

Demande d'accès à des données administratives à des fins statistiques au titre de l'article 7bis de la loi n°51-7111 du 7 juin 1951 modifiée

1. Date de l'examen de la demande en commission

Examen par la commission « Services publics et services aux publics » du 23 septembre 2025.

2. Service demandeur

Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche -
Direction de l'évaluation, de la prospective et de la performance (DEPP)

3. Source administrative demandée :

Lors de la Journée Défense et Citoyenneté (JDC), obligatoire pour tous les jeunes français entre 16 et 25 ans, un test de détection des difficultés de lecture (DDL) est organisé. La DEPP est déjà destinataire des résultats de ce test à partir desquels elle publie annuellement [une note d'information](#). La demande porte sur la transmission de données identifiantes supplémentaires, de sorte à pouvoir appairer les données scolaires des élèves avec leur résultat à ce test.

4. Organisme détenteur des données demandées

Ministère des Armées - Direction du service national et de la jeunesse (DSNJ).

5. Données demandées :

Pour pouvoir enrichir les données scolaires des élèves avec leur résultat au test de lecture, la DEPP a besoin de disposer de **nouvelles données identifiantes sur les participants (nom, prénom, UAI du dernier établissement scolaire fréquenté)**, en plus de celles dont elle dispose déjà (sexe, date de naissance, commune de résidence).

La DEPP demande ces données identifiantes **pour les années 2020 à 2025**. L'étude principale, visant à étudier le lien entre les résultats à l'évaluation nationale exhaustive de début de sixième et les résultats au test de lecture de la JDC (cf. §7.1), portera sur les élèves scolarisés en sixième en 2017. Ces élèves sont principalement nés en 2006 (pour 88 % d'entre eux), en 2005 (9 %) et en 2007 (3 %). Or 53 % des élèves passent la JDC l'année de leurs 17 ans, 37 % l'année de leurs 18 ans, 6 % l'année de leurs 19 ans, 2 % l'année de leurs 20 ans et 2 % au-delà. L'appariement avec les millésimes 2020 à 2025 du test DDL devrait permettre de couvrir environ 95 % des élèves français scolarisés en sixième en 2017.

En plus de ces données identifiantes, la DEPP souhaiterait, de façon régulière, disposer de davantage d'informations sur la situation scolaire / professionnelle du jeune, figurant sur la fiche de renseignements complétée par le jeune avant de passer le test :

- Si le jeune est toujours scolarisé : son type d'établissement (public, privé, centre de formation agricole, etc.), la filière et la classe fréquentées ;
- Si le jeune n'est plus scolarisé : son type de contrat, sa profession, la dernière classe suivie et son année de départ.

6. Besoin de données nominatives ou identifiantes

La demande porte principalement sur la transmission de données identifiantes. Ces données serviront à retrouver l'Identifiant National Elève (INE) des participants au test, en appariant les données identifiantes des jeunes avec celles présentes dans les bases scolarité.

Une fois l'INE des jeunes retrouvé, les résultats du test DDL pourront servir à enrichir plusieurs sources détenues par la DEPP (cf. §9).

7. Finalité des traitements prévus avec les données demandées, notamment place par rapport au système d'information existant :

L'appariement entre les données des tests de détection des difficultés de lecture et les données sur les résultats et les parcours des élèves donnera lieu à plusieurs études. Les études envisagées à ce stade sont les suivantes :

1. Etudier dans quels domaines des lacunes en sixième sont prédictives de l'illettrisme à la JDC. Cela permettrait de déterminer des profils « à risque » et de cibler, sur ces élèves, des actions de remédiation dès la sixième.
2. Etudier la trajectoire scolaire et le profil des jeunes « décrocheurs », c'est-à-dire sortis du système scolaire sans diplôme plus élevé que le diplôme national du brevet, de sorte à comprendre les déterminants du décrochage.
A quel âge ces jeunes sortent-ils du système scolaire ? Dans quelles filières étaient-ils inscrits ? Avaient-ils redoublé ? Quels étaient leurs résultats aux dernières évaluations qu'ils ont passées ? Quel est leur milieu social (appréhendé par l'indice de position sociale de la DEPP) ? Quelles différences territoriales ?
3. Etablir une correspondance entre les profils de lecteurs déterminés lors de la JDC et les profils des élèves de première année de CAP en littérature, établis à partir de leurs résultats au test de positionnement de début d'année.

8. Nature des travaux statistiques prévus et description des traitements

Pour retrouver l'INE des participants au test DDL, un appariement sur données identifiantes sera effectué.

Les données de la JDC utilisées seront le nom, le prénom, la date de naissance, le sexe, le département de résidence et l'identifiant établissement (UAI) du dernier établissement scolaire fréquenté par le jeune.

Les données DEPP utilisées seront issues d'extractions de la base SYSCA (système d'information statistiques consolidée académique), mobilisées sur plusieurs années et plusieurs niveaux.

Un appariement exact sera d'abord réalisé, puis des contraintes seront relâchées successivement. Pour s'assurer qu'un appariement ne soit pas rejeté en raison de différences orthographiques dans les noms propres, des appariements probabilistes mobilisant des calculs de distances de Levenshtein pourront être réalisés.

9. Description des appariements prévus le cas échéant

Les données du test DDL pourront être appariées avec les sources suivantes, détenues par la DEPP :

- Les évaluations nationales exhaustives de 6^e (millésimes 2017 à 2020)
- Les évaluations nationales exhaustives de début de seconde, GT et pro (millésimes 2020 à 2024)
- Les tests de positionnement de début de première année de CAP (millésimes 2020 à 2024)
- Les données des examens (millésimes 2020 à 2024 pour le DNB, 2023 à 2025 pour le bac)
- Les données du panel CP 2011. La quasi-totalité des entrants en CP en 2011 ont eu 17 ans en 2022 et 18 ans en 2023.

10. Périodicité de la transmission

La DEPP souhaite une transmission :

- Ponctuelle pour les données identifiantes des participants aux tests DDL
- Mensuelle pour les données sur la situation scolaire / professionnelle du jeune. Ces nouvelles données viendraient compléter les données déjà transmises mensuellement par la DSNJ à la DEPP.

11. Output final : nom de la source ou du système d'information

Pour les années pour lesquelles les données identifiantes sont demandées (2020 à 2025), l'output final est la table des résultats au test DDL enrichie de l'INE, pour tous les participants qui auront pu être retrouvés dans les bases SYSCA de la DEPP.

Ces tables pourront ainsi facilement être appariées avec les données DEPP décrites au paragraphe 9.

12. Publications prévues le cas échéant

Les résultats des études réalisées à partir de ces appariements seront valorisés dans divers publications de la DEPP visant à décrire la trajectoire des jeunes en difficulté de lecture.

13. Diffusion de la base de données individuelles en output :

Les résultats au test DDL des élèves du panel CP 2011 pourront venir enrichir les données d'évaluation du panel. Ces données sont mises à la disposition des chercheurs dans le cadre de Quetelet-Progedo Diffusion.

14. Mots-clés :

Journée Défense et Citoyenneté, JDC, test de Détection des difficulté de lecture , Illettrisme, Jeunes