



# LES ÉPREUVES STANDARDISÉES

## Élément-clé du pilotage du système éducatif luxembourgeois

---

### Christophe Dierendonck

Université du Luxembourg,  
Unité de recherche *Education, Culture, Cognition and Society* (ECCS),  
Institut *Lifelong Learning and Guidance* (LLLG)

### Amina Kafai

Ministère de l'Éducation nationale, de l'Enfance et de la Jeunesse (Luxembourg),  
Service de Coordination de la Recherche et de l'Innovation pédagogiques et technologiques (SCRIPT),  
Agence pour le développement de la qualité scolaire (ADQS)

### Antoine Fischbach, Romain Martin et Sonja Ugen

Université du Luxembourg, *Luxembourg Centre for Educational Testing* (LUCET)

---

Jusqu'au début des années 2000, le pilotage de l'école luxembourgeoise s'opérait uniquement en référence aux *inputs* investis dans le système. Depuis lors, on assiste à une transformation progressive vers un pilotage par les *outputs* atteints par le système, basé sur la conduite d'évaluations externes des acquis des élèves et sur la mise en projet de développement de la qualité scolaire de tous les établissements scolaires. Dans ce contexte, depuis l'année scolaire 2008-2009, des évaluations externes des acquis des élèves, appelées « Épreuves Standardisées » (ÉpStan), sont conduites dans toutes les classes de grade 3 (CE2 en France) et de grade 9 (troisième en France), et le dispositif tend à s'élargir.

Ces épreuves sont commanditées par le ministère de l'Éducation nationale, de l'Enfance et de la Jeunesse (MENJE) et élaborées par le *Luxembourg Centre for Educational Testing* (LUCET) de l'université du Luxembourg. Par ailleurs, l'Agence pour le Développement de la Qualité Scolaire (ADQS) a été créée en 2009 au sein du MENJE pour accompagner les établissements scolaires dans la définition et la mise en œuvre de leur projet de développement scolaire. Dans cet article, l'accent est mis sur la présentation des objectifs et de la méthodologie des ÉpStan ainsi que sur l'utilisation, par le LUCET et par l'ADQS, des résultats de ces épreuves à des fins de pilotage du système éducatif et de développement de la qualité scolaire. L'article se termine sur une présentation des défis futurs du LUCET et de l'ADQS.

**A**u Luxembourg, comme dans d'autres pays, la comparaison des systèmes éducatifs effectuée par l'enquête PISA 2000 a fait office d'électrochoc<sup>1</sup> et a permis une véritable prise de conscience quant à l'état de l'École. Progressivement, l'assurance et le développement de la qualité scolaire ont été placés au centre des préoccupations du ministère de l'Éducation nationale, de l'Enfance et de la Jeunesse (MENJE) et de son Service de Coordination de la Recherche et de l'Innovation pédagogiques et technologiques (SCRIPT). Le pilotage de l'école luxembourgeoise, qui s'opérait jusqu'alors uniquement en référence aux *inputs* investis dans le système (formation des enseignants, infrastructures à disposition, ressources financières, matériels, horaires et programmes, etc.), a ainsi initié sa transformation en direction d'un pilotage par les *outputs* atteints par le système (mesure des performances scolaires, vérification de l'atteinte des socles de compétence, publication des taux de redoublement, de *diplômation*, de décrochage scolaire, etc.).

En 2007, le MENJE a officialisé sa volonté d'élaborer des instruments d'évaluation externe des acquis des élèves, avec l'objectif de produire, de manière récurrente, des données fiables sur les compétences des élèves aux moments-clés du parcours scolaire [SCRIPT, 2007]. C'est dans ce contexte que depuis l'année scolaire 2008-2009, des évaluations externes<sup>2</sup> des acquis des élèves sont organisées. Appelées « Épreuves Standardisées » (ÉpStan), elles sont commanditées par le Ministère et élaborées par le *Luxembourg Centre for Educational Testing* (LUCET) de l'université du Luxembourg. On signalera, comme l'indique le **tableau 1**, qu'à côté des ÉpStan conduites actuellement aux grades 1, 3 et 9, existent deux autres types d'épreuves nationales d'évaluation des acquis des élèves qui sont élaborées au sein même du MENJE sans la collaboration de l'université : les épreuves communes de cycle 4.2. (grade 6) qui servent à orienter les élèves en fin d'école fondamentale et les épreuves communes de V<sup>e</sup>ES /9<sup>e</sup>EST (grade 9)<sup>3</sup> qui remplissent une fonction de bilan sommatif.

► **Tableau 1** Épreuves nationales d'évaluation des acquis des élèves organisées à l'heure actuelle au Luxembourg

Épreuves nationales	Niveaux d'études testés	
	Selon la nomenclature internationale	Selon la nomenclature française
ÉpStan de cycle 2.1.	Grade 1	CP
ÉpStan de cycle 3.1.	Grade 3	CE2
Épreuves commune de cycle 4.2.	Grade 6	Sixième
ÉpStan de V <sup>e</sup> ES /9 <sup>e</sup> EST	Grade 9	Troisième
Épreuves commune de V <sup>e</sup> ES /9 <sup>e</sup> EST		

1. Dans les trois domaines évalués par PISA, le Luxembourg s'est classé à l'antépénultième position du classement des pays.

2. Au sens strict, on pourrait argumenter que seules les ÉpStan de grade 9 (administrées sur ordinateur) sont véritablement des épreuves externes puisqu'elles sont élaborées, administrées et corrigées de manière externe. Ce statut d'« épreuve externe » peut cependant également être conféré aux ÉpStan de grade 1 et 3, car même si l'enseignant se voit confier la responsabilité d'administrer les épreuves et de les corriger, il a pu être montré empiriquement que des formats de question fermés corrigés à partir d'une grille de correction extrêmement détaillée (voir partie « la passation » et note de bas de page n° 12 p. 68) assurent une fidélité inter-correcteurs extrêmement élevée qui garantit au maximum la standardisation de la correction.

3. La nomenclature des années d'étude varie selon le type d'enseignement fréquenté (ES = enseignement secondaire de type général ; EST = Enseignement secondaire technique).

En 2009, une loi a ancré l'assurance de la qualité de l'enseignement dans les écoles fondamentales<sup>4</sup> et secondaires comme une des trois missions du SCRIPT. Ce texte légal précise que l'assurance de la qualité de l'enseignement doit être fondée, d'une part, sur des évaluations internes menées par le Ministère et, d'autre part, sur des évaluations externes du système éducatif conduites par des instituts universitaires (ex. : PISA, ÉpStan). Pour poursuivre cette mission et coordonner les actions en la matière, l'Agence pour le développement de la qualité scolaire<sup>5</sup> (ADQS) a été créée au sein du SCRIPT. Cette agence a pour tâche principale d'accompagner les écoles fondamentales (appelées écoles) et les écoles secondaires (appelées lycées) dans les processus de développement de la qualité scolaire qu'elles sont obligées (au fondamental) ou invitées (au secondaire) à mettre en place et à renouveler tous les trois ans. Concrètement, il est demandé à chaque établissement de rédiger un plan de réussite scolaire (PRS au fondamental) ou un plan de développement scolaire (PDS au secondaire) qui comprend quatre étapes : 1. analyse de la situation de l'école, 2. définition des objectifs à atteindre ; 3. mise en œuvre des actions et 4. suivi et bilan des actions menées. L'ADQS offre aux écoles un soutien scientifique pour mieux comprendre leur environnement scolaire (compilation de données relatives aux écoles du pays et à chaque école en particulier, préparation de questionnaires, aide à l'interprétation de données issues d'enquêtes ou d'évaluations nationales ou internationales, etc.) ainsi que des outils méthodologiques (cadre de référence pour le développement de la qualité scolaire, etc.) dans leur démarche de définition d'objectifs, de mise au point d'actions concrètes et d'évaluation de ces actions.

Dans cet article, l'accent est mis sur la présentation des objectifs et de la méthodologie d'un seul des trois types d'épreuves nationales (les ÉpStan) ainsi que sur l'utilisation, par le LUCET et par l'ADQS, des résultats ÉpStan à des fins de pilotage du système éducatif et de développement de la qualité scolaire. L'article se termine sur une présentation des défis futurs du LUCET et de l'ADQS.

## LE DISPOSITIF DES ÉPSTAN

### Objectif principal, objectifs secondaires et description générale du dispositif

L'objectif principal du dispositif des ÉpStan est le *monitoring* du système scolaire et non son pilotage, qui relève de la responsabilité du MENJE. Les termes « monitoring » et « pilotage » ne sont en effet pas synonymes [voir à ce sujet, DIERENDONCK et MARTIN, 2008]. Le terme anglais suppose de collecter des informations fiables sur le fonctionnement et les résultats du système scolaire alors que le terme français implique, outre la prise d'informations, une action en vue de modifier une situation donnée en direction d'une situation souhaitée. Les prises d'information annuelles conduites au travers du dispositif ÉpStan autorisent non seulement des analyses

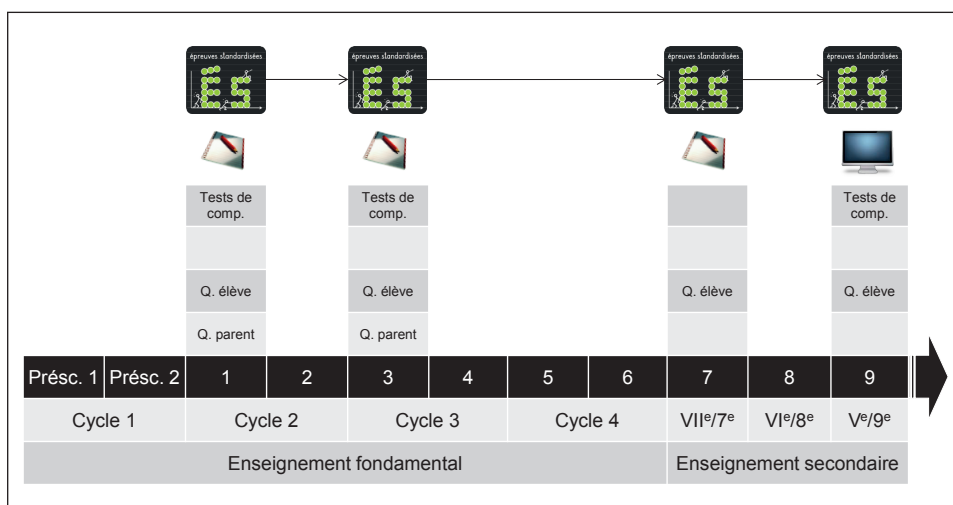
4. Les écoles fondamentales accueillent les enfants entre 4 et 12 ans (âges théoriques).

5. L'ADQS (<https://portal.education.lu/qualitescolaire/Accueil.aspx>), composée de 14 personnes (statisticienne, pédagogues, enseignants, psychologues), est une des trois divisions du SCRIPT. Les deux autres divisions du SCRIPT sont la cellule de compétence pour l'innovation pédagogique et technologique (INNO) et l'institut de formation continue (IFC).

transversales fondées sur des données fiables, mais également une comparaison des performances scolaires des élèves à travers le temps. Ces analyses, reprises dans un rapport national, sont censées fonder les décisions en matière de pilotage du système éducatif.

Actuellement, le dispositif concerne l'ensemble des élèves de l'enseignement public et privé (subventionné) en début des grades 1, 3, 7 et 9 ▶ **Figure 1**. Il comprend, selon l'année d'études considérée, des **tests de compétence** en compréhension de l'oral et de l'écrit (luxembourgeois au grade 1, allemand au grade 3, allemand et français au grade 9) et en mathématiques<sup>6</sup> présentés en version papier-crayon ou sur ordinateur, un **questionnaire bilingue (allemand/français) destiné aux élèves** portant sur les caractéristiques familiales (langue parlée à la maison, origine sociale, etc.) et sur des dimensions motivationnelles spécifiques (concept de soi, anxiété, intérêt, climat de classe, etc.) et un **questionnaire trilingue<sup>7</sup> (allemand, français, portugais) complété par les parents** (niveau d'études, profession, pays de naissance des parents). L'élaboration des tests de compétence et des questionnaires sont sous la responsabilité des chercheurs du LUCET. Les données récoltées sont ensuite mises à la disposition des acteurs concernés. Les informations portant sur les caractéristiques familiales sont principalement utilisées pour calculer les intervalles de performance attendue, mais elles peuvent également être utilisées à des fins de recherche. Le **tableau 2** donne une vue exhaustive des composantes et des particularités du dispositif. Les résultats obtenus par les élèves aux tests de compétence n'entrent pas en ligne de compte pour déterminer le parcours scolaire. À ce titre, ces tests sont à considérer comme des évaluations à faibles enjeux pour les élèves. Les objectifs secondaires du dispositif, qui ont émergé progressivement au fil du temps, notamment pour répondre à certains besoins exprimés par les acteurs de

▶ **Figure 1** Vue synthétique du dispositif ÉpStan actuel



6. Dans le cadre du test de mathématiques au grade 9, les élèves ont en permanence la possibilité de voir les items en version allemande ou en version française.

7. À côté des trois langues officielles du pays (luxembourgeois, allemand et français), d'autres langues sont très utilisées au Luxembourg, notamment le portugais qui est parlé par plus de 20 % de la population.

terrain, sont de l'ordre du pilotage et du développement de la qualité au niveau des écoles, au niveau des classes et au niveau des élèves. Il s'agit notamment de mettre à la disposition des différents acteurs des rapports synthétisant les données issues des ÉpStan. Ces rapports présentent, pour chaque niveau d'agrégation possible des données (national, type d'enseignement, école, classe, élève), une comparaison normée qui tient compte des caractéristiques sociodémographiques des élèves. L'objectif de cette démarche multi-niveaux est triple : transmettre des signaux aux acteurs pour qu'ils amorcent des processus de développement scolaire au sein des écoles et des classes, renforcer la capacité diagnostique des enseignants en cernant les besoins individuels des élèves en termes d'apprentissage et amener les élèves et

► **Tableau 2** Vue exhaustive des composantes et des particularités du dispositif ÉpStan actuel

	Enseignement fondamental		Enseignement secondaire	
Grades	1	3	7	9
Nombre approximatif d'élèves concernés	5 100	5 100	5 500	6 600
Support	Papier-crayon	Papier-crayon	Papier-crayon	Ordinateur
Tests de compétence	<ul style="list-style-type: none"> <li>– Mathématiques</li> <li>– Compréhension de l'oral (luxembourgeois)</li> <li>– Précurseurs de la lecture (par ex. conscience phonologique)</li> </ul>	<ul style="list-style-type: none"> <li>– Mathématiques</li> <li>– Compréhension de l'oral (allemand)</li> <li>– Compréhension de l'écrit (allemand)</li> </ul>		<ul style="list-style-type: none"> <li>– Mathématiques</li> <li>– Compréhension de l'écrit (allemand)</li> <li>– Compréhension de l'écrit (français)</li> </ul>
Questionnaire élève	<ul style="list-style-type: none"> <li>– Langues parlées à la maison</li> <li>– Concept de soi global</li> <li>– Concepts de soi dans les branches évaluées</li> <li>– Intérêt scolaire global</li> <li>– Intérêt pour les domaines scolaires évalués</li> <li>– Anxiété scolaire globale</li> <li>– Anxiété pour les branches évaluées</li> <li>– Besoin de cognition</li> <li>– Climat de classe</li> <li>– Attitude envers l'école</li> </ul>	<ul style="list-style-type: none"> <li>– Langues parlées à la maison</li> <li>– Concept de soi global</li> <li>– Concepts de soi dans les branches évaluées</li> <li>– Intérêt scolaire global</li> <li>– Intérêt pour les domaines scolaires évalués</li> <li>– Anxiété scolaire globale</li> <li>– Anxiété pour les branches évaluées</li> <li>– Besoin de cognition</li> <li>– Climat de classe</li> <li>– Attitude envers l'école</li> </ul>	<ul style="list-style-type: none"> <li>– Langues parlées à la maison</li> <li>– Pays de naissance</li> <li>– Concept de soi global</li> <li>– Concepts de soi dans les branches évaluées</li> <li>– Intérêt scolaire global</li> <li>– Intérêt pour les domaines scolaires évalués</li> <li>– Anxiété scolaire globale</li> <li>– Anxiété pour les branches évaluées</li> <li>– Besoin de cognition</li> <li>– Climat de classe</li> <li>– Attitude envers l'école</li> <li>– Profession des parents</li> </ul>	<ul style="list-style-type: none"> <li>– Langues parlées à la maison</li> <li>– Pays de naissance</li> <li>– Concept de soi global</li> <li>– Concepts de soi dans les branches évaluées</li> <li>– Intérêt scolaire global</li> <li>– Intérêt pour les domaines scolaires évalués</li> <li>– Anxiété scolaire globale</li> <li>– Anxiété pour les branches évaluées</li> <li>– Besoin de cognition</li> <li>– Climat de classe</li> <li>– Attitude envers l'école</li> <li>– Profession des parents</li> </ul>
Questionnaire parents	<ul style="list-style-type: none"> <li>– Profession</li> <li>– Niveau d'études</li> <li>– Pays de naissance</li> </ul>	<ul style="list-style-type: none"> <li>– Profession</li> <li>– Niveau d'études</li> <li>– Pays de naissance</li> </ul>		

leurs parents à réfléchir et, le cas échéant, réagir aux résultats individuels obtenus. Dans des pays plus grands que le Luxembourg, ces différents objectifs sont souvent poursuivis dans le contexte de prises de données distinctes, avec notamment des évaluations par échantillon pour le niveau système. Or, au Luxembourg, étant donné la taille du pays, la participation de l'ensemble des élèves pour les différents niveaux est requise pour des raisons méthodologiques. Il a donc été décidé que les différents objectifs évoqués précédemment s'appuient sur une seule et même collecte de données.

À long terme, le dispositif devrait concerner les grades 1, 3, 5, 7 et 9 pour ce qui est des tests de compétence et des questionnaires. Ceci devrait permettre d'opérer un suivi longitudinal de cohortes d'élèves et d'étudier ainsi les trajectoires scolaires. Les effets à moyen et à long terme de projets pilotes et d'interventions spécifiques pourront également être évalués à partir de ce dispositif.

### Communication autour du dispositif

La communication mise en place autour du dispositif est primordiale et se fait en étroite collaboration entre le LUCET et le ministère de l'Éducation pour faciliter l'acceptation des ÉpStan et la compréhension de leurs objectifs par les acteurs sur le terrain. En début d'année scolaire, le ministère de l'Éducation annonce de manière officielle les ÉpStan auprès des responsables des établissements scolaires, des enseignants et des parents d'élèves des grades concernés. Des réunions régionales d'information sont ensuite organisées par le Ministère et le LUCET pour permettre aux enseignants concernés d'interagir avec les mandataires et les concepteurs du dispositif. Le LUCET s'occupe de l'élaboration (et prochainement de la diffusion<sup>8</sup>) des informations concrètes autour de la passation des ÉpStan (documents explicatifs, manuels de test et de correction, tests de compétence, questionnaires, rapports, etc.). Pour garantir une transparence maximale, toutes les communications ponctuelles sont reprises sur un site web bilingue ([www.epstan.lu](http://www.epstan.lu)) qui comprend en outre une série de vidéos explicatives, des exemples de questions et les rapports des années précédentes.

### Élaboration des épreuves et des documents d'accompagnement

Les tests de compétence, composés d'une trentaine d'items par domaine scolaire évalué, sont élaborés par des groupes de travail dirigés par des chercheurs du LUCET spécialisés en psychologie, psychométrie, éducation ou didactique. Un groupe de travail compte, outre le chercheur responsable, de cinq à sept membres, en majorité des enseignants actifs<sup>9</sup> dans les grades et les branches scolaires concernés par le dispositif ÉpStan. Les réunions de travail se font en moyenne toutes les deux semaines. Une plateforme informatique permet aux membres des groupes de travail de partager, dans un environnement sécurisé, les documents utilisés et les items produits.

<sup>8</sup>. La diffusion des lettres d'information et du matériel de *testing* incombe à l'ADQS depuis 2008-2009, mais il est prévu qu'à terme, le LUCET soit responsable de cette tâche également.

<sup>9</sup>. Chaque année, le ministère de l'Éducation et le LUCET lancent un appel pour recruter des enseignants intéressés par le développement des épreuves. Ce travail est rémunéré en tant qu'heures supplémentaires ou par une décharge d'enseignement. Les enseignants reçoivent une formation portant sur le développement d'items (par ex. format de questions) et sur le cadre conceptuel défini pour élaborer les ÉpStan.

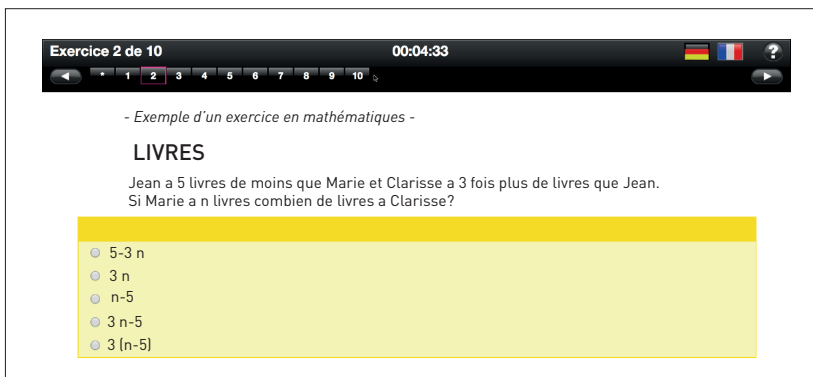
Sur la base des documents de référence officiels spécifiant les standards à atteindre à la fin de chaque cycle<sup>10</sup>, un cadre conceptuel d'élaboration (*test framework*) a été développé pour chaque test de compétence. Ce document spécifie le nombre d'items nécessaires par types de compétences et par niveau de difficulté. Les différents groupes de travail développent et classifient les items élaborés en fonction de ce cadre conceptuel. Les items sont modifiés jusqu'à ce que les membres d'un groupe de travail tombent d'accord sur le contenu et les caractéristiques de chacun des items. La finalisation d'un item au niveau fondamental inclut l'illustration et la présentation dans le carnet de test ainsi que l'enregistrement des textes, consignes et items pour le test de compréhension de l'oral. Au niveau secondaire, les items finalisés sont transcrits sur la plateforme informatique OASYS (*Online Assessment SYStem*) développée par le LUCET ► **Figure 2**.

L'élaboration des items prend place durant toute l'année scolaire et aboutit, généralement en mai, à la conduite d'un pré-test du matériel. À l'issue de ce pré-test, les groupes de travail discutent à nouveau des items et les modifient si nécessaire en fonction des données empiriques récoltées. Les analyses statistiques utilisées lors du pré-test sont les mêmes que lors de la collecte de données principale.

Au niveau des épreuves de grade 9, trois versions sont élaborées pour chaque test de compétence. Ces versions présentent des degrés de difficulté différents qui correspondent aux différences de niveau scolaire existant entre les trois types d'enseignement au Luxembourg<sup>11</sup>. Pour que les résultats restent comparables entre les types d'enseignement, les versions de chaque test comprennent des items communs et des items spécifiques.

La compilation finale du test principal s'opère non seulement à partir du cadre conceptuel, mais également en tenant compte d'autres critères comme l'ordre des items, l'alternance des formats de réponse ou la place des items d'ancrage longitudinal et

► **Figure 2** Capture d'écran d'un item de mathématiques présenté aux élèves de grade 9 sur la plateforme de *testing* OASYS



<sup>10</sup>. Au niveau du secondaire, ces documents de référence restent au niveau descriptif des contenus à aborder (standards de contenus), mais ils ne définissent pas encore les critères minimaux à atteindre (standards de performance).

<sup>11</sup>. Chaque test se décline en une version pour l'enseignement secondaire (ES), une version pour l'enseignement secondaire technique orientation technique et polyvalente (EST) et une version pour l'enseignement secondaire technique orientation pratique et modulaire (EST-PR).

d'ancrage transversal. Pour garantir une comparabilité longitudinale des résultats, une sélection d'items est reprise d'une année à l'autre selon des critères statistiques spécifiques. Cet objectif de suivi longitudinal empêche la publication d'un grand nombre d'items.

Les chercheurs responsables des différents groupes de travail ont des échanges réguliers au sein d'une instance nommée « Conseil des développeurs d'items », dont la finalité est de garantir que les différents tests de compétence se structurent de manière similaire et que les items élaborés sont de qualité élevée. Cette instance permet également aux chercheurs de réfléchir aux développements potentiels du dispositif (de nouveaux types d'items par exemple) et d'échanger à propos des avancées scientifiques dans le domaine de l'évaluation des compétences scolaires.

### Les préparatifs logistiques

Les préparatifs logistiques du dispositif ÉpStan s'opèrent en étroite concertation entre l'ADQS et le LUCET. Ils comprennent l'impression, l'envoi, la réception et l'archivage de tout le matériel nécessaire. Au niveau fondamental, le matériel inclut pour chaque élève : des carnets de test (en général, un carnet par test de compétence), un questionnaire motivationnel, un questionnaire parent, la lettre d'informations pour les parents ainsi qu'une feuille sur laquelle l'enseignant reporte le codage des réponses données. Chaque enseignant au fondamental reçoit un manuel de passation pour les tests, un manuel de passation pour le questionnaire élève et un manuel de correction pour chaque test de compétence. À titre d'illustration, environ 5 100 élèves du grade 3 répartis en 380 classes ont participé en 2013-2014 aux ÉpStan. Un carnet de test contenant entre 24 à 28 pages, la quantité de matériel imprimée peut être estimée à environ 840 000 pages pour le grade 3. Au grade 9, ce nombre est réduit étant donné que la passation se fait par ordinateur. Les établissements secondaires reçoivent cependant eux aussi des manuels de passation, des lettres d'information pour les parents et des lettres d'information pour les enseignants des branches concernées (au total environ 16 000 pages).

### La passation

Les ÉpStan sont administrées au début de l'année scolaire en novembre<sup>12</sup> et évaluent les compétences acquises au cycle précédent. Puisque le retour aux acteurs de terrain se fait relativement rapidement (fin janvier), ceci permet aux titulaires des classes d'utiliser le retour d'informations fourni afin d'ajuster leurs interventions pédagogiques. Au fondamental, la passation et la correction se font par l'enseignant de la classe au départ d'un protocole d'administration du test et d'une grille de correction très détaillés supposés garantir une standardisation élevée<sup>13</sup>. Les tests de compétence sont répartis sur quatre matinées à des dates fixes. La durée de chaque test est de 50 minutes au grade 3 et de 35 minutes au grade 1.

Au secondaire, les trois tests de compétence sont administrés sur ordinateur lors d'une seule matinée de testing. La passation de l'épreuve est organisée dans chaque

12. Ce n'est qu'à partir de début octobre que l'on dispose des bases de données ministérielles permettant de définir de manière précise les participants aux ÉpStan. Les tests sont donc organisables, au plus tôt, en novembre.

13. En 2010-2011, une étude de fidélité inter-correcteurs a montré une correspondance très élevée (Kappa = 0,95) entre la correction réalisée par les enseignants et la correction effectuée par des correcteurs spécifiquement formés.



établissement par un coordinateur qui envoie les horaires de *testing* de chaque classe au LUCET afin de permettre un bon déroulement au niveau technique. Les jours de passation peuvent être choisis par chaque établissement à condition de se dérouler au mois de novembre. Chaque test de compétence dure 50 minutes, avec une pause de 15 minutes entre chaque test. Le dernier test est suivi du questionnaire élève dont la passation prend 20 minutes. L'ordre de présentation des trois tests de compétence est aléatoire. Pour soutenir les établissements, des assistants formés par l'équipe des ÉpStan sont présents lors de la première passation. Durant toute la période de *testing*, le LUCET assure une assistance continue par courriel et *via* une ligne téléphonique spécifique. L'ADQS est également contactée en cas de questions ou de problèmes rencontrés par les acteurs concernés par le dispositif ÉpStan.

Durant la période de *testing*, l'ADQS procède à des observations et à des entretiens dans quelques écoles afin de contrôler le respect des procédures et des consignes définies et de recueillir des informations en lien avec la passation proprement dite (conditions de passation dans la salle de classe, durée des épreuves, comportement des élèves pendant les épreuves, attitudes et perceptions des enseignants par rapport aux ÉpStan). Ce recueil d'informations qualitatives a pour objectif de contribuer à l'amélioration continue du dispositif.

Après réception du matériel de *testing* à l'université, les feuilles de codage et les questionnaires pour le fondamental sont scannés (*via* le logiciel Teleform) afin de permettre le traitement informatisé des données. Les bases de données sont ensuite vérifiées et nettoyées.

## Analyse des données<sup>14</sup>

### Scaling et ancrage longitudinal

Afin de permettre la comparabilité des résultats dans le temps et la conduite d'analyses de tendance, il est nécessaire de rendre compte des résultats issus des différents tests sur une échelle de performance commune. Deux cohortes d'élèves sont considérées chaque année : la cohorte 1 des élèves de l'année du test et la cohorte 0 composée de l'ensemble des élèves testés les années précédentes. La procédure adoptée dans le cadre des ÉpStan s'inspire de celle décrite par NAGY et NEUMANN [2010]. Cette procédure comprend cinq étapes successives : conformité au modèle de Rasch, analyse descriptive du fonctionnement différentiel des items (FDI), catégorisation des FDI selon la typologie définie par l'*Educational Testing Service* (ETS), analyse de sensibilité et estimation finale des paramètres de compétence des sujets.

Lors de l'étape 1, seules les données de la cohorte 1 sont analysées. Les items sont calibrés de manière libre et la contrainte d'identification est placée sur les sujets [Wu, ADAMS *et alii*, 2007]. On procède ensuite à la sélection des items jugés suffisamment discriminants (items qui présentent une corrélation satisfaisante avec le score global et qui permettent de distinguer les élèves ayant un score global élevé de ceux ayant un score global faible [EBEL et FRISBIE, 1986]. On s'intéresse également à l'ajustement

14. Pour plus de détails, se référer au rapport technique disponible sur <http://www.epstan.lu>.

des items au modèle de RASCH : seuls les items dont le comportement est bien prédit par le modèle sont conservés [BOND et FOX, 2010 ; GUSTAFSSON, 1980 ; MARTIN-LÖF, 1974 ; WRIGHT, LINACRE *et alii*, 1994].

L'étape 2 est une étape descriptive lors de laquelle les paramètres de difficulté des items d'ancrage longitudinal potentiels (c'est-à-dire les items communs aux cohortes 1 et 0) sont comparés graphiquement. Concrètement, les indices de difficulté estimés au départ de la cohorte 1 ainsi que les intervalles de confiance (à 95 %) qui y sont associés sont tracés en regard de ceux obtenus à partir de la cohorte 0. Ceci permet d'établir un diagnostic rapide de la robustesse des paramètres de difficulté des items au travers des cohortes et peut aider à identifier des items susceptibles de présenter un problème de fonctionnement différentiel.

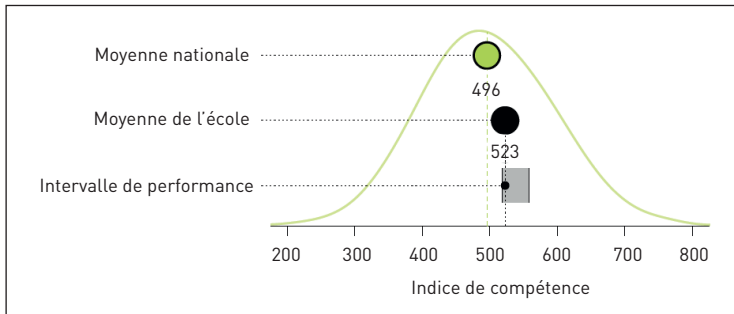
Durant l'étape 3, on s'intéresse au fonctionnement différentiel des items (FDI) d'ancrage potentiels. Une classification de ces items est établie selon la magnitude du FDI en utilisant la typologie définie par l'*Educational Testing Service* (ETS) et reprise dans NAGY et NEUMANN [2010]. Trois types d'items sont ainsi distingués : les items de type ETS A (sans FDI), les items de type ETS B (avec un FDI modéré) et les items de type ETS C (avec un FDI important). Lors de cette étape, les données des deux cohortes sont mises en relation et l'on crée un tableau de données ancrées à partir duquel les compétences des élèves sont estimées [scores WLE, *Weighted Likelihood Estimation*, décrits dans WARM, 1989]. Les scores WLE ainsi qu'une variable indicatrice de cohorte (1 ou 0) sont ensuite introduites, pour chaque item, dans un modèle de régression logistique. Le coefficient de régression associé à la variable renseignant la cohorte permet d'apprécier le degré auquel la relation d'appartenance à une cohorte influence la probabilité de donner une bonne réponse à cet item, et par conséquent, est utilisé en tant que mesure d'intensité du FDI pour obtenir la classification mentionnée précédemment. Une autre classification, plus fine, basée également sur ce coefficient permet de distinguer 4 groupes d'items d'intensité croissante de FDI. En somme, ces classifications permettent de dégager 7 groupes d'items d'ancrage potentiels définissant chacun un scénario d'ancrage possible.

L'étape 4 consiste à mener une analyse de sensibilité sur la base des 7 scénarios d'ancrage définis lors de l'étape précédente, le but étant de conserver un maximum d'items d'ancrage tout en minimisant l'impact sur les compétences estimées des élèves de la cohorte 1. Au terme de cette étape, un ensemble d'items d'ancrage optimal est retenu. Enfin, lors de l'étape 5, on procède à l'estimation finale de la compétence des élèves de la cohorte 1 en prenant soin de fixer les paramètres des items d'ancrage retenus à cet effet lors de l'étape précédente.

### Calcul des intervalles de performance attendue pour disposer d'une comparaison équitable

Pour permettre une comparaison plus équitable des performances des écoles et des classes, des intervalles de performance attendue sont déterminés au départ de modèles de régression dans lesquels la performance agrégée (au niveau école ou au niveau classe) est prédite par les caractéristiques sociodémographiques agrégées des élèves (au niveau école ou au niveau classe) ► **Figure 3**. Les caractéristiques individuelles retenues sont la filière fréquentée (implicitement), le sexe, la ou les langues parlées à la maison, le statut socioéconomique (estimé au départ de l'indice de richesse à la maison, du nombre de livres dans le foyer et de l'occupation professionnelle des parents), l'année de naissance et le parcours scolaire antérieur effectué ou non au Luxembourg.

► **Figure 3** Exemple de graphique renseignant l'intervalle de performance attendue pour une école en particulier



### Calcul des seuils de coupure pour déterminer les niveaux de compétence

L'analyse des données issues des ÉpStan ne se limite pas à offrir une comparaison normée qui tient compte des caractéristiques des populations scolaires. Elle fournit également une comparaison critériée<sup>15</sup> qui permet de situer chaque élève par rapport aux attentes du programme d'études. Ainsi, pour chaque test, les items sont répartis en différents niveaux de compétence. Cette répartition des items est réalisée d'abord de manière théorique au départ des référentiels, puis sur la base empirique suite à l'analyse des données du pré-test. À partir des paramètres de difficulté des items, un paramètre de difficulté médian est défini pour chaque niveau de compétence. C'est ce paramètre qui permet de calculer le score de coupure pour chaque niveau. Concrètement, si un élève est assigné à un niveau de compétence particulier, cela signifie qu'il y a une probabilité élevée que cet élève réussisse au moins la moitié des items classés dans ce niveau de compétence. Cette « probabilité élevée » a été opérationnalisée à 0,62, à l'instar de ce qui se fait dans le programme PISA [OCDE, 2009, p. 300].

## L'UTILISATION DES DONNÉES ISSUES DES ÉPSTAN

### Utilisation des données par l'université

Une fois que les données des ÉpStan ont été analysées, différents rapports sont mis à la disposition des acteurs concernés<sup>16</sup> sur une plateforme informatique du MENJE. Sur le site [www.epstan.lu](http://www.epstan.lu), outre des exemples de rapports, on peut trouver deux documents explicatifs : un document technique exhaustif qui détaille la méthodologie du dispositif d'évaluation externe et un document explicatif destiné aux directions et aux enseignants.

<sup>15</sup>. Décrire ce dont un élève ou un étudiant est capable, sans égard à la performance des autres, tel est le but visé avec la mesure à interprétation critériée [SCALLON, 2004, p. 2].

<sup>16</sup>. Grade 1 et 3 : environ 15 300 rapports élève (1 rapport par domaine scolaire testé), 360 rapports classe et 160 rapports école. Grade 7 : résultats du questionnaire uniquement présentés dans le rapport national. Grade 9 : environ 19 000 rapports élève (un rapport par domaine scolaire testé), 1 000 rapports classe et 35 rapports école.

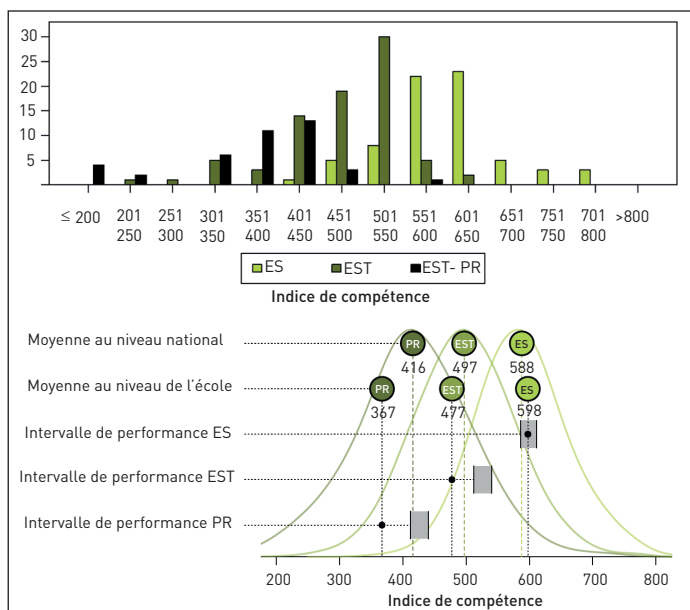
### Le « rapport national » – pour les décideurs et pour le public

Le « rapport national », publié tous les trois ans, est rédigé par le LUCET et rendu public sur le site [www.epstan.lu](http://www.epstan.lu). Il présente les objectifs et la méthodologie des ÉpStan ainsi que plusieurs types d'analyse des données permettant la comparaison, mais de manière anonyme : une analyse globale des résultats observés, une analyse des résultats qui tient compte des facteurs socio-économique et migratoire et une analyse comparant les résultats obtenus dans les filières d'enseignement. Le « rapport national » se termine par une analyse synthétique de la situation actuelle et par la présentation de différentes pistes d'action. Le rapport 2011-2014 présente, pour la première fois, une analyse des tendances observées ces trois dernières années scolaires.

### Le « rapport école » – pour les comités d'écoles fondamentales et les directions des écoles secondaires

Un « rapport école », envoyé aux comités d'écoles fondamentales, à chaque direction d'établissement secondaire et à l'ADQS, présente de manière globale les résultats des élèves de l'école concernée et autorise une comparaison anonyme avec les résultats observés ailleurs dans le pays et au sein des différentes filières d'enseignement (ES, EST, EST-PR) ▶ **Figure 4**. Un intervalle de performance attendue est calculé pour chaque école. Cet intervalle permet une comparaison qui tient compte des caractéristiques principales de la population scolaire accueillie dans l'école (niveau socio-économique et origine migratoire). Le « rapport école » renseigne aussi les indicateurs moyens observés au niveau école pour les différentes dimensions motivationnelles étudiées au travers du questionnaire (concept de soi, anxiété, intérêt, climat de classe, etc.) et permet une comparaison avec les moyennes observées sur le plan national.

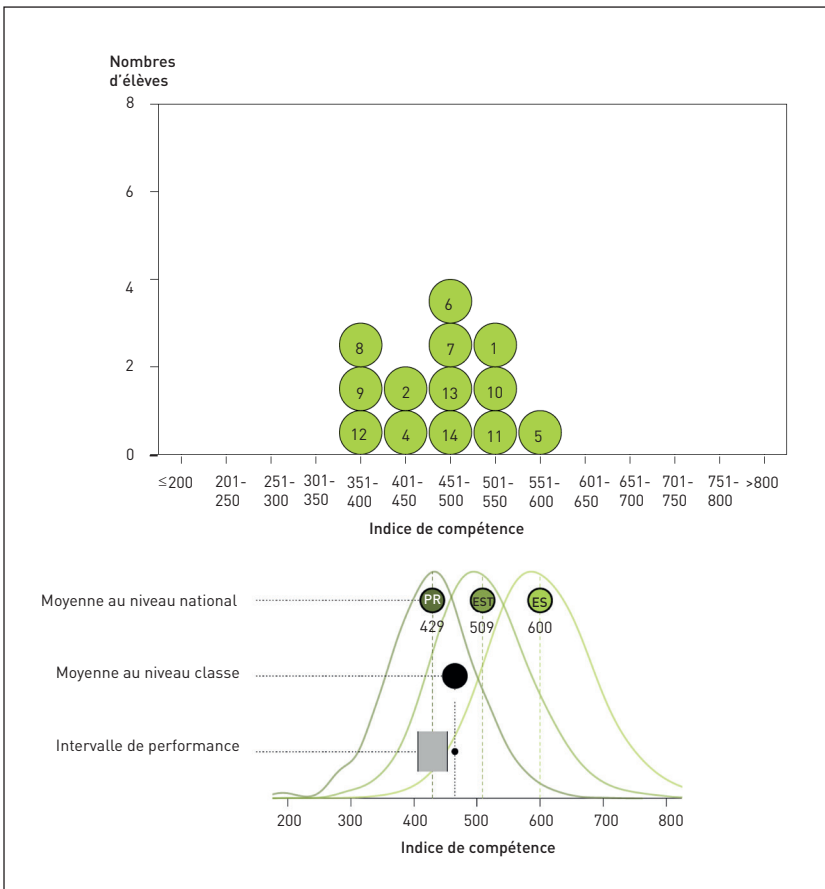
▶ **Figure 4** Principaux graphiques issus du rapport adressé à un établissement secondaire en particulier



### Le « rapport classe » pour les enseignants

Un « rapport classe » est mis à la disposition des enseignants concernés par les ÉpStan sur une plateforme de téléchargement du MENJE avec code d'accès personnalisé ► **Figure 5**. Ce rapport présente, notamment, la distribution des résultats des élèves de la classe (l'enseignant dispose de la clé de codage permettant d'identifier précisément ses élèves) et permet de situer ceux-ci par rapport aux résultats observés au niveau du pays et au niveau des différentes filières d'enseignement. Un intervalle de performance attendue est calculé pour chaque classe. Cet intervalle permet une comparaison qui tient compte des caractéristiques principales de la population scolaire accueillie dans la classe (niveau socio-économique et origine migratoire). Le « rapport classe » renseigne également les moyennes observées au niveau classe pour les différentes dimensions motivationnelles étudiées au travers du questionnaire (concept de soi, anxiété, intérêt, climat de classe, etc.) et permet une comparaison avec les indicateurs moyens nationaux.

► **Figure 5** Principaux graphiques issus du rapport adressé à une classe de secondaire en particulier



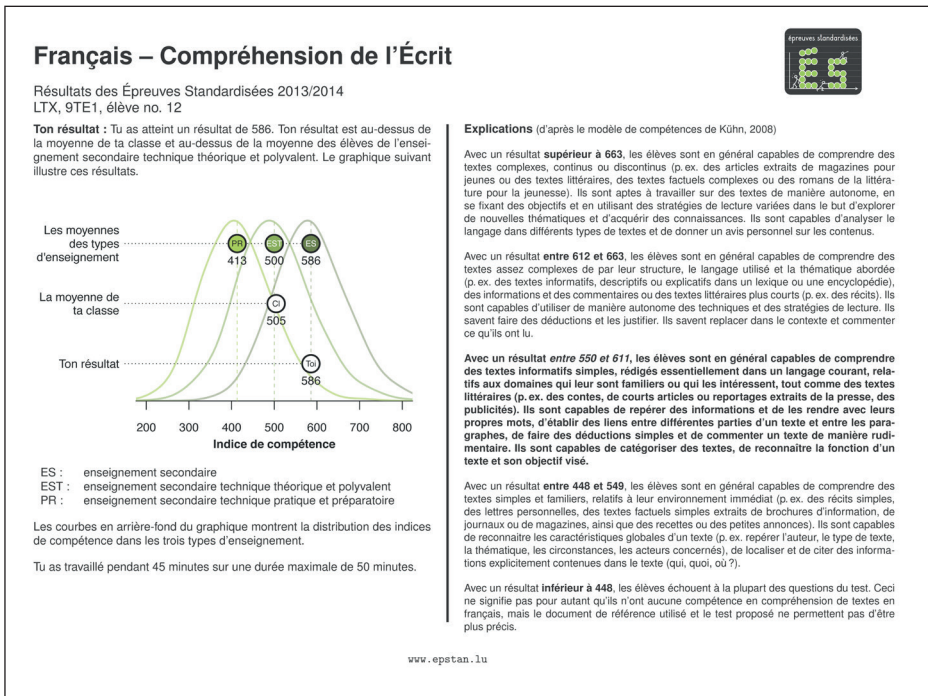
## Le « rapport élève » – pour les élèves et leurs parents

Un « rapport élève » est rédigé, pour chaque élève, dans les trois domaines de compétence évalués par les ÉpStan ▶ **Figure 6**. Les enseignants téléchargent ces rapports individualisés en même temps que le « rapport classe ». Ils ont la responsabilité de mettre ces rapports à disposition des élèves et des parents. Ces rapports individuels décrivent la performance individuelle des élèves et permettent une double comparaison (normative et critériée). La comparaison normative consiste à renseigner le score obtenu par l'élève en le situant sur un graphique par rapport à la moyenne de la classe et par rapport aux distributions et scores moyens observés dans chaque filière d'enseignement. La dimension critériée apparaît dans la partie droite du rapport élève. Le score obtenu par l'élève permet de situer celui-ci dans un niveau de compétence qui décrit le type de questions ou le type de tâches face auxquelles les élèves donnent généralement une réponse correcte. Enfin, le rapport renseigne le temps que les élèves ont effectivement passé à faire le test.

## Utilisation des données par l'ADQS

Pour rappel, l'ADQS est chargée d'accompagner les établissements scolaires dans la démarche de développement de la qualité scolaire. Le travail de l'ADQS se décline essentiellement en trois types d'actions : accompagner les écoles dans l'élaboration,

▶ **Figure 6** Exemple de rapport élève (compréhension de l'écrit en français, enseignement secondaire)



la conduite et l'évaluation de leur plan de réussite scolaire (PRS) ou de leur plan de développement scolaire (PDS), documenter les écoles en leur envoyant les données quantitatives et qualitatives qui les concernent et intervenir sur demande des établissements scolaires pour soutenir ou encadrer des projets spécifiques.

Seules les données produites au niveau école à l'issue des ÉpStan sont utilisées par l'ADQS dans le cadre de la documentation des écoles sur leurs situations respectives. Concrètement, chaque école compile – dans un « classeur école » ou dans un « classeur lycée » structuré en quatre parties (données démographiques, performances scolaires, processus scolaires, perceptions des acteurs) – les informations que l'ADQS lui envoie dès que celles-ci sont disponibles. Les résultats obtenus par l'établissement scolaire lors des ÉpStan sont insérés dans la partie du « classeur école » ou du « classeur lycée » consacrée aux performances scolaires de l'établissement.

## LES DÉFIS POUR LE FUTUR

### Défis pour l'université

#### Élargir le dispositif

Actuellement, les ÉpStan sont administrées au grade 1, au grade 3 et au grade 9. Au grade 7, seul le questionnaire motivationnel est soumis aux élèves. Un modèle « 1-3-5-7-9 » complet (tests de compétence et questionnaire) est actuellement proposé aux décideurs politiques pour disposer, à long terme, en début de chaque cycle d'apprentissage, d'une évaluation des acquis des élèves développés lors du cycle précédent et permettre un suivi longitudinal de cohortes d'élèves. Afin de disposer de profils de compétence plus larges, il est aussi question d'évaluer d'autres domaines disciplinaires (les sciences notamment) et d'autres processus (le raisonnement notamment). Face à ces ambitions partagées par le LUCET et le MENJE, la question de la faisabilité d'un tel dispositif se pose car le dispositif actuel, conduit annuellement rappelons-le, nécessite déjà une quinzaine de chercheurs engagés à temps plein. En outre, les données collectées actuellement pourraient être exploitées davantage (ex. : données comportementales recueillies lors des passations, etc.), voire partagées, sous certaines conditions, avec des chercheurs externes au LUCET.

#### Questionner la perception, la compréhension et l'utilisation réelle des rapports ÉpStan

Au fil du temps, les routines de communication, d'organisation, de collecte, d'analyse des données et de rédaction des rapports produits dans le cadre des ÉpStan ont progressivement été affinées et la plateforme informatique OASYS utilisée pour le *testing* dans les écoles secondaires apparaît très fiable et très flexible.

Quant aux rapports envoyés aux différents acteurs, ils connaissent chaque année des développements plus ou moins marqués, tant au niveau de la forme que des contenus produits. Ainsi, la perspective initiale quasi-exclusivement psychométrique ne permettant pas d'identifier les élèves et leurs lacunes (jusqu'en 2009-2010) s'est transformée progressivement en une perspective mixte (depuis 2010-2011) recherchant un équilibre entre comparaisons normatives et informations plus

diagnostiques faisant davantage référence au curriculum scolaire et autorisant une identification des élèves en difficulté. C'est aussi en 2010-2011 que le rapport élève apparaît pour la première fois. Cette transformation progressive a été décidée à la demande du terrain pour augmenter l'utilité des épreuves, des données récoltées et des rapports envoyés aux enseignants et aux élèves<sup>17</sup>. Sur ce point, nous n'avons que très peu d'informations fiables sur la façon dont les acteurs se saisissent des rapports mis à leur disposition.

Ces questions liées à la perception du dispositif, à la compréhension et à l'utilisation des rapports par les acteurs sont des questions difficiles car, jusqu'ici, peu d'informations étaient disponibles à ce sujet. Il a donc été décidé de conduire, durant le premier semestre 2014, une enquête par questionnaire et des discussions en groupe auprès des différents acteurs concernés par les ÉpStan (enseignants, directions, coordinateurs, parents, élèves), ceci afin d'examiner dans quelle mesure les rapports mis à disposition sont lus, compris et utilisés par les différents publics-cibles. Les informations récoltées sont en cours d'analyse et seront discutées en 2015 dans le cadre d'une démarche qualité menée en interne. Plus largement, l'étude de l'impact réel du dispositif ÉpStan dans le cadre du pilotage du système scolaire luxembourgeois semble constituer un défi majeur pour le futur.

### Développer de nouveaux types d'items d'évaluation

Les ÉpStan sont actuellement uniquement constituées d'items à choix multiple et de questions ouvertes à réponse courte, ceci afin de garantir la plus grande objectivité lors de la correction des réponses des élèves. Mais il est question de profiter des possibilités offertes par la plateforme informatique de *testing* OASYS pour expérimenter de nouveaux formats d'items mettant en œuvre des processus ou des comportements spécifiques (prise de notes durant la lecture, surligner des passages, glisser-déplacer avec la souris, etc.).

### Assurer la validité des constats dressés

Les ÉpStan ont clairement une fonction informative puisque les performances des élèves à ces tests ne sont pas prises en compte pour influencer ou déterminer leur parcours scolaire. On cherche en outre à réduire au maximum l'anxiété que ces tests pourraient provoquer chez les élèves.

Face à ce genre d'évaluations « à faibles enjeux » pour les élèves, on formule généralement l'hypothèse implicite que les performances observées lors des tests ont été réalisées avec un niveau d'effort élevé [WOLF et SMITH, 1995 ; ZERPA, HACHEY *et alii*, 2011]. À ce sujet, les résultats de la recherche sont contradictoires : certaines études ont mis en évidence que les élèves sont raisonnablement motivés à donner le meilleur d'eux-mêmes quand le test est à faibles enjeux [BAUMERT et DEMMICH, 2001] tandis que

<sup>17</sup>. On soulignera que cette transformation a également engendré un paradoxe puisque l'objectif principal du monitoring du système empêche d'aller au bout de la logique au niveau de l'élève et au niveau de l'enseignant, en ne rendant publics que quelques items utilisés dans les tests. Cela dit, le nombre d'items rendus publics s'accroît chaque année et une banque d'items illustrant les différents niveaux de compétence sera rapidement constituée.



d'autres ont conclu que les enjeux du test exerçaient effectivement une influence significative positive sur la motivation et la performance [WOLF et SMITH, 1995]. Les élèves seraient donc plus ou moins motivés et pourraient dès lors s'impliquer différemment. Il est donc possible de questionner la validité [au sens de MESSICK, 1995] des constats dressés à l'issue des évaluations « à faibles enjeux », puisqu'on peut faire l'hypothèse que, lors de ce type d'épreuves dépourvues de conséquences scolaires pour les élèves, l'effort consenti par certains n'est pas maximal et que, *de facto*, leurs performances ne peuvent véritablement être considérées comme des indicateurs valides de leurs acquis. Plus globalement, les scores moyens observés aux évaluations externes à faibles enjeux seraient dès lors sous-estimés et ne reflèteraient donc pas ce dont les élèves sont réellement capables, mais plutôt « ce qu'ils démontrent avec un effort minimal » [O'NEIL, SUGRUE, BAKER, 1996].

Pour examiner dans quelle mesure les ÉpStan sont touchées par ce problème éventuel de validité, des données de motivation initiale pour les tests et d'effort consenti durant ceux-ci [voir à ce sujet KESKPAIK et ROCHER, 2012 et DIERENDONCK, SONNLEITNER *et alii*, 2013] sont recueillies depuis deux années. Ces données font actuellement l'objet d'une analyse minutieuse.

## Défis pour l'ADQS

### De la méfiance à la confiance, en passant par l'indifférence

La dynamisation des établissements scolaires *via* les PRS et les PDS et la création de l'ADQS en 2009 ont initialement été perçues par une partie du personnel enseignant comme des tentatives de contrôle du travail accompli dans les écoles ou d'évaluation des enseignants eux-mêmes<sup>18</sup>. Mais peu à peu, par un travail intense de simplification des procédures et de communication/explicitation des objectifs et des rôles de chacun mené auprès de chaque établissement scolaire, les tensions se désamorcent et les projets de développement se mettent en place, certes pas partout avec la même facilité, le même entrain ou les mêmes ambitions, loin de là, mais toutes les écoles au Luxembourg sont « en mouvement ». Le défi de l'ADQS consiste à présent à trouver les moyens d'un accompagnement différencié, chevillé aux besoins particuliers des écoles. L'idée est d'offrir de nouvelles formations, de nouveaux outils et de rechercher des personnes-relais (comme les inspecteurs et les instituteurs-ressources à l'école fondamentale) pouvant jouer un rôle démultiplicateur, car l'ADQS demeure une petite structure de 14 personnes qui est chargée d'accompagner plus de 150 écoles fondamentales et 35 établissements secondaires.

### À la recherche d'une complémentarité entre toutes les évaluations (internes et externes, nationales et internationales)

À côté des évaluations faites par les enseignants (devoirs en classe, bilans intermédiaires) et des enquêtes internationales comme PIRLS et PISA, il y a actuellement, comme énoncé en introduction, trois types d'évaluations « externes » nationales des

<sup>18</sup>. Ce sentiment a été exacerbé par le contexte politique du projet de réforme de la fonction publique et de l'évaluation des fonctionnaires (dont les enseignants font partie).

acquis des élèves au Luxembourg coordonnées par l'ADQS : les épreuves communes de cycle 4.2 (grade 6)<sup>19</sup>, les épreuves communes de V<sup>e</sup>/9<sup>e</sup> (grade 9)<sup>20</sup> et les ÉpStan (grades 1, 3 et 9). Alors que les deux premiers types d'évaluations remplissent une fonction spécifique (orientation des élèves pour les épreuves communes de grade 6 et bilan des acquis pour les épreuves communes de grade 9), les ÉpStan tentent de conjuguer des fonctions différentes et de poursuivre des objectifs multiples, parfois difficilement conciliables. Toutes ces évaluations sont en outre construites par des acteurs différents avec des cadres de référence<sup>21</sup> et d'élaboration différents, ce qui engendre une certaine confusion chez les enseignants. Le défi de l'ADQS serait de clarifier la situation et de travailler à l'articulation et à la complémentarité des différentes évaluations des acquis des élèves présentes au sein du système scolaire, quitte à repenser les fonctions et les caractéristiques de certains dispositifs.

### Amener les établissements scolaires à utiliser le « classeur école » et le « classeur lycée » pour définir leur PRS ou leur PDS

L'ADQS a pu constater que les informations qu'elle met à la disposition de chaque équipe pédagogique dans le cadre du « classeur école » ou du « classeur lycée » demeuraient sous-exploitées au moment de définir leur projet trisannuel de développement scolaire. En particulier, pour identifier d'éventuels problèmes et formuler des pistes de développement possible, il n'y aurait que très peu de tentatives d'analyse et de mise en relation des informations rassemblées au sein des quatre parties de l'outil (données démographiques, performances scolaires, processus scolaires, perceptions des acteurs). Il s'agit donc pour l'ADQS de mener une analyse de l'outil proposé et d'envisager des pistes pour faciliter la lecture et la mise en relation des données mises à disposition. L'idée est aussi de compléter l'outil actuel avec d'autres données (de contexte et de performance) susceptibles de mettre en évidence la « valeur ajoutée » de chaque établissement scolaire.

### Ne pas laisser les enseignants seuls avec des constats

Si la mise à disposition de données comparatives fiables est un élément essentiel du pilotage et du développement de la qualité du système scolaire, il ne suffit pas d'établir des constats pour permettre aux acteurs de les dépasser. Sur le plan des pratiques pédagogiques par exemple, on peut supposer que des pistes didactiques, formulées par des experts disciplinaires (de la cellule de l'innovation du SCRIPT et/ou de l'université) au départ des résultats des évaluations externes, seraient bien accueillies par

**19.** Ces épreuves ont lieu au courant du deuxième trimestre de l'année scolaire, à des dates déterminées, en disposant d'un temps bien défini et avec un document reprenant des consignes de passation. Les types d'items sont variables d'une année à l'autre et sont corrigés par l'enseignant. Les résultats permettent de comparer chaque élève à l'ensemble des élèves ayant passé l'épreuve. Ils sont utilisés dans la procédure d'orientation des élèves vers l'enseignement secondaire.

**20.** Ces épreuves, élaborées par le MENJE, portent exclusivement sur le contenu des programmes des branches concernées et traitées en classe et n'exigent, dès lors, aucune préparation complémentaire par rapport à une évaluation en classe ordinaire. Les copies des élèves sont corrigées uniquement par le titulaire de la classe et les notes obtenues comptent au même titre qu'une évaluation en classe. Les résultats sont utilisés par les écoles pour situer leurs classes au sein du lycée et au niveau national.

**21.** À ce sujet, il faut rappeler que contrairement au référentiel de l'enseignement fondamental, les référentiels pour l'enseignement secondaire ne présentent, à l'heure actuelle, que des « standards de contenus » et pas encore de « standards de performance », ce qui rend encore très subjective l'appréciation des performances scolaires des élèves par les enseignants, mais également par les concepteurs des épreuves externes d'évaluation.

les enseignants, comme c'est le cas dans d'autres systèmes scolaires. Sans préjuger de leurs compétences et de leur liberté pédagogique, les enseignants placés devant le constat que leurs élèves font « moins bien qu'attendu » ou qu'une proportion élevée d'entre eux n'atteint pas les exigences du programme disposeraient ainsi d'une opportunité ou d'une aide pour réagir aux constats dressés.

### **Optimiser les ressources humaines et financières pour le développement scolaire**

Les compressions budgétaires accrues qui touchent actuellement l'éducation ont conduit à une réduction des ressources humaines affectées aux écoles. Le ministère de l'Éducation est évidemment conscient de ce contexte particulier et le fait que certains établissements parviennent, à partir de leurs propres ressources, à entreprendre et à mener à bien des projets de développement ne peut justifier que rien ne soit fait pour les établissements qui n'y parviennent pas. C'est pourquoi le Ministère entend à l'avenir examiner les procédures d'allocation des ressources aux établissements dans le but d'encourager ceux-ci à s'engager pleinement dans le développement de leur qualité. À l'heure actuelle, les écoles fondamentales qui en ont fait la demande ont déjà reçu entre 2 et 12 heures par semaine afin de définir leur PRS. Les écoles secondaires ont quant à elles la possibilité de dégager 7 heures par semaine pour travailler à leur PDS. Mais cela reste insuffisant et l'effort doit à présent être dirigé vers l'accompagnement des équipes pédagogiques qui en font la demande.

## BIBLIOGRAPHIE

BAUMERT J., DEMMRICH A., 2001, "Test motivation in the assessment of student skills – The effects of incentives on motivation and performance", *European Journal of Psychology of Education*, No. 16, Springer, p. 441-462.

BOND T., FOX C., 2010, *Applying the Rasch Model – Fundamental measurement in the human sciences*, 2<sup>nd</sup> ed., New York, London, Routledge.

BURTON R., MARTIN R., 2008, « L'orientation scolaire au Luxembourg : "Au-delà de l'égalité des chances... le gâchis d'un potentiel humain" », in MARTIN R., DIERENDONCK C., MEYERS C., NOESEN M., *La place de l'école dans la société luxembourgeoise de demain*, Bruxelles, De Boeck, p.165-186.

DIERENDONCK C., SONNLEITNER P., UGEN S., KELLER U., FISCHBACH A., MARTIN R., 2013, *La mesure de la motivation et de l'effort des élèves dans le cadre des épreuves standardisées au Luxembourg*, Actes du congrès de l'actualité de la recherche en éducation et formation (AREF-AECSE), Montpellier.

DIERENDONCK C., MARTIN R., 2008, « Le pilotage des systèmes éducatifs », in MARTIN R., DIERENDONCK C., MEYERS C. NOESEN M., *La place de l'école dans la société luxembourgeoise de demain*, Bruxelles, De Boeck, p. 429-476.

EBEL R., FRISBEE D., 1986, *Essentials of educational measurement*, 4<sup>th</sup> ed., Toronto, Prentice Hall.

GUSTAFSSON J.-E., 1980, "Testing and obtaining fit of data to the Rasch model", *British Journal of Mathematical and Statistical Psychology*, No. 33, vol. 2, The British Psychological society, p. 205-233.

KESKPAIK S., ROCHER T., 2012, *Les évaluations à faibles enjeux : quel rôle joue la motivation ? Une expérience à partir de PISA*, Communication dans le cadre du 24<sup>e</sup> colloque de l'Admée-Europe, Luxembourg.

MARTIN-LÖF P., 1974, "The Notion of Redundancy and Its Use as a Quantitative Measure of the Discrepancy between a Statistical Hypothesis and a Set of Observational Data", *Scandinavian Journal of Statistics*, vol. 1, No. 1, p. 3-18.

MESSICK S., 1995, "Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning", *American Psychologist*, vol. 50, No. 9, p. 741-749.

NAGY G., NEUMANN M., 2010, „Psychometrische Aspekte des Tests zu den voruniversitären Mathematikleistungen in TOSCA-2002 und TOSCA-2006: Unterrichtsvalidität, Rasch-Homogenität und Messäquivalenz“, in TRAUTWEIN U., NEUMANN M., NAGY G., LÜDTKE O., MAAZ K., *Schulleistungen von Abiturienten – Die neu geordnete gymnasiale Oberstufe auf dem Prüfstand*, Wiesbaden, Springer VS, p. 281-306.

OCDE, 2009, *PISA 2006, Technical report*, Paris, OECD Publishing.

O'NEIL H. , SUGRUE B., ABEDI J., BAKER E., GOLAN S., 1996, *Final Report of Experimental Studies on Motivation and NAEP Test Performance*, CSE Technical Report 427, Los Angeles, CRESST, University of California.

SCALLON G., 2004, *L'évaluation des apprentissages dans une approche par compétences*, Bruxelles, De Boeck.

SCRIPT, 2007, « Courrier de l'Éducation Nationale – Die Steuerung des Luxemburger Schulwesens », n° spécial, MENFP.

WARM T., 1989, "Weighted Likelihood Estimation of Ability in Item Response Theory", *Psychometrika*, vol. 54, No. 3, Springer, p. 427-450.

WOLF L., SMITH J., 1995, "The Consequence of Consequence: Motivation, Anxiety and Test Performance", *Applied Measurement in Education*, vol. 8, No. 3, p. 227-242.

WRIGHT B., LINACRE J. , GUSTAFSSON J.-E., MARTIN-LÖF P., 1994, "Reasonable mean-square fit values", *Rasch Measurement Transactions*, vol. 8, No. 3, p. 370.

WU M., ADAMS R., WILSON M., HALDANE S., 2007, *ACER ConQuest. Version 2.0: Generalised item response modelling software*, Camberwell, ACER Press.

ZERPA C., HACHEY K., VAN BARNEVELD C., SIMON M., 2011, *Modeling Student Motivation and Students' Ability Estimates From a Large-Scale Assessment of Mathematics*, SAGE Open.

