

Rapport d'activité final

Projet COMETE

Résumé « Grand Public »

Environ 30% des enfants présenteraient des difficultés d'apprentissage de l'écriture et pour 10%, celles-ci persisteront. L'apprentissage de l'écriture est déterminant pour la réussite scolaire et l'estime de soi. Actuellement, l'identification des difficultés d'écriture est tardive et repose initialement sur les observations des enseignants ou des parents. Puis, des bilans cliniques standardisés peuvent être réalisés pour évaluer les caractéristiques statiques de l'écriture (p. ex., BHK) et les difficultés associées (p. ex., coordination). Cependant, les résultats à ces bilans varient entre les évaluateurs et ne sont pas répétables, limitant le suivi des progrès des enfants. Les tablettes numériques permettent de mesurer de nouveaux paramètres comme le mouvement du tracé et la pression sur l'écran. Plusieurs études ont proposé des évaluations de l'écriture avec des tablettes, identifiant les paramètres les plus pertinents en comparaison à des bilans cliniques ou des questionnaires remplis par les enseignants. Ces comparaisons limitent l'évaluation à des enfants scripturaux (sachant écrire). De plus, bien que les paramètres issus des tablettes ne dépendent pas de l'évaluateur, leur intérêt dans le suivi des enfants est inexploré. Enfin, en utilisant des formes graphiques ces évaluations pourraient s'étendre à des enfants préscripturaux (ne sachant pas encore écrire).

L'objectif principal de cette étude est de proposer une nouvelle méthode d'évaluation de l'écriture sur tablette validée en comparaison aux évaluations des thérapeutes et des enseignants. Les objectifs secondaires sont d'étudier chez les enfants scripturaux si entre deux évaluations les paramètres mesurés (1) sont les mêmes (2) identifient les difficultés et (3) chez des enfants préscripturaux le lien entre les paramètres mesurés en reproduisant des formes et la coordination.

Valider une évaluation de l'écriture auprès de différents intervenants autour de l'enfant faciliterait leur communication. L'utilisation de listes de mots aléatoires adaptées à l'âge de l'enfant permettrait le suivi des enfants.

Pour cela, cent enfants scripturaux seront recrutés (50 sans difficulté d'écriture et 50 avec), et 50 enfants préscripturaux.

Les enfants scripturaux (CP à CE2) recopieront aléatoirement deux listes de mots lors de deux évaluations de 5 minutes. Les enfants préscripturaux (grande section) recopieront aléatoirement des formes pendant environ 5 minutes.

La coordination de tous les enfants sera évaluée par les parents. Pour les enfants scripturaux l'écriture sera évaluée par l'enseignant avec un questionnaire. Pour les enfants scripturaux pris en soin pour ces difficultés d'écriture, le score du BHK sera demandé au thérapeute.

Chez les enfants scripturaux, une comparaison sera faite avec l'évaluation de l'enseignant et du thérapeute, et la répétabilité des paramètres mesurés par la tablette seront évaluées statistiquement. Chez les enfants préscripturaux, une corrélation sera faite entre les variables issues de la tablette et la coordination.

Les perspectives seraient de proposer un apprentissage de l'écriture individualisé reposant sur la caractérisation de l'écriture de chaque enfant.

1. Etat d'avancement

Retour sur le planning prévisionnel du projet

Période	Prévisionnel	Etat d'avancement
Septembre - octobre 2023	Contact sera établi avec les écoles et les centres pour mettre en œuvre la récolte de données	Contact établi avec les écoles De nouveaux centres sont sollicités
Septembre 2023 - juin 2024	Récolte de donnée	Fini dans les écoles
Juillet - aout 2024	Compléter l'analyse des données	L'ensemble des données acquises ont été traitées Seules les données des enfants en CP-CE ont été analysées
Juillet - septembre	Écriture de deux articles scientifiques	Un article est en cours d'écriture (Soumission prévue début 2025)

Etat d'avance des tâches du dernier reporting

Dates - Délais	Tache	Barrières/Solutions	Avancement
Décembre 2024 – Juin 2025	Finalisation de la récolte de données (Enfants avec des difficultés d'écriture)	Partenariat contractuel avec le CRTA Sollicitation du réseau de FRISBEE & de l'ANFE	Fait
Janvier - aout 2025	Compléter l'analyse des données avec les données acquises	Stagiaire Ingénieur en Master 2	En cours
Janvier - aout 2025	Analyses des données des enfants en Grande Section		Reporté
Janvier - aout 2025	Finalisation du premier article et rédaction du second	RAS	En cours

Un retard a été pris face à la difficulté de recrutement.

Le travail mené par l'étudiante ingénieure en Master 2 (Elise Ourceau), entre février et aout 2024, sur les données acquises a tout de même permis d'obtenir des résultats prometteurs qui ont fait l'objet d'une présentation à la SOFMER 2024 et à l'EACD 2025.

Le traitement des données des formes prescripturales s'est avéré plus complexe que prévue et très différente de l'écriture, et nécessitera un temps dédié pour identifier les variables d'intérêt.

Un article en cours d'écriture sur la reproductibilité et la validité de l'évaluation sur tablette numérique.

Les tâches à venir

Dates - Délais	Tache	Barrières/Solutions
Juillet - septembre 2025	Compléter l'analyse avec les données acquises	RAS
Juillet - décembre 2025	Finalisation du premier article et rédaction du second	RAS

2. Hypothèses

Les hypothèses sont :

- Une évaluation du tracé, lors de l'écriture de listes de mots, permettrait d'identifier, de manière spécifique et sensible, les difficultés d'écriture de l'enfant avec un nombre réduit de variables statiques, cinématiques et de pression, à l'école, à la maison ou lors d'évaluations par des professionnels de réadaptation.
- L'utilisation de différentes listes de mots adaptées à l'âge de l'enfant et ayant les mêmes caractéristiques permettrait l'évaluation fiable et reproductible des difficultés d'écriture.
- L'utilisation de tâches de tracé de formes prescripturales pourrait permettre l'identification précoce des enfants à risque de présenter des troubles de l'écriture.

3. Protocole

3.1. Population

Pour les enfants en CP et CE1, 48 enfants ont été recrutés dans des 3 écoles et 21 dans 4 structures (Tableau 1).

Tableau 1 : Effectif des enfants recrutés en CP-CE1

		Nombre d'enfants	HPSQ > 14	HPSQ > 20
Ecoles	CP	24	12	5
	CE1	24		
Structures	CP	7	21	21
	CE1	14		

Pour les grandes sections, 50 enfants ont été recrutés dans 3 écoles.

Critères d'inclusion :

- Des enfants, garçons et filles, en grande section, CP, CE1 ou début CE2.

Critères de non-inclusion :

- Enfants présentant des troubles aigus neurologiques ou musculo-squelettiques associés.
- Enfants présentant une déficience intellectuelle diagnostiquée (Déclaration des parents).

Critères d'exclusion :

- Les enfants suivants seront retirés de l'étude, si l'évaluation est incomplète ou si l'enfant n'arrive pas à se familiariser avec la tablette (doit effacer pour réécrire plus de la moitié des mots).

3.2. Matériel

Le *Handwriting Proficiency Screening Questionnaire* (HPSQ) consiste en une série de 10 questions qui permettent à l'enseignant d'évaluer l'écriture de l'enfant. Les enfants sont considérés avec des difficultés d'écriture si le score au HPSQ était supérieur à 14.

Le *Brave Handwriting Kinder* (BHK) [23] est une évaluation clinique de l'écriture réalisée en pratique courante par les thérapeutes (c.-à-d., ergothérapeutes, orthophonistes, neuropsychologues, psychomotriciens). Les enfants sont considérés avec des difficultés d'écriture si leur score au BHK est inférieur à -1 écart-type.

Afin d'évaluer la coordination, il était demandé aux parents de remplir le questionnaire d'identification du trouble développemental de la coordination (DCDQ-FE) [41]. Le DCDQ-FE explore 15 habilités de la vie quotidienne de l'enfant, dont les activités « papier-crayon ».

Les évaluations sur tablettes numériques étaient réalisées sur le module évaluation de l'application Kaligo (Learn&Go). L'application a été installée sur des tablettes (Samsung Galaxy S6 Light). Les tablettes étaient placées dans un support afin qu'il n'y ait pas de différence de niveau entre la tablette et l'avant-bras de l'enfant. Les évaluations sur tablettes ont été réalisées avec une personne de l'équipe de recherche.

3.3. Recueil de données

3.3.1. Enfants scripturaux

Pour les enfants scripturaux, au niveau CP et CE1, une évaluation de 5 minutes, consistant à recopier des blocs de mots adaptés à leur niveau scolaire, était réalisée deux fois (10 minutes au total). Les Blocs (composés de 5 listes de 8 mots) utilisés lors de ces deux évaluations étaient construits pour avoir un maximum de critères (fréquence et type de lettre) en commun [42] mais avec des mots différents. L'ordre des Blocs, listes et mots était aléatoire. L'enfant aura 1 minute par liste pour recopier une liste de huit mots, si l'enfant ne finit pas la liste en une minute, il passera tout de même à la suivante afin que le test dure 5 minutes.

Afin que les observations des enseignants puissent être comparables, il leur était demandé de compléter le *Handwriting Proficiency Screening Questionnaire* (HPSQ). Dans les structures accueillant des enfants avec des difficultés d'écriture, le score du BHK, réalisé en pratique courante, était recueilli auprès du thérapeute ayant effectué l'évaluation. En parallèle, il était demandé aux parents de remplir le questionnaire d'identification du trouble développemental de la coordination (DCDQ-FE) [41] pour évaluer la coordination.

3.3.2. Enfants préschrifturaux

Pour les enfants en Grande Section l'évaluation de 5 minutes consistait à recopier des formes et a été réalisée une fois. Les formes à réaliser reprenaient des formes retrouvées dans des exercices réalisés à ce niveau scolaire. L'enfant devait réaliser deux types de formes dont l'ordre était randomisé : des formes individuelles à répéter et des formes continues. Pour les formes continues, il existait deux versions de formes continues dont l'utilisation sera aussi randomisée.

En parallèle, il était demandé aux parents de remplir le questionnaire d'identification du trouble développemental de la coordination (DCDQ-FE) [41] pour évaluer la coordination.

3.4. Traitement des données

3.4.1. Pré-traitement

Les données étaient directement sauvegardées sur les tablettes.

Les fichiers extraits (.inkml) ont été traités avec un script codé sur Python. Ce code python permettait de filtrer le signal avec un filtre *butterworth* de 4^{ème} ordre sans déphasage avec un *cut-off* à 15hz.

3.4.2. Algorithmes de classification

A partir des données des variables ont été calculées :

- Vitesse moyenne
- Fréquence médiane du spectre de la vitesse
- Accélérations horizontales

- Ratio temps en l'air/sur la feuille
- Pression médiane
- 60ème percentile de la pression
- Nombre de pics de changement de la vitesse
- Fréquence médiane du spectre de vitesse de la pression
- Densité
- Taille
- Forme (Parcours de Graham)
- Concordance avec le modèle

A partir de la librairie Python *scikit*, quatre algorithmes de classification ont été utilisés pour identifier les variables les plus pertinentes. Les quatre algorithmes de classification ont été : *Support Vector Machine*, *Extra Tree*, *Random Forest*, et *Logistic Regression* (régression logistique).

3.4.3. Réseaux de neurones

A partir de la librairie Python *scikit*, deux réseaux de neurones ont aussi été utilisés.

Le premier ResNet18 s'appuyait sur des images du tracé de l'enfant. Les images représentaient le tracé avec des points afin d'avoir la position et l'aspect temporel (distance entre les points), et la pression en niveau de gris (Figure 1).



Figure 1 : Image analysée par le réseau de neurones ResNet18

Le second réseau de neurones à 12 couches (*lstm*) utilisait les données de position (x, y), temporelles et de pression issue du pré-traitement (sans calcul de variable).

3.5. Analyse des données

3.5.1. Validité

A partir des résultats du HPSQ les enfants étaient identifiés sans difficulté d'écriture (>14) ou avec des difficultés d'écriture (≥ 14). Ce questionnaire a servi de référence pour les tableaux de contingence avec les algorithmes de classification et les réseaux de neurones.

Pour chacune des méthodes, la justesse (accuracy), la précision (precision), la sensibilité et la spécificité ont été calculées.

3.5.2. Fiabilité

Afin d'évaluer le test retest, la fiabilité des variables les plus pertinentes issues des algorithmes de classification a été évaluée entre les deux évaluations (Blocs) en calculant le coefficient de corrélation intraclasse (CCI), à l'aide du module *SimplyAgree* sur Jamovi (2.4.11).

Pour chacune des variables, le coefficient de variation ($CV = \text{Ecart-Type} / \text{Moyenne}$) a été calculé.

4. Résultats

4.1. Algorithmes de classification

4.1.1. Validité

Les méthodes *Extra Tree* et *Random Forest* n'ont pas montré des résultats supérieurs au hasard. Seuls la régression logistique et le SVM ont montré des résultats supérieurs au hasard (Figure 2). La régression logistique présente une justesse de 0.68, une précision de 0.59, une sensibilité de 0.76 et une spécificité de 0.62. Quant au SVM, il présente une justesse de 0.66, une précision de 0.65, une sensibilité de 0.49 et une spécificité de 0.78.

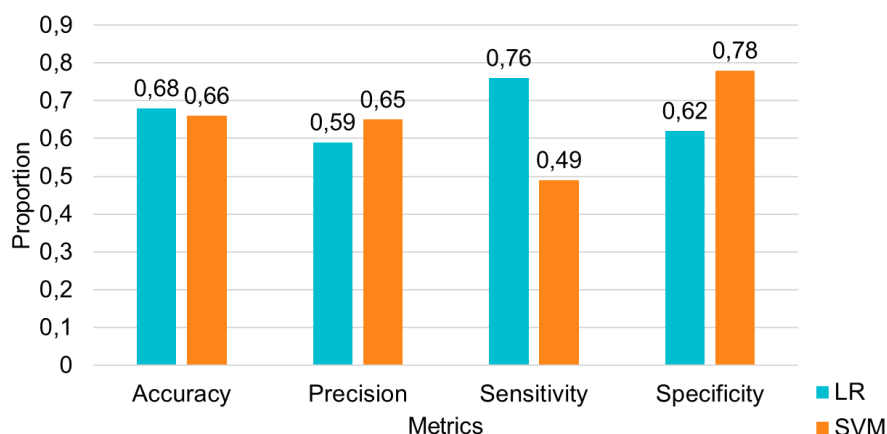


Figure 2: Validité des algorithmes de classification (LR : Régression logistique ; SVM : Support Vector Machine)

4.1.2. Fiabilité

La fiabilité des variables issues des algorithmes de classification est bonne ($ICC > 0.7$) voir excellente ($ICC > 0.9$). Les variables de pression et cinématiques sont celles qui présentent une majorité (respectivement, 2/3 et 5/8) de variables avec une excellente fiabilité.

Le coefficient de variation des variables oscille entre 2.46% et 20.84%.

Tableau 2 : Fiabilité des variables issues des algorithmes de classification - ICC : Coefficient de Corrélation Intraclasse ; CV : Coefficient de variation

Variables		ICC	CV
Pressure	Median pressure	0,923	10,6
	60th percentile	0,924	9,93
	Median frequency	0,85	4,98
Kinematic	Ration on paper/in air time	0,951	12,21
	Total time	0,929	16,29
	Stroke speed	0,882	20,84
	Total speed	0,916	18,66
	Number of speed inversion	0,945	14,5
	Speed median frequency	0,924	2,46
	Horizontal jerk	0,894	25,75
	Vertical acceleration	0,772	19,51
Static	Density	0,893	19,34
	Length	0,905	15,53
	Height	0,871	13,66
	Width	0,864	18,88

4.2. Validité des réseaux de neurones

Les réseaux de neurones montrent une validité inférieure aux algorithmes de classification. En effet, le réseau de neurones *ResNet18* présente une justesse de 0.56, une précision de 0.40, une sensibilité de 0.58 et une spécificité de 0.55. Quant au *Istm*, il présente une justesse de 0.55, une précision de 0.61, une sensibilité de 0.27 et une spécificité de 0.83.

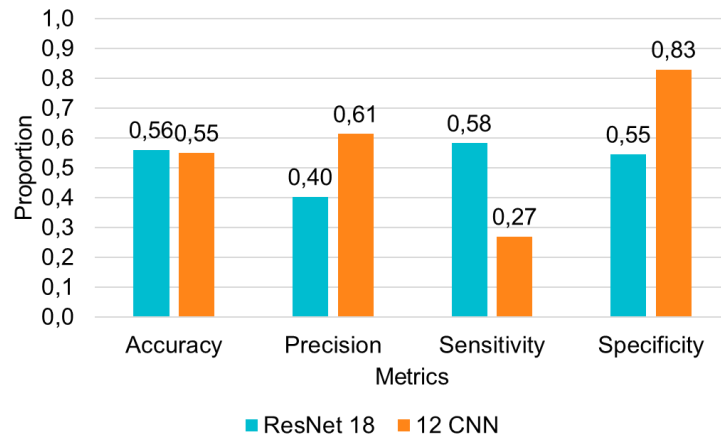


Figure 3 : Validité des réseaux de neurones (12 CNN = *Istm*)

5. Discussion

L'hypothèse principale était qu'une évaluation du tracé, lors de l'écriture de listes de mots, permettrait d'identifier, de manière spécifique et sensible, les difficultés d'écriture de l'enfant avec un nombre réduit de variables statiques, cinématiques et de pression, à l'école, à la maison ou lors d'évaluations par des professionnels de réadaptation.

Les résultats sont mitigés, surtout pour l'utilisation des réseaux de neurones, malgré une spécificité de 0.83 avec *Istm*. Les résultats des algorithmes de classification (régression logistique et SVM) présentent la meilleure justesse et précision. Cependant, leur sensibilité et leur spécificité ne sont pas identiques et semblent complémentaires. En effet, la régression logistique présente une meilleure sensibilité (0.76) que le SVM (0.49), ce qui pourrait en faire un outil diagnostique. En revanche, le SVM présente une meilleure sensibilité (0.78) que la régression logistique (0.62) ce qui permettrait son utilisation dans une optique de dépistage.

Cette différence entre les algorithmes de classification et les réseaux de neurones peut s'expliquer par la base de données. En effet, l'utilisation de mots différents crée une grande variabilité dans les images permettant l'entraînement et la validation des réseaux de neurones. Ce manque de données s'est matérialisé par une impossibilité de faire converger les réseaux de neurones avec des modèles complexes. C'est pour cela que des modèles simples ont été utilisés (*ResNet* : 18 couches et *Istm* : 12 couches).

Si ces résultats sont mitigés en termes de validité pure, cette étude demeure une des premières à utiliser une tablette commerciale [1] et des réseaux de neurones [2]. De plus, en comparaison avec la précédente étude ayant utilisé une tablette commerciale, hormis la taille de l'échantillon (390 enfants sans difficulté et 58 enfants avec des difficultés sévères), les caractéristiques de la population étudiée semblent pouvoir influencer les résultats. En effet, dans l'étude menée, les enfants avec des difficultés d'écriture ne présentaient pas forcément des difficultés sévères à l'inverse des études antérieures [1,2], ce qui rend plus difficile la catégorisation des enfants. Cette explication a été testée pour les réseaux de neurones en fixant un seuil au HPSQ supérieur à 20. Les résultats ont été grandement améliorés avec une justesse, une précision et une sensibilité supérieure à 0.7 (Figure 4).

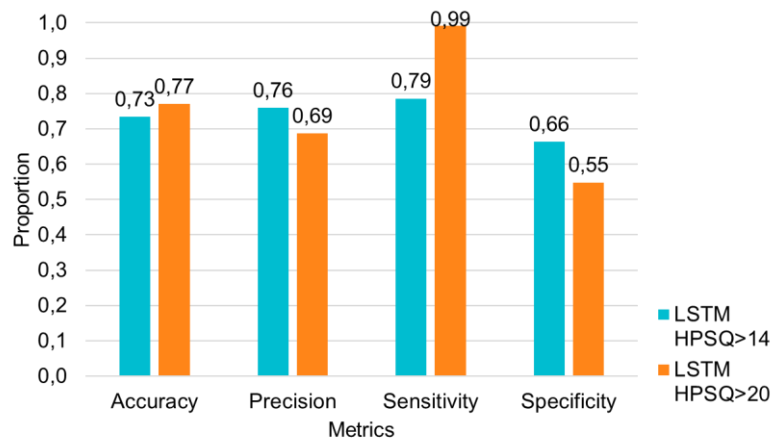


Figure 4 : Validité des réseaux de neurones avec un seuil supérieur à 20 au HPSQ (12 CNN = lstm)

Enfin, le fait que les évaluations utilisent des mots différents engendre une variabilité dans les données, ce qui pourrait expliquer le besoin d'une base de données plus conséquente pour permettre aux réseaux de neurones de converger plus facilement et utiliser des réseaux de neurones plus complexes. En comparaison aux précédentes études qui ont aussi mis en avant de meilleures performances avec les régressions logistiques et le SVM, l'augmentation de la taille d'échantillon serait aussi bénéfique pour les algorithmes de classifications.

La seconde hypothèse était que l'utilisation de différentes listes de mots adaptées à l'âge de l'enfant et ayant les mêmes caractéristiques permettrait l'évaluation fiable et reproductible des difficultés d'écriture.

Malgré des listes de mot différentes, la fiabilité des variables issues des algorithmes de classification était bonne voir excellente. Ce niveau de fiabilité serait adapté pour un suivi scolaire des enfants et acceptable pour une évaluation clinique. Cette étude reste la première à évaluer la fiabilité d'une évaluation sur tablette numérique et à proposer une méthode pour une évaluation régulière.

La fiabilité est en proportion meilleure pour les variables de pression et cinématiques (respectivement, 2/3 et 5/8 sont excellentes [ICC>0.9]) que pour les variables statiques (1/4 est excellente). Cependant, même si pour 13 variables sur 15 le coefficient de variation reste en dessous de 20%, il demeure au-dessus de 10% pour 10 variables sur 13. Même si les variables sont globalement fiables, le coefficient de variations sera à prendre en compte dans le suivi des enfants pour identifier de réels progrès.

La troisième hypothèse était que l'utilisation de tâches de tracé de formes prescripturales pourrait permettre l'identification précoce des enfants à risque de présenter des troubles de l'écriture.

Le traitement des données des formes prescripturales s'est avéré plus complexe que prévu et très différent de l'écriture. Ces difficultés sont retrouvées dans de précédentes études (DUI & Devilaïne) qui ont dû définir des variables plus précises (1^{er} arrêt après telle partie de la forme prescriptive). Cette analyse nécessitera un temps dédié pour identifier les variables d'intérêt.

De plus, les travaux se sont actuellement centrés sur les comparaisons des variables issues de l'évaluation sur tablette avec le HPSQ. Or chez les enfants en grande section la comparaison serait avec le DCDQ-FE ce qui nécessiterait dans un premier temps de comparer les données d'écriture des enfants en CP-CE1 avec le HPSQ.