



LA ROBUSTESSE EN PÉDAGOGIE

Ses implications dans les pratiques d'évaluation



europé
admeee

Mai 2026

PENSER LA ROBUSTESSE EN ÉVALUATION

01

Un contexte signifiant

- a. La notion de polycrise
- b. L'impact des crises sur l'enseignement supérieur

02

La robustesse : un changement de paradigme

- a. Des robustesses disciplinaires
- b. La notion de robustesse
- c. Robustesse / Performance / Suroptimisation
- d. 7 critères – un choix assumé

03

La robustesse au travers des pratiques évaluatives

04

Le contrat social de l'université : une réalité

CONTEXTUALISATION

01

Un contexte signifiant

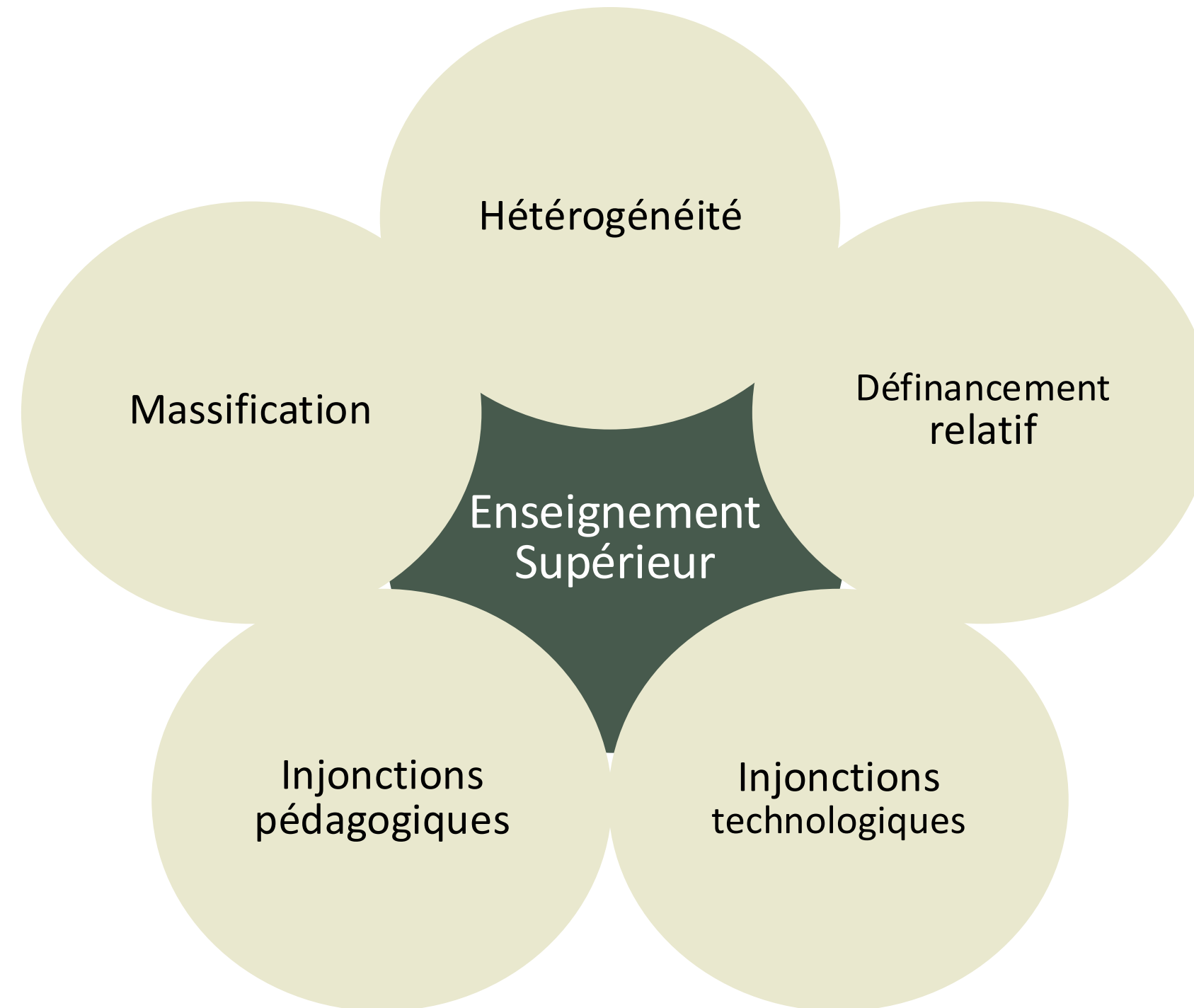
- a. La notion de polycrise
- b. L'impact des crises sur l'enseignement supérieur

LA NOTION DE POLYCRISE :

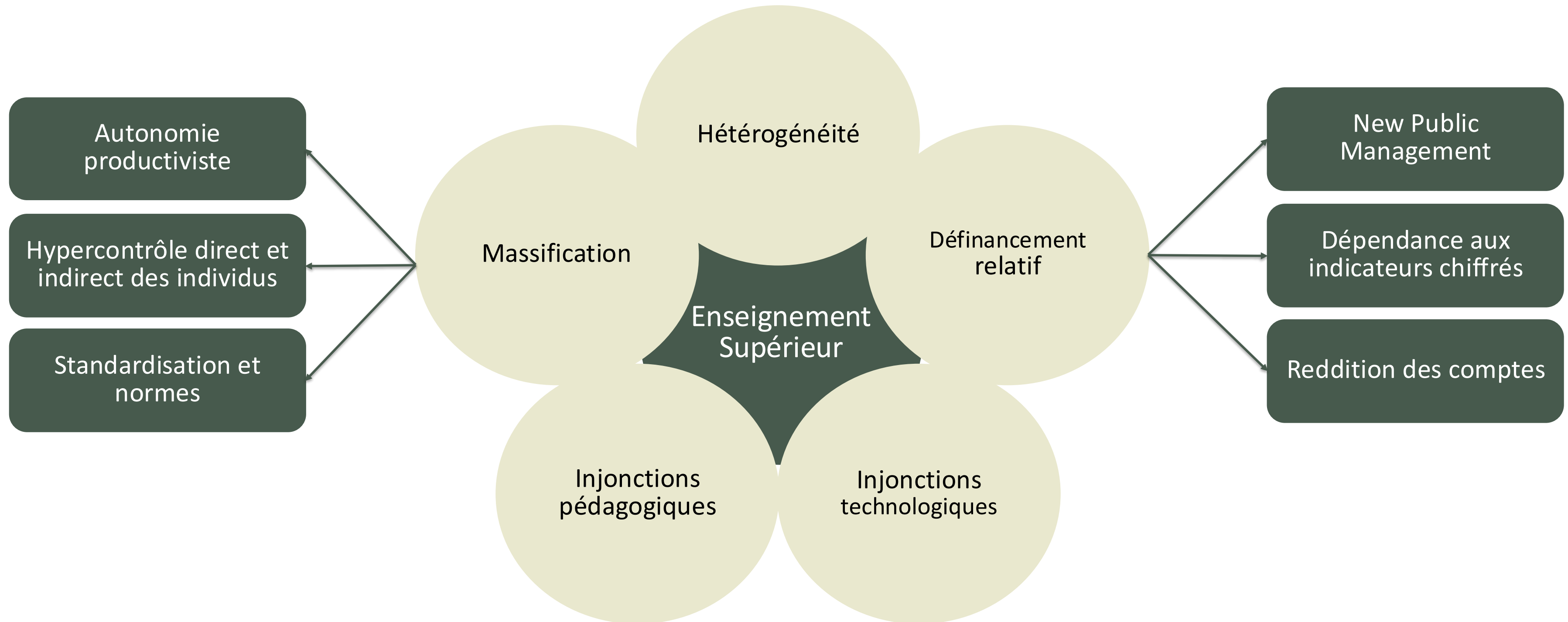
La **polycrise** est désignée par Lawrence et al. (2024) comme « l'entrelacement causal de crises affectant plusieurs systèmes globaux de manière à dégrader significativement les perspectives de l'humanité » (p. 3).

Là où le risque systémique désigne la probabilité qu'un système s'effondre, la polycrise désigne l'instant où cet effondrement se produit simultanément dans plusieurs systèmes interdépendants. (Liu & Renn, 2025)

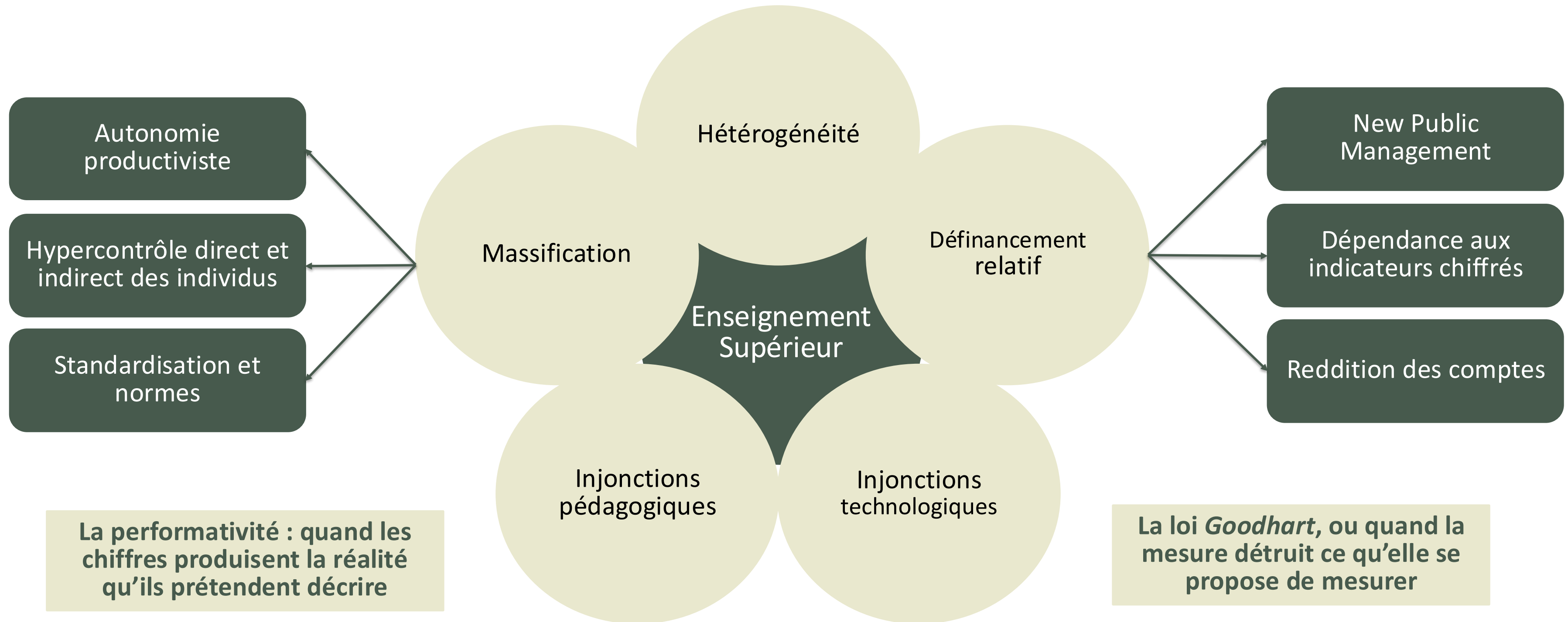
Cela nécessite : une pensée complexe, attentive aux interdépendances, boucles de rétroaction et à l'incertitude (Morin, 1999)



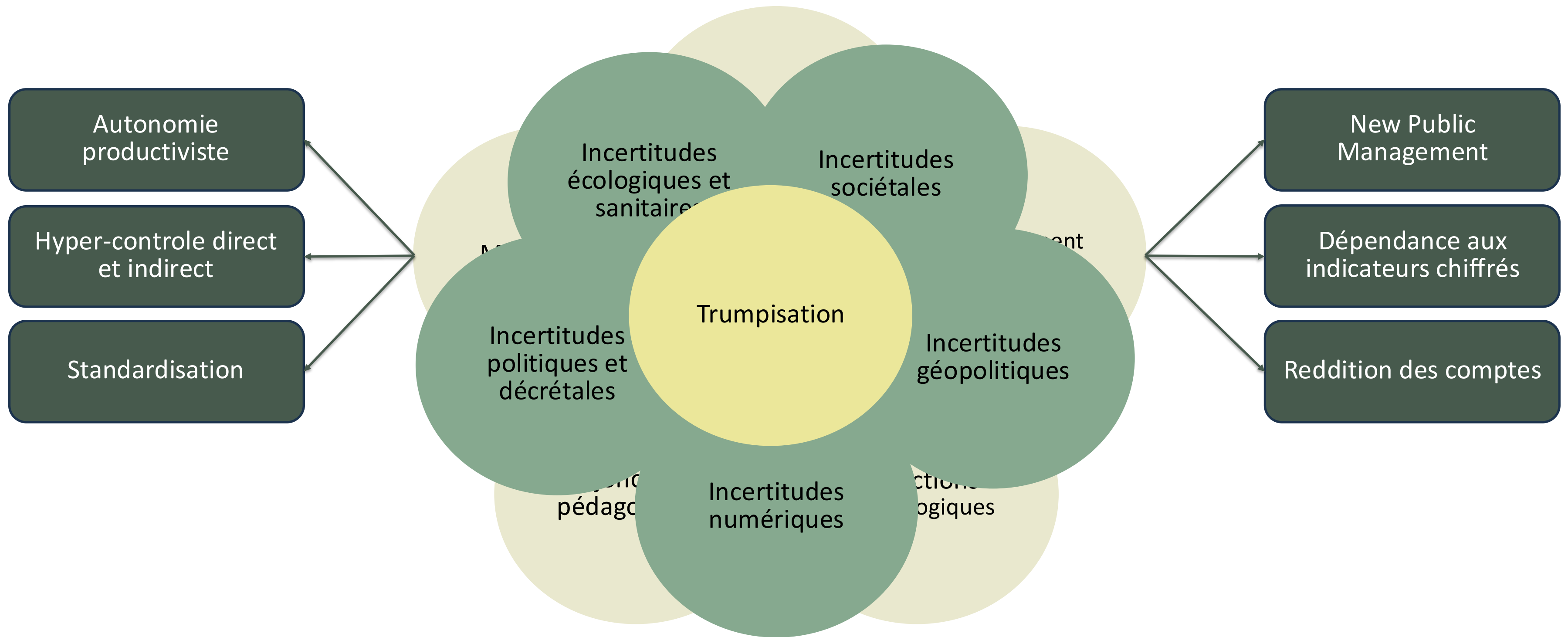
« Toujours plus d'étudiants, de publications, de projets... avec toujours moins de moyens. »



Osciller sans se rompre : la tentation du contrôle



Osciller sans se rompre (à court terme en tout cas)

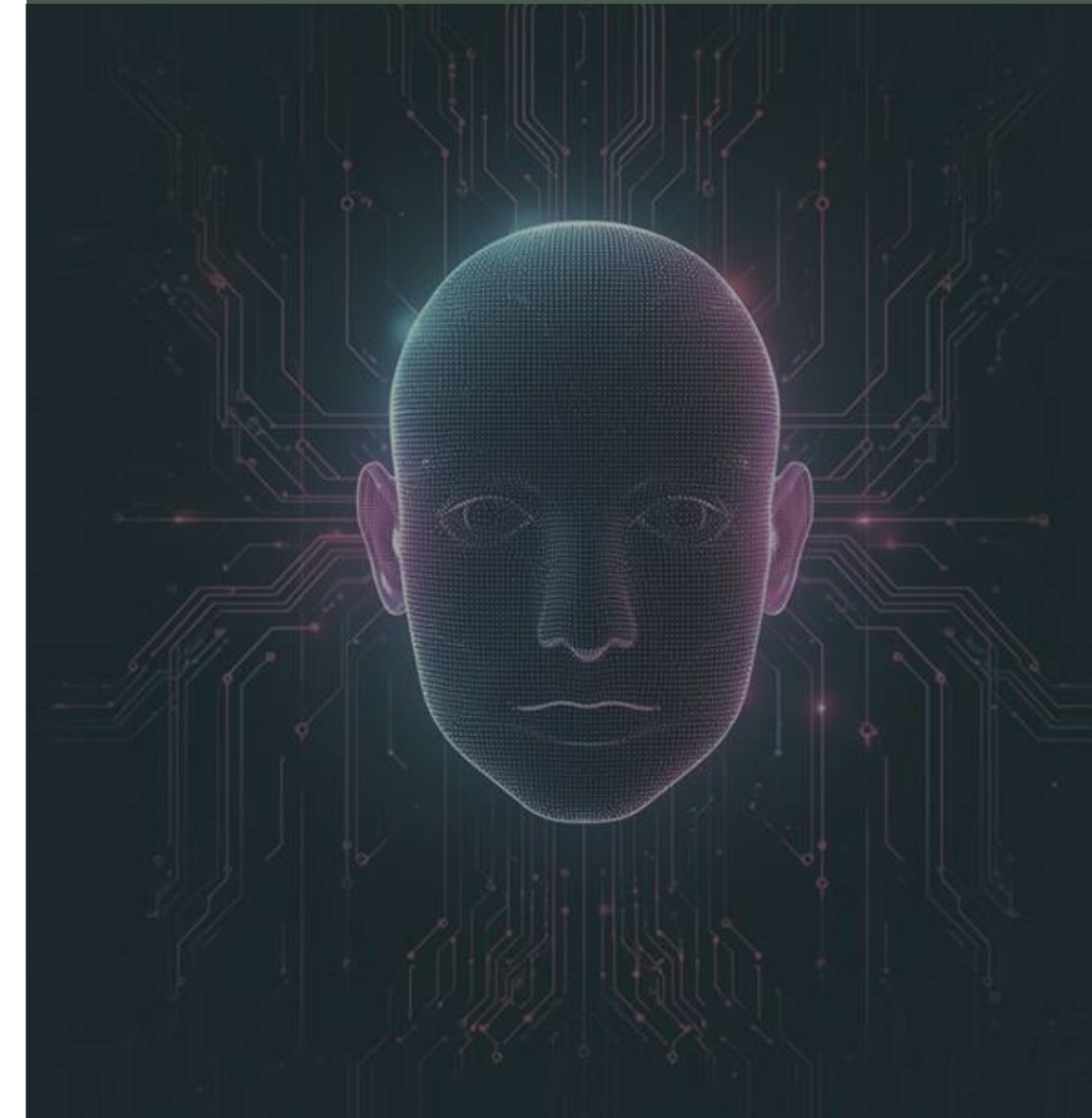


Les outils de stabilité deviennent des **facteurs de fragilité**

L'IA REDÉFINIT LE TRAVAIL ÉTUDIANT (ET L'ÉVALUATION)

« J'AI PARFOIS L'IMPRESSION DE NE PLUS SAVOIR APPRENDRE, DE
NE PLUS SAVOIR TRAVAILLER OU LIRE UN TEXTE SEULE. »
(GRAVELEAU, 2025)

- Entre 53 % et 73 % des étudiants utilisent l'IA dans leurs travaux
(Paquelin et al., 2025)
- L'IA reconfigure le rapport au savoir et complexifie l'évaluation des
apprentissages
 - Une étude récente de Microsoft intitulée « Working with AI:
Measuring the Occupational Implications of Generative AI » (Tomlinson et
al., 2025) démontre que les métiers les plus exposés à l'automatisation
par l'IA ne sont plus seulement les emplois manuels ou routiniers, mais
également des professions hautement qualifiées, nécessitant des
compétences cognitives complexes.



LA ROBUSTESSE

02

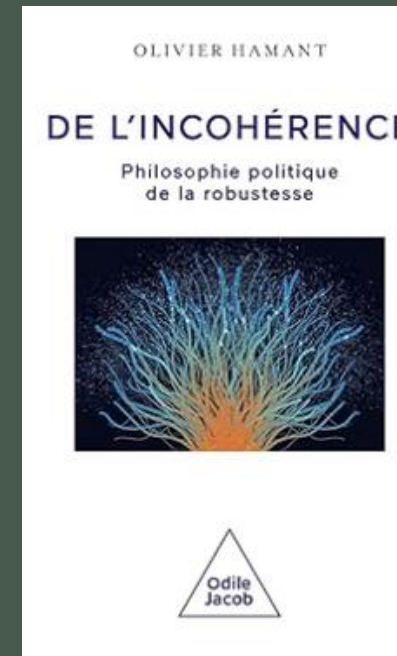
LA ROBUSTESSE : UN CHANGEMENT DE PARADIGME

- a. Des robustesses disciplinaires
- b. La notion de robustesse
- c. Robustesse / Performance / Suroptimisation
- d. 7 critères – un choix assumé

<h2>Biologie des systèmes</h2> <p>(Stelling et al., 2004)</p> <p>« La robustesse — la capacité à maintenir les performances face aux perturbations et à l'incertitude, est une propriété fondamentale des systèmes vivants. »</p>	<h2>Économie</h2> <p>(Hansen & Sargent, 2008)</p> <p>« Une règle de décision robuste est celle qui performe bien sous une variété de modèles probabilistes — en évaluant la pire performance d'une règle de décision sur un ensemble de modèles alternatifs. »</p>
<h2>IA</h2> <p>(Braiek & Khomh, 2024)</p> <p>« La robustesse d'un modèle désigne sa capacité à maintenir des performances prédictives stables face aux variations et aux changements dans les données d'entrée. »</p>	<h2>Ingénierie du contrôle</h2> <p>(Zames, 1981)</p> <p>« La capacité d'un système de contrôle à maintenir la stabilité et les performances en présence d'incertitudes est désignée par le terme robustesse. L'incertitude de modélisation est quantifiée et la performance optimisée. »</p>

LA ROBUSTESSE

Olivier Hamant (2022, 2024)



« LA CAPACITÉ D'UN SYSTÈME À MAINTENIR SA STABILITÉ MALGRÉ LES FLUCTUATIONS QU'IL SUBIT »

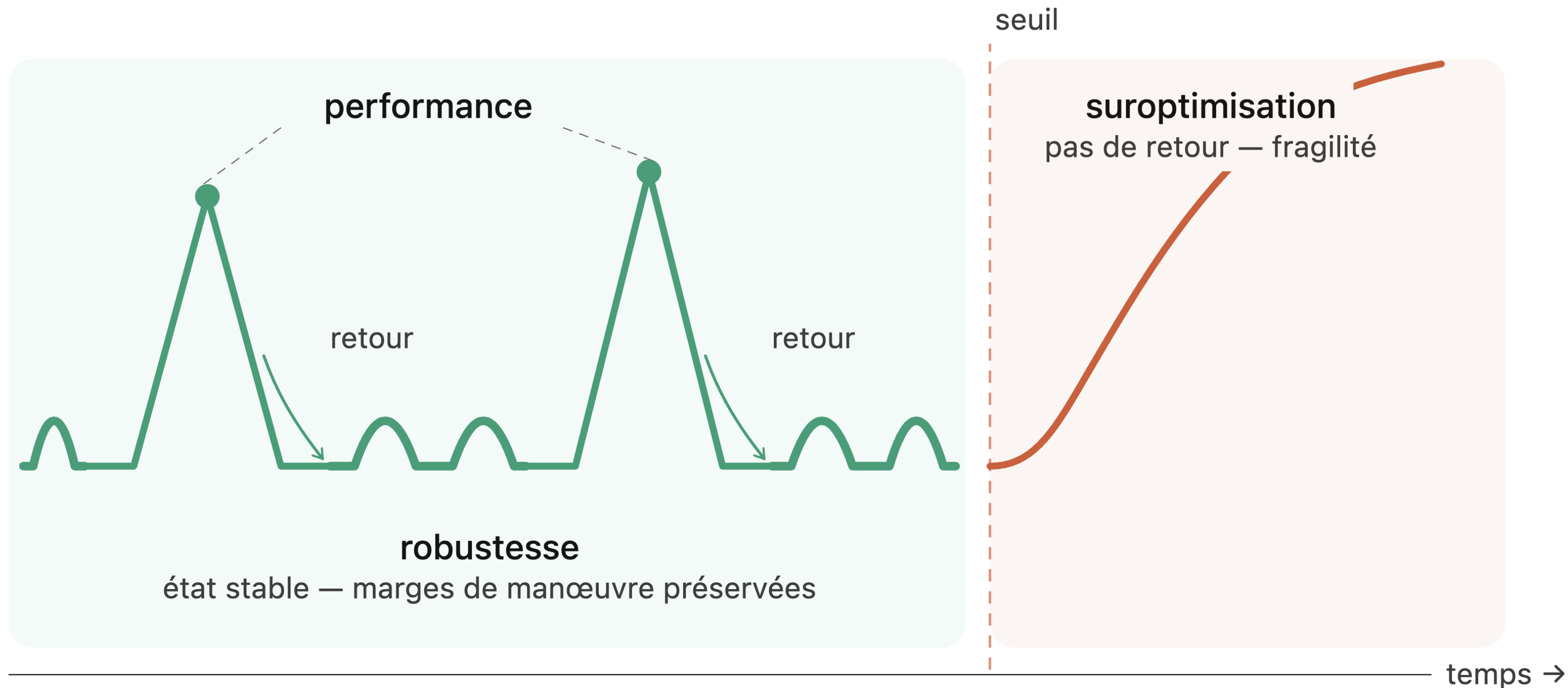
La performance est à redéfinir et la suroptimisation à combattre.

Pas un dogme, mais
un cadre de réflexion
stimulant

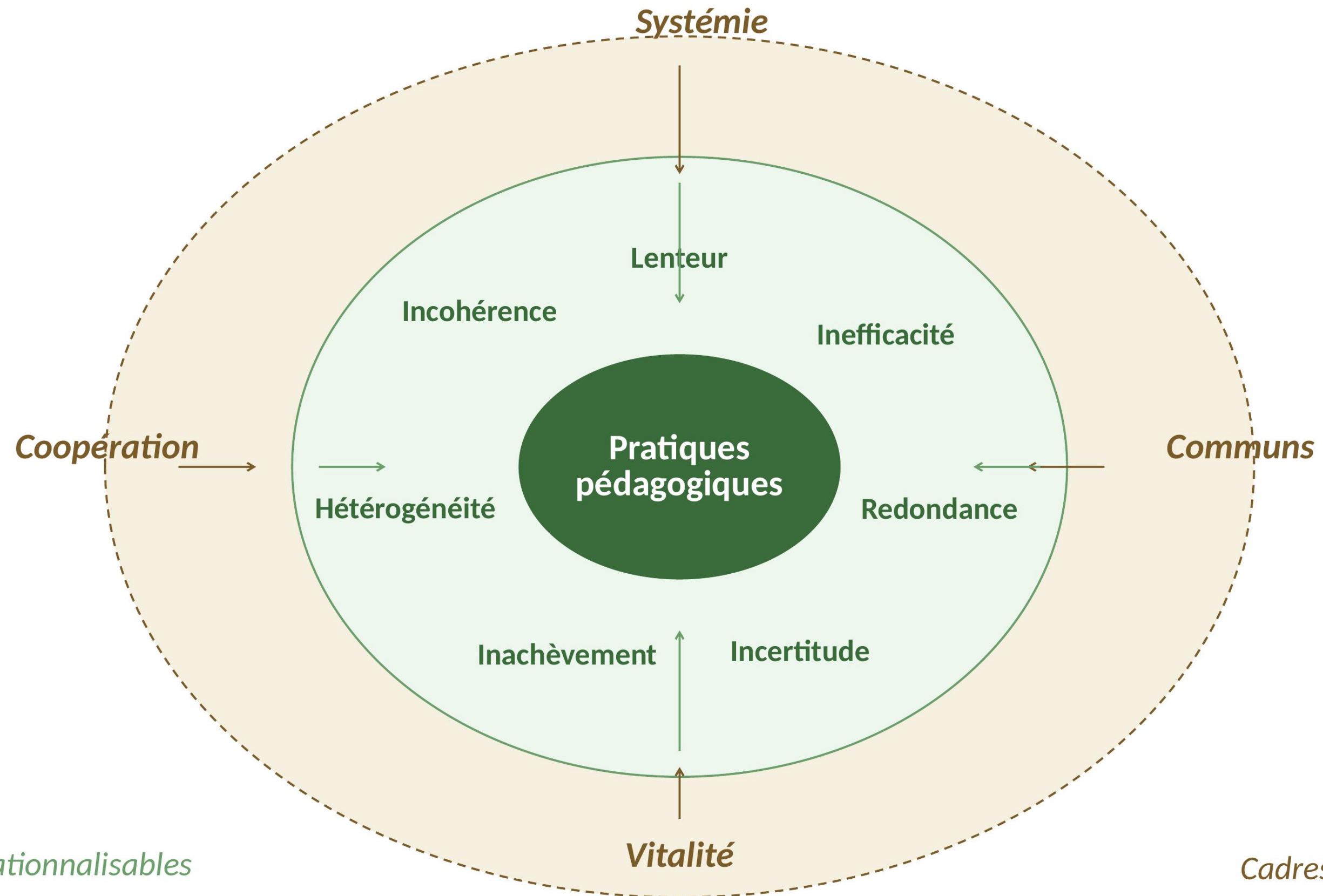
Les 7 caractéristiques de la robustesse (Hamant)

→ La robustesse se traduit par des propriétés souvent perçues comme des défauts — lenteur, inefficacité, redondance, incertitude, inachèvement, hétérogénéité et incohérence — mais qui, dans le vivant, sont des ressources d'adaptabilité.

→ Transposées à la pédagogie, elles offrent des pistes concrètes pour concevoir des dispositifs capables de soutenir les apprentissages dans un contexte incertain.



Est-ce réellement la performance qu'il faut opposer à la robustesse ?



EN PRATIQUE

03

LA ROBUSTESSE AU TRAVERS DES PRATIQUES
ÉVALUATIVES

LENTEUR · Portfolio réflexif

Le portfolio s'inscrit dans « *une perspective réflexive et expérientielle de l'apprentissage et de l'évaluation ainsi que dans une visée de responsabilisation de l'étudiant vis-à-vis de son processus d'apprentissage et d'évaluation.* » (Fischer et al., 2020, p. 3)

« *Les processus réflexifs se constituent pour tous sur la durée et sur la régularité des échanges.* » (Garnier et Marchand, 2012 ; Verslype et al., 2017 ; Naccache et al., 2006, cités par Boucher et al., 2018)

Sur le plan empirique, « *les dimensions les plus marquées par le dispositif dès la mi-parcours sont la réflexivité et la prise de responsabilité dans les apprentissages* », auxquelles s'ajoutent le sentiment de maîtrise de la construction des savoirs et les capacités d'autoévaluation (Ferrier , 2024). La pensée réflexive « *est provoquée par un événement qui induit un doute, une perplexité ou une incertitude et qui conduit un individu* » à retravailler sa compréhension. (Naccache et al., 2006 cités par Boucher et al., 2018)

Le portfolio, de par le temps long qu'il induit, développe chez l'étudiant des compétences qui tendent vers la robustesse : la capacité à habiter l'incertitude, à réguler son propre apprentissage dans la durée, et à construire un sentiment de maîtrise qui ne dépend pas uniquement d'une performance ponctuelle.

LENTEUR

INEFFICACITÉ · Le statut de l'erreur

« La robustesse résulte de procédés intrinsèquement et localement inefficaces et inefficients, c'est-à-dire opérant contre la performance. » (Hamant, 2022, p. 9)

Bonniol et Vial (1997) présentent deux approches de la régulation, qui divergent selon la place accordée à l'erreur. Dans une perspective cybernétique, la régulation fonctionne en circuit fermé, focalisée sur la performance : l'erreur est perçue comme un écart à corriger. À l'inverse, la régulation systémique adopte une boucle ouverte, considérant l'erreur comme un « surplus de significations » révélant le fonctionnement global du système, et donc comme une ressource à valoriser plutôt qu'à éliminer.

L'évaluation formative traduit cette seconde régulation puisqu'elle « intègre l'erreur et l'approximation comme règles afin de pouvoir rectifier n'importe quand dans le processus », là où un modèle performatif chercherait à les éliminer. (Perrenoud, 1991)

En intégrant l'erreur comme levier plutôt que comme faute à sanctionner, elle développe chez l'étudiant la tolérance à l'imperfection, la capacité de régulation de son propre apprentissage et une relation non anxieuse au sentiment d'inefficacité.

INEFFICACITÉ

REDONDANCE · Constance des critères et de l'instrument

La « répliquabilité ou stabilité (fidélité) d'une mesure est sa stabilité dans le temps ou entre corrections » : plus une évaluation est répétée sur le même objet, plus le jugement qu'elle produit gagne en fiabilité (Leclercq, 2006).

La redondance opère cependant à une autre échelle, celle du temps long et de la stabilité du cadre. Les grilles d'évaluation descriptives générales évaluent des dimensions transversales et non des tâches spécifiques. Elles donnent à l'étudiant un référentiel qui ne change pas d'un cours à l'autre et qui servent de support au dialogue avec l'enseignant.

Le *Progress Test* mis en œuvre à Maastricht repose sur ce principe. Tous les étudiants, du bachelier au master, passent à intervalles réguliers des tests équivalents calibrés sur le niveau de fin de cursus, avec des seuils de réussite différenciés selon l'année. La stabilité de l'instrument rend la progression visible à l'étudiant lui-même (Schuwirth & Van der Vleuten, 2012).

La redondance ne sert pas à accumuler des notes. Elle fait du cadre évaluatif un repère stable : c'est la constance des critères et de l'instrument dans la durée qui rend la progression de l'étudiant lisible pour lui et discutabile avec l'enseignant.

REDONDANCE

INCERTITUDE · Évaluation écologique

L'évaluation écologique engage une approche holistique pour tenter d'appréhender les conditions de milieu et de sens. Elle ne cherche pas à standardiser les situations d'évaluation, mais à prendre en compte les « caractéristiques et dynamiques propres à chaque milieu » (p.8) ainsi que ses dimensions disciplinaires, sociales et technologiques (Younes, 2020)

Renoncer à la standardisation, c'est aussi renoncer à l'idée qu'une seule réponse soit attendue. Certains dispositifs poussent ce principe jusqu'au cœur de l'évaluation. Le Test de concordance de script est conçu pour évaluer le raisonnement face à des problèmes mal définis, où les médecins expérimentés divergent considérablement dans leur cheminement vers la solution (Charlin et al., 2002). Le score ne mesure pas la conformité à une bonne réponse, mais le degré de similitude entre le raisonnement de l'étudiant et celui d'un panel d'experts dont chacune des réponses « reflète une opinion valide qui devrait être prise en compte » (Charlin et al., 2002, p. 139).

En confrontant l'étudiant à des contextes complexes, situés et non standardisés, elle développe la capacité à mobiliser ses ressources face à l'inconnu face à un monde dont les fluctuations ne peuvent pas être anticipées.

INCERTITUDE

INACHÈVEMENT · Évancipation

L'évaluation est traditionnellement un acte de clôture : elle fige un jugement dans une norme stable, à reproduire et à atteindre. À ce titre, l'inachèvement reste peu présent dans les pratiques évaluatives.

L'évancipation - néologisme forgé à partir d'*émancipation* et d'*évaluation* - s'inscrit dans un dialogue mouvant et non immuable autour de l'évaluation (Marcel & Gremion, 2025). Elle refuse de figer les critères du jugement dans une norme immuable et engage une démarche en mouvement permanent, inspirée de la boucle récursive de Morin : remettre en cause l'ordre existant pour s'assurer qu'il reste en phase avec le système, dans une perpétuelle remise en question.

Cette posture implique une tension : contrôle ou soutien au développement, externe ou interne, subir ou décider. Elle vise in fine à développer une « capacité d'*empowerment* » — apprendre à se conformer tout en restant critique, à désobéir en gardant la posture la plus éthique possible, à développer un éthos professionnel (Marcel & Gremion, 2025).

L'évancipation développe chez l'étudiant la capacité à habiter un référentiel en mouvement plutôt qu'à atteindre un état définitif de conformité et à se reconnaître comme acteur ayant un choix à faire face aux injonctions paradoxales.

INACHÈVEMENT

HÉTÉROGÉNÉITÉ · Multiplication des modalités d'évaluation

Aucune méthode d'évaluation unique ne peut mesurer tous les niveaux de compétence (Mirbahai & Adie, 2020). Chaque méthode possède en effet ses forces et ses faiblesses propres (Schuwirth & Van der Vleuten, 2018).

C'est pourquoi la multiplication des modalités s'impose : les défauts des unes étant comblés par les qualités des autres, elle permet de couvrir des niveaux taxonomiques variés à savoir des savoirs déclaratifs aux compétences complexes sans réduire le jugement évaluatif à un seul format (Detroz, Malay & Crahay, 2020)

En refusant de réduire le jugement à un format unique, la multiplication des modalités reconnaît la pluralité des manières d'apprendre et de démontrer ses compétences. Elle développe chez l'étudiant une capacité à se mobiliser dans des registres variés.

HÉTÉROGÉNÉITÉ

INCOHÉRENCE · Évaluation par les pairs et autoévaluation

L'autoévaluation est définie comme « *un processus par lequel un sujet est amené à porter un jugement sur la qualité de son cheminement, de son travail ou de ses acquis en regard d'objectifs prédéfinis* » (Legendre, 2005), tandis que l'évaluation par les pairs consiste en une « *interaction entre des individus appelés à évaluer la quantité, la valeur, la qualité et le succès des productions ou de l'apprentissage de leurs pairs* » (Topping, 1998, cité par Roy & Michaud, 2018, p. 55)

Ces dispositifs produisent des effets empiriquement documentés. En effet, ils diminuent l'anxiété des étudiants, développent leur autonomie et leur esprit critique, favorisent la réflexion et pour 84 % des étudiants dans une étude de référence, la démarche a été jugée bénéfique car elle les a rendus plus critiques et a favorisé une meilleure réflexion » (Orsmond et al., 2000, cités dans Roy & Michaud, 2018, p. 56)

L'évaluation par les pairs et l'autoévaluation introduisent délibérément des regards multiples et parfois divergents sur un même travail. Cette multiplicité de jugements, potentiellement contradictoires, est précisément ce qui génère la régulation la plus riche. Elle développe chez l'étudiant la capacité à naviguer dans l'ambiguïté et à construire son propre jugement

Incohérence

CONTRAT SOCIAL

04 | LE CONTRAT SOCIAL DE L'UNIVERSITÉ : UNE RÉALITÉ

TENSIONS : ENTRE LOGIQUE DE CERTIFICATION ET COMPÉTENCES COMPLEXES

L'université est tenue par un contrat social qui lui impose d'attester du niveau de maîtrise atteint par ses diplômés (Allal, 2013 ; Romainville, 2013). Cette exigence certificative repose sur la fiabilité et la validité des dispositifs d'évaluation, deux qualités que les outils standardisés garantissent plus aisément.

Or, les objectifs de formation universitaire vont bien au-delà de ce que ces dispositifs peuvent attester. C'est là que réside la tension : les tâches ouvertes et complexes donc celles qui développent et attestent les compétences universitaires réelles résistent à la standardisation et donc à la fidélité et à la comparabilité des mesures. Il devient alors pertinent de trouver un équilibre entre la rigueur de la mesure et la capacité des dispositifs à certifier l'acquisition de compétences complexes.

La robustesse ne remet donc pas en cause la certification. Elle questionne les formes qu'elle prend, en particulier lorsqu'elle se réduit à une logique de suroptimisation du jugement.

3 PÔLES POUR FAIRE DISCUSSION (*Detroz, 2026*)

Pour évaluer les forces et les limites de tout dispositif, trois pôles et sept continuums permettent d'organiser la réflexion :

Pôle A — Intégrité : garantir la note

1. Qualité de la mesure : validité, fidélité, comparabilité
2. Résilience IA : capacité à attribuer formellement la compétence à l'étudiant, y compris dans un contexte où les tâches peuvent être automatisées

Pôle B — Sens pédagogique : garantir le sens

3. Complexité des acquis : étendue et profondeur des compétences évaluées
4. Valeur d'apprentissage : capacité du dispositif à agir comme levier réflexif

Pôle C — Soutenabilité : garantir le maintien

5. Soutenabilité enseignant : temps de conception, de correction, charge cognitive
6. Soutenabilité étudiant : charge de travail, effets cumulatifs
7. Soutenabilité logistique : infrastructure, locaux, gestion des flux

4 PRINCIPES POUR RÉDUIRE LA TENSION (*Detroz, 2026*)

Pour tenter de réduire cette tension, une réflexion articulée autour de quatre principes peut guider la conception des dispositifs d'évaluation :

- ① **Double mouvement — validité et complexité** Assurer la rigueur de la mesure tout en garantissant une complexité suffisante des tâches.
- ② **Faire en fonction des enjeux** Tous les cours ne nécessitent pas le même niveau d'instrumentalisation certificative. La question est de distinguer ce qui exige une grande rigueur de la mesure standardisée de ce qui peut s'en affranchir — en fonction des objectifs de chaque cours au regard du programme de formation.
- ③ **Variation structurée des modalités** Organiser une diversité de dispositifs aux propriétés complémentaires, de manière à ce que les forces des uns compensent les limites des autres car aucune méthode d'évaluation unique ne peut mesurer tous les niveaux de compétence.
- ④ **Évaluation à l'échelle du programme** Organiser l'évaluation au-delà du cours individuel, en répartissant les fonctions et les exigences évaluatives dans le temps et entre les différentes unités d'enseignement.

01. La robustesse et la certification sont-elles foncièrement incompatibles ou la tension entre elles est-elle précisément ce qui peut faire avancer nos pratiques évaluatives ?

02. Si la robustesse suppose de renoncer à l'optimisation du jugement, doit-on questionner la note pure comme horizon indépassable de la certification universitaire ?

03. Modifier le paradigme évaluatif chez les enseignants et au sein des institutions conduira-t-il à un autre type d'engagement des étudiants face à l'évaluation et donc face à leurs apprentissages ?

- Braiek, H. B., & Khomh, F. (2024). *Machine Learning Robustness : A Primer* (arXiv:2404.00897). arXiv. <https://doi.org/10.48550/arXiv.2404.00897>
- Bonniol, J.-J., & Vial, M. (1997). *Les modèles de l'évaluation : textes fondateurs avec commentaires*. Bruxelles : De Boeck Université.
- Boucher, C. L., Pentecouteau, H., & Lameul, G. (s. d.). *Développer la réflexivité en formation avec un eportfolio de formation*.
- Charlin, B., Gagnon, R., Sibert, L., & Van der Vleuten, C. (2002). Le test de concordance de script, un instrument d'évaluation du raisonnement clinique. *Pédagogie Médicale*, 3(3), 135-144. <https://doi.org/10.1051/pmed:2002022>
- Detroz, P, Malay, L, & Crahay, V. (2020). Une démarche structurée pour définir quelques conseils en vue de limiter l'impact de la pandémie sur l'évaluation de nos étudiants, Numéro Hors-série, 1, 97-110.
- Fischer, L., Girardet, C., Lopez, L. M., Romainville, M., & Tessaro, W. (2020). Le portfolio dans l'enseignement supérieur : Perceptions des étudiants et intentions des enseignants. *e-JIREF*, 6(1), 3-30. <https://doi.org/10.48782/ew2yw095>
- Graveleau, S. (2025, septembre 2). « Avec ChatGPT, j'ai parfois l'impression de ne plus savoir apprendre... » : Comment motiver les étudiants à l'heure de l'IA ? https://www.lemonde.fr/campus/article/2025/09/02/avec-chatgpt-j-ai-parfois-l-impression-de-ne-plus-savoir-apprendre-comment-motiver-les-etudiants-a-l-heure-de-l-ia_6638099_4401467.html
- Hamant, O. (2022). *La troisième voie du vivant*. Éditions Odile Jacob.
- Hamant, O. (2024). *De l'incohérence*. Éditions Odile Jacob.
- Hansen, L. P., & Sargent, T. J. (2008). *Robustness* (STU-Student edition). Princeton University Press. <http://www.istor.org/stable/j.ctt1dr35gx>
- Kozanitis, A. (2021). Évaluation des apprentissages en formation à distance : Les situations authentiques à la rescousse. *Revue internationale des technologies en pédagogie universitaire / International Journal of Technologies in Higher Education*, 18(1), 276-290. <https://doi.org/10.18162/ritpu-2021-v18n1-24>
- Lawrence, M., Homer-Dixon, T., Janzwood, S., Rockström, J., Renn, O., & Donges, J. F. (2024). Global polycrisis : The causal mechanisms of crisis entanglement. *Global Sustainability*, 7, e6. <https://doi.org/10.1017/sus.2024.1>
- Leclercq, D. (2006). L'évolution des QCM. In G. Figari & L. Mottier-Lopez (dir.), *Recherches sur l'évaluation en éducation*. Paris : L'Harmattan, pp. 139-146.
- Legendre, R. (2005). *Dictionnaire actuel de l'éducation* (3e éd.). Montréal, QC : Guérin.
- Liu, H., & Renn, O. (2025). Polycrisis and Systemic Risk : Assessment, Governance, and Communication. *International Journal of Disaster Risk Science*, 16(4), 526-549. <https://doi.org/10.1007/s13753-025-00636-3>
- Marcel, J.-F. et Gremion, C. (dir.) (2025). *Évancipation dans l'institution. Oser le rapprochement entre émancipation et évaluation*. Cépaduès ÉDITIONS.

Mirbahai L, Adie JW (2020) Applying the utility index to review single best answer questions in medical education assessment. *Arch Epid Pub Health* 2: DOI: 10.15761/AEPH.1000113.

Morin, E. (avec Internet Archive). (1999). *Homeland earth : A manifesto for the new millenium*. Cresskill, N.J. : Hampton Press. <http://archive.org/details/homelandearthman0000mori>

Paquelin, D., Tremblay, C., Hains, Y., Lison, C., Cayouette-Guilloteau, V., Kropf, V., LaBillois, D., Philion, R., Meyer, F., Gravelle, F., & Blanchet Garneau, A. (2025). De l'habitude au risque, de l'individu au collectif : Dynamique transformative des pratiques pédagogiques. *Revue internationale de pédagogie de l'enseignement supérieur*, 41(2). <https://doi.org/10.4000/14kwc>

Perrenoud, P. (1991). Pour une approche pragmatique de l'évaluation formative. *Mesure et évaluation en éducation*, 13(4), 49-81.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249-255.

Roy, M., & Michaud, N. (2018). L'autoévaluation et l'évaluation par les pairs en enseignement supérieur : promesses et défis. *Formation et profession (Sainte-Foy)*, 26(2), 54. <https://doi.org/10.18162/fp.2018.458>

Schuwirth, L. W. T., & Van der Vleuten, C. P. M. (2012). The use of progress testing. *Perspectives on Medical Education*, 1(1), 24-30. <https://doi.org/10.1007/s40037-012-0007-2>

Schuwirth, L. W. T., & van der Vleuten, C. P. M. (2019). How 'Testing' Has Become 'Programmatic Assessment for Learning'. *Health Professions Education*, 5(3), 177-184. <https://doi.org/10.1016/j.hpe.2018.06.005>

Stelling, J., Sauer, U., Szallasi, Z., Doyle, F. J., & Doyle, J. (2004). [Rev. of *Robustness of Cellular Functions*]. *Cell*, 118(6), 675–685. <https://doi.org/10.1016/j.cell.2004.09.008>

Tomlinson, K., Jaffe, S., Wang, W., Counts, S., & Suri, S. (2025). *Working with AI : Measuring the Applicability of Generative AI to Occupations*. <https://www.microsoft.com/en-us/research/publication/working-with-ai-measuring-the-occupational-implications-of-generative-ai/>

Younes, N. (2020). *Vers une évaluation écologique dans l'enseignement supérieur : dispositifs et transformations en jeu*.

Zames, G. (1981). Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses. *IEEE Transactions on Automatic Control*, 26(2), 301–320. <https://doi.org/10.1109/TAC.1981.1102603>

MERCI POUR VOTRE ATTENTION

Larcin.C – ADMEE Suisse | 2026