



La compréhension des savoirs transmis est primordiale à leur assimilation

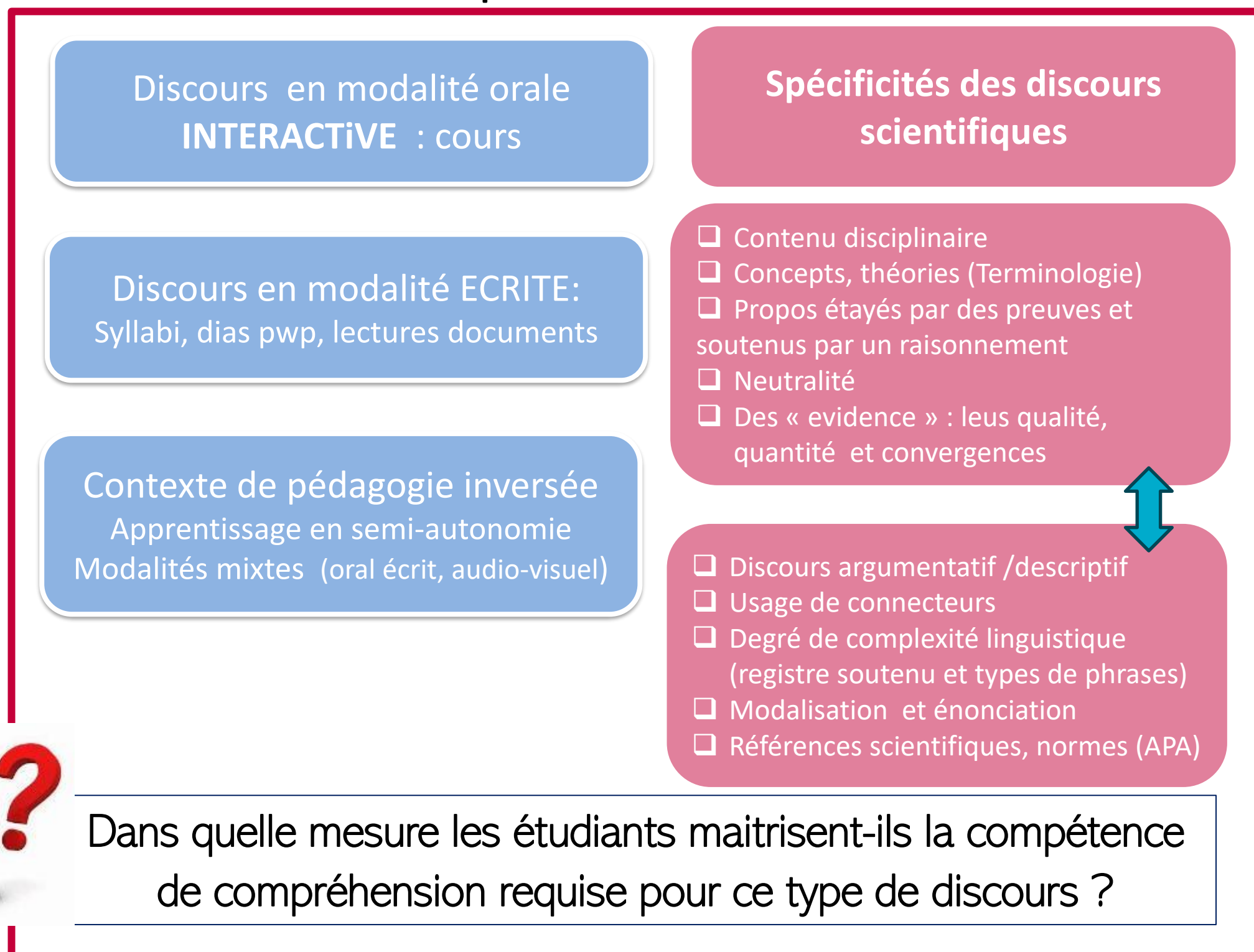
Constats issus de la littérature

Les évaluations linguistiques à l'aide de la technologie (TMLA) ont attiré de plus en plus d'attention au cours des dernières années (Barrot, 2023 ; Ockey & Neiriz, 2021), Des chercheurs ont souligné que l'*Automatic item generation* (AIG) est une technique qui présente plusieurs avantages par rapport aux méthodes d'évaluations traditionnelles. Or, si l'IA s'est montrée plutôt efficace dans divers domaines, son application ne semble pas des plus appropriée pour des tâches difficiles à modéliser comme l'est la **compréhension écrite** (Attali et al., 2022).

Notre question : CHAT GPT 4 peut-il fournir une assistance à la construction d'épreuves d'évaluation de compréhension écrite ; plus spécifiquement, est-il à même de générer des textes et des questions de niveau universitaire en respectant un certain nombre de critères de qualité (exactitude du contenu, aspects linguistiques et psychométriques) ? **Si oui, à quelles conditions ?**

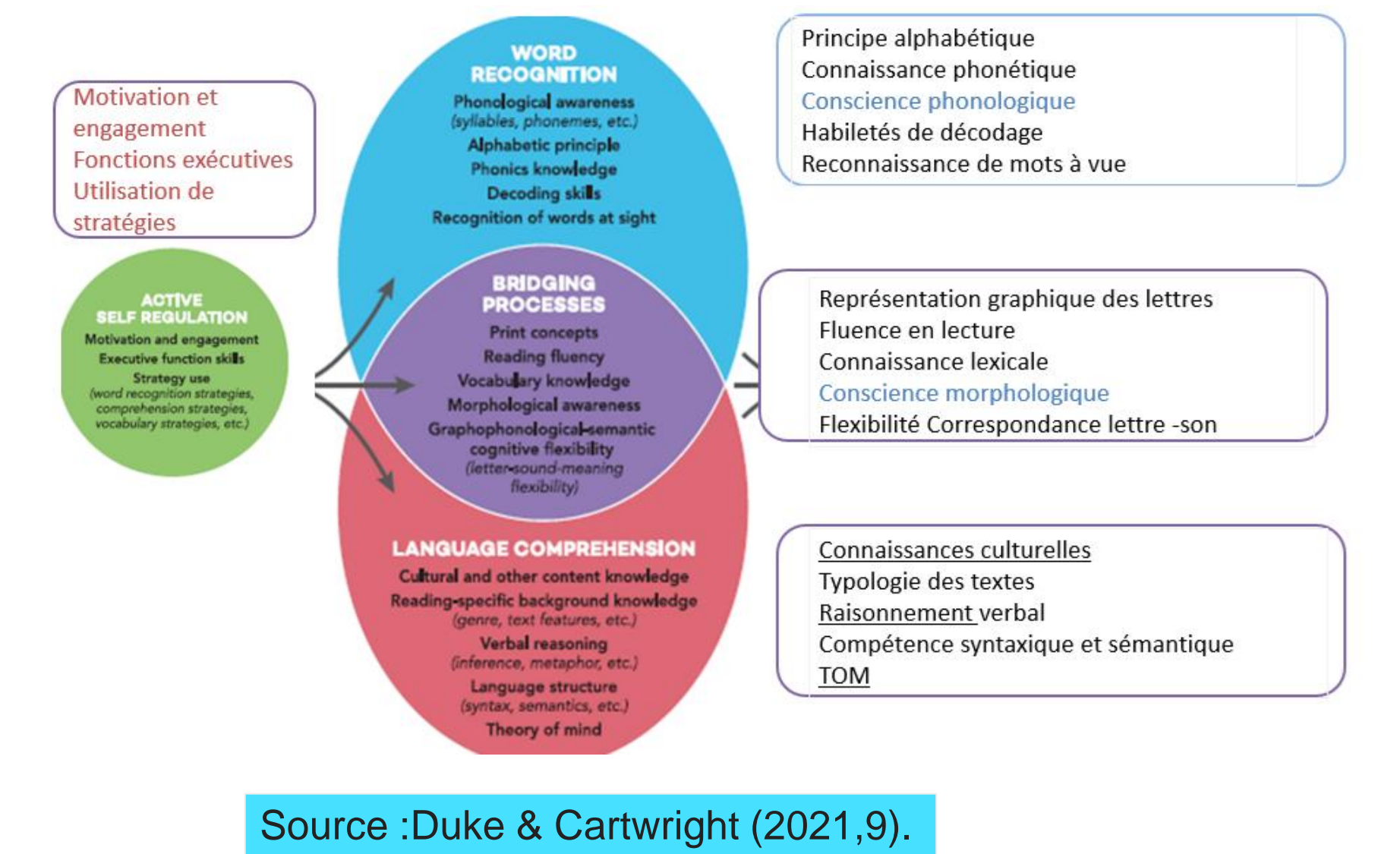
1. Pédagogie et littératie universitaires

1.1. Modalités discursives de transmission des savoirs scientifiques



1.2. Qu'est-ce que comprendre ?

La compréhension (orale et écrite) est une habileté reconnue comme complexe car mobilisant de nombreux savoirs et savoir-faire. Le modèle illustré ci-dessous (Duke & Cartwright, 2021) le montre à souhait.

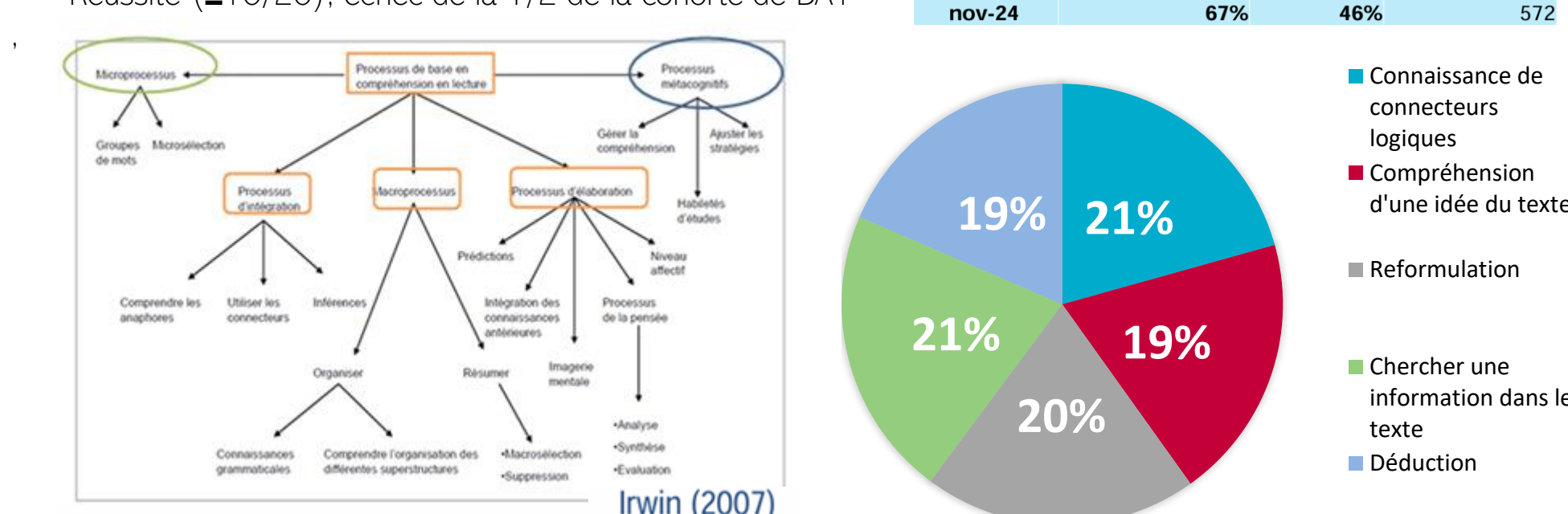


Source : Duke & Cartwright (2021,9).

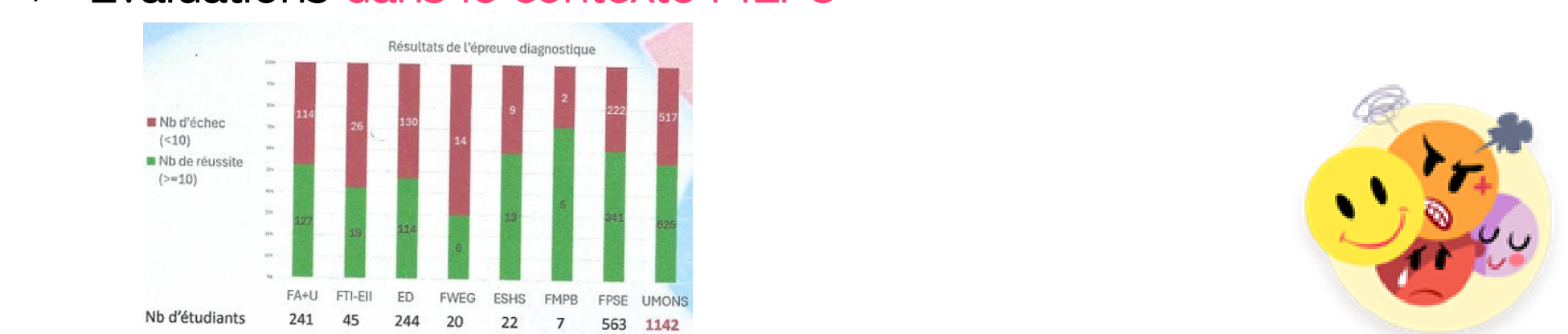
1.3. Evaluation des niveaux de compétence des primo-arrivants

❖ Evaluations dans le contexte des épreuves diagnostiques à la FPSE

- Epreuves « homemade » selon un cahier des charges inspiré des modèles (construct-driven approach, Embretson and Yang, 2006)
- Calibrage qui permet la construction d'épreuves parallèles,
- Trois épreuves élaborées chaque année académique
- Réussite (≥10/20) : échec de la 1/2 de la cohorte de BA1



❖ Evaluations dans le contexte MLFU



2. Génération de textes

Choix du texte initial

- 4 Sources :
- 1. Ouvrage didactique de niveau universitaire
- 2. Article scientifique
- 3. Rapport d'enquête
- 4. Site médical sur internet

- Profilage et calibrage du support textuel (template-based method, Kurdi et al., 2020)
- Identification des éléments à évaluer
- Entraînement sur des bases textuelles diversifiées
- Evaluations adaptatives ou parallèles

Objectif : Opérer une transformation du texte initial en vue de sa conformité aux standards rédactionnels de l'écrit académique.(cf . 1.1) et de son adéquation au public

Processus itératif de requêtes afin de construire à chaque étape un prc permettant à CHAT GPT de :

- générer une analyse linguistique du texte support**
 - Registre soutenu (types de phrases et de termes)
 - Lexique scientifique
 - Structure discursive : (contre)-arguments, exemples
 - Modalisation et connecteurs variés...
- générer une analyse du texte support au niveau des critères de type scientifique**
 - Exemples :Présence de sources scientifiques sûres, couverture plus exhaustive des positions théoriques, etc
- dresser la liste, sur base des profil linguistique et scientifique dressés, des ajustements requis et de les réaliser**

- Exemples de support textuel (avant et après transformation)
 - « Amitié et santé : ce que disent les recherches »
 - « Le renoncement aux soins »
- Insertion de requêtes processuelles ou de vérification
 - À quoi as-tu été attentif quand j'ai demandé de simplifier/complexifier le texte ?
 - Peux-tu expliquer comment tu as fait pour identifier les références scientifiques ? Quelles sources as-tu consultées?

Les explicitations du processus de traitement de ChatGPT4 deviennent critères analyse supplémentaires

3. Génération d'items d'évaluation (QCM)

Objectif : à partir d'un processus itératif de requêtes, en adoptant une « *computational psychometrics approach* » (Von Davier et al. 2021), mettre au point un éventail de Prompts afin :

- De préciser le Type de QCM souhaité
- D'investiguer la nature et nombre de sous-compétences à cibler (construct coverage)
- De vérifier la conformité aux principes de conception d'un bon QCM
- De parvenir à construire des items et distracteurs efficaces et fiables

Comparaison phase initiale vs phase postérieure

- Texte : les rêves
- Extrait du texte : Sigmund Freud (1900/1953), père de la psychanalyse, postule que les rêves constituent un exutoire aux désirs inconscients refoulés

Phase initiale avec un Prompt basique:

Selon Sigmund Freud, quelle est la fonction principale des rêves ?

- Résoudre des problèmes mathématiques
- Compter les moutons
- Servir d'exutoire aux désirs inconscients refoulés
- Améliorer la mémoire

Phase initiale avec un Prompt basique:

Texte : les rêves

Quelle est la conclusion principale du texte concernant les théories des rêves ?

- Les théories psychodynamiques sont les seules valides
- Les théories neurobiologiques sont les seules valides
- Les théories psychodynamiques et neurobiologiques pourraient être complémentaires
- Les rêves sont à interpréter

Mutuellement exclusives

Phase postérieure après processus itératif

QCM 2 – Compétence : Comprendre un détail explicitement énoncé

Selon Freud, à quoi servent les rêves ?

- À évacuer les souvenirs du jour
- À s'adapter aux changements environnementaux
- À exprimer des désirs inconscients refoulés
- À déclencher l'activation du cortex visuel

Phase postérieure après processus itératif

QCM 6 – Compétence : Résumer l'idée principale d'un passage

Quel est le message principal de la section "Une convergence des hypothèses" ?

- Les rêves ont une origine exclusivement neurologique
- Les théories psychanalytiques sont dépassées
- Une approche intégrative permet de mieux comprendre les rêves
- Les rêves n'ont pas de fonction précise

4. Comment rendre ChatGPT plus intelligent ?

Prompt I : Que ne permet pas de déduire cet extrait : « le circuit de récompense entretenu par cette hormone du plaisir [la dopamine] est aussi ce que recherche à produire la nouvelle économie capitaliste numérique avec ses algorithmes : son objectif de croissance la pousse en effet à contrôler nos usages d'Internet et des réseaux sociaux pour en extraire nos données de consommation »

a. GpT : Voici ce que l'extrait ne permet pas de déduire, c'est-à-dire ce qu'il n'énonce pas explicitement et ne laisse pas raisonnablement supposer

b. Formule une question de compréhension type QCM en insérant des distracteurs à partir de fausses déductions

c. Distracteurs (fausses déductions) :

- Faux : le texte n'indique pas si cette exploitation du circuit de la dopamine est consciente ou scientifiquement fondée.
- Faux : il n'y a aucune mention de légalité ou de cadre juridique dans le passage.
- Faux : le texte parle de contrôle, mais ne nie pas toute forme de libre arbitre.
- Faux et absurde : il ne s'agit pas d'un acte médical ou chimique, mais d'un processus psychologique déclenché par l'expérience utilisateur.

Prompt II : As-tu bien pris en compte les principes de sélection des distracteurs ?



- GpT : Voici une relecture des distracteurs pour les QCM que je t'ai proposés, avec un focus sur la qualité et la pertinence des distracteurs selon les 4 principes
- Procède à une relecture critique de l'ensemble des distracteurs pour renforcer encore ces principes

QCM 1 – Sens d'un mot (paréidolie)	
Bonne réponse : C. Une tendance à percevoir du sens dans des formes vagues	
Option	Analyse
A. Pour irrationnelle liée au sommeil	Non plausible : trop éloigné du contexte, pas une erreur courante.
B. Illusion due à un trouble visuel	Plausible et proche (proximité conceptuelle), confusion possible avec hallucination.
C. Bonne réponse	-
D. Méthode d'interprétation des rêves	Plausible pour un étudiant qui confond paréidolie avec symbolisme onirique.
★ À ajuster : remplacer A par une option plus subtile, ex. « Une confusion entre perception et réalité sensorielle ».	

- Horizons captivants** : efficacité pour la génération de textes, réactivité ; rapidité ; suggestions spontanées (classement des questions selon la taxonomie de Bloom, version élève, correctif avec explication de la bonne et mauvais réponse, grille de correction, fichier pour implémentation sur Moodle) ; correction et FB automatisé ; diversité des compétences évaluées et des formats textuels
- Pavés glissants** : fiabilité psychométrique des distracteurs ; difficulté à évaluer la nuance et l'implicite ; jugement insuffisant de la pertinence pédagogique

L'IA offre certes des avancées majeures pour l'évaluation des compétences de compréhension ; toutefois, elle doit encore progresser et être encadrée et complétée par l'expertise humaine.