être comparées d'une année scolaire à l'autre et, le cas échéant, d'un type d'enseignement à l'autre, même si les tests de compétence comportent des exercices différents d'une année à l'autre. La seule condition est que certains exercices aient déjà été utilisés dans les versions précédentes du test.

Les scores de compétences ont été normalisés de manière à ce que la moyenne de tous les élèves au Luxembourg enseignement confondus) soit de 500 points. L'écart-type (c'est-à-dire l'écart moyen des valeurs du test par rapport à la moyenne) a été fixé à 100 points. Pour chaque cycle d'apprentissage, il existe une année de référence. Pour le cycle 2.1, il s'agit de l'année scolaire 2014/15, pour le cycle 3.1 de l'année scolaire 2011/12 et pour le cycle 4.1 de l'année scolaire 2017/18. Pour les classes de 7<sup>e</sup><sup>1</sup>, c'est l'année scolaire 2018/19 et pour la 5<sup>e</sup>, c'est l'année scolaire 2011/12. Cette année scolaire constitue donc le point de départ et la première année de référence jusqu'à laquelle les résultats ÉpStan de l'année en question peuvent être comparés. Ainsi, si le score ÉpStan moyen d'un groupe d'élèves donné change d'une année à l'autre, un changement réel des compétences dans ce groupe s'est produit.

#### 3.2 Qu'est-ce que l'intervalle de résultats attendus ?

De nombreux enseignant•e•s savent d'emblée que les résultats de leur classe seront très probablement inférieurs ou supérieurs à la moyenne nationale en raison de sa

---

<sup>1</sup> Il faut noter que seuls deux tiers environ de tous les élèves de 7e ont participé aux ÉpStan. Lors de l'interprétation des résultats, il faut donc tenir compte du fait que les élèves testés ne représentent pas l'ensemble des classes de 7e et que le seuil pour déterminer le niveau Socle pourrait encore évoluer dans les années à venir.

composition socioculturelle et socioéconomique. Du coup, pourquoi explorer les performances si les résultats semblent de toute façon déjà connus ?

De nombreuses études internationales et nationales (par exemple PISA, PIRLS, ÉpStan) ont montré que les caractéristiques telles que le sexe, le contexte socio-économique et la langue maternelle sont clairement liées aux performances aux tests de compétence. Au Luxembourg, par exemple, les jeunes dont la langue maternelle est le luxembourgeois ou l'allemand, qui ont un statut socioéconomique élevé et/ou qui sont de sexe féminin, obtiennent une performance moyenne plus élevée au niveau de la compréhension de l'écrit en allemand. D'autre part, les jeunes de langue maternelle portugaise ou française, de statut socioéconomique inférieur et/ou de sexe masculin obtiennent en moyenne des performances plus faibles dans ce domaine de compétence. Or, la composition socioculturelle et socioéconomique de la population scolaire au Luxembourg varie considérablement d'une école à l'autre et d'une classe à l'autre, de sorte qu'on ne peut pas simplement comparer les performances de différentes classes données.

Essayons d'expliquer cela à l'aide d'un exemple (très simplifié) : imaginons une classe d'école fictive qui a passé un test d'allemand. Les élèves ont obtenu en moyenne 430 points à ce test (l'indication de points se réfère à la métrique ÉpStan de la section 3.1) et se situent donc nettement en dessous de la moyenne nationale. Supposons en outre que la classe est principalement fréquentée par des garçons issus de milieux socioéconomiquement défavorisés et parlant le portugais ou le français à la maison. Dans ce contexte, les résultats de cette classe, inférieurs à la moyenne, ne sont pas surprenants. Une question plus intéressante serait donc la suivante : quel est le rapport entre le résultat réel de la classe et le résultat auquel on aurait pu s'attendre compte tenu de la composition de la population scolaire ? Afin de pouvoir répondre à cette question, nous calculons des "intervalles de résultats attendus" à l'aide d'un modèle statistique (appelé modèle de régression).

Pour ce faire, nous procédons de la manière suivante : il faut d'abord établir le lien entre les performances des élèves aux tests dans les domaines de compétence mesurés d'une part, et les caractéristiques telles que sexe, milieu socioéconomique, langue maternelle, contexte de migration, type d'enseignement (au niveau du secondaire) et données relatives au parcours scolaire, d'autre part. Ces caractéristiques sont appelées variables de contexte. Pour chacune de ces variables de contexte, le modèle de régression fournit une pondération, appelée paramètre, qui indique dans quelle mesure cette caractéristique influence la performance au test.

Dans un deuxième temps, nous utilisons ces paramètres pour calculer, pour chaque classe, un intervalle de résultats attendus qui devrait comporter le résultat de la classe si les relations représentées dans le modèle de régression reflètent la composition réelle de la population scolaire au Luxembourg. L'amplitude de cet intervalle de résultats attendus reflète la précision de la mesure des tests utilisés, la précision de prédiction du modèle de régression ainsi que l'incertitude due au manque de données. Si maintenant une classe obtient un résultat supérieur à l'intervalle de résultats attendus, cela signifie qu'elle a obtenu de meilleurs résultats que ceux qui auraient été statistiquement attendus sur la base des caractéristiques socio-économiques et socioculturelles des élèves en question. En revanche, un résultat inférieur à l'intervalle de résultats attendus indique un résultat moins bon que celui auquel on aurait pu s'attendre statistiquement.

Les ÉpStan permettent donc de mettre en relation le profil de performance d'une classe donnée avec des classes de composition socioculturelle et socioéconomique comparable. Ainsi, les enseignant•e•s qui travaillent de manière efficace dans des conditions "difficiles", par exemple, peuvent avoir un retour positif et scientifiquement fondé sur le niveau de performance de leur classe. En principe, les intervalles de résultats attendus sont toutefois intéressants pour *tous les enseignant•e•s* qui souhaitent avoir un aperçu équitable des compétences de leurs élèves.

## 4 Explication sur la motivation à apprendre

En plus des tests de performance, les élèves répondent à un questionnaire. Celui-ci porte sur les domaines suivants :

- Motivation à apprendre dans les différents domaines de compétence visés par les ÉpStan (motivation à apprendre par branche scolaire)
- Motivation générale à apprendre
- Climat au niveau de la classe et au niveau de l'école

La motivation des élèves à apprendre est évaluée à l'aide d'affirmations telles que « Je m'intéresse à la plupart des branches scolaires ». Les élèves indiquent leur degré d'accord avec ces affirmations en cochant la case correspondante. Pour ce faire, une échelle de réponse à quatre niveaux est proposée, avec des catégories de réponse allant de « pas d'accord » à « d'accord ». Les réponses aux différentes affirmations sont ensuite regroupées par thème et présentées en valeurs d'échelle (par exemple pour une échelle d'intérêt général pour l'école). Au niveau secondaire, les élèves peuvent répondre au questionnaire en allemand ou en français, selon leur préférence.

### 4.1 Motivation à apprendre par branche scolaire

La motivation à apprendre par branche scolaire est évaluée à l'aide de quatre échelles : concept de soi, intérêt, anxiété et intérêt pour la lecture (ce dernier uniquement pour les branches linguistiques).

- L'échelle relative au concept de soi renseigne sur l'évaluation que font les élèves de leurs propres performances et capacités scolaires dans la branche concernée. Pour mesurer le concept de soi spécifique à une branche scolaire, les élèves prennent position sur des affirmations telles que : « Je suis bon / bonne en mathématiques ».
- L'échelle relative à l'intérêt indique dans quelle mesure les élèves s'intéressent à une branche scolaire et y prennent plaisir. Voici un exemple d'affirmation à ce sujet : « Je m'intéresse aux mathématiques ».
- L'échelle relative à l'anxiété indique dans quelle mesure cette branche scolaire fait peur aux élèves, leur cause des soucis ou les rend nerveux•ses lors des examens. Voici un exemple d'énoncé : « J'ai peur des mathématiques ».

- L'échelle relative à l'intérêt pour la lecture indique dans quelle mesure les élèves aiment lire. Voici un exemple d'énoncé : « J'aime lire en allemand ».

Les valeurs moyennes se lisent comme suit : plus elles sont élevées, plus le concept de soi des élèves, leur intérêt pour la branche en question ou encore leur intérêt pour la lecture - mais aussi leur anxiété dans une branche scolaire donnée - sont élevés.

## 4.2 Motivation générale à apprendre

La motivation générale à apprendre est évaluée à l'aide des trois échelles suivantes :

- Concept de soi (p.ex. « Je suis bon / bonne dans la plupart des branches scolaires »)
- Intérêt (p.ex. « La plupart des branches scolaires me font plaisir. »)
- Anxiété (p.ex. « J'ai peur de la plupart des branches scolaires. »)

Ici, des moyennes plus élevées représentent un concept de soi plus positif, un plus grand intérêt pour l'école, mais aussi une plus grande anxiété liée à l'école.

## 4.3 Climat au niveau de la classe et au niveau de l'école

La perception qu'ont les élèves de leur école ou de leur classe est évaluée à l'aide de trois échelles :

- Relation enseignant-élève (p.ex. « Dans ma classe, les enseignant(e)s me donnent une aide supplémentaire lorsque j'en ai besoin. »)
- Climat de classe (p. ex. « Dans ma classe, tout le monde s'entend bien. »)
- Tendance à perturber (p. ex. « Dans ma classe, nous perturbons parfois intentionnellement le cours. »)

Ici, des moyennes plus élevées représentent une relation enseignant-élève plus positive, un meilleur climat de classe ou, au contraire, une tendance plus marquée à la perturbation.

## 5 Références

- MENFP (2007). *Français. Socle de compétences. Niveau fin 6<sup>e</sup> ES*. Luxembourg: MENFP.
- MENFP (2008). *L'approche par compétences. Enseignement primaire. Cycle 2*. Luxembourg: MENFP.
- MENFP (2011). *Plan d'études. École fondamentale*. Luxembourg: MENFP.

- MENFP (2008a). *Bildungsstandards Sprachen*. Luxembourg: MENFP.
- MENFP (2008b). *Mathématiques. Division inférieure de l'enseignement secondaire. Compétences disciplinaires*. Luxembourg: MENFP.
- MENFP (2008c). *Mathématiques. Division inférieure de l'enseignement secondaire technique. Compétences disciplinaires*. Luxembourg: MENFP.