

Auteurs :

Christophe Dierendonck

Marianne Milmeister

Paul Milmeister

Christiane Weis

Sonja Ugen

Antoine Fischbach

Romain Martin

1 Vers une démarche d'assurance-qualité pour les Épreuves Standardisées

05

Depuis la première édition des Épreuves Standardisées (ÉpStan) durant l'année scolaire 2008-2009, le dispositif a considérablement évolué sur les plans conceptuel, organisationnel et technologique, mais également en termes de rapports adressés aux différents acteurs concernés par ces évaluations externes des acquis scolaires. Ces transformations successives ont été opérées en fonction des contraintes et des opportunités rencontrées au fil du temps, mais sans référence explicite à une démarche d'assurance-qualité globale. ●

1.1 CONTEXTE ET APERÇU

Le dispositif étant stabilisé depuis 2010-2011, le moment semblait idéal pour prendre du recul et initier une démarche d'assurance-qualité. En 2013, il a donc été décidé de collecter des données auprès des acteurs concernés par les ÉpStan, tant en externe au niveau des acteurs du terrain scolaire (directions, coordinateurs ÉpStan dans les lycées, enseignants du fondamental et du secondaire, inspecteurs, présidents des comités d'école, élèves, parents et syndicats actifs dans le domaine de l'éducation) qu'en interne au niveau de l'ancienne Agence pour le développement de la qualité scolaire (ADQS)¹ du Service de Coordination de la Recherche et de l'Innovation pédagogiques et technologiques (SCRIPT) du Ministère de l'Éducation nationale, de l'Enfance et de la Jeunesse (MENJE) et du *Luxembourg Centre for Educational Testing* (LUCET) de l'Université du Luxembourg (UL). Le but de cette collecte de données était de permettre la conduite d'une démarche d'assurance-qualité en fondant les réflexions et les éventuelles décisions sur des constats tirés de données empiriques.

L'étude s'est articulée autour de trois questions principales qui ont progressivement émergé lors de réunions de travail de l'équipe ÉpStan, lors de plusieurs échanges informels avec des enseignants et suite à certaines observations menées dans les classes durant les tests :

- (1) Dans quelle mesure le principe d'une évaluation externe des acquis scolaires des élèves est-il soutenu par les acteurs scolaires ?
- (2) Dans quelle mesure les élèves font-ils les tests ÉpStan sérieusement ?
- (3) Dans quelle mesure les rapports envoyés aux acteurs dans le cadre des ÉpStan sont-ils compris et utilisés ?

Pour répondre à ces questions, quatre types de données ont été récoltées. Les opinions de tous les types d'acteurs concernés par le dispositif ÉpStan ont été sondées par des *focus groups* (a). Les enseignants du cycle 3.1 et de V^e/9^e ont en outre été invités à compléter un questionnaire à large échelle (b). Au niveau des élèves de V^e/9^e, deux autres types de données ont été recueillies : des données auto-rapportées de motivation et d'effort ont été collectées au moyen de questions posées au début et à la fin de chacun des trois tests ÉpStan (c) et des données comportementales ont été enregistrées par la plate-forme de testing lors de la passation des tests par les élèves (d).

Le présent texte a pour but de synthétiser les résultats de recherche et de tirer un bilan intermédiaire de la démarche d'assurance-qualité entreprise. Il est structuré en fonction des trois questions de recherche présentées ci-dessus. Dans chaque partie sont repris deux types de propos : d'une part, des constats formulés par les auteurs de l'étude à partir des données récoltées et, d'autre part, des clarifications sur l'outil des ÉpStan présentées sous la forme d'encadrés explicatifs rédigés par l'équipe ÉpStan.

Avant de poursuivre, que les représentants du terrain soient remerciés pour leur précieuse collaboration. Sans leur participation au questionnaire et l'échange franc lors des *focus groups*, cette étude n'aurait pas été possible. ●

¹ Suite à la restructuration du SCRIPT en 2017, le service comprend actuellement six divisions. La division « du traitement de données sur la qualité de l'encadrement et de l'offre scolaire et éducative » reprend notamment la partie monitoring du portfolio de l'ancienne « Agence pour le Développement de la Qualité Scolaire » (ADQS).

1.2 MÉTHODOLOGIE

Les méthodologies utilisées pour les différentes collectes de données, de type quantitatif et qualitatif, ont été décrites de façon précise dans le chapitre 7 du rapport national 2011-2013 des ÉpStan (Dierendonck et al., 2015, p.99-109).

Seules quelques informations essentielles sont reprises ici pour permettre au lecteur de comprendre la nature des données récoltées :

a) En 2014, douze **focus groups** réunissant au total 77 participants ont été organisés. Ces discussions en groupe ont rassemblé différents participants : des directeurs de lycée, des coordinateurs ÉpStan des lycées, des enseignants du cycle 3.1 et de V^e/9^e, des inspecteurs, des élèves de V^e/9^e, des représentants de syndicats actifs dans le domaine de l'éducation, des membres de l'ancienne ADQS (SCRIPT) et des membres du LUCET. Faute de volontaires, deux *focus groups* prévus initialement n'ont pas pu avoir lieu : ceux avec les parents d'élèves du cycle 3.1 et avec les présidents des comités d'école.

b) Les enseignants du cycle 3.1 et les professeurs d'allemand, de français et de mathématiques de V^e/9^e ont été sollicités en 2014 pour participer à l'étude par **questionnaire**.

Sur l'ensemble des sollicitations, 152 enseignants du cycle 3.1 et 131 enseignants de V^e/9^e se sont connectés à la plate-forme de sondage. Les participants avec un taux d'omissions aux questions de plus de 25% ont été retirés des échantillons. Au final, l'échantillon pour le cycle 3.1 est constitué de 150 enseignants (taux de participation d'environ 41%) et celui pour la V^e/9^e se compose de 121 enseignants (taux de participation d'environ 13%) dont 43 professeurs de mathématiques, 34 professeurs d'allemand et 44 professeurs de français. Il est difficile de spécifier le nombre exact d'enseignants sollicités, même en se basant sur les listes des classes ayant participé aux ÉpStan : au secondaire, certains enseignants exercent dans plusieurs classes de V^e/9^e et au fondamental, des enseignants travaillent en *team teaching* ou à temps partiel. Le nombre d'enseignants sollicités a été estimé autour de 360 pour

le fondamental et autour de 1010 pour le secondaire. Les taux de participation mentionnés ci-dessus sont par conséquent aussi à considérer comme une estimation de la participation des enseignants.

c) Les **données auto-rapportées de motivation et d'effort** ont été recueillies lors des ÉpStan de V^e/9^e. Au début de chacun des trois tests, les élèves ont été invités à estimer, sur une échelle à quatre modalités, leur degré de motivation pour le test. En fin de test, ils ont dû estimer l'effort consenti pour le test et l'effort qu'ils auraient consenti si le test avait compté comme note scolaire. L'instrument utilisé pour mesurer l'effort des élèves (composé d'échelles de 1 à 10) est une adaptation du « thermomètre d'effort » développé dans le cadre de PISA (*Programme for International Student Assessment*) de l'OCDE (Dierendonck et al., 2013).

d) Afin de disposer d'un autre indicateur du sérieux des élèves lors des tests ÉpStan, il a semblé intéressant d'exploiter les potentialités du testing par ordinateur en enregistrant une série de **données de type comportemental**. Actuellement, il est possible d'exploiter les informations rendant compte du temps durant lequel chaque stimulus² a été affiché à l'écran. En totalisant les temps d'affichage de tous les stimuli, on dispose d'une variable rendant compte du temps total *a priori* consacré au traitement des stimuli par les élèves. Il faut nuancer qu'il est impossible d'assimiler avec certitude le temps d'affichage des tâches au temps investi effectivement par les élèves pour les traiter : on peut en effet envisager la possibilité que lors de l'affichage des questions, les élèves regardent (ou non) l'écran, sans véritablement traiter les tâches et les questions proposées. Cette variable est, à l'heure actuelle, le meilleur indicateur à disposition pour tenter d'estimer objectivement le degré de sérieux avec lequel les élèves ont pris part au test. ●

² Les épreuves sont constituées de plusieurs stimuli (par exemple, différents textes dans le test d'allemand) et plusieurs questions sont associées à chaque stimulus.