

Rapport d'évaluation de la perception des élèves-ingénieurs de la qualité des enseignements à SUP'COM

Synthèse consolidée sur 3 années universitaires (2022–2025)

Octobre 2026



Table des matières

Table des matières	2
1. Introduction	3
2. Objectifs spécifiques de l'étude	3
3. Méthodologie	4
3.1 Données mobilisées	4
3.2 Indicateur global de pédagogie perçue	4
3.3 Priorisation des leviers d'amélioration	5
3.4 Rôle des algorithmes IA utilisés	5
4. Résultats et plan d'action prescriptif	6
4.1 Typologie des modules	6
4.2 Benchmark interne	6
4.3 Plan d'action prioritaire	7
4.4 Fiabilité du système de priorisation	9
5. Hétérogénéité des perceptions étudiantes et analyse contextuelle	9
5.1 Niveau d'accord entre étudiants	9
5.2 Croisement du consensus avec le contexte pédagogique	10
6. Ce que disent les commentaires libres	11
6.1 Analyse longitudinale sur les huit périodes	11
6.2 Analyse thématique	12
6.3 Limites et biais d'interprétation	12
7. Précautions d'interprétation	12
8. Synthèse des principales contributions de l'étude	13
9. Vers un plan d'action	14
9.1 Axe 1 : Interventions d'urgence et accompagnement pédagogique prioritaire	14
9.2 Axe 2 : Harmonisation des pratiques et convergence des perceptions	14
9.3 Axe 3 : Gouvernance, restitution et suivi de l'évolution	14
9.4 Axe 4 : Recommandations méthodologiques et évolution de l'ingénierie d'évaluation	15
10. Conclusion	15

1. Introduction

Le rapport d'évaluation des enseignements 2024-2025, disponible sur le site web de SUP'COM dans la rubrique Qualité, avait dressé un bilan descriptif détaillé des appréciations des élèves-ingénieurs. Le présent document prolonge cette démarche d'écoute et de suivi, mais en élargit la finalité : il ne s'agit plus seulement de décrire les résultats d'une campagne d'évaluation, mais de consolider les données de trois années universitaires (2022-2023, 2023-2024 et 2024-2025) afin d'identifier des profils de modules, de hiérarchiser les besoins d'amélioration et de fournir des éléments d'aide au pilotage de la qualité pédagogique.

L'analyse porte sur trois dimensions complémentaires de l'expérience d'apprentissage : la pertinence du contenu pour la formation, la qualité perçue des méthodes pédagogiques et la pertinence des activités pratiques pour approfondir les connaissances théoriques. À ces trois dimensions s'ajoute une appréciation globale de l'enseignement, exprimée sous la forme d'une note en étoiles allant de 1 à 5.

Dans cette étude, plusieurs méthodes complémentaires sont utilisées : regroupement des modules par profil, construction d'un benchmark interne, calcul d'un potentiel théorique d'amélioration, modèle prédictif de contrôle, mesure du consensus entre étudiants et analyse des commentaires libres.

Le document reste volontairement non technique dans sa présentation des méthodes d'intelligence artificielle. Il présente les données utilisées, le rôle de chaque méthode, la nature des résultats obtenus, leur interprétation ainsi que les pistes d'amélioration qui peuvent en être déduites. Les modèles d'intelligence artificielle sont considérés comme des outils d'aide à la décision et non comme des substituts à l'analyse pédagogique et au jugement des responsables concernés.

Par ailleurs, conformément à la réglementation tunisienne relative à la protection des données à caractère personnel et aux principes de confidentialité applicables aux données issues des évaluations pédagogiques, un protocole d'anonymisation a été appliqué à l'ensemble du jeu de données exploité.

Bien que les enquêtes menées auprès des élèves-ingénieurs soient intrinsèquement anonymes, le croisement de scores d'évaluation avec des intitulés précis d'éléments d'enseignement peut conduire, dans certains contextes, à une identification indirecte des équipes pédagogiques concernées. En conséquence, les résultats destinés à la diffusion dans le présent document ne permettent pas d'associer directement un score individuel à un enseignant. Les informations contextuelles nécessaires à l'analyse — notamment les périodes académiques et les domaines d'enseignement — sont conservées à un niveau agrégé permettant de préserver la valeur analytique de l'étude tout en limitant les risques de ré-identification.

2. Objectifs spécifiques de l'étude

Cette étude vise à répondre aux objectifs suivants :

- Consolider trois années d'évaluations des enseignements dans une base de données unique et exploitable.
- Harmoniser les informations issues des différentes campagnes autour de trois dimensions analytiques communes : pertinence, pédagogie et pratique.
- Construire un critère composite de pilotage permettant de disposer d'une vision synthétique de l'appréciation pédagogique tout en conservant les dimensions élémentaires pour le diagnostic.
- Établir une typologie des modules selon leur profil multidimensionnel de satisfaction, à partir de la pertinence, de la pédagogie, de la pratique et du score global.
- Construire un benchmark interne à partir des modules présentant les profils les plus favorables, afin d'estimer les écarts à combler pour les autres modules.
- Prioriser, pour chaque module nécessitant une amélioration, la dimension présentant le potentiel théorique d'amélioration le plus important.
- Quantifier le potentiel théorique d'amélioration dans un scénario où la dimension prioritaire atteindrait le niveau du benchmark interne.
- Vérifier la cohérence des résultats de priorisation à l'aide de modèles prédictifs complémentaires et mesurer le niveau d'accord ou de dispersion des réponses étudiantes.
- Croiser les résultats quantitatifs avec l'analyse des commentaires libres afin de compléter l'interprétation des résultats chiffrés.

3. Méthodologie

3.1 Données mobilisées

L'étude repose sur un jeu de données d'évaluations pédagogiques réelles agrégées sur trois années universitaires : 2022-2023, 2023-2024 et 2024-2025. La base initiale comprend 8 255 observations individuelles, correspondant aux retours de 956 participants sur 12 périodes d'enseignement (les 2 premières années du cycle ingénieur à SUP'COM). Après application des règles de nettoyage et de contrôle de cohérence définies dans le pipeline d'analyse, 8 209 observations ont été retenues pour les analyses principales, soit 46 observations exclues.

Le périmètre analytique final couvre 62 modules d'enseignement, chacun disposant d'un effectif supérieur à 20 observations, ce qui permet de limiter les situations dans lesquelles une moyenne serait fortement influencée par un très faible nombre de réponses. Les observations couvrent tous les domaines d'enseignement, les différentes modalités pédagogiques et modalités d'évaluation.

Afin de disposer d'une structure homogène sur les trois années, trois dimensions ont été retenues autour de 3 questions posées lors des enquêtes :

- Q1 – Pertinence : le contenu de l'élément d'enseignement était-il pertinent pour la formation ?
- Q2 – Pédagogie : l'enseignant a-t-il utilisé des méthodes pédagogiques adaptées afin d'expliquer clairement les concepts ?
- Q3 – Pratique : les travaux pratiques, projets ou travaux dirigés étaient-ils pertinents pour approfondir les connaissances théoriques ?

Les réponses à trois valeurs Non / Pas sûr / Oui sont codées respectivement 1 / 3 / 5, afin de disposer d'une échelle numérique commune pour les analyses quantitatives.

La satisfaction globale, exprimée sous la forme d'une note de 1 à 5 étoiles, est conservée comme variable complémentaire. Elle n'est pas confondue avec les trois dimensions précédentes.

Sur l'ensemble des observations disponibles, les réponses favorables représentent environ 60,3 % pour Q1, 54,7 % pour Q2 et 50,9 % pour Q3. La dimension pratique présente ainsi le niveau de satisfaction favorable le plus faible des trois dimensions.

La satisfaction globale mesurée par le score en étoiles présente une moyenne de 3,13/5, avec un écart-type de 1,35, traduisant une dispersion importante des perceptions étudiantes.

Ces résultats descriptifs constituent le point de départ du travail de consolidation et de modélisation présenté dans les sections suivantes.

3.2 Indicateur global de pédagogie perçue

Les trois dimensions Q1, Q2 et Q3 apportent chacune une information spécifique. Pour le pilotage de la qualité, il est néanmoins utile de disposer parallèlement d'un indicateur synthétique permettant de *i)* comparer les profils de modules sur une échelle commune, *ii)* mesurer un écart global par rapport à un niveau de référence, *iii)* quantifier un potentiel théorique d'amélioration et *iv)* hiérarchiser les actions lorsque plusieurs dimensions présentent simultanément des écarts.

Le Score Composite d'Appréciation Pédagogique (SCAP) est donc introduit comme un critère composite de pilotage venant compléter, et non remplacer, les trois indicateurs Q1, Q2 et Q3. Il est calculé à partir des moyennes des trois dimensions :

$$SCAP = 0,35 Q1 + 0,40 Q2 + 0,25 Q3$$

Le score obtenu reste exprimé sur une échelle de 1 à 5.

Les pondérations ont été définies *a priori* comme des pondérations de pilotage interne :

- Q2 – Pédagogie : 40 %, afin d'accorder une priorité légèrement supérieure à la qualité perçue de la transmission et des méthodes pédagogiques ;
- Q1 – Pertinence : 35 %, afin de donner un poids important à l'adéquation des contenus avec la formation ;
- Q3 – Pratique : 25 %, afin de prendre en compte la contribution des activités pratiques et de mise en application.

Ces coefficients ne constituent pas une affirmation selon laquelle la pédagogie serait universellement plus importante que la pertinence ou la pratique. Ils traduisent un choix de pilotage propre à cette étude, donnant une légère priorité à Q2 tout en maintenant une contribution substantielle des deux autres dimensions.

La construction du SCAP permet de distinguer clairement deux niveaux d'analyse :

- Diagnostic : Q1, Q2 et Q3 permettent d'identifier précisément la dimension sur laquelle un module présente une fragilité.
- Pilotage : le SCAP fournit une mesure synthétique permettant de comparer les situations, d'estimer les écarts et de suivre l'évolution globale.

3.3 Priorisation des leviers d'amélioration

Afin de transformer cette mesure globale en information directement exploitable pour l'amélioration, l'analyse est ensuite déclinée au niveau des trois dimensions qui composent le SCAP. Cette étape vise à identifier, pour chaque module, la dimension présentant le plus grand potentiel théorique de rapprochement du niveau de qualité visé.

Pour chaque module n'appartenant pas au profil « Excellent », les écarts entre les moyennes observées et les benchmarks internes sont calculés pour Q1, Q2 et Q3. Le levier prioritaire correspond à la dimension présentant le plus grand écart à combler. Cette approche répond en fait à une question opérationnelle : sur quelle dimension faudrait-il prioritairement agir pour rapprocher ce module du niveau de référence interne visé ?

Le potentiel théorique d'amélioration du SCAP associé au levier prioritaire est calculé selon sa pondération dans le SCAP :

$$Gain_{SCAP} = w_j \times (Cible_j - Actuel_j),$$

où w_j représente le poids de la dimension prioritaire. Le SCAP projeté dans ce scénario est :

$$SCAP_{projeté} = SCAP_{actuel} + Gain_{SCAP}.$$

Ce résultat doit être interprété comme un scénario de benchmark : il estime ce que deviendrait le SCAP si la dimension prioritaire atteignait le niveau de référence interne, toutes choses égales par ailleurs. Il ne représente donc pas un gain effectivement garanti par une action pédagogique.

3.4 Rôle des algorithmes IA utilisés

Des algorithmes d'intelligence artificielle ont été utilisés pour répondre à plusieurs finalités :

- Regroupement des modules par profil. Un premier algorithme IA prend en entrée les notes moyennes de chaque module selon les 4 dimensions (Q1, Q2, Q3 et score global) et regroupe automatiquement les modules qui se ressemblent. Il produit en sortie 3 groupes, ou « profils » : modules fragiles, modules modérés et modules excellents. Aucun seuil n'est fixé à l'avance pour décider, c'est l'algorithme IA qui détecte lui-même les regroupements les plus cohérents dans les données.
- Priorisation des leviers d'action. Un second algorithme prend en entrée le profil de chaque module fragile ou modéré et le compare à celui des modules jugés excellents. En sortie, il indique quel aspect parmi les 3 (le contenu, la pédagogie ou la mise en pratique) présente l'écart le plus important à combler, ainsi qu'une estimation du gain de satisfaction attendu si cet écart était comblé.
- Contrôle de cohérence par un modèle prédictif. Un troisième algorithme a été utilisé comme vérification indépendante : à partir du profil de réponses d'un module et de son contexte (domaine d'enseignement, modalité d'enseignement, mode d'évaluation), il ré-estime la note moyenne que ce module devrait obtenir, puis la compare à sa note réelle. Plus l'estimation se rapproche de la réalité, plus cela confirme que les profils et priorités identifiés plus haut reposent sur des critères cohérents.
- Mesure du niveau d'accord entre étudiants. Au-delà de la moyenne, il est utile de savoir si les étudiants répondent globalement de la même façon ou si les avis sont partagés. Un indicateur de « consensus » a été calculé pour chaque question : plus il est élevé, plus les étudiants sont d'accord entre eux ; plus il est bas, plus les avis sont dispersés, ce qui peut masquer des situations contrastées derrière une moyenne d'apparence.
- Analyse des commentaires libres. Les commentaires laissés librement par les étudiants ont été analysés automatiquement pour en extraire un ton général (plutôt positif, négatif ou très négatif) afin de compléter la lecture chiffrée par un éclairage qualitatif.

4. Résultats et plan d'action prescriptif

4.1 Typologie des modules

Les modules sont regroupés à partir de leurs profils moyens selon quatre dimensions : Q1, Q2, Q3 et score global en étoiles. Une méthode de regroupement automatique de type k-means est utilisée afin d'identifier trois profils de modules : fragile, modéré et excellent. La simulation fournit 16 modules fragiles (25,8 %), 24 modules modérés (38,7 %) et 22 modules excellents (35,5 %).

La figure 1 positionne chaque module selon ses moyennes en Q1, Q2 et Q3 : chaque point est un module, sa couleur indique le profil auquel il a été rattaché (rouge : fragile, orange : modéré, vert : excellent). La 4ème dimension (nombre d'étoiles) n'a pas été omise de l'étude mais la visualisation directe en 4D n'est pas réalisable. La séparation entre profils est cohérente : le profil 1 (fragile) est systématiquement le plus faible sur les 4 dimensions, le profil 3 (excellent) est le plus élevé, ce qui confirme la cohérence de ce regroupement.

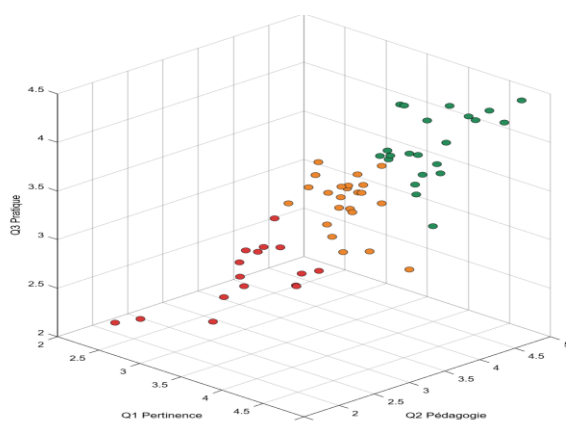


Figure 1 : positionnement des modules selon leur profil de satisfaction (Q1, Q2, Q3).

La figure 2 compare les profils moyens des 3 typologies sur les 4 dimensions évaluées. Elle montre que la pertinence est la dimension la mieux notée par l'ensemble des trois groupes, ce qui montre un consensus sur l'intérêt. En revanche, l'écart entre les profils s'élargit sur la pédagogie, la pratique et le score en étoiles, révélant que ce sont les aspects méthodologiques et pratiques qui différencient le plus le niveau de satisfaction des répondants.

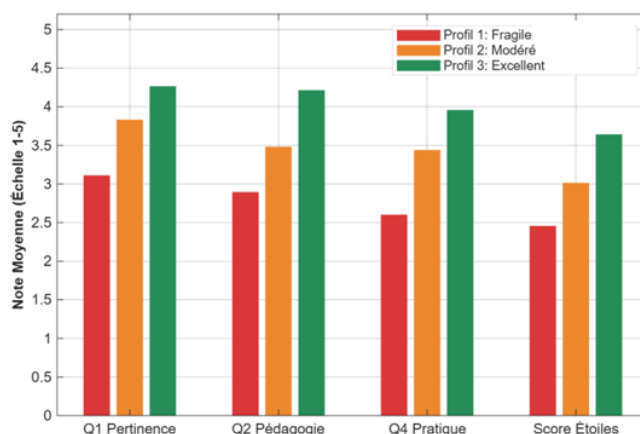


Figure 2 : signature moyenne des 3 typologies de modules sur les 4 dimensions évaluées.

4.2 Benchmark interne

Le groupe « Excellent » sert ensuite de référence interne. Pour chacune des trois dimensions, sa moyenne constitue une valeur cible à atteindre par les autres modules fragiles et modérés. Ces valeurs ne constituent ni des normes externes, ni des seuils réglementaires, ni des niveaux universellement optimaux. Elles représentent un point de comparaison interne permettant d'estimer les écarts observés entre les modules. Les valeurs de référence obtenues sont listées dans le tableau 1.

Tableau 1 : Valeurs moyennes des profils d'excellence utilisées comme références internes.

Dimension	Benchmark interne
Q1 – Pertinence	4,265
Q2 – Pédagogie	4,187
Q3 – Pratique	3,959
SCAP	4,157

4.3 Plan d'action prioritaire

L'analyse précédente permet désormais de passer du diagnostic à la hiérarchisation des priorités d'amélioration. Pour chaque module n'appartenant pas au profil « Excellent », le levier présentant le plus grand écart au benchmark interne est retenu comme priorité d'action. Le potentiel associé est exprimé par le Gain_SCAP (section 3.3).

Sur les 62 modules analysés, 40 nécessitent une action prioritaire. La répartition des leviers est la suivante :

- Pédagogie : 23 modules (57,5 %) ;
- Pratique : 10 modules (25,0 %) ;
- Pertinence : 7 modules (17,5 %).

La pédagogie constitue ainsi le principal axe d'amélioration identifié dans l'ensemble des modules concernés.

Le Gain_SCAP moyen théorique est de +0,387, avec une médiane de +0,335. Dans le scénario de benchmark retenu, le SCAP moyen des 40 modules concernés passerait de 3,307 à 3,694.

Le tableau 2 présente l'ensemble des 40 modules nécessitant une action, classés par Gain_SCAP décroissant. Le levier prioritaire correspond à la dimension présentant le plus grand écart au benchmark interne.

Tableau 2 – Principales priorités d'action et leurs performances escomptées.

Module	Profil	Levier prioritaire	SCAP actuel	Gain_SCAP	SCAP projeté
	Fragile	Pédagogie	2,263	+0,907	3,170
	Fragile	Pédagogie	2,124	+0,889	3,014
	Fragile	Pédagogie	2,864	+0,655	3,519
	Fragile	Pédagogie	2,731	+0,647	3,378
	Fragile	Pédagogie	2,959	+0,557	3,516
	Fragile	Pédagogie	2,861	+0,548	3,409
	Modéré	Pédagogie	3,430	+0,490	3,920
	Fragile	Pédagogie	3,184	+0,468	3,653
	Fragile	Pratique	2,609	+0,468	3,077
	Fragile	Pédagogie	3,055	+0,460	3,515
	Modéré	Pédagogie	3,291	+0,455	3,746
	Fragile	Pertinence	3,008	+0,454	3,463
	Fragile	Pertinence	2,895	+0,451	3,346
	Modéré	Pédagogie	3,500	+0,405	3,905
	Fragile	Pratique	3,111	+0,392	3,503
	Modéré	Pédagogie	3,716	+0,369	4,085
	Fragile	Pratique	3,120	+0,356	3,476
	Modéré	Pédagogie	3,429	+0,355	3,784
	Fragile	Pratique	3,266	+0,344	3,610
	Fragile	Pratique	3,176	+0,338	3,513
	Modéré	Pertinence	3,539	+0,333	3,871
	Modéré	Pertinence	3,426	+0,332	3,759
	Modéré	Pédagogie	3,588	+0,323	3,911
	Modéré	Pédagogie	3,598	+0,322	3,920
	Modéré	Pédagogie	3,685	+0,304	3,990
	Modéré	Pédagogie	3,568	+0,302	3,870
	Modéré	Pédagogie	3,569	+0,299	3,868
	Modéré	Pratique	3,724	+0,292	4,016
	Modéré	Pédagogie	3,667	+0,288	3,955

	Modéré	Pratique	3,436	+0,282	3,718
	Modéré	Pertinence	3,512	+0,279	3,790
	Fragile	Pratique	3,135	+0,275	3,410
	Modéré	Pédagogie	3,531	+0,269	3,799
	Modéré	Pertinence	3,617	+0,254	3,871
	Modéré	Pratique	3,575	+0,246	3,820
	Modéré	Pédagogie	3,712	+0,243	3,954
	Modéré	Pédagogie	3,756	+0,239	3,995
	Modéré	Pertinence	3,632	+0,209	3,842
	Modéré	Pédagogie	3,851	+0,203	4,054
	Modéré	Pratique	3,582	+0,161	3,743

Le Gain_SCAP constitue un indicateur de priorisation et non une prédiction de l'effet réel d'une action. La mise en œuvre des actions doit donc être complétée par une analyse pédagogique du contexte de chaque module et, lorsque cela est pertinent, par un échange avec les équipes concernées.

La Figure 3 isole les 10 modules présentant les plus forts potentiels d'impact (gains compris entre +0,47 et +0,92). Sur ce périmètre d'urgence haute, le levier « Pédagogie » concentre 9 des 10 recommandations. Cela démontre que pour les cours les plus en crise, c'est la façon d'enseigner — plutôt que la structure du cours ou les activités pratiques — qui constitue le vecteur décisif de redressement de la satisfaction apprenant.

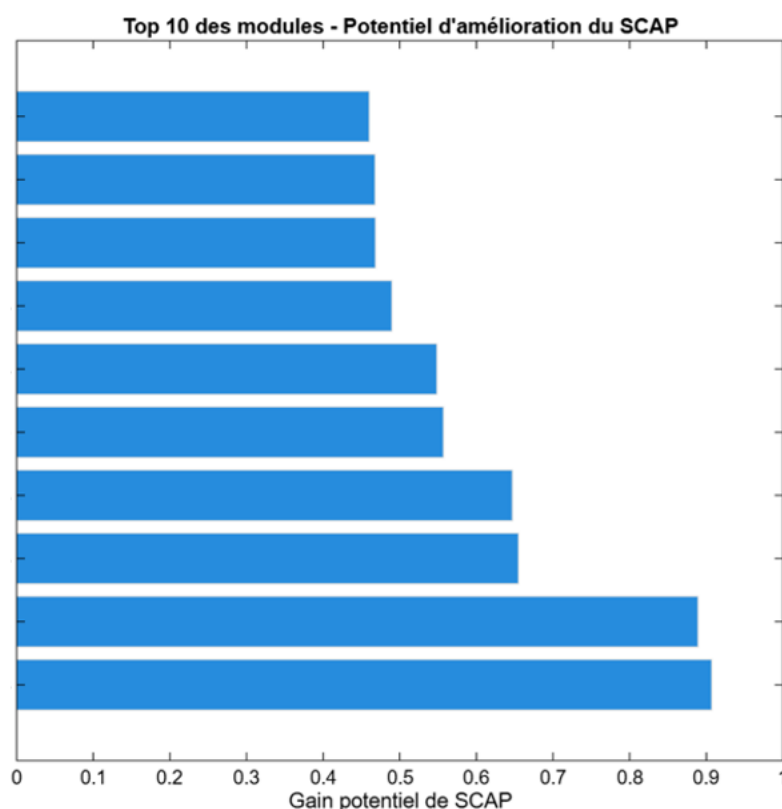


Figure 3 : gain de score estimé pour les 10 modules prioritaires.

Ces résultats permettent d'orienter les plans d'action vers trois catégories principales :

- Pédagogie : Examiner prioritairement les méthodes de transmission, la clarté des explications et les modalités d'accompagnement.
- Pratique : Examiner les TD, TP, projets et l'articulation entre connaissances théoriques et mise en application.
- Pertinence : Examiner l'alignement des contenus avec les objectifs et les compétences de la formation.

4.4 Fiabilité du système de priorisation

Afin d'évaluer la robustesse des profils et la légitimité des priorités d'action identifiées précédemment, un modèle prédictif supervisé, indépendant du processus de classification, a été développé à titre de contrôle complémentaire. Ce modèle est entraîné directement au niveau du module. Il cherche à estimer le score en étoiles moyen attribué par les apprenants à partir de ses caractéristiques intrinsèques suivantes :

- Le profil moyen sur les trois piliers pédagogiques (pertinence, pédagogie et pratique) ;
- Les variables contextuelles (modalité pédagogique, domaine d'enseignement, mode d'évaluation).

La Figure 4 illustre la confrontation entre la note moyenne observée (réelle) et la note moyenne estimée par l'algorithme sous un protocole rigoureux de validation. La répartition des 62 modules (représentés par les points colorés) le long de la ligne pointillée montre un ajustement très fidèle :

- Zone basse / modules fragiles (points rouges, < 2,5) : Le modèle identifie correctement les cours en difficulté et estime leur note dans des valeurs basses concordantes.
- Zone médiane (points orange, 3,0 à 3,8) : Les prédictions sont particulièrement resserrées autour de la ligne idéale, traduisant une excellente précision pour les modules de performance intermédiaire.
- Zone haute / modules d'excellence (points verts, > 4,0) : Les cours les mieux évalués sont fidèlement projetés dans les hauts scores.

On note aussi que le modèle parvient à expliquer 69 % de la variance des notes moyennes des modules. Cette performance élevée confirme que la satisfaction globale ne relève pas d'un jugement aléatoire des étudiants, mais repose sur une structure sous-jacente prédictible et parfaitement captée par l'apprentissage automatique.

Toutefois, cette analyse doit être interprétée comme un outil de vérification et de pilotage stratégique, et non comme une démonstration causale directe. Si le modèle confirme que l'amélioration des leviers identifiés (notamment Q2et Q3) coïncide avec une hausse de l'évaluation globale, les gains prédits restent conditionnés à la qualité de la mise en œuvre pédagogique sur le terrain.

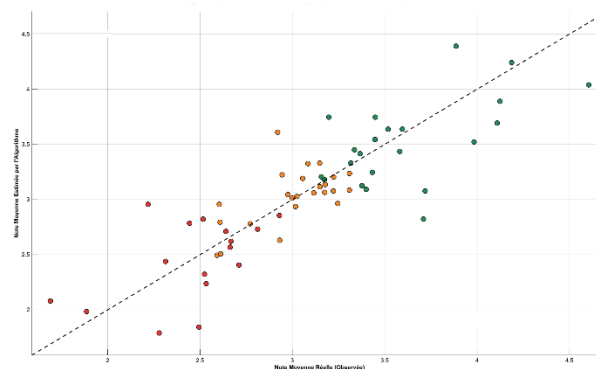


Figure 4 : note moyenne estimée par l'algorithme vs note moyenne réelle, par module (chaque point = un module ; plus un point est proche de la ligne pointillée, plus l'estimation est fidèle).

5. Hétérogénéité des perceptions étudiantes et analyse contextuelle

5.1 Niveau d'accord entre étudiants

Une note moyenne peut parfois masquer de fortes divergences au sein d'une même population étudiante. Deux enseignements obtenant une moyenne identique peuvent correspondre à des situations différentes : une appréciation relativement homogène autour de cette valeur, ou au contraire une distribution plus dispersée des opinions entre étudiants satisfaits et insatisfaits.

Dans cette étude, cette dimension est appréhendée au moyen d'un Indice de Consensus Apprenants (ICP) construit à partir de l'entropie de Shannon :

$$ICP = \left(1 - \frac{H}{\log_2(N)}\right) \times 100,$$

où H représente l'entropie de la distribution des réponses et $\log_2(N)$ est utilisée pour normaliser par l'entropie maximale (N est le cardinal).

Une valeur élevée de l'ICP traduit une plus forte concentration des réponses, et donc un consensus plus marqué ; une valeur faible traduit au contraire une distribution plus dispersée des réponses.

Les résultats obtenus sur les 8 209 observations retenues montrent une différence très nette entre les critères détaillés et la note globale en étoiles. Les ICP des trois critères Q1, Q2 et Q3 se situent respectivement à 14 %, 9,3 % et 6,81 %, tandis que l'ICP du score global en étoiles n'est que de 1,3 %.

Le résultat est particulièrement surprenant : les réponses aux trois critères analytiques présentent une concentration sensiblement supérieure à celle observée pour la note globale en étoiles mais restent faibles, ce qui montre des perceptions individuelles différentes.

Ce constat justifie de ne pas utiliser la seule note globale comme indicateur de pilotage de la qualité pédagogique. L'analyse des trois piliers Q1, Q2 et Q3 permet au contraire de distinguer les dimensions sur lesquelles les perceptions convergent davantage ou, au contraire, restent dispersées. Le score composite SCAP constitue ensuite un moyen de synthétiser ces dimensions selon les pondérations retenues dans l'étude.

5.2 Croisement du consensus avec le contexte pédagogique

La Figure 5 propose une lecture bidimensionnelle des perceptions étudiantes en croisant la satisfaction moyenne (note en étoiles) et l'Indice de Consensus Apprenants (ICP). L'axe vertical représente la satisfaction globale et l'axe horizontal la concentration des réponses. Les deux seuils de cadrage correspondent aux médianes observées sur le périmètre analysé : seuil de satisfaction médian est $Y=3,06$ étoiles et le seuil de consensus médian est $ICP=19,2$ %.

La cartographie proposée distingue quatre cadrans d'analyse :

- Zone I (Gris) — Satisfaction élevée / Consensus faible ($Y \geq 3,06$, $ICP < 19,2\%$) : Appréciation globale favorable mais tirée par des avis hétérogènes.
- Zone II (Vert) — Satisfaction élevée / Consensus élevé ($Y \geq 3,06$, $ICP \geq 19,2\%$) : Zone d'excellence soutenue par une adhésion homogène des étudiants.
- Zone III (Jaune) — Faible satisfaction / Consensus faible ($Y < 3,06$, $ICP < 19,2\%$) : Performance insuffisante accompagnée d'une dispersion ou polarisation des perceptions.
- Zone IV (Rouge) — Faible satisfaction / Consensus élevé ($Y < 3,06$, $ICP \geq 19,2\%$) : Zone de rejet consensuel, signalant un dysfonctionnement structurel généralisé.

Notons que les qualificatifs « élevé » (ou fort) et « faible » s'entendent au sens relatif par rapport au contexte de l'établissement, la médiane de 19,2 % servant de point de bascule.

Analyse selon la modalité d'enseignement

La modalité « Cours TD TP » constitue le cas le plus critique, enregistrant une moyenne de 2,52 étoiles pour un ICP de 15,20%. Se situant en Zone II (faible satisfaction, consensus faible), elle traduit un mécontentement partagé de manière mitigée au sein des effectifs. La modalité « TP » (2,77 étoiles ; 15,80 %) et le format « Cours AT » (2,72 étoiles ; 14,70 %) s'inscrivent dans un schéma similaire, tous regroupés en Zone II. À l'inverse, les modalités « Cours TD AT » (3,38 étoiles ; 25,90 %), « Cours TD » (3,26 étoiles ; 22,20 %), « Cours TP » (3,20 étoiles ; 23,20 %) et « AT » (3,10 étoiles ; 19,60 %) combinent une satisfaction élevée et un consensus fort, se positionnant nettement en Zone III. Les zones IV et I ne contiennent aucun élément.

Analyse selon le mode d'évaluation

Le mode d'évaluation « TP DS EX » enregistre la satisfaction la plus basse (2,52 étoiles ; 15,10 %), se plaçant au cœur de la Zone II (faible satisfaction, consensus faible). Le mode « AT Examen » se trouve en Zone IV (satisfaction élevée, consensus faible) en frôlant le seuil de satisfaction à 3,09 étoiles pour un ICP de 18,30 %. À l'opposé, le mode « DS Examen » s'impose comme le format le plus performant et le plus consensuel avec 3,37 étoiles pour un ICP de 26,10 %, largement ancré en Zone III. Il est accompagné dans cette Zone III par le mode « TP Examen » (3,07 étoiles ; 18,90 %). Les zones I et IV (hors AT Examen) restent vides. Ces résultats montrent que la structure d'évaluation constitue un levier déterminant dans le ressenti étudiant.

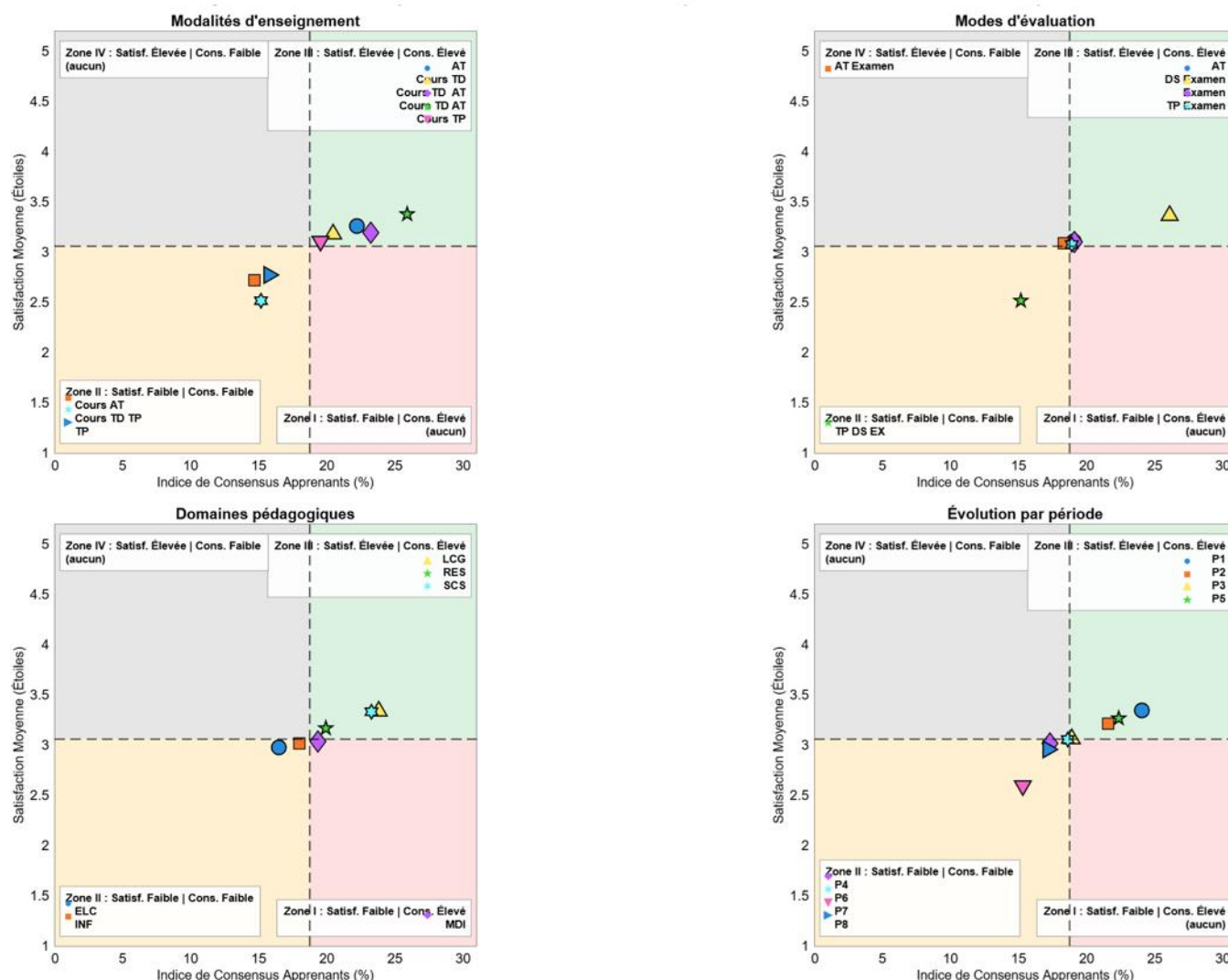
Analyse temporelle

La décomposition par périodes met en évidence une dynamique temporelle marquée. La période P7 constitue le point bas du parcours avec une moyenne de 2,59 étoiles pour un ICP de 15,20 %, la positionnant seule au bas de la Zone II. Ce résultat indique un décrochage marqué du ressenti. Les périodes P4 (2,98 étoiles ; 16,90 %), P6 (2,96 étoiles ; 17,20 %) et P8 (3,03 étoiles ; 18,60 %) occupent également la Zone II. À l'inverse, les périodes P1 (3,35 étoiles ; 24,10 %), P2 (3,22 étoiles ; 22,10 %), P5 (3,27 étoiles ; 20,10 %) et P3 (3,08 étoiles ; 18,70 %) se situent dans la Zone III d'excellence (satisfaction élevée et consensus fort). L'analyse qualitative des commentaires textuels vient confirmer ces tendances, notamment sur P7 où les remarques ciblent des difficultés d'organisation récurrentes.

Analyse par domaine pédagogique

L'analyse disciplinaire fait apparaître des profils très contrastés. Les domaines des Sciences Sociales SCS (3,34 étoiles ; 23,10 %) et Langues et Culture Générale LCG (3,34 étoiles ; 24,10 %) s'affirment en Zone III comme de véritables références consensuelles, suivis par les Réseaux RES (3,16 étoiles ; 19,70 %). À l'autre extrémité, le domaine Électronique de

Communication ELC (2,98 étoiles ; 18,80 %) et l'Informatique INF (2,92 étoiles ; 16,50 %) présentent les satisfactions les plus modestes en Zone II. Enfin, le domaine Mathématiques de l'Ingénieur MDI constitue une exception singulière : situé en Zone I, il associe une faible satisfaction (2,98 étoiles) à un fort consensus (29,80 %), traduisant une insatisfaction partagée par une très large majorité d'étudiants.



6. Ce que disent les commentaires libres

Si les questions fermées fournissent des indicateurs quantitatifs permettant de mesurer de manière structurée différents aspects de l'enseignement, les commentaires libres offrent une information complémentaire particulièrement riche. Ils permettent en effet aux apprenants d'exprimer spontanément leurs appréciations, leurs difficultés, leurs attentes et leurs propositions d'amélioration, sans être limités par les modalités de réponse prédéfinies du questionnaire.

Une analyse des commentaires libres a ainsi été conduite à l'aide de techniques de traitement automatique du langage naturel. Le retour des apprenants, comprend 461 commentaires exploitables. Une première dimension de l'analyse porte sur la polarité des commentaires. Les verbatims ont été classés selon trois catégories :

- Positive, en nombre de 38, lorsque le commentaire exprime principalement une appréciation favorable ;
- Négative, en nombre de 159, lorsqu'il traduit principalement une insatisfaction, une difficulté ou une critique ;
- Neutre ou mixte, en nombre de 264, lorsque le commentaire est essentiellement descriptif, factuel, ambigu ou comporte simultanément des éléments positifs et négatifs.

6.1 Analyse longitudinale sur les huit périodes

L'analyse temporelle permet d'apporter une dimension supplémentaire à l'interprétation des résultats. Le tableau 3 résume, pour chaque période, le ton général des commentaires libres. Cette analyse met en évidence une variabilité importante de la polarité. Les périodes P1 et P6 présentent notamment les volumes de commentaires les plus importants, P6 se distingue par une proportion particulièrement élevée de commentaires négatifs et P7 est la période la moins bien notée, celle où le désaccord entre étudiants est le plus faible, et celle dont le ton des commentaires est le plus négatif, avec des mots-clés explicites (« Inutile », « Démotivation »). La convergence des 3 lectures indépendantes — score, désaccord et commentaires — renforce la fiabilité de ce signal.

Période	Commentaires substantiels	Positifs n (%)	Neutres / mixtes n (%)	Négatifs n (%)
P1	122	6 (4,9 %)	78 (63,9 %)	38 (31,1 %)
P2	72	3 (4,2 %)	55 (76,4 %)	14 (19,4 %)
P3	40	1 (2,5 %)	26 (65,0 %)	13 (32,5 %)
P4	54	3 (5,6 %)	43 (79,6 %)	8 (14,8 %)
P5	46	5 (10,9 %)	29 (63,0 %)	12 (26,1 %)
P6	77	6 (7,8 %)	35 (45,5 %)	36 (46,8 %)
P7	25	0 (0,0 %)	17 (68,0 %)	8 (32,0 %)
P8	25	1 (4,0 %)	20 (80,0 %)	4 (16,0 %)
Total	461	25 (5,4 %)	303 (65,7 %)	133 (28,9 %)

Tableau 3 : ton des commentaires libres, par période.

6.2 Analyse thématique

Au-delà de la polarité, une analyse thématique a été réalisée afin d'identifier les principaux sujets abordés spontanément par les apprenants. Plusieurs dimensions récurrentes ont été mises en évidence.

La pédagogie et la clarté des enseignements constituent de loin la thématique la plus fréquemment rencontrée, apparaissant dans environ 65 % des commentaires substantiels. Cette catégorie regroupe notamment les remarques relatives à la qualité des explications, à la compréhension des notions, à la maîtrise pédagogique des contenus et au caractère plus ou moins clair des enseignements.

La deuxième dimension importante concerne les programmes et les contenus pédagogiques, présents dans environ 44 % des commentaires. Les apprenants évoquent notamment le contenu des modules, la pertinence ou le volume des notions abordées et leur adéquation avec leurs besoins ou leurs attentes.

Les enseignants et les compétences perçues apparaissent également dans environ 43 % des commentaires. Cette catégorie comprend aussi bien les appréciations relatives à la compétence, à la disponibilité et à l'investissement des enseignants que les critiques portant sur certaines pratiques pédagogiques.

La charge et le volume de travail constituent une autre préoccupation majeure, retrouvée dans environ 40 % des commentaires. Les remarques portent notamment sur la quantité de contenu, la charge de travail, le temps nécessaire à la réalisation des projets ou encore la répartition des activités.

Les travaux pratiques et la dimension appliquée des enseignements représentent environ 26 % des commentaires, ce qui souligne l'importance accordée par les apprenants à l'articulation entre connaissances théoriques et mise en pratique.

D'autres thématiques apparaissent également, mais avec une fréquence moindre : les modalités d'évaluation et les exigences de présence (16 %), l'organisation et les emplois du temps (11 %) ainsi que les ressources et supports pédagogiques (11 %).

Ces pourcentages ne sont pas exclusifs : un même commentaire peut aborder plusieurs dimensions simultanément. Ils doivent donc être interprétés comme des indicateurs de fréquence des thématiques et non comme une répartition exclusive des opinions.

6.3 Limites et biais d'interprétation

L'interprétation des commentaires libres nécessite toutefois une précaution méthodologique importante liée au biais de sélection ou biais de volontariat.

Contrairement aux réponses aux questions fermées, qui peuvent être renseignées systématiquement dans le cadre du questionnaire, le commentaire libre repose sur une démarche volontaire. Les apprenants particulièrement insatisfaits peuvent être davantage motivés à exprimer une critique afin de signaler une difficulté ou de demander une amélioration. À l'inverse, un apprenant satisfait peut considérer qu'il n'est pas nécessaire de formuler un commentaire supplémentaire, notamment lorsque son appréciation positive est déjà exprimée par les questions fermées.

Le corpus de commentaires peut donc présenter une surreprésentation des expériences négatives ou problématiques par rapport à la population globale des répondants. Cette situation explique notamment pourquoi la proportion de commentaires négatifs ne doit pas être assimilée au taux d'insatisfaction des apprenants. On parlerait dans ce cas d'étude de « 38 % de commentaires substantiels analysés présentant une polarité négative » que de « 38 % d'apprenants insatisfaits ».

Une seconde limite concerne la nature même du langage utilisé dans les commentaires. Les formulations peuvent être très courtes, ironiques, ambiguës, grammaticalement approximatives ou comporter simultanément plusieurs opinions. C'est la raison pour laquelle la catégorie « neutre/mixte » a été conservée et pourquoi les résultats doivent être considérés comme une aide à l'analyse plutôt que comme une vérité absolue.

7. Précautions d'interprétation

L'analyse consolidée mobilise plusieurs approches complémentaires — statistiques descriptives, mesure de concentration des réponses, classification, modèles d'apprentissage automatique et analyse des commentaires libres. Cette combinaison permet

d'obtenir une lecture plus riche des données, mais les résultats doivent être interprétés en tenant compte de plusieurs limites méthodologiques.

- Le niveau d'analyse des modèles doit être distingué du niveau individuel. Les performances des modèles statistiques et d'apprentissage automatique sont principalement évaluées sur des données agrégées au niveau des modules. Elles renseignent donc sur la cohérence des relations observées entre les caractéristiques pédagogiques et les indicateurs de satisfaction, mais ne signifient pas que les modèles peuvent prédire avec précision la réponse ou le comportement d'un étudiant particulier.
- Le consensus est une mesure de concentration et non un taux d'accord. L'indice de consensus pédagogique retenu est construit à partir de la concentration des réponses aux trois dimensions Q1, Q2 et Q3. Un consensus élevé signifie que les réponses sont davantage concentrées autour de certaines modalités ; il ne signifie pas que tous les étudiants ont donné exactement la même réponse. Inversement, un consensus faible traduit une plus grande dispersion des réponses. Cet indicateur doit donc être interprété comme une mesure comparative de l'homogénéité des perceptions et non comme une proportion d'étudiants satisfaits ou insatisfaits.
- Les indicateurs de consensus doivent être interprétés conjointement avec les effectifs. Certains résultats portant sur des modalités d'enseignement, des modes d'évaluation ou des sous-groupes peuvent reposer sur un nombre limité de réponses. Une concentration élevée ou faible des réponses peut alors être plus sensible aux fluctuations d'échantillonnage. Les résultats portant sur des groupes suffisamment représentés et confirmés par plusieurs dimensions doivent être considérés comme plus robustes que les observations isolées.
- Les seuils utilisés pour les analyses en quadrants sont des seuils descriptifs internes. Les médianes utilisées pour distinguer les niveaux de satisfaction et de consensus permettent de structurer la lecture des résultats au sein du corpus étudié. Elles ne constituent pas des normes externes de qualité et ne doivent pas être interprétées comme des seuils absolus de satisfaction pédagogique.
- Le score SCAP constitue un indicateur interne de pilotage. Le score composite SCAP repose sur une pondération définie pour cette étude entre les dimensions Pertinence, Pédagogie et Pratique. Il constitue un outil de comparaison et de priorisation au sein du dispositif de qualité, mais ne peut être assimilé à une mesure absolue de la qualité pédagogique d'un enseignement.
- Les profils issus du clustering sont relatifs au corpus analysé. Les catégories « Fragile », « Modéré » et « Excellent » correspondent à des regroupements statistiques obtenus sur les données étudiées. Elles ne constituent pas des classifications pédagogiques absolues. En particulier, l'étiquette « Excellent » doit être comprise comme désignant un profil relatif présentant les caractéristiques les plus favorables dans le corpus, et non comme une certification de qualité pédagogique.
- Les recommandations issues du modèle constituent des scénarios de priorisation. Les gains potentiels associés aux actions proposées correspondent à des simulations construites à partir des relations observées dans les données. Ils ne constituent pas des effets causaux garantis. Une amélioration théorique du score SCAP après modification d'une dimension ne signifie donc pas qu'une action donnée produira nécessairement le même gain dans la réalité.
- Les modèles de prédiction et de validation doivent être distingués. Les performances obtenues sur les données d'apprentissage ne peuvent pas être assimilées à une validation indépendante. Les résultats obtenus sur les données de test et par validation croisée fournissent une estimation plus pertinente de la capacité de généralisation du modèle. Les valeurs de performance doivent donc être présentées en précisant systématiquement le protocole d'évaluation utilisé.
- Le modèle utilisé fournit une référence comparative et non une recommandation automatique. L'identification d'un module présentant un profil proche d'un autre module permet de faciliter la comparaison entre situations pédagogiques similaires. Elle ne signifie pas que les pratiques du module de référence doivent être reproduites automatiquement. Une analyse qualitative du contexte reste nécessaire avant toute transposition.
- L'analyse des commentaires libres est soumise à un biais de sélection. Contrairement aux questions fermées, le commentaire libre repose sur une démarche volontaire. Les apprenants rencontrant une difficulté ou souhaitant signaler une insatisfaction peuvent être davantage susceptibles de rédiger un commentaire. La distribution des tonalités observées dans les commentaires ne peut donc pas être assimilée à la distribution de la satisfaction dans l'ensemble de la population répondante.
- L'analyse automatique du ton et des thèmes constitue une aide à l'interprétation. Les commentaires peuvent être courts, ambigus, ironiques, incomplets ou comporter plusieurs opinions simultanément. Les résultats de l'analyse NLP doivent donc être considérés comme des indicateurs complémentaires. Les périodes ou thématiques identifiées comme particulièrement sensibles doivent faire l'objet d'une vérification qualitative avant toute décision importante.
- La convergence de plusieurs sources constitue un argument de robustesse, mais pas une preuve causale. Lorsqu'une même période ou un même profil est identifié simultanément par plusieurs indicateurs — satisfaction moyenne, concentration des réponses, caractéristiques des modules et tonalité des commentaires — la convergence constitue un signal plus robuste qu'une observation isolée. Elle ne permet toutefois pas, à elle seule, d'identifier la cause du phénomène observé.
- Le cas de P7 doit ainsi être considéré comme un signal prioritaire à investiguer. La convergence observée pour cette période justifie une analyse complémentaire portant notamment sur l'organisation des enseignements, la charge de travail, le calendrier, les modalités d'évaluation, les caractéristiques des modules concernés et les éventuels événements particuliers. Il serait prématuré d'attribuer directement les résultats observés à une cause pédagogique déterminée sans cette analyse complémentaire.
- Enfin, l'analyse longitudinale devra être poursuivie sur plusieurs campagnes successives. La consolidation des données sur trois ans permet déjà d'identifier des tendances et des profils récurrents, mais un suivi continu permettra de distinguer plus clairement les effets structurels des variations ponctuelles. La reconduction annuelle du même dispositif, associée au suivi des actions engagées, constituera progressivement une base plus solide pour l'évaluation de l'évolution de la qualité pédagogique.

8. Synthèse des principales contributions de l'étude

Cette synthèse consolidée sur trois ans confirme et affine les constats du rapport annuel précédent : la satisfaction des élèves-ingénieurs de SUP'COM apparaît principalement associée à la qualité perçue de la pédagogie, davantage qu'au domaine, à la période ou aux modalités d'enseignement. L'apport de cette analyse est de traduire ce constat en un plan d'action priorisé,

module par module, en identifiant les situations dans lesquelles une amélioration ciblée des dimensions pédagogiques pourrait être associée au gain de satisfaction le plus important.

Les analyses statistiques, les profils de modules et les modèles d'apprentissage utilisés dans cette étude convergent vers l'importance particulière de la dimension « Pédagogie ». Cette convergence doit toutefois être interprétée comme un faisceau d'indices statistiques et non comme la démonstration d'une relation causale.

L'analyse du consensus pédagogique, fondée sur la concentration des réponses aux dimensions Q1, Q2 et Q3, apporte une lecture complémentaire de la satisfaction moyenne. Elle permet de distinguer les situations dans lesquelles une satisfaction faible ou élevée s'accompagne d'une plus ou moins grande homogénéité des réponses. Cet indicateur doit être interprété comme une mesure relative de concentration des réponses et non comme un pourcentage d'étudiants partageant exactement la même opinion.

La mise en regard des résultats quantitatifs avec l'analyse des commentaires libres fait ressortir un signal particulièrement intéressant pour la période P7 : satisfaction plus faible, concentration importante des réponses et tonalité négative des commentaires. Cette convergence de plusieurs sources constitue un signal d'alerte prioritaire pour le pilotage de la qualité, tout en nécessitant une vérification qualitative des circonstances propres à cette période avant toute conclusion définitive.

Les pistes complémentaires évoquées — repérage de modules présentant des pratiques intéressantes, suivi longitudinal et mise en place progressive d'un système de pilotage continu de la qualité pédagogique — constituent des perspectives pertinentes. Leur exploitation devra toutefois s'appuyer sur des données longitudinales suffisamment riches et sur la reconduction régulière des mêmes indicateurs.

9. Vers un plan d'action

Pour transformer le diagnostic quantitatif et qualitatif en une démarche effective d'amélioration continue, le plan d'action proposé s'articule autour de quatre axes stratégiques.

9.1 Axe 1 : Interventions d'urgence et accompagnement pédagogique prioritaire

- Investigation approfondie sur la période P7 : Mandater une cellule d'analyse, formée par le directeur des études et les directeurs des départements, dédiée pour étudier spécifiquement les circonstances de la période P7 (organisation, charge de travail cumulée, événements contextuels, ...) afin d'identifier les causes racines de la baisse brutale de satisfaction et de la convergence des commentaires négatifs.
- Déploiement d'un plan d'accompagnement sur le levier « Pédagogie » : Mettre en place un dispositif d'appui pédagogique ciblé pour les 40 modules identifiés comme prioritaires (et en particulier le *Top 10* des modules à fort potentiel de gain SCAP). Les actions porteront prioritairement sur :
 - La clarté des explications et l'adaptation des méthodes d'enseignement aux prérequis étudiants ;
 - La révision des supports de cours et la médiatisation des contenus ;
 - La meilleure articulation entre les concepts théoriques et les applications pratiques.

9.2 Axe 2 : Harmonisation des pratiques et convergence des perceptions

Afin de réduire la dispersion des avis étudiants (faible ICP) et de remonter la satisfaction globale dans les modules à perception hétérogène, les leviers suivants peuvent être activés :

- Clarification du contrat pédagogique : Expliciter dès le début du module les objectifs d'apprentissage, les prérequis attendus et les critères d'évaluation pour réduire les écarts d'interprétation des apprenants.
- Harmonisation inter-intervenants : Établir des réunions de coordination obligatoires pour les modules partagés entre plusieurs enseignants afin de garantir une cohérence des exigences, des supports et des modalités d'évaluation.
- Lissage de la charge de travail et des évaluations : Mieux répartir les travaux évalués (DS, TP, projets) au cours du semestre afin d'éviter la saturation perçue par les élèves-ingénieurs sur certaines périodes critiques.

9.3 Axe 3 : Gouvernance, restitution et suivi de l'évolution

- Restitution ciblée et collaborative : Partager les résultats du benchmark et les recommandations (Tableau 2) avec les responsables de domaines pédagogiques (SCS, LCG, RES, INF, ELC, MDI), via les directeurs de départements, pour co-construire des feuilles de route d'amélioration adaptées aux spécificités disciplinaires.
- Institutionnalisation de la boucle de rétroaction annuelle : Reconduire l'analyse synthétique selon le même protocole chaque année académique afin d'évaluer l'impact des actions correctives engagées et d'observer la trajectoire des modules (passage du profil *Fragile* vers *Modéré* ou *Excellent*).

9.4 Axe 4 : Recommandations méthodologiques et évolution de l'ingénierie d'évaluation

- Optimisation de l'outil de mesure et du questionnaire :
 - Expliciter formellement les ancrages descriptifs de l'échelle globale en étoiles (1 à 5) pour réduire la variance d'interprétation individuelle.
 - Conserver et renforcer les trois questions analytiques (*Q1 Pertinence*, *Q2 Pédagogie*, *Q3 Pratique*) qui constituent le cœur explicatif du diagnostic.
- Enrichissement du modèle de pilotage : Étudier l'opportunité d'intégrer progressivement de nouvelles dimensions analytiques (ex. régularité du travail, qualité du feedback donné aux étudiants) et d'adapter le rythme de collecte sans générer de fatigue de réponse chez les apprenants.

10. Conclusion

En consolidant trois années d'évaluations, cette étude transforme l'écoute des élèves-ingénieurs en un véritable levier de pilotage stratégique de la qualité pédagogique à SUP'COM. Grâce à la priorisation des axes d'amélioration — au premier rang desquels figure la pédagogie — et à un plan d'action structuré, l'établissement dispose d'une feuille de route claire pour accompagner les équipes enseignantes, réduire l'hétérogénéité des perceptions et inscrire l'excellence académique dans une dynamique d'amélioration continue durable.