Actes de CORIA-TALN-RJCRI-RECITAL 2025
Marseille, France, 30 juin au 4 juillet 2025

SEPT : Détecter les difficultés des étudiants à travers le clustering de leurs trajectoires émotionnelles et physique lors d'évaluations en ligne sur Moodle

Edouard Nadaud[1, 2] Antoun Yaacoub[1] Bénédicte Legrand[2] Lionel Prevost[1, 2]
(1) ESIEAlab, équipe Learning Data Robotic (LDR), ESIEA, 74 bis Av. Maurice Thorez, 94200 Ivry-sur-Seine, France
(2) Centre de Recherche en Informatique (CRI), Paris 1 panthéon Sorbonne, 31, rue Baudricourt 75013 Paris, France
edouard.nadaud@esiea.fr, antoun.yaacoub@esiea.fr,
benedicte.le-grand@univ-paris1.fr, lionel.prevost@esiea.fr

RÉSUMÉ

Imaginez une salle de classe où les difficultés et réussites des étudiants s'expriment non par des mots, mais par l'expression de leurs visages et mouvements, captés en temps réel pendant un quiz. Les méthodes d'enseignement dans le supérieur se font de plus en plus hybride et à distance. Les interactions directes sont réduites, rendant difficile la détection des moments de décrochage. Pour y remédier, nous introduisons le concept de Trajectoires Émotionnelles et Physiques Étudiantes (SEPT). Grâce aux webcams de 89 étudiants de première année de Master, nous avons enregistré et analysé chaque seconde leurs expressions faciales (valence, arousal selon le modèle de Russell) et états physiques (orientation de la tête, distance à l'écran). Les séries temporelles ainsi obtenues révèlent des motifs distincts selon que les difficultés soient individuelles ou liées aux questions. SEPT offres des perspectives pour des systèmes intelligents de suivi affectif en contexte éducatif numérique.

ABSTRACT

SEPT : Uncovering Student Difficulties through Emotional and Physical Trajectories during Online Assessments

Imagine a classroom where students' difficulties and successes are expressed not by words, but by their facial expressions and movements, captured in real time during a quiz. Teaching methods in higher education are increasingly hybrid and remote. Direct interactions are reduced, making it difficult to detect moments of disengagement. To address this, we introduce the concept of Student Emotional and Physical Trajectories (SEPT). Using the webcams of 89 first year Master's students, we recorded and analyzed every second their facial expressions (valence, arousal according to the Russell model) and physical states (head orientation, distance from the screen). The resulting time series reveal distinct patterns depending on whether the difficulties are individual, or question related. SEPT offers perspectives for intelligent systems for affective monitoring in a digital educational context.

MOTS-CLÉS : Informatique affective, trajectoires émotionnelles, applications de l'apprentissage automatique, IA éthique, analytique de l'apprentissage..

KEYWORDS: Affective Computing, Emotion Trajectories, Machine-Learning Applications, Ethical AI, Learning analytics..

ARTICLE : **Accepté à** IA-ÉDU@CORIA-TALN 2025.

Actes de l'atelier Intelligence Artificielle générative et ÉDUcation :
Enjeux, Défis et Perspectives de Recherche 2025 (IA-ÉDU), pages 111–125.

1 Introduction

Les interactions traditionnelles en présentiel entre enseignants et étudiants deviennent moins fréquentes à l'ère de l'enseignement hybride et à distance (Shi *et al.*, 2024). Cette évolution offre certes une flexibilité accrue et un accès plus large à l'apprentissage, mais soulève une préoccupation majeure : comment prévenir l'abandon scolaire lorsque les enseignants ne peuvent pas directement observer les signes de difficultés chez les étudiants ? Sans indices visuels et interpersonnels immédiats, il devient difficile pour les enseignants de détecter lorsque les étudiants sont confus, frustrés ou désengagés (Frenkel *et al.*, 2024). Des recherches antérieures ont établi que les difficultés académiques sont souvent liées aux états émotionnels des étudiants (Frenkel *et al.*, 2024). Notre approche s'appuie sur des travaux précédents qui soulignent les limites des mesures émotionnelles statiques et agrégées pour prédire les résultats académiques (Nadaud *et al.*, 2024). Nous proposons une nouvelle approche nommée « Trajectoires Émotionnelles et Physiques des Étudiants (SEPT) », afin de surveiller continuellement les signaux émotionnels et comportementaux des étudiants durant les activités d'apprentissage. L'objectif est de réintroduire une forme de rétroaction « visuelle » pour les enseignants dans les environnements d'apprentissage à distance, permettant ainsi d'identifier rapidement les difficultés des étudiants et d'offrir un soutien en temps opportun. À partir de ces motivations, nous formulons les hypothèses clés suivantes :

H1. La valence et l'arousal seuls sont insuffisants : Bien que la valence et l'arousal offrent une mesure basique de l'affect, elles ne capturent pas complètement l'état émotionnel de l'étudiant. Nous supposons que l'intégration d'états physiques additionnels (comme les mouvements physiques, faciaux et la direction du regard) améliore la précision et la richesse de l'évaluation émotionnelle.

H2. Les émotions varient à l'échelle des questions individuelles : L'état émotionnel n'est pas statique tout au long d'une évaluation, mais fluctue dynamiquement au sein de chaque question du quiz. Capturer ces variations fines est essentiel pour comprendre en temps réel la frustration et l'état mental des étudiants.

H3. Les trajectoires émotionnelles dépendent du contexte, et ne sont pas constantes pour un étudiant donné : Nous faisons l'hypothèse qu'un même étudiant peut présenter des réponses émotionnelles très différentes selon la charge cognitive de chaque question. Cela remet en question l'idée d'un état émotionnel stable propre à l'étudiant, suggérant ainsi la nécessité de considérer le contexte.

H4. Existence de modèles communs pour les questions difficiles : Certains modèles SEPT pourraient réapparaître chez différents étudiants confrontés à une même question difficile. Autrement dit, des défis cognitifs similaires pourraient provoquer des trajectoires SEPT comparables chez plusieurs étudiants, révélant ainsi des points de difficulté communs.

Pour tester ces hypothèses, nous avons développé une approche capturant continuellement les expressions faciales et les mouvements physiques et faciaux des étudiants via leurs webcams durant les quiz. Notre système suit les états émotionnels via les métriques de valence et d'arousal, et intègre des indicateurs physiques afin d'offrir un portrait complet et temporellement détaillé du comportement des étudiants. En examinant les émotions et les comportements au niveau granulaire des questions individuelles, nous cherchons à identifier les moments critiques de difficulté, fournissant ainsi des informations susceptibles de permettre des interventions pédagogiques opportunes. Finalement, notre approche réintroduit un élément crucial de l'expérience en classe : la rétroaction émotionnelle en temps réel dans les contextes d'apprentissage hybrides et à distance en exploitant les progrès de

l'informatique affective. Ce travail rapproche l'informatique affective de la pédagogie, offrant aux enseignants des outils pour comprendre et soutenir les apprenants d'une manière jusqu'alors impossible dans les environnements distants.

La suite de l'article est organisée comme suit : la Section 2 présente les travaux connexes, la Section 3 décrit notre méthodologie, la Section 4 expose les résultats obtenus, et enfin la conclusion traite les limites de notre approche et propose des pistes pour des recherches futures.

2 Travaux connexes

Au cours des dernières années, les interactions entre les émotions des étudiants, leurs indicateurs physiques et leurs performances académiques ont suscité un intérêt croissant, notamment en raison du passage des salles de classe traditionnelles vers des environnements d'apprentissage en ligne et hybrides (Spitzer & Moeller, 2023).

Évolution des environnements d'apprentissage : L'expansion rapide des plateformes d'apprentissage numérique, particulièrement accélérée par la pandémie de COVID-19, a considérablement remodelé l'enseignement supérieur (Chaubey & Bhattacharya, 2015). Garrison *et al.* (Garrison *et al.*, 1999) soulignent qu'un apprentissage efficace nécessite la présence cognitive, pédagogique et sociale ; cette dernière étant souvent réduite dans les contextes à distance. Cette perte d'interaction directe peut créer une « distance émotionnelle » entre enseignants et étudiants, rendant plus difficile pour les enseignants la détection des difficultés rencontrées par les étudiants. Blikstein et Worsley (Blikstein & Worsley, 2016) précisent que ce manque de rétroaction est particulièrement problématique lors des évaluations, où les possibilités d'intervention pédagogique en temps réel sont limitées. Ces défis ont encouragé les chercheurs à explorer des méthodes technologiques permettant de combler l'écart entre la perception des enseignants et le soutien aux étudiants.

Émotions et résultats d'apprentissage : De nombreuses preuves démontrent que les émotions jouent un rôle central dans les résultats d'apprentissage. Par exemple, D'Mello et Graesser (Graesser & D'Mello, 2012) ont montré que les états affectifs transitoires tels que l'engagement, la confusion et la frustration peuvent influencer directement le traitement cognitif et la rétention des connaissances. Les états émotionnels positifs sont associés à une plus grande motivation et au plaisir d'apprendre (Harley *et al.*, 2015), alors que les états négatifs comme l'ennui ou l'anxiété peuvent entraver les progrès d'apprentissage (Arroyo *et al.*, 2014). Ces observations ont encouragé l'intégration de l'informatique affective dans les environnements éducatifs, dans le but de suivre les expériences émotionnelles des étudiants et d'y répondre de manière à améliorer l'apprentissage.

Détection des émotions et trajectoires en contexte éducatif : Les premières méthodes de détection des émotions des étudiants reposaient sur des auto-évaluations ou des évaluations par des observateurs (Baker *et al.*, 2010). Ces méthodes ont été critiquées en raison de leur subjectivité et du risque de perturber le processus d'apprentissage. Les progrès dans la détection sensorielle et la vision par ordinateur ont permis d'adopter des approches plus objectives. Des capteurs physiologiques (par exemple, des moniteurs de fréquence cardiaque ou des dispositifs de conductance cutanée) ont été utilisés pour mesurer l'arousal et les niveaux de stress (Jiang *et al.*, 2018), et des techniques basées sur des caméras peuvent suivre le regard et les expressions faciales pour inférer l'attention et les émotions (Arroyo *et al.*, 2014; Bosch *et al.*, 2016). Cependant, beaucoup de ces approches nécessitent un équipement spécialisé ou des conditions contrôlées, limitant ainsi leur extensibilité dans des

salles de classe réelles (Paquette *et al.*, 2014). Pour modéliser les émotions détectées, la plupart des études précédentes utilisaient soit les catégories émotionnelles discrètes d'Ekman (Perry, 2014), soit traitaient l'affect selon des dimensions continues moyennées sur une activité (par exemple, valence et arousal moyens) (Jiang *et al.*, 2018). (La valence réfère au caractère positif ou négatif de l'émotion, et l'arousal correspond au niveau d'activation ou d'alerte.) Des recherches récentes par Harley *et al.* (Harley *et al.*, 2017) et Silva *et al.* (Silva *et al.*, 2014) affirment que ces représentations statiques ne parviennent pas à capturer les fluctuations rapides et contextuelles des émotions pendant les tâches d'apprentissage. Ceci a conduit à un intérêt croissant pour l'analyse temporelle des données affectives, examinant comment les émotions évoluent au fil du temps plutôt qu'en considérant seulement des moyennes globales. Dans une étude antérieure (Nadaud *et al.*, 2024), une émotion par question, représentant le centroïde de toutes les valeurs émotionnelles enregistrées pendant la réponse à cette question, a été calculée. Ces centroïdes ont été connectés pour former des trajectoires tout au long du quiz. Les auteurs avaient émis l'hypothèse que des notes faibles correspondraient à des centroïdes émotionnels négatifs, mais cette méthode n'a pas montré de corrélation significative entre les états émotionnels moyennés et les performances.

Intégration d'indicateurs physiques : Les expressions faciales seules offrent une vue partielle de l'état des étudiants. Les états physiques tels que la posture et les mouvements reflètent également l'effort cognitif (Arroyo *et al.*, 2014), et leur intégration avec les données faciales peut améliorer la prédiction de la frustration et des performances (Whitehill *et al.*, 2014; Bosch & D'Mello, 2017). Les systèmes multimodaux surpassent systématiquement les approches unimodales (D'mello & Kory, 2015), bien que la fusion temporelle de ces signaux reste difficile en raison de leur asynchronisme.

Techniques avancées d'analyse temporelle : Des études récentes ont appliqué des méthodes analytiques basées sur les séquences aux données étudiantes, révélant des modèles comportementaux nuancés. Par exemple, Moreno-Marcos *et al.* (Moreno-Marcos *et al.*, 2020) ont utilisé le clustering pour identifier des trajectoires distinctes d'engagement dans des cours en ligne, constatant que certaines trajectoires temporelles d'interaction sont associées à de meilleurs taux d'achèvement. Rodrigo *et al.* (Rodrigo *et al.*, 2012) ont utilisé l'analyse de séquences sur les états affectifs des étudiants (par exemple, confusion → insight → confusion), et ont associé des séquences émotionnelles spécifiques à des gains d'apprentissage ou au désengagement. Piot *et al.* (Piot *et al.*, 2019) explorent « l'effet eureka » (confusion → surprise → joie). Zhou *et al.* (Zhao & Itti, 2016) ont introduit l'utilisation du Dynamic Time Warping (DTW) pour aligner et comparer les séquences temporelles des comportements des apprenants, en tenant compte des différences individuelles de rythme. Malgré ces progrès, des lacunes importantes subsistent. Paquette *et al.* (Paquette *et al.*, 2014) ont noté que peu d'efforts intégraient les indicateurs physiques (comme la posture corporelle ou le regard) avec les données émotionnelles en une trajectoire multidimensionnelle unique.

En s'appuyant sur une première exploration des trajectoires émotionnelles (Nadaud *et al.*, 2024), cet article présente les SEPT comme un avancement complet. Les SEPT étendent le modèle bidimensionnel valence-arousal à un cadre à sept dimensions, incorporant des indicateurs physiques (orientation de la tête, distance du visage à l'écran) en complément des caractéristiques émotionnelles. Les SEPT capturent des fluctuations continues, seconde par seconde, en exploitant le DTW et le clustering afin d'identifier des motifs parmi les trajectoires à un niveau de granularité fin.

3 Méthodologie

Pour tester nos hypothèses, nous avons mené une expérimentation contrôlée en plusieurs phases. Nous décrivons ci-dessous précisément le cadre expérimental et chaque étape de notre méthodologie.

3.1 Cadre expérimental

Nous avons réalisé notre étude auprès de 89 étudiants de première année de Master dans une école d'ingénieurs française. Tous les participants avaient une formation en technologies de l'information et étaient à l'aise avec l'utilisation d'ordinateurs portables et de webcams. L'expérience s'est déroulée durant un cours en présentiel où chaque étudiant utilisait son propre ordinateur portable muni d'une webcam intégrée afin de passer un quiz en ligne. Afin de maintenir un environnement naturel, aucune consigne spécifique concernant leur posture ou la position de la caméra n'a été donnée ; les étudiants interagissaient normalement avec le quiz. Le quiz, proposé au début d'un cours sur l'apprentissage automatique, comportait 6 à 7 questions de différents formats : tâches de codage, questions à choix multiples et exercices de type glisser-déposer. Chaque question ne permettait qu'une seule tentative, sans retour en arrière, assurant une progression linéaire. Les étudiants disposaient de 12 minutes au maximum pour terminer toutes les questions, après quoi le quiz se fermait automatiquement et était noté. Durant le quiz, les étudiants étaient tenus de rester sur la page sans possibilité de navigation externe. Nous avons utilisé l'extension Moodle Proctoring (adaptée de sa fonctionnalité initiale de vérification d'identité) afin de capturer des images de chaque étudiant à intervalles réguliers tout au long du quiz. Ce dispositif a permis une observation continue des expressions faciales et comportements des étudiants tout en préservant une expérience naturelle de passation de quiz.

3.2 Collecte des données

Notre dispositif expérimental réduit les besoins matériels à un simple ordinateur portable équipé d'une webcam, évitant ainsi intentionnellement l'utilisation de dispositifs coûteux tels que des bracelets capteurs ECG/EDM (Spitzer & Moeller, 2023; Nandi *et al.*, 2021). Les données collectées consistent en des images capturées toutes les secondes. L'interface utilisée est strictement identique à celle d'un quiz standard sans surveillance vidéo, préservant ainsi une expérience utilisateur fluide et minimisant toute influence sur les performances au quiz (Lee, 2020). Simultanément, nous avons collecté des données complètes provenant du système de gestion de l'apprentissage (LMS), incluant les notes obtenues, le temps passé par question, et les niveaux de difficulté des questions.

3.3 Conformité RGPD et considérations éthiques de l'étude

Cette étude a été menée en conformité stricte avec les directives éthiques et le Règlement Général sur la Protection des Données (RGPD) (Union, 2018) afin de protéger la vie privée des participants. Plusieurs mesures ont été mises en œuvre pour garantir l'éthique de la recherche :

— **Consentement éclairé :** La participation était entièrement volontaire. Les étudiants ont été pleinement informés des objectifs du projet, des données collectées (images périodiques de la webcam et données de performance) et de leur utilisation exclusive à des fins de recherche. Un consentement écrit explicite a été obtenu auprès de chaque participant avant la collecte

des données. Les étudiants étaient libres de refuser ou de se retirer de l'étude à tout moment, sans aucune conséquence académique.

— **Préparation et transparence :** Avant le quiz noté, un quiz d'entraînement a été organisé avec le même dispositif de surveillance pour que les participants puissent s'y familiariser. Les étudiants refusant de participer passaient le quiz dans les mêmes conditions à l'exception de la capture vidéo, préservant ainsi l'équité des conditions d'examen.
— **Approbation éthique :** Le protocole de recherche a été revu et approuvé par le comité d'éthique de l'école, garantissant la conformité avec les standards éthiques établis pour la recherche impliquant des participants humains.
— **Conformité juridique et sécurité des données :** Nous avons consulté un avocat spécialisé en protection des données pour concevoir nos formulaires de consentement et notre plan de gestion des données, en conformité totale avec le RGPD. Toutes les données collectées (images et données LMS) ont été anonymisées et stockées sur un serveur sécurisé de l'école, accessible uniquement par l'équipe de recherche autorisée.
— **Préservation de la confidentialité :** Les images webcam ont été utilisées exclusivement pour extraire automatiquement les expressions faciales et les mouvements de la tête, sans reconnaissance d'identité.

3.4 Prétraitement des données

Avant l'analyse émotionnelle, nous avons appliqué une série d'étapes de prétraitement afin d'améliorer la qualité des données. En conditions réelles de classe, certaines images capturées peuvent être impropres à la reconnaissance émotionnelle (visage partiellement visible, mauvaise luminosité, etc.). Un pipeline de filtrage a été implémenté pour ne conserver que des images exploitables. Après avoir évalué plusieurs algorithmes de détection faciale (Haar Cascade, dlib, MTCNN, YOLOv5), nous avons retenu le modèle YOLOv5 (Ieamsaard *et al.*, 2021) pour sa précision et sa robustesse. Les images sans visage détecté ont subi une amélioration (correction gamma, égalisation d'histogramme), suivie d'une seconde tentative de détection. Les images restantes non valides ont été exclues. Ce processus a conduit à la sélection finale de 82 étudiants (sur les 89 initiaux), assurant ainsi une meilleure qualité des données.

3.5 Méthodes d'analyse des données

Après nettoyage, nous avons analysé les SEPT via des techniques de vision par ordinateur et apprentissage automatique. L'analyse comportait deux étapes principales : (1) prédiction continue des états émotionnels (valence, arousal) à partir des images, et (2) intégration avec des indicateurs physiques pour former les SEPT.

Prédiction des émotions. Pour inférer avec précision les états affectifs à partir d'environ 99 000 images capturées et les cartographier selon le modèle d'affect de Russell, nous avons abandonné les méthodes d'annotation traditionnelles en raison de leur nature gourmande en ressources et de leur subjectivité, qui entraînent souvent une faible concordance entre annotateurs et des délais prolongés (Graesser & D'Mello, 2012). Le volume et la complexité de notre ensemble de données d'images ont rendu les approches conventionnelles d'apprentissage automatique inadéquates, car elles peinent à traiter des données de grande dimension sans traitement manuelle significative des caractéristiques. Nous avons adopté une approche d'apprentissage profond pour surmonter les limites

des modèles de prédiction des émotions 2D existants. En exploitant les ensembles de données AffectNet (Mollahosseini *et al.*, 2019) et Affwild (Liu & Kollias, 2019), largement reconnus et couramment utilisés dans la recherche en informatique affective, nous avons entraîné deux modèles à prédire la valence et l'arousal.

Modèle	**Valence**		**Arousal**	
	CCC	*RMSE*	*CCC*	*RMSE*
CAGE 2024 (Wagner *et al.*, 2024)	**0.716**	**0.331**	0.642	**0.305**
VGG-F 2021 (Bulat *et al.*, 2022)	0.710	0.356	0.629	0.326
Nos modèles	0.714	0.339	**0.644**	0.323

TABLE 1 – Comparaison des performances des modèles sur les dimensions de valence et d'arousal.

Le tableau 1 compare nos modèles aux données de pointe en matière de prédiction de l'arousal de valence. Chaque modèle comprend cinq couches convolutives et de sous-échantillonnage séquentiel (Conv2D, Conv2D, MaxPooling), suivies de deux couches denses, totalisant environ 2 millions de paramètres. Cette architecture a permis des prédictions émotionnelles précises, avec une perte de validation inférieure à 0,1, permettant une représentation détaillée des états émotionnels des étudiants, conforme au modèle de Russell (voir Fig. 1 - droite). La représentation circulaire 2D obtenue capture efficacement les expressions faciales actives, les expressions les plus prononcées étant positionnées plus à droite ou à gauche du cercle.

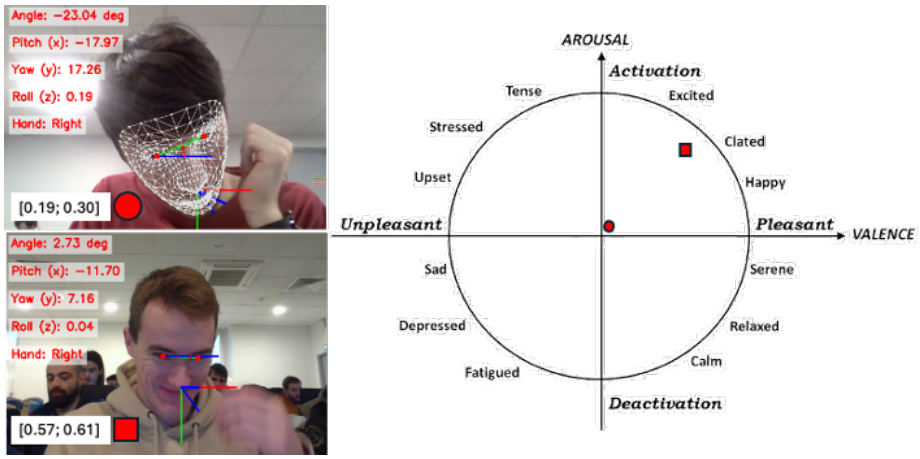

FIGURE 1 – Exemple de capture de données 7D pour la construction de la trajectoire SEPT

SEPT. Outre la valence et l'arousal, nous avons extrait l'orientation tridimensionnelle de la tête (tangage, lacet et roulis) et une estimation de la distance de l'étudiant à l'écran pour chaque image (basée sur la taille du cadre de délimitation du visage comme indicateur). Ainsi, chaque image est représentée par un vecteur de caractéristiques à 7 dimensions : [valence, arousal, tangage, lacet, roulis, angle d'inclinaison du regard, distance à l'écran] (voir Fig. 1 - gauche). Ensemble, ces caractéristiques capturent l'expression émotionnelle et physique de l'étudiant à chaque seconde. Comme chaque étudiant a consacré un temps différent à chaque question, la longueur (nombre d'images) de la trajectoire varie selon la paire étudiant-question. Pour comparer les trajectoires entre les étudiants, nous avons utilisé la méthode DTW pour les aligner temporellement. Nous avons ainsi calculé les distances par paire entre les trajectoires, ce qui nous indique le degré de similarité de deux SEPT de réponse émotionnelle/physique. Le passage d'un seul point de données moyenné par question à cette

série chronologique à 7 dimensions constitue une avancée méthodologique majeure pour l'analyse du décrochage scolaire. En résumé, le comportement de chaque étudiant lorsqu'il répond à chaque question est désormais représenté sous la forme d'un SEPT : une séquence alignée dans le temps capturant les changements émotionnels et physiques seconde par seconde.

Visualisation. Après normalisation des données de chaque dimension dans l'intervalle [-1,1], nous avons visualisé le SEPT afin d'en extraire des informations analytiques. Compte tenu de la nature multidimensionnelle inhérente de ces trajectoires, nous avons sélectionné la valence, l'arousal et la distance face à l'écran comme axes clés pour visualiser le SEPT. Chaque point de la visualisation correspond à une image enregistrée, liée aux expressions faciales et physique de l'étudiant. Nous avons appliqué l'Agglomerative Hierarchical Clustering avec complete linkage et une approche par matrice de distance précalculée en 7D basée sur des mesures de similarité DTW, ce qui nous a permis de regrouper les étudiants présentant une dynamique émotionnelle et une posture comportementale comparables. Dans la figure 2, le code couleur représente les groupes identifiés, capturant des schémas communs d'évolution des émotions et de la position de la tête tout au long de la session de test. Cette visualisation fournit une représentation temporelle des fluctuations émotionnelles des étudiants et de leur influence potentielle sur les performances cognitives. En analysant ces schémas, nous cherchons à identifier des indicateurs comportementaux corrélés aux résultats scolaires.

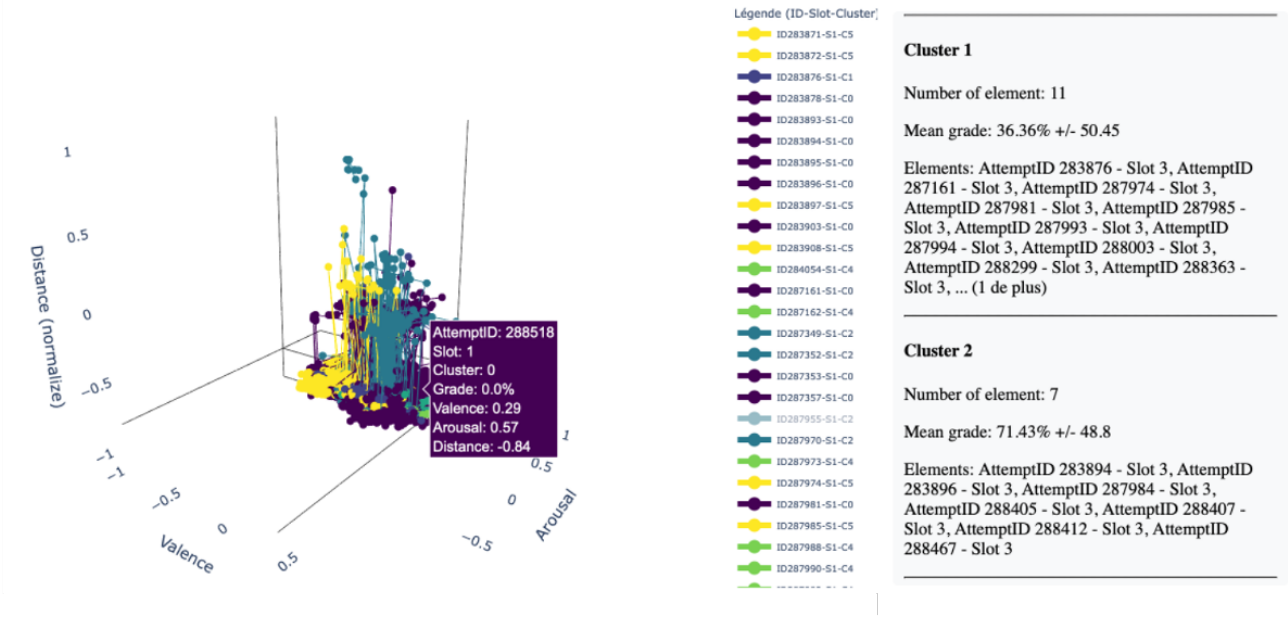

FIGURE 2 – Visualisation 3D du SEPT pour la question 7 du quiz, utilisant la valence, l'arousal et la distance à l'ecran. Les points représentent les données horodatées des étudiants, codées par couleur par l'Agglomerative Hierarchical Clustering dans l'espace DTW 7D. Les résumés des clusters à droite indiquent le nombre d'éléments, la note moyenne et les tentatives associées.

4 Résultats et discussion

Afin d'évaluer la pertinence et le pouvoir discriminant du SEPT, nous avons réalisé des analyses de clustering à deux niveaux, parmi les étudiants et entre eux pour chaque quiz. L'objectif était de déterminer si les dynamiques émotionnelles et physiques observées s'expliquaient mieux par des schémas individuels ou par la difficulté des questions, et de valider nos quatre hypothèses (H1–H4).

H1. La valence et l'arousal seuls sont insuffisants. Des observations antérieures révèlent une déconnexion entre la valence/l'arousal émotionnel et les résultats de performance, remettant en question la suffisance de ces dimensions en tant qu'indicateurs autonomes. Plusieurs étudiants ayant obtenu de faibles scores ont présenté des trajectoires de valence systématiquement positive ou neutre, et de même, les étudiants performants n'ont pas toujours affiché des niveaux élevés d'arousal ou de valence. Pour valider statistiquement ces résultats, nous avons réalisé une ACP sur des caractéristiques comportementales multimodales et visualisé leurs contributions aux deux principales composantes. En prenant le questionnaire 4231, question 2, comme exemple représentatif, la figure 3 montre que la valence et l'arousal présentent une forte corrélation et des vecteurs relativement courts dans le cercle de corrélation, ce qui signifie qu'ils expliquent moins la variance des données que d'autres caractéristiques comme la rotation de la tête ou la distance à l'écran. Ces dimensions ont peu contribué à la formation de groupes d'étudiants. Cela confirme l'hypothèse H1 : si la valence et l'arousal apportent des informations, elles ne constituent pas à elles seules des prédicteurs fiables de la difficulté perçue. Un ensemble plus large d'indicateurs, incluant la posture physique, doit être pris en compte.

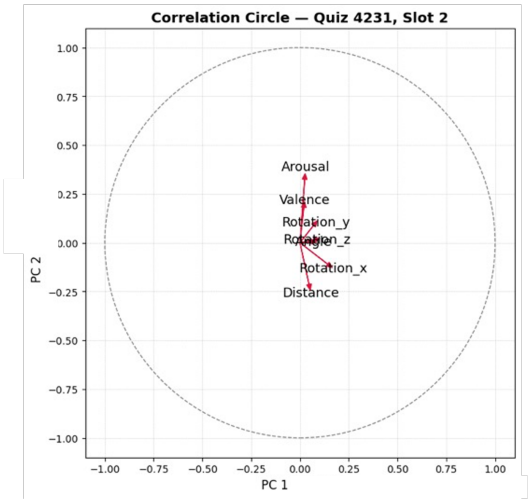

FIGURE 3 – Cercle de corrélation des variables comportementales (espace ACP), Quiz 4231 - question 2. La longueur et la direction de chaque flèche représentent la corrélation d'une caractéristique avec les deux composantes principales.

H2. Les émotions varient selon les questions. Pour tester H2, nous avons appliqué un regroupement par question sur SEPT, extrait pour chaque paire (quiz, question). L'objectif était de déterminer si les étudiants ayant des performances similaires, notamment ceux en difficulté, présentaient également des schémas affectifs similaires pour la même question. Sur les 13 questions analysées, 7 ont montré une nette distinction entre les étudiants performants et les étudiants peu performants, avec des profils émotionnels distincts.

Par exemple :

— Le quiz 4233, question 3 (p-value = 0,051) a montré une nette distinction : un groupe d'étudiants obtenant seulement 37,5 % (Cluster 0), un autre 83,3 % (Cluster 1) et un troisième 50 % (Cluster 2). (Fig. 4.) Ce contraste marqué, combiné au regroupement affectif/physique, suggère que les étudiants en difficulté ont réagi avec des schémas émotionnels similaires, reflétant probablement de la frustration ou une déconcentration. De même, la question 4 (p-value = 0,348), bien que moins prononcée, a révélé un groupe d'étudiants obtenant des résultats systématiquement inférieurs (61,8 %), confirmant la présence d'une convergence affective induite par la tâche. Ces résultats renforcent l'hypothèse H2 en confirmant que les émotions varient selon la question et que les étudiants peu performants présentent des trajectoires affectives comparables lorsqu'ils sont mis au défi.

— Quiz 4231, question 5 (p-value = 0,092) : Un groupe d'étudiants peu performants a montré une forte activation. Ces schémas reflètent une mobilisation émotionnelle intense, renforçant l'idée que la difficulté déclenche des réponses physiologiques et affectives communes.

— Quiz 4233, question 1 (p-value = 0,039) a révélé deux grands groupes d'étudiants obtenant des résultats inférieurs à 50 %, et un groupe atypique composé d'un seul étudiant très performant. Bien que ce regroupement mette en évidence une difficulté générale chez la plupart des étudiants, le test statistique ne confirme pas l'existence d'une structure affective significative entre les groupes de performance. Néanmoins, la prévalence de faibles scores peut néanmoins indiquer un stress émotionnel partagé en réponse au contenu de la question.

Dans ces exemples, les réponses émotionnelles ne sont pas idiosyncrasiques : les étudiants en difficulté ont tendance à adopter des comportements convergents face à la même question. Ces résultats valident l'hypothèse H2, montrant que les réponses affectives varient non seulement d'un étudiant à l'autre, mais aussi d'une question à l'autre, avec des schémas cohérents émergeant en réponse à la difficulté perçue ou à la charge cognitive.

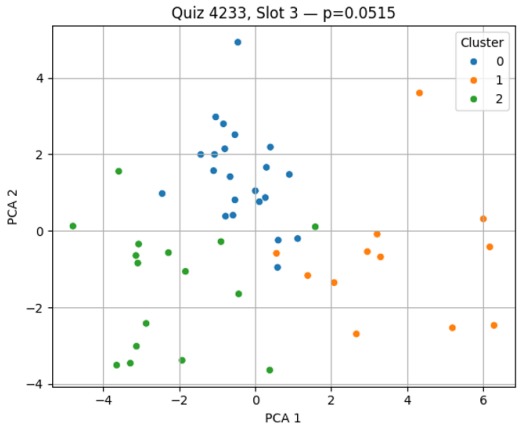

FIGURE 4 – Projection PCA des trajectoires des étudiants pour le questionnaire 4233, question 3. Trois groupes distincts émergent (p-value = 0,0515), révélant des états émotionnels et physiques cohérents parmi les étudiants peu performants, à l'appui de H2.

H3. Les trajectoires émotionnelles dépendent du contexte et ne sont pas des constantes propres à chaque étudiant. Les résultats du regroupement par étudiant ont également révélé que les étudiants ne présentaient pas un état affectif unique et récurrent d'une question à l'autre. Au contraire, les

trajectoires étaient façonnées par la complexité spécifique de chaque question. L'état émotionnel et physique d'un étudiant lors d'une question à choix multiples facile différait sensiblement de son état lors d'une tâche de codage complexe, même réalisée au cours de la même séance. Cela contredit la notion de profils affectifs fixes et conforte l'hypothèse H3 : les réponses affectives dépendent du contexte plutôt qu'elles ne sont intrinsèques à l'apprenant.

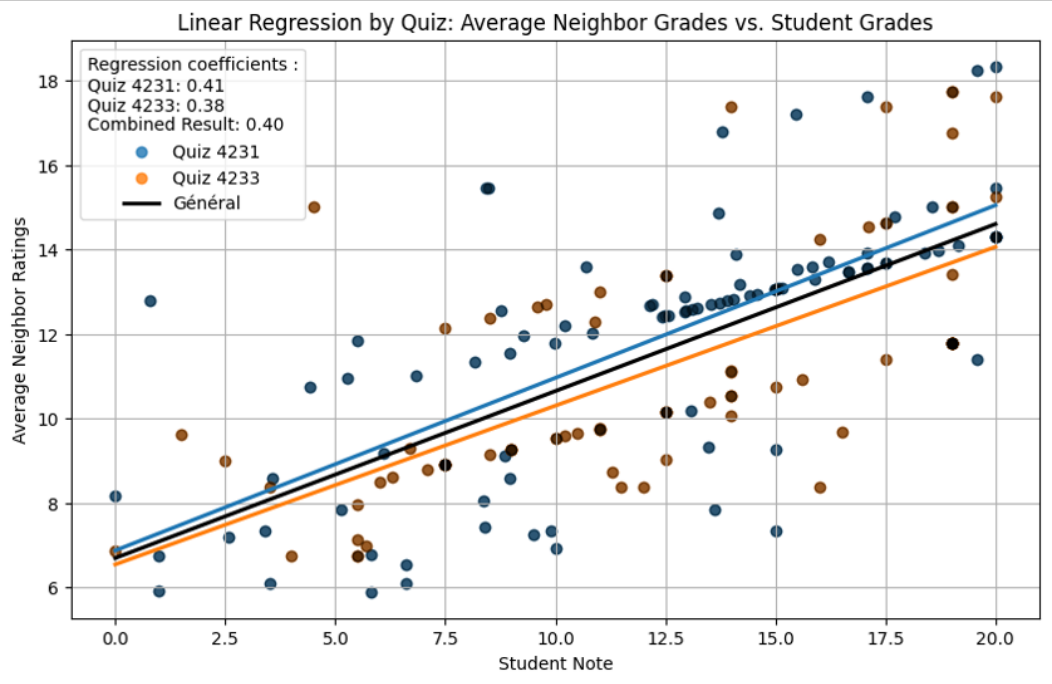

FIGURE 5 – Nuage de points illustrant la corrélation entre les scores individuels des étudiants et les scores moyens de leurs pairs du groupe pour les quiz 4231 et 4233.

H4. Schémas communs aux questions difficiles. Afin de tester la convergence interindividuelle face à des défis cognitifs partagés, nous avons appliqué un regroupement par question : le regroupement des trajectoires de tous les étudiants sur une question spécifique. Une structure beaucoup plus claire est apparue. Les groupes composés d'étudiants peu performants présentaient systématiquement des caractéristiques de trajectoire communes, telles qu'une valence décroissante, des mouvements fréquents de la tête et une proximité réduite avec l'écran, tandis que les groupes très performants présentaient des schémas affectifs et physiques plus stables. Quantitativement, ces groupes étaient significativement alignés avec la performance : pour le quiz 4231, la corrélation de Pearson entre la note d'un étudiant et le score moyen de ses pairs était de $r = 0{,}41$ (p-value $< 0{,}001$) ; pour le quiz 4233, $r = 0{,}38$ (p-value $< 0{,}001$). Sur l'ensemble des données du quiz, cette corrélation est restée stable autour de $r \approx 0{,}40$, indiquant que les étudiants d'un même groupe de trajectoire avaient tendance à obtenir des scores similaires. Ces regroupements cohérents et alignés sur les performances (Fig. 5) corroborent l'hypothèse H4 : confrontés à la même question difficile, les étudiants ont tendance à présenter des réactions émotionnelles et physiques similaires, qui se manifestent par des groupes homogènes dans l'espace SEPT. Ainsi, SEPT capture non seulement les réactions individuelles, mais aussi des schémas de difficulté communs, offrant ainsi un moyen évolutif de détecter les contenus problématiques en temps réel.

5 Conclusion

Cette étude a présenté le SEPT comme un nouveau cadre pour la capture et l'analyse de données comportementales et affectives fines lors des évaluations en ligne. Grâce à des trajectoires dynamiques et multivariées combinant la valence, l'arousal, l'orientation de la tête et la distance face à l'écran, nous avons démontré que le comportement et la difficulté des étudiants peuvent être déduits au niveau de chaque question. Nos analyses de clustering ont validé les hypothèses clés : (H1) la valence et l'arousal seuls ne suffisaient pas à saisir l'état physique des étudiants, ce qui enrichissait l'interprétation de la confusion ou de l'effort cognitif ; (H2) les réponses émotionnelles et physiques variaient significativement d'une question à l'autre, confirmant que l'émotion est un processus dynamique et contextuel plutôt qu'un phénomène stable à l'échelle de l'évaluation ; (H3) l'absence de signatures individuelles stables au cours d'un quiz a confirmé que ces trajectoires sont davantage façonnées par les exigences spécifiques à la question que par les traits intrinsèques des étudiants ; et (H4) l'émergence de schémas comportementaux communs en réponse aux questions difficiles a révélé que les étudiants en difficulté sur une même question présentaient souvent des profils SEPT similaires, correspondant à des performances faibles et homogènes au sein des groupes. Bien que ces résultats confirment le potentiel discriminant du SEPT, plusieurs limites doivent être reconnues. Premièrement, la fusion de données multimodales émotionnelles et posturales reste techniquement difficile, notamment pour synchroniser et pondérer des entrées hétérogènes. Les travaux futurs devraient explorer les architectures de fusion basées sur l'apprentissage profond, en utilisant des mécanismes d'attention temporelle ou des réseaux neuronaux récurrents, afin d'affiner l'interprétabilité et la robustesse (D'mello & Kory, 2015). Deuxièmement, nos résultats sont actuellement limités à un domaine académique spécifique (apprentissage automatique) et à une population étudiante homogène. Une validation plus large entre les matières, les institutions et les contextes culturels est nécessaire, ainsi que des études longitudinales pour déterminer si les trajectoires du SEPT prédisent les résultats d'apprentissage à long terme (Rodrigo *et al.*, 2012). Pour remédier à ces limites et améliorer le cadre du SEPT, nous identifions quatre axes de recherche futurs. Premièrement, nous cherchons à dériver un indicateur d'engagement en temps réel en transformant le mouvement de la tête, la distance visage-écran et la dynamique faciale en un score d'engagement composite. Des travaux récents utilisant des réseaux convolutifs de graphes spatio-temporels ont montré des résultats prometteurs dans la capture de l'engagement à partir de repères faciaux (Mangaroska *et al.*, 2021). Deuxièmement, nous proposons d'étendre le SEPT pour mesurer la concentration cognitive, en intégrant les changements d'attention via la posture de la tête et le suivi du regard comme indicateurs de concentration ou de distraction pendant la résolution de problèmes (Abedi & Khan, 2024). Troisièmement, pour enrichir la diversité des signaux, le SEPT pourrait intégrer des sources de données multimodales telles que les signaux physiologiques (par exemple, la variabilité du rythme cardiaque) et le comportement numérique (par exemple, la dynamique de la souris ou du clavier), qui se sont avérés améliorer la précision de la reconnaissance des affects (Gaudi *et al.*, 2022). Enfin, des mécanismes de rétroaction en temps réel devraient être développés pour fournir des informations exploitables aux instructeurs ou aux systèmes de tutorat intelligents, en adaptant dynamiquement les stratégies pédagogiques en fonction des trajectoires des étudiants.

Références

ABEDI A. & KHAN S. S. (2024). Engagement Measurement Based on Facial Landmarks and Spatial-Temporal Graph Convolutional Networks. arXiv :2403.17175 [cs], DOI : 10.48550/arXiv.2403.17175.

ARROYO I., WOOLF B. P., BURELSON W., MULDNER K., RAI D. & TAI M. (2014). A Multimedia Adaptive Tutoring System for Mathematics that Addresses Cognition, Metacognition and Affect. *International Journal of Artificial Intelligence in Education*, **24**(4), 387–426. DOI : 10.1007/s40593-014-0023-y.

BAKER R. S. J. D., D'MELLO S. K., RODRIGO M. M. T. & GRAESSER A. C. (2010). Better to be frustrated than bored : The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, **68**(4), 223–241. DOI : 10.1016/j.ijhcs.2009.12.003.

BLIKSTEIN P. & WORSLEY M. (2016). Multimodal Learning Analytics and Education Data Mining : using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, **3**(2), 220–238. Number : 2, DOI : 10.18608/jla.2016.32.11.

BOSCH N. & D'MELLO S. (2017). The affective experience of novice computer programmers. *International Journal of Artificial Intelligence in Education*, **27**(1), 181–206. Place : Germany Publisher : Springer, DOI : 10.1007/s40593-015-0069-5.

BOSCH N., D'MELLO S. K., BAKER R. S., OCUMPAUGH J., SHUTE V., VENTURA M., WANG L. & ZHAO W. (2016). Detecting student emotions in computer-enabled classrooms. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, p. 4125–4129, New York, New York, USA : AAAI Press.

BULAT A., CHENG S., YANG J., GARBETT A., SANCHEZ E. & TZIMIROPOULOS G. (2022). Pre-training strategies and datasets for facial representation learning. arXiv :2103.16554 [cs] version : 2, DOI : 10.48550/arXiv.2103.16554.

CHAUBEY A. & BHATTACHARYA B. (2015). Learning Management System in Higher Education. *IJSTE - International Journal of Science Technology & Engineering* |, **2**, 158–162.

D'MELLO S. K. & KORY J. (2015). A Review and Meta-Analysis of Multimodal Affect Detection Systems. *ACM Comput. Surv.*, **47**(3), 43 :1–43 :36. DOI : 10.1145/2682899.

FRENKEL J., CAJAR A., ENGBERT R. & LAZARIDES R. (2024). Exploring the impact of nonverbal social behavior on learning outcomes in instructional video design. *Scientific Reports*, **14**. DOI : 10.1038/s41598-024-63487-w.

GARRISON D. R., ANDERSON T. & ARCHER W. (1999). Critical Inquiry in a Text-Based Environment : Computer Conferencing in Higher Education. *The Internet and Higher Education*, **2**(2), 87–105. DOI : 10.1016/S1096-7516(00)00016-6.

GAUDI G., KAPRALOS B., COLLINS K. & URIBE QUEVEDO A. (2022). Affective computing : An introduction to the detection, measurement, and current applications. DOI : 10.1007/978-3-030-80571-5_3.

GRAESSER A. C. & D'MELLO S. (2012). Chapter Five - Emotions During the Learning of Difficult Material. In B. H. ROSS, Éd., *Psychology of Learning and Motivation*, volume 57 de *The Psychology of Learning and Motivation*, p. 183–225. Academic Press. DOI : 10.1016/B978-0-12-394293-7.00005-4.

HARLEY J. M., BOUCHET F., HUSSAIN M. S., AZEVEDO R. & CALVO R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior*, **48**, 615–625. DOI : 10.1016/j.chb.2015.02.013.

HARLEY J. M., LAJOIE S. P., FRASSON C. & HALL N. C. (2017). Developing Emotion-Aware, Advanced Learning Technologies : A Taxonomy of Approaches and Features. *International Journal of Artificial Intelligence in Education*, **27**(2), 268–297. DOI : 10.1007/s40593-016-0126-8.

IEAMSAARD J., CHAROENSOOK S. N. & YAMMEN S. (2021). Deep Learning-based Face Mask Detection Using YoloV5. In *2021 9th International Electrical Engineering Congress (iEECON)*, p. 428–431. DOI : 10.1109/iEECON51072.2021.9440346.

JIANG Y., BOSCH N., BAKER R. S., PAQUETTE L., OCUMPAUGH J., ANDRES J. M. A. L., MOORE A. L. & BISWAS G. (2018). Expert Feature-Engineering vs. Deep Neural Networks : Which Is Better for Sensor-Free Affect Detection ? In C. PENSTEIN ROSÉ, R. MARTÍNEZ-MALDONADO, H. U. HOPPE, R. LUCKIN, M. MAVRIKIS, K. PORAYSKA-POMSTA, B. MCLAREN & B. DU BOULAY, Éds., *Artificial Intelligence in Education*, p. 198–211, Cham : Springer International Publishing. DOI : 10.1007/978-3-319-93843-1_15.

LEE J. W. (2020). Impact of Proctoring Environments on Student Performance : Online vs Offline Proctored Exams.

LIU M. & KOLLIAS D. (2019). Aff-Wild Database and AffWildNet. arXiv :1910.05318 [cs], DOI : 10.48550/arXiv.1910.05318.

MANGAROSKA K., SHARMA K., GASEVIC D. & GIANNAKOS M. (2021). Exploring students' cognitive and affective states during problem solving through multimodal data : Lessons learned from a programming activity. *Journal of Computer Assisted Learning*, **38**. DOI : 10.1111/jcal.12590.

MOLLAHOSSEINI A., HASANI B. & MAHOOR M. H. (2019). AffectNet : A Database for Facial Expression, Valence, and Arousal Computing in the Wild. *IEEE Transactions on Affective Computing*, **10**(1), 18–31. arXiv :1708.03985 [cs], DOI : 10.1109/TAFFC.2017.2740923.

MORENO-MARCOS P. M., MUÑOZ-MERINO P. J., MALDONADO-MAHAUAD J., PÉREZ-SANAGUSTÍN M., ALARIO-HOYOS C. & DELGADO KLOOS C. (2020). Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs. *Computers & Education*, **145**, 103728. DOI : 10.1016/j.compedu.2019.103728.

NADAUD E., YAACOUB A., HAIDAR S., GRAND B. & PREVOST L. (2024). Emotion Trajectory and Student Performance in Engineering Education : A Preliminary Study. DOI : 10.1007/978-3-031-59465-6_25.

NANDI A., XHAFA F., SUBIRATS L. & FORT S. (2021). Real-Time Emotion Classification Using EEG Data Stream in E-Learning Contexts. *Sensors*, **21**(5), 1589. Number : 5 Publisher : Multidisciplinary Digital Publishing Institute, DOI : 10.3390/s21051589.

PAQUETTE L., DE CARVALHO A. M. & BAKER R. S. (2014). Towards Understanding Expert Coding of Student Disengagement in Online Learning : 36th Annual Meeting of the Cognitive Science Society, CogSci 2014. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society, CogSci 2014*, p. 1126–1131. Publisher : The Cognitive Science Society.

PERRY, RAYMOND P. R. P. (2014). Control-Value Theory of Achievement Emotions. In *International Handbook of Emotions in Education*. Routledge. Num Pages : 22.

PIOT M., ALABARBE T., GONZALEZ J., LE BAIL C., PREVOST L., BOURDEAU J., BERNARD F. X., BAKER M. & DETIENNE F. (2019). Joint analysis of verbal and nonverbal interactions in collaborative E-learning. In *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, p. 1–5. DOI : 10.1109/ACIIW.2019.8925033.

RODRIGO M., BAKER R., AGAPITO J., NABOS J., REPALAM M., REYES JR S. & SAN PEDRO C. (2012). The Effects of an Interactive Software Agent on Student Affective Dynamics while Using ;an Intelligent Tutoring System. *IEEE Transactions on Affective Computing*, **3**, 224–236. DOI : 10.1109/T-AFFC.2011.41.

SHI G., CHEN S., LI H., TIAN S. & WANG Q. (2024). A study on the impact of COVID-19 class suspension on college students' emotions based on affective computing model. *Applied Mathematics and Nonlinear Sciences*, **9**(1).

SILVA L. C., DE M.B. OLIVEIRA F. C., DE OLIVEIRA A. C. & DE FREITAS A. T. (2014). Introducing the JLoad : A Java Learning Object to Assist the Deaf. In *2014 IEEE 14th International Conference on Advanced Learning Technologies*, p. 579–583. ISSN : 2161-377X, DOI : 10.1109/ICALT.2014.169.

SPITZER M. W. H. & MOELLER K. (2023). Performance increases in mathematics during COVID-19 pandemic distance learning in Austria : Evidence from an intelligent tutoring system for mathematics. *Trends in Neuroscience and Education*, **31**, 100203. DOI : 10.1016/j.tine.2023.100203.

UNION E. (2018). General Data Protection Regulation (GDPR) – Official Legal Text.

WAGNER N., MÄTZLER F., VOSSBERG S. R., SCHNEIDER H., PAVLITSKA S. & ZÖLLNER J. M. (2024). CAGE : Circumplex Affect Guided Expression Inference. arXiv :2404.14975 [cs] version : 1, DOI : 10.48550/arXiv.2404.14975.

WHITEHILL J., SERPELL Z., LIN Y.-C., FOSTER A. & MOVELLAN J. (2014). The Faces of Engagement : Automatic Recognition of Student Engagementfrom Facial Expressions. *Affective Computing, IEEE Transactions on*, **5**, 86–98. DOI : 10.1109/TAFFC.2014.2316163.

ZHAO J. & ITTI L. (2016). shapeDTW : shape Dynamic Time Warping. arXiv :1606.01601 [cs], DOI : 10.48550/arXiv.1606.01601.