

CHAIRE DE DYNAMIQUE SOCIALE (C.D.S)

MOUVEMENTS & ENJEUX SOCIAUX

Revue Internationale
des Dynamiques Sociales



WWW.MESRIDS.ORG
JUILLET - AOÛT 2026
KINSHASA-R.D.C

LITTERATIE EN INTELLIGENCE ARTIFICIELLE CHEZ LES ETUDIANTS DE L'UNIVERSITE DE KINSHASA : *structure de mesure, profil de compétences et facteurs associés*

Par

Teddy BOFULA LOBEBE

Université de Kinshasa,
République démocratique du Congo

Résumé

La présente étude poursuit trois objectifs : examiner les qualités psychométriques d'une échelle de littératie en IA à cinq dimensions en contexte congolais, décrire le profil de compétences déclarées des étudiants, et identifier les facteurs qui y sont associés. Pour atteindre ces objectifs, un questionnaire a été administré à un échantillon non probabiliste de 825 étudiants de l'Université de Kinshasa.

Le score global de onze items de l'échelle de littératie en IA est fiable ($\alpha = 0,83$; $\omega = 0,83$), mais les cinq scores dimensionnels, qui reposent sur deux items sauf le premier, ont une cohérence interne faible (0,36 à 0,70) et l'analyse parallèle n'extraît que deux facteurs : un facteur de capacité d'application et un facteur regroupant les huit items de jugement. L'échelle se distingue de l'auto-efficacité académique (rapport hétérotrait-monotrait = 0,64) mais non de la perception de l'IA, composée des items de facilité et d'utilité perçues (0,90), ce qui montre qu'un score incorporant ces croyances serait dominé par une disposition évaluative indifférenciée. L'invariance configurale est retenue selon le genre mais non selon le champ disciplinaire. Le niveau global dépasse de peu le point milieu de l'échelle ($M = 2,57$; IC 95 % [2,52 ; 2,61]), et le profil n'est pas graduel : la capacité d'application est la seule dimension située sous le point milieu ($M = 2,36$; 54,5 % des étudiants en dessous) et elle se détache des quatre autres, qui s'échelonnent sur un intervalle de 0,10 point. L'écart intra-individuel entre l'auto-efficacité déclarée à l'égard de l'IA et la capacité d'application atteint 0,30 point ($d_z = 0,33$) et concerne 56,1 % des étudiants. Le champ disciplinaire est le seul facteur sociodémographique d'ampleur appréciable ($\omega^2 = 0,064$), les sciences sociales et humaines se situant en deçà des deux autres champs (d de 0,50 à 0,58), avec un écart maximal sur la capacité d'application ($d = 0,72$). Le genre et l'âge n'ont pas d'effet détectable, et la littératie déclarée est plus élevée à l'entrée dans le cursus qu'aux niveaux supérieurs. Enfin, la relation entre usages et littératie n'est pas linéaire : les étudiants d'usage intermédiaire déclarent la compétence la plus faible, en deçà des utilisateurs occasionnels comme des utilisateurs intensifs, résultat qui subsiste après contrôle des variables sociodémographiques. Les indicateurs d'usage n'ajoutent que 0,9 % à la variance du score global, contre 25,4 % pour l'auto-efficacité académique générale.

Ces résultats invitent à distinguer la compétence déclarée de la compétence démontrée, à cibler l'apprentissage par la pratique disciplinaire plutôt que

l'exposition aux outils, et à compléter les échelles auto-rapportées par des épreuves de performance.

Mots-clés : *littératie en intelligence artificielle ; validation d'échelle ; auto-évaluation des compétences ; enseignement supérieur.*

Abstract

This study pursues three aims: to examine the psychometric properties of a five-dimension AI literacy scale in the Congolese context, to describe students' self-reported competence profile, and to identify associated factors. A questionnaire was administered to a non-probability sample of 825 students at the University of Kinshasa. Analyses use continuous scores, robust estimation, effect sizes with confidence intervals, and false-discovery-rate correction.

The eleven-item global score is reliable ($\alpha = .83$; $\omega = .83$), but the five dimension scores, based on two items each except the first, have weak internal consistency (.36 to .70) and parallel analysis extracts only two factors: an application-ability factor and a factor grouping the eight judgement items. The scale discriminates from academic self-efficacy (HTMT = .64) but not from AI perception, made up of the perceived-ease and perceived-usefulness items (.90), showing that a score incorporating these beliefs would be dominated by an undifferentiated evaluative disposition. Configural invariance holds across gender but not across academic field. The global level barely exceeds the scale midpoint ($M = 2.57$, 95 % CI [2.52, 2.61]), and the profile is not graded: application ability is the only dimension below the midpoint ($M = 2.36$; 54.5 % of students below it) and it separates from the other four, which span a 0.10-point range. The within-person gap between self-reported AI self-efficacy and application ability reaches 0.30 points ($d_z = 0.33$) and concerns 56.1 % of students. Academic discipline is the only sociodemographic factor of appreciable magnitude ($\omega^2 = .064$), with social sciences and humanities below the other two fields ($d = 0.50$ to 0.58) and the largest gap on application ability ($d = 0.72$). Gender and age show no detectable effect, and self-reported literacy is higher at entry than at later stages of the curriculum. Finally, the usage-literacy relation is not linear: intermediate users report the lowest competence, below both occasional and intensive users, a result that survives adjustment for sociodemographic covariates. Usage indicators add only 0.9 % to the variance of the global score, against 25.4 % for general academic self-efficacy.

These findings call for distinguishing self-reported from demonstrated competence, for targeting discipline-embedded practice rather than mere exposure to tools, and for complementing self-report scales with performance-based tests.

Keywords: *artificial intelligence literacy; scale validation; self-assessment of competence and higher education.*

INTRODUCTION

L'essor récent de l'intégration des outils numériques dans l'enseignement supérieur transforme non seulement les modalités de transmission de savoir,

mais également les pratiques d'apprentissage, d'évaluation et de production académique. Cette transformation s'est accélérée avec l'arrivée des outils d'IA génératifs tels que ChatGPT, qui offrent aux étudiants des possibilités peu coûteuses et immédiates de réaliser leurs travaux académiques, de rechercher des informations ou encore de résoudre des problèmes complexes. L'adoption rapide de ces outils par les étudiants africains ne s'accompagne pas nécessairement d'un encadrement et d'une compréhension suffisante des biais, des limites ainsi que des implications éthiques (African Union, 2024), alors que la politique nationale du numérique fait de la formation aux technologies un axe explicite (République démocratique du Congo, 2021).

En effet, ces usages massifs des outils d'intelligence artificielle sont susceptibles de masquer une faible maîtrise des principes fondamentaux qui sous-tendent ces technologies. Selon Long et Magerko (2020), une utilisation efficace de l'IA requiert non seulement des compétences opérationnelles, mais également une capacité d'analyse critique des résultats produits par ces systèmes. Dans ce même ordre d'idées, les travaux de synthèse convergent pour faire de la littératie en IA une compétence transversale de l'enseignement supérieur (Ng et al., 2021 ; Laupichler et al., 2022), et les cadres institutionnels appellent à une éducation à l'IA fondée sur des principes éthiques, afin de favoriser une utilisation responsable et inclusive de ces technologies dans les systèmes éducatifs (UNESCO, 2021, 2024 ; OECD, 2021).

Malgré la multiplicité récente des instruments d'évaluation, certaines dimensions psychométriques et la structure multidimensionnelle de la littératie en IA restent encore insuffisamment documentées et peu stabilisées, notamment en ce qui concerne l'invariance entre groupes et la validité interculturelle (Lintner, 2024). En outre, les travaux évaluant la littératie en IA recourent majoritairement à des mesures auto-rapportées, alors que les instruments fondés sur des performances objectives demeurent minoritaires. Or la littérature disponible sur l'auto-évaluation des compétences établit que la corrélation entre compétence perçue et compétence démontrée est modeste et qu'elle est la plus faible chez les moins compétents (Kruger et Dunning, 1999 ; Dunning, 2011 ; Zell et Krizan, 2014). Ces limites sont particulièrement importantes dans le contexte des universités africaines, où les instruments développés ailleurs sont encore peu examinés.

La littérature empirique disponible renseigne sur deux limites qui motivent la présente étude.

La première est conceptuelle. La littératie en IA est fréquemment mesurée par des instruments qui agrègent, sous une même étiquette, des compétences et des croyances d'acceptation. La facilité d'utilisation perçue et l'utilité perçue, issues du modèle d'acceptation des technologies (Davis, 1989), sont ainsi parfois traitées comme des dimensions de la littératie. Or ces deux

construits ne mesurent pas ce qu'une personne sait faire : ils mesurent ce qu'elle anticipe du système. Les traiter comme des composantes de la compétence gonfle mécaniquement le niveau estimé de littératie, puisque les jugements d'utilité sont massivement favorables dans les populations étudiantes, et rend le score global non interprétable comme indicateur de maîtrise.

La seconde limite est méthodologique. Les études recourent presque exclusivement à des mesures auto-rapportées, sans contrôle de la calibration des jugements. Or la recherche sur l'auto-évaluation des compétences établit que la corrélation entre compétence perçue et compétence démontrée est modeste et qu'elle est la plus faible chez les moins compétents (Kruger et Dunning, 1999 ; Dunning, 2011 ; Zell et Krizan, 2014). Une échelle auto-rapportée de littératie en IA mesure donc une compétence déclarée, dont la relation à la compétence effective reste à établir. A ce sujet, la conséquence pratique est directe : un score global élevé ne suffit pas à conclure à une maîtrise, et la structure interne du profil déclaré devient l'information la plus utile.

En République démocratique du Congo, les travaux disponibles décrivent l'adoption des outils numériques et de l'IA, notamment pour la recherche d'information et la rédaction académique (Bofula Lobebe et al., 2023 ; Target Sarl, 2025), sans documenter les compétences sous-jacentes. À notre connaissance, aucune étude empirique n'a porté sur la littératie en IA des étudiants de l'Université de Kinshasa.

La présente étude poursuit trois objectifs et leur associe cinq questions de recherche. Le premier objectif est de mesurer : examiner la structure factorielle, la fiabilité et l'invariance de l'échelle utilisée (Q1). Le deuxième est de décrire : établir quelles dimensions sont les mieux et les moins maîtrisées, et quantifier l'écart entre confiance déclarée et capacité d'application déclarée (Q2). Le troisième est d'expliquer : estimer les différences selon le genre, le champ disciplinaire et la promotion d'études (Q3), la forme de la relation entre indicateurs d'usage et littératie (Q4), et la part de variance que ces facteurs expliquent conjointement (Q5).

2. CADRE THEORIQUE

2.1. La littératie en IA comme compétence, et ce qu'elle n'est pas

Long et Magerko (2020) définissent la littératie en IA comme un ensemble de compétences permettant à un individu d'évaluer de manière critique les technologies d'IA, de communiquer et de collaborer efficacement avec elles, et de les utiliser comme outils dans la vie quotidienne et professionnelle. Ng et al. (2021) en proposent une organisation en quatre facettes : connaître et comprendre l'IA, l'utiliser et l'appliquer, l'évaluer et créer avec elle, et enfin

en saisir les enjeux éthiques. Le référentiel de l'UNESCO (2024) pour les élèves reprend cette architecture en y adjoignant explicitement une dimension de conception centrée sur l'humain.

Ces cadres partagent une caractéristique décisive pour la mesure : ils décrivent des capacités d'action, non des dispositions favorables. C'est ce qui les distingue du modèle d'acceptation des technologies. Chez Davis (1989), la facilité d'utilisation perçue et l'utilité perçue sont des croyances antécédentes de l'intention d'usage ; elles expliquent pourquoi une personne adopte un système, et non ce qu'elle est capable d'en faire. Un étudiant peut juger un assistant conversationnel très utile et très simple, tout en étant incapable d'évaluer la fiabilité de ses productions. Les deux familles de construits ne sont donc pas des composantes d'un même ensemble : les croyances d'acceptation sont, au mieux, des corrélats ou des antécédents de la littératie. Cette distinction est ici testable, car le questionnaire comporte une échelle distincte de perception de l'IA formée d'items de facilité et d'utilité perçues ; la section 4.1 montre que cette échelle ne se sépare pas empiriquement de celle de littératie, ce qui donne à l'argument conceptuel une conséquence de mesure directe.

Une distinction voisine s'impose entre littératie et auto-efficacité. L'auto-efficacité désigne la croyance d'un individu en sa capacité à réussir une tâche donnée (Bandura, 1986, 1997). Elle favorise l'engagement, la persévérance et l'exploration, mais elle ne garantit pas la maîtrise : c'est une croyance sur la compétence, non la compétence. Ce point est particulièrement sensible dans les échelles auto-rapportées de littératie, où les items de compétence prennent la forme d'énoncés d'accord et se rapprochent, dans leur formulation même, d'items d'auto-efficacité. Une échelle de littératie auto-rapportée est en pratique une mesure de compétence perçue, et la variance qu'elle partage avec les échelles générales d'auto-efficacité constitue une information à documenter plutôt qu'à ignorer (Podsakoff et al., 2003).

2.2. Les cinq dimensions mesurées

L'échelle employée ici est celle de Nong et al. (2024), construite pour évaluer la littératie en IA du grand public à partir de cinq dimensions : capacité d'application, capacité cognitive, moralité, esprit critique et auto-efficacité. Ces dimensions sont mesurées par onze items, trois pour la capacité d'application et deux pour chacune des autres. Cette architecture recoupe les facettes de Ng et al. (2021) : la capacité d'application correspond à la facette « utiliser et appliquer », la capacité cognitive à « connaître et comprendre », l'esprit critique à « évaluer », et la moralité à la facette éthique. La dimension d'auto-efficacité, qui porte sur la confiance de l'étudiant dans sa capacité à mobiliser ces outils, n'appartient pas au registre des compétences au sens strict, mais elle est conservée telle que l'instrument la définit, ce qui autorise

précisément la comparaison intra-individuelle entre confiance et capacité d'application déclarée.

Cette organisation permet une prédiction structurée. Les dimensions cognitive, critique et éthique portent sur des jugements généraux, formulables sans expérience approfondie des outils : un étudiant peut souscrire à l'idée qu'il faut vérifier les productions d'un système d'IA sans jamais avoir effectué cette vérification. La capacité d'application, en revanche, engage un rapport à des tâches concrètes et se prête moins à une adhésion de principe. On s'attend donc à ce que cette dimension se détache des autres par le bas.

H1. La capacité d'application obtient le score le plus faible du profil et se sépare significativement des autres dimensions, lesquelles ne se différencient pas nettement entre elles.

H2. Il existe un écart intra-individuel positif entre l'auto-efficacité déclarée à l'égard de l'IA et la capacité d'application déclarée.

2.3. Champ disciplinaire, avancement dans le cursus et genre

Les formations scientifiques et de santé intègrent davantage les outils numériques, l'analyse de données et les applications de l'IA, offrant des occasions plus nombreuses de développer des compétences instrumentales (Zawacki-Richter et al., 2019 ; Chan et Hu, 2023). Si la littératie se construit par l'exposition à des tâches disciplinaires, l'écart entre champs devrait être maximal sur la dimension la plus instrumentale.

H3. Le champ disciplinaire est associé au niveau de littératie, et l'écart entre champs est plus marqué sur la capacité d'application que sur les dimensions de jugement.

L'effet de l'avancement dans le cursus est théoriquement ambivalent. Si les compétences se développent avec l'expérience (Bandura, 1997), la littératie devrait croître avec la promotion d'études. Mais si elle dépend d'occasions d'apprentissage spécifiques plutôt que de l'ancienneté (Long et Magerko, 2020), une progression régulière n'est pas attendue en l'absence de formation structurée. Quant au genre, plusieurs travaux récents concluent à une réduction des écarts de compétences numériques dans l'enseignement supérieur (Laupichler et al., 2022 ; Chiu et al., 2023). Une conclusion d'absence d'effet exige toutefois davantage qu'un test non significatif : elle demande de vérifier que l'intervalle de confiance exclut les effets d'intérêt, ce qui relève des tests d'équivalence (Lakens, 2017).

2.4. Usages de l'IA : exposition ou pratique délibérée ?

Les référentiels de littératie postulent que la pratique régulière contribue au développement des connaissances, compétences et attitudes liées à l'IA (Ng et al., 2021 ; Carolus et al., 2023). Cette proposition est cependant plus

faible qu’il n’y paraît. La recherche sur l’acquisition de l’expertise établit que c’est la pratique structurée et rétroactive, non le temps d’exposition, qui produit la compétence (Ericsson et al., 1993). Appliquée à l’IA, la distinction sépare l’étudiant qui délègue une tâche à un assistant de celui qui contrôle, corrige et confronte la production obtenue.

Deux conséquences en découlent pour la mesure. D’une part, familiarité, fréquence et durée des sessions ne sont pas des indicateurs interchangeables d’un même construit d’« usage » : une durée de session longue peut aussi bien signaler un travail approfondi qu’une difficulté à obtenir un résultat exploitable. Les agréger en un score moyen suppose une cohérence interne qu’il faut vérifier avant, et non après, l’analyse (Bollen et Lennox, 1991 ; Diamantopoulos et Winklhofer, 2001). D’autre part, la relation entre usage et compétence déclarée n’a aucune raison d’être monotone. Si l’usage modifie la calibration du jugement autant que la compétence elle-même, les utilisateurs occasionnels, faute de points de comparaison, peuvent se déclarer aussi compétents que les utilisateurs experts, alors que les utilisateurs intermédiaires, confrontés aux limites des outils sans les maîtriser, se jugeraient plus sévèrement.

H4. Les trois indicateurs d’usage ne forment pas une échelle homogène et doivent être traités séparément.

H5. La relation entre intensité d’usage et littératie déclarée n’est pas linéaire.

3. METHODOLOGIE

3.1. Contexte et participants

L’étude a été réalisée à l’Université de Kinshasa (UNIKIN), qui regroupe treize facultés. Pour les analyses comparatives, celles-ci ont été regroupées en trois champs disciplinaires : sciences de la santé, sciences et technologie, sciences sociales et humaines. La population cible était constituée des étudiants régulièrement inscrits au cours de l’année académique 2025-2026. Un échantillon non probabiliste de 825 étudiants a été constitué (tableau 1) ; l’ensemble des questionnaires exploitables a été retenu, sans donnée manquante sur les variables analysées.

Tableau 1. Caractéristiques des participants (N = 825)

Caractéristique	Modalité	n	%
Genre	Masculin	457	55,4
	Féminin	368	44,6
Champ disciplinaire	Sciences de la santé	153	18,5
	Sciences et technologie	212	25,7
	Sciences sociales et humaines	460	55,8
Promotion d’études	L2 LMD	418	50,7
	L3 LMD	207	25,1
	Master 1	85	10,3

Caractéristique	Modalité	n	%
Tranche d'âge	Master 2	115	13,9
	Moins de 20 ans	328	39,8
	20-24 ans	427	51,8
	25-29 ans	70	8,5

3.2. Instruments

3.2.1. Littératie en IA.

L'échelle de littératie en IA utilisé dans cette étude est celle de Nong et al. (2024) comportant onze items répartis en cinq dimensions. Chaque item est coté de 1 (totalement en désaccord) à 4 (totalement en accord) ; le point milieu théorique de l'échelle vaut 2,5. Les scores de dimension sont les moyennes non arrondies des items correspondants, et le score global la moyenne des onze items. Le nombre d'items par dimension, réduit à deux pour quatre d'entre elles, borne d'emblée la fiabilité atteignable au niveau dimensionnel ; la présente étude en tire les conséquences analytiques.

3.2.2. Perception de l'IA.

Quatre items mesurent la facilité d'utilisation perçue (deux items) et l'utilité perçue (deux items), également cotés de 1 à 4. Ces items ne sont pas intégrés au score de littératie : ils constituent une échelle distincte ($\alpha = 0,680$), utilisée ici pour tester empiriquement la séparation entre compétence déclarée et croyances d'acceptation. Le construit de littératie évalué est celui des cinq dimensions ci-dessus, et lui seul.

3.2.3. Auto-efficacité académique générale.

Dix items adaptés de l'échelle d'auto-efficacité généralisée de Schwarzer et Jerusalem (1995), également cotés de 1 à 4, mesurent la persévérance, la gestion des tâches et la confiance en soi dans le travail académique. Elle est utilisée comme covariable, afin d'estimer la part de la variance du score de littératie attribuable à une disposition auto-évaluative générale.

3.2.4. Indicateurs d'usage de l'IA.

Trois questions à modalités ordinales décrivent le rapport aux outils : niveau de familiarité (faible, moyen, élevé), fréquence d'utilisation (faible, moyenne, élevée) et durée d'une session d'utilisation (moins de 30 minutes, 30 à 60 minutes, plus d'une heure). Ces indicateurs sont analysés séparément, et l'homogénéité de leur agrégation est testée explicitement (H4).

3.2.5. Variables sociodémographiques.

Quatre variables sociodémographiques, à savoir : Genre, champ disciplinaire, promotion d'études et tranche d'âge ont été pris en compte.

3.3. Préparation des données

Toutes les analyses rapportées dans cette étude portent sur les scores continus et des indicateurs de qualité de réponse ont été calculés sur

l'ensemble des items du questionnaire : écart-type intra-sujet, plus longue suite de réponses identiques et distance de Mahalanobis sur les scores dimensionnels. Trente-neuf participants (4,7 %) présentent un patron de réponse invariant ou une suite prolongée de réponses identiques. Ils sont conservés dans les analyses principales, et l'ensemble des résultats est répliqué sans eux (N = 786) (Section 4.6). La structure de mesure a été examinée sur l'échantillon complet, l'analyse exploratoire et l'analyse confirmatoire portant sur les mêmes observations.

3.4. Stratégie d'analyse

L'analyse factorielle exploratoire a été conduite après vérification de l'adéquation des données (indice KMO, test de sphéricité de Bartlett), le nombre de facteurs étant déterminé par l'analyse parallèle de Horn (1965) plutôt que par le seul critère de Kaiser. Les modèles confirmatoires ont été comparés par le CFI, le TLI et le RMSEA, et par le test de différence de χ^2 pour les modèles emboîtés. La fiabilité est rapportée par l'alpha de Cronbach et l'omega de McDonald avec intervalles de confiance bootstrap, l'omega étant l'estimateur approprié lorsque les saturations sont inégales (McDonald, 1999 ; Dunn et al., 2014 ; Hayes et Coutts, 2020). Pour les échelles à deux items, où ni l'alpha ni l'omega ne sont appropriés, le coefficient rapporté est celui de Spearman-Brown (Eisinga et al., 2013). La validité discriminante a été évaluée par le rapport hétérotrait-monotrait (Henseler et al., 2015), et l'invariance de mesure par comparaison de modèles emboîtés selon le critère de variation du CFI (Cheung et Rensvold, 2002).

Les comparaisons de dimensions au sein des mêmes individus reposent sur le test de Friedman puis sur des contrastes appariés, avec d_z et intervalle bootstrap. Les comparaisons de groupes utilisent le test t de Welch et l'analyse de variance de Welch, qui ne supposent pas l'homogénéité des variances, avec l'omega carré comme taille d'effet et les comparaisons post hoc de Games-Howell (1976). Les équivalents non paramétriques (Mann-Whitney, Kruskal-Wallis, delta de Cliff) sont rapportés systématiquement. La conclusion d'absence d'effet du genre est examinée par un test d'équivalence bilatéral avec bornes fixées à $\pm 0,20$ écart-type (Lakens, 2017). Les valeurs p d'une même famille de tests sont ajustées par la procédure de Benjamini et Hochberg (1995).

Les modèles multivariés sont des régressions par blocs à erreurs types robustes à l'hétéroscédasticité (HC3), l'apport de chaque bloc étant testé par un test de Wald robuste et quantifié par la variation de R^2 . La forme fonctionnelle des relations avec les indicateurs d'usage est examinée en comparant une spécification linéaire, une spécification quadratique et une spécification saturée en indicatrices.

4. RESULTATS

4.1. Structure et fiabilité de l'échelle

L'échelle de littératie administrée compte onze items et les données se prêtent à l'analyse factorielle ($KMO = 0,862$; test de sphéricité de Bartlett : $\chi^2(55) = 2\,232,19$; $p < 0,001$). L'analyse parallèle de Horn retient deux facteurs, et non cinq : les valeurs propres des deux premiers facteurs (4,038 et 1,183) dépassent les seuils issus de 500 permutations (1,234 et 1,170), la troisième restant en deçà. Le deuxième facteur ne franchit toutefois ce seuil que de justesse, ce qui interdit de surinterpréter la séparation. Après rotation varimax, le premier facteur réunit les trois items de capacité d'application (saturations de 0,560 à 0,740) et le second les huit items de jugement (0,366 à 0,620), les deux items de moralité présentant une saturation secondaire sur le premier facteur (0,329 et 0,365). La structure empirique oppose donc l'application au reste de l'échelle.

Les analyses confirmatoires (tableau 2) indiquent que modèle à cinq dimensions corrélées s'ajuste correctement et surpasse nettement le modèle unifactoriel, dont le CFI tombe sous le seuil de 0,90 ; la différence de χ^2 est franche, $\chi^2(10) = 189,95$, $p < 0,001$. Les distinctions théoriques entre dimensions ont donc une réalité latente, alors même que l'information portée par les items se résume pour l'essentiel à deux blocs. Ces deux constats ne s'opposent pas : les cinq dimensions ne sont pas interchangeables, mais elles ne fournissent pas cinq indicateurs empiriquement séparables dans cet échantillon.

Tableau 2. Comparaison des modèles de mesure de la littératie en IA (N = 825)

Modèle	χ^2	ddl	CFI	TLI	RMSEA
Cinq dimensions corrélées (structure publiée)	184,39	34	0,931	0,889	0,073
Un facteur général	374,34	44	0,849	0,812	0,096

Note. Comparaison des deux modèles : $\chi^2(10) = 189,95$; $p < 0,001$.

Les indices de fiabilité (tableau 3) fixent la portée de ce résultat. Le score global à onze items est fiable ($\alpha = 0,825$; $\omega = 0,826$). La capacité d'application, seule dimension à trois items, atteint le seuil conventionnel de 0,70. Les quatre autres dimensions ne comportant que deux items, le coefficient rapporté est celui de Spearman-Brown, approprié pour les échelles à deux items (Eisinga et al., 2013) : il vaut de 0,542 à 0,580 pour la moralité, l'auto-efficacité à l'égard de l'IA et la capacité cognitive, et seulement 0,360 pour l'esprit critique. Cette dernière dimension ne peut pas être interprétée isolément.

Tableau 3. Statistiques descriptives et fiabilité des scores de littératie (N = 825)

Échelle	k	M	ET	Coefficient	Estimation [IC 95 %]	omega [IC 95 %]
Capacité d'application	3	2,361	0,843	alpha de Cronbach	0,704 [0,662 ; 0,739]	0,713 [0,674 ; 0,747]
Moralité et éthique	2	2,658	0,931	Spearman-Brown	0,565 [0,497 ; 0,625]	—
Esprit critique	2	2,582	0,832	Spearman-Brown	0,360 [0,259 ; 0,450]	—
Auto-efficacité à l'égard de l'IA	2	2,658	0,868	Spearman-Brown	0,542 [0,473 ; 0,606]	—
Capacité cognitive	2	2,684	0,883	Spearman-Brown	0,580 [0,506 ; 0,640]	—
Score global	11	2,568	0,645	alpha de Cronbach	0,825 [0,804 ; 0,844]	0,826 [0,806 ; 0,844]

Note. k = nombre d'items. Les intervalles de confiance sont obtenus par bootstrap.

La validité discriminante indique que la littératie se distingue de l'auto-efficacité académique générale : le rapport hétérotrait-monotrait vaut 0,640, en deçà du seuil de 0,85, pour une corrélation observée de 0,541 (IC 95 % [0,491 ; 0,587]). En revanche, elle ne se distingue pas de l'échelle de perception de l'IA, composée des items de facilité d'utilisation perçue et d'utilité perçue : le rapport hétérotrait-monotrait atteint 0,901, la corrélation entre facteurs latents 0,905, et la corrélation observée 0,672 [0,633 ; 0,708]. Les deux instruments mesurent empiriquement presque la même chose. Ce résultat ne réhabilite pas les croyances d'acceptation comme dimensions de la compétence : il montre au contraire qu'un score qui les incorporerait serait dominé par une disposition évaluative indifférenciée, sans que rien ne permette d'en isoler la part de compétence.

L'invariance de mesure a été examinée pour les deux comparaisons de groupes utilisées ensuite, et les deux conclusions diffèrent. Selon le genre, l'invariance configurale est retenue (CFI minimal = 0,905 ; variation du CFI = 0,013) : les moyennes des hommes et des femmes sont comparables. Selon le champ disciplinaire, elle n'est pas retenue (CFI minimal = 0,856 ; variation du CFI = 0,066 ; variation du RMSEA = 0,027) : la structure factorielle n'est pas identique d'un champ à l'autre, et les comparaisons doivent être lues comme des différences de score observé, sans garantie qu'elles correspondent à des différences de niveau latent.

Quatre conséquences en découlent. Le score global est utilisable comme indicateur principal. La capacité d'application, seule dimension à la fois fiable et empiriquement séparée du reste, peut être suivie à part. Les trois dimensions de jugement autres que l'esprit critique sont utilisables à des fins descriptives et de comparaison intra-individuelle, mais non comme variables d'un modèle inférentiel. L'esprit critique, dont la fiabilité est trop faible, n'est rapporté que pour mémoire.

4.2. Profil des dimensions et écart entre confiance et application

Le score global de littératie dépasse le point milieu de l'échelle, mais faiblement : M = 2,568 (ET = 0,645 ; IC 95 % [2,524 ; 2,612]), soit un écart au point milieu de 0,068 point, $t(824) = 3,02$, $p = 0,003$, $d = 0,11$. Près d'un

étudiant sur deux (45,5 %) se situe en dessous de 2,5. Le niveau global n'est donc pas satisfaisant : il est à peine supérieur à la position d'indécision de l'échelle.

Le profil des cinq dimensions n'est pas graduel (figure 1). Le test de Friedman établit l'existence de différences intra-individuelles, $\chi^2(4) = 125,21$, $p < 0,001$, et la structure de ces différences est pour l'essentiel dichotomique. La capacité d'application ($M = 2,361$) est la seule dimension située sous le point milieu, avec 54,5 % des étudiants en dessous de 2,5 et une majorité de réponses de désaccord (52,0 % contre 48,0 % d'accord). Elle se sépare des quatre autres dimensions avec des tailles d'effet appariées homogènes : $d_z = -0,330$ [-0,399 ; -0,259] face à la moralité, $-0,223$ [-0,289 ; -0,159] face à l'esprit critique, $-0,329$ [-0,398 ; -0,257] face à l'auto-efficacité à l'égard de l'IA et $-0,342$ [-0,419 ; -0,273] face à la capacité cognitive, toutes les valeurs p ajustées étant inférieures à 0,001.

Les six contrastes entre les quatre autres dimensions ne sont pas tous nuls, trois demeurent significatifs après ajustement, et tous trois impliquent l'esprit critique : moralité contre esprit critique ($d_z = 0,077$; p ajusté = 0,038), esprit critique contre auto-efficacité ($-0,085$; p ajusté = 0,025) et esprit critique contre capacité cognitive ($-0,112$; p ajusté = 0,003). Les trois autres sont non significatifs ($|d_z| \leq 0,031$; p ajusté $\geq 0,463$). L'esprit critique se situe donc légèrement en retrait des trois dimensions de jugement, mais l'ampleur de cet écart est négligeable et son coefficient de fiabilité (0,360) interdit de lui accorder une signification substantielle. L'hypothèse H1 est soutenue dans son premier volet, et dans son second sous réserve de cette exception d'ampleur négligeable : l'application est en retrait de 0,22 à 0,34 écart-type intra-sujet, alors que les quatre autres dimensions s'échelonnent sur un intervalle de 0,10 point.

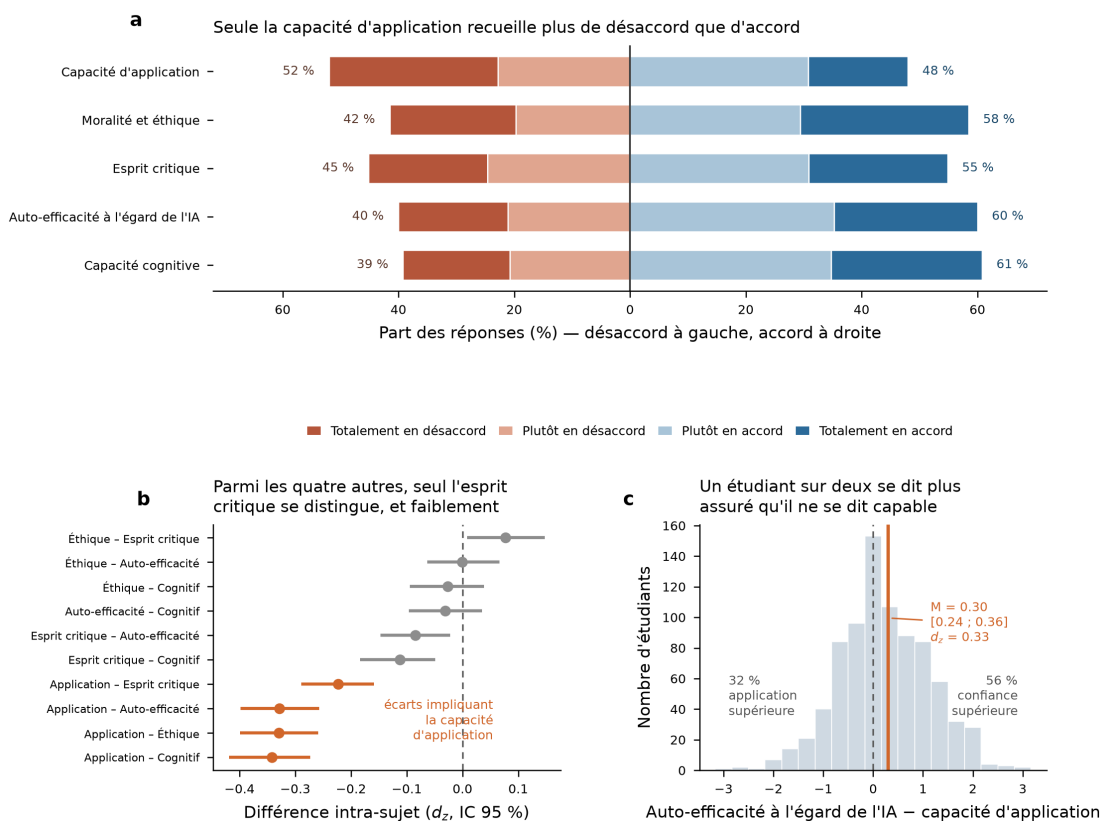


Figure 1. Profil des cinq dimensions de la littératie en IA (N = 825). (a) Répartition des réponses aux onze items, agrégée par dimension ; les pourcentages indiquent la part cumulée de désaccord (à gauche) et d'accord (à droite). (b) Contrastes intra-sujets entre paires de dimensions, exprimés en d_z avec intervalles de confiance bootstrap à 95 % (1 000 répliquations) ; valeurs p ajustées par la procédure de Benjamini-Hochberg. (c) Distribution de l'écart intra-individuel entre auto-efficacité à l'égard de l'IA et capacité d'application ; la ligne verticale indique la moyenne.

Le rapprochement de ces constats fournit le résultat central de cette section. Les dimensions les plus élevées du profil sont la capacité cognitive ($M = 2,684$) et, à égalité, la moralité et l'auto-efficacité à l'égard de l'IA ($M = 2,658$ chacune) ; la plus basse est la capacité d'application ($M = 2,361$). Calculé individu par individu, l'écart entre l'auto-efficacité à l'égard de l'IA et la capacité d'application atteint 0,297 point ($ET = 0,905$; IC 95 % $[0,236 ; 0,359]$; $d_z = 0,329$; $t(824) = 9,43$; $p < 0,001$). Il est positif chez 56,1 % des étudiants, nul chez 11,6 % et négatif chez 32,2 % : un étudiant sur deux se déclare plus assuré de pouvoir mobiliser l'IA qu'il ne se déclare capable de l'appliquer à des situations concrètes. L'hypothèse H2 est soutenue.

4.3. Facteurs sociodémographiques

Le tableau 4 rassemble les comparaisons de groupes sur le score global et sur les cinq dimensions.

Tableau 4. Comparaisons de groupes sur la littératie en IA (N = 825)

Variable dépendante	Facteur	Test	Statistique	p ajusté	Taille d'effet [IC 95 %]
Score global	Genre	t de Welch	$t(798,1) = -1,56$	0,237	$d = -0,109 [-0,242 ; 0,021]$
Score global	Champ disciplinaire	ANOVA de Welch	$F(2 ; 356,7) = 29,01$	< 0,001	$\omega^2 = 0,064 [0,036 ; 0,101]$
Score global	Promotion d'études	ANOVA de Welch	$F(3 ; 261,7) = 5,94$	0,004	$\omega^2 = 0,018 [0,004 ; 0,044]$
Score global	Tranche d'âge	ANOVA de Welch	$F(2 ; 187,8) = 1,07$	0,747	$\omega^2 = 0,000 [-0,002 ; 0,014]$
Capacité d'application	Champ disciplinaire	ANOVA de Welch	$F(2 ; 347,9) = 29,11$	< 0,001	$\omega^2 = 0,061 [0,033 ; 0,096]$
Auto-efficacité à l'égard de l'IA	Champ disciplinaire	ANOVA de Welch	$F(2 ; 368,5) = 20,79$	< 0,001	$\omega^2 = 0,043 [0,020 ; 0,074]$
Moralité éthique	Champ disciplinaire	ANOVA de Welch	$F(2 ; 359,2) = 19,32$	< 0,001	$\omega^2 = 0,043 [0,021 ; 0,075]$
Capacité cognitive	Champ disciplinaire	ANOVA de Welch	$F(2 ; 366,0) = 13,35$	< 0,001	$\omega^2 = 0,029 [0,011 ; 0,057]$
Esprit critique	Champ disciplinaire	ANOVA de Welch	$F(2 ; 362,1) = 8,72$	< 0,001	$\omega^2 = 0,018 [0,005 ; 0,043]$

Note. Valeurs p ajustées par la procédure de Benjamini-Hochberg au sein de chaque famille de tests (un facteur, six variables dépendantes). Les tests non paramétriques correspondants conduisent aux mêmes conclusions pour l'ensemble des comparaisons rapportées.

Comparaison en fonction du genre

L'écart entre femmes ($M = 2,607$; $ET = 0,630$) et hommes ($M = 2,537$; $ET = 0,656$) n'est pas significatif, t de Welch($798,1$) = $-1,56$, $p = 0,119$, p ajusté = $0,237$, $d = -0,109 [-0,242 ; 0,021]$, et le test de Mann-Whitney conduit à la même conclusion ($p = 0,121$; $\text{delta de Cliff} = -0,063$). Aucune des cinq dimensions ne montre de différence après ajustement (p ajusté $\geq 0,226$). Ce résultat doit toutefois être énoncé avec précision : le test d'équivalence avec bornes à $\pm 0,20$ écart-type n'est pas concluant ($p = 0,097$). Les données sont donc compatibles avec l'absence d'effet, mais elles ne permettent pas d'exclure un léger avantage féminin de l'ordre d'un cinquième d'écart-type. Par conséquent, aucune différence de genre n'est détectée.

Comparaison en fonction du champ disciplinaire.

L'effet est significatif et d'ampleur modérée, F de Welch($2 ; 356,7$) = $29,01$, $p < 0,001$, $\omega^2 = 0,064 [0,036 ; 0,101]$. Le test de Levene n'est pas significatif pour le score global ($p = 0,714$), mais il l'est pour la capacité d'application ($p = 0,006$) et l'auto-efficacité à l'égard de l'IA ($p < 0,001$), ce qui justifie le recours à des procédures ne supposant pas l'homoscédasticité. Les comparaisons de Games-Howell (tableau 5) montrent que le patron n'est pas une gradation à trois niveaux : les sciences de la santé ($M = 2,787$) et les sciences et technologie ($M = 2,729$) ne diffèrent pas entre elles (différence = $0,058$; $p = 0,660$), et toutes deux se situent au-dessus des sciences sociales et humaines ($M = 2,421$), avec des écarts de $0,367 [0,226 ; 0,508]$ ($d = 0,58$) et $0,308 [0,189 ; 0,428]$ ($d = 0,50$). Seul ce dernier groupe se situe en deçà du point milieu de l'échelle.

Conformément à la réserve établie à la section 4.1, ces écarts sont des différences de score observé et non de niveau latent.

L'analyse dimension par dimension précise ce résultat (figure 2b) : l'effet du champ est significatif sur les cinq dimensions, avec une variance expliquée maximale sur la capacité d'application ($\omega^2 = 0,061$), intermédiaire sur l'auto-efficacité à l'égard de l'IA et la moralité (0,043 chacune) et la capacité cognitive (0,029), et minimale sur l'esprit critique (0,018). Sur la capacité d'application, l'écart entre sciences de la santé et sciences sociales et humaines atteint 0,563 point ($d = 0,72$), et les sciences de la santé dépassent également les sciences et technologie ($d = 0,40$). C'est la seule dimension sur laquelle les deux champs scientifiques se distinguent l'un de l'autre. L'hypothèse H3 est soutenue.

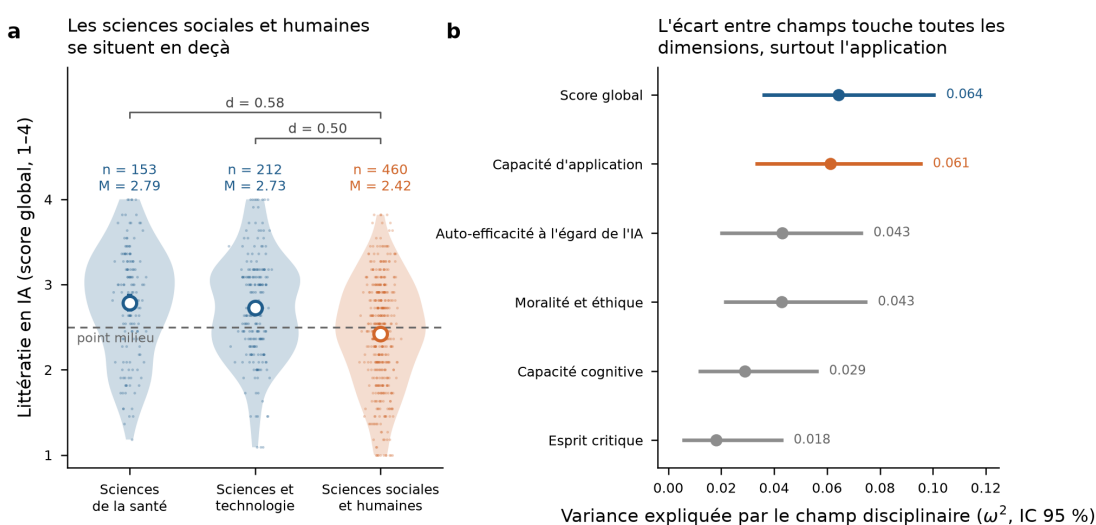


Figure 2. Littératie en IA selon le champ disciplinaire (N = 825). (a) Distribution du score global par champ ; les points sont les observations individuelles avec bruitage horizontal, le cercle blanc la moyenne et le segment épais son intervalle de confiance à 95 % ; les d de Cohen proviennent des comparaisons de Games-Howell. (b) Variance expliquée par le champ disciplinaire sur le score global et sur chaque dimension, avec intervalles de confiance bootstrap à 95 % (1 000 répliques).

Comparaison en fonction de la promotion d'études.

L'effet est significatif mais faible, F de Welch(3 ; 261,7) = 5,94, $p < 0,001$, p ajusté = 0,004, $\omega^2 = 0,018$ [0,004 ; 0,044], et il est décroissant plutôt que croissant : les étudiants de L2 obtiennent le score le plus élevé (M = 2,656) et dépassent ceux de L3 (M = 2,506 ; $d = 0,23$; $p = 0,034$), de Master 1 (M = 2,388 ; $d = 0,41$; $p = 0,006$) et de Master 2 (M = 2,491 ; $d = 0,26$; $p = 0,034$), ces trois derniers groupes ne se distinguant pas entre eux. La littératie déclarée ne croît donc pas avec l'avancement dans le cursus ; elle est au contraire plus élevée

à l'entrée. La tranche d'âge, enfin, n'a aucun effet détectable sur le score global ($\omega^2 = 0,000$).

Tableau 5. Comparaisons post hoc de Games-Howell selon le champ disciplinaire

Variable dépendante	Comparaison	Différence [IC 95 %]	ddl	p	d de Cohen
Score global	Santé vs sciences et technologie	0,058 [-0,100 ; 0,216]	315,8	0,660	0,09
Score global	Santé vs sciences sociales et humaines	0,367 [0,226 ; 0,508]	251,5	< 0,001	0,58
Score global	Sciences et technologie vs sciences sociales et humaines	0,308 [0,189 ; 0,428]	417,1	< 0,001	0,50
Capacité d'application	Santé vs sciences et technologie	0,350 [0,138 ; 0,562]	347,0	< 0,001	0,40
Capacité d'application	Santé vs sciences sociales et humaines	0,563 [0,387 ; 0,739]	252,5	< 0,001	0,72
Capacité d'application	Sciences et technologie vs sciences sociales et humaines	0,214 [0,045 ; 0,383]	360,3	0,009	0,26

4.4. Usages de l'IA : un indice inutilisable et une relation non linéaire

Le premier résultat de cette section porte sur la mesure de l'usage. Les trois indicateurs ne constituent pas une échelle : leur alpha vaut 0,011 (IC 95 % [-0,112 ; 0,117]), intervalle qui contient zéro et des valeurs négatives. Les corrélations inter-indicateurs expliquent ce résultat : la familiarité et la fréquence sont faiblement liées ($\rho = 0,134$; $p < 0,001$), mais la durée d'une session est négativement liée à la fréquence ($\rho = -0,099$; $p = 0,005$) et sans lien avec la familiarité ($\rho = -0,053$; $p = 0,130$). Un étudiant qui consulte souvent un assistant le fait par sessions courtes ; un étudiant qui y consacre de longues sessions le consulte rarement. Moyenner ces trois indicateurs produirait un composite sans validité de construit. L'hypothèse H4 est soutenue, et les trois indicateurs sont traités séparément.

Analysés séparément, ils ne se comportent pas de la même façon (tableau 6). La familiarité est le seul indicateur associé positivement au score global de manière robuste ($\rho = 0,136$ [0,068 ; 0,202] ; p ajusté = 0,001), et son association est plus forte avec la capacité d'application ($\rho = 0,179$ [0,112 ; 0,245]). La fréquence d'utilisation n'entretient aucune association monotone significative avec le score global ($\rho = 0,049$; p ajusté = 0,359), alors que le test de Kruskal-Wallis sur les trois niveaux est hautement significatif ($H = 39,13$; $p < 0,001$). Cette dissociation entre une corrélation nulle et un test de comparaison de groupes très significatif est la signature d'une relation non monotone. La durée d'une session, enfin, est sans lien avec le score global ($\rho = -0,027$; p ajusté = 0,647) mais négativement associée à la capacité d'application ($\rho = -0,124$ [-0,190 ; -0,056] ; p ajusté = 0,002).

Tableau 6. Associations entre indicateurs d'usage et littératie en IA (N = 825)

Indicateur	Variable dépendante	rho de Spearman [IC 95 %]	p ajusté	H de Kruskal-Wallis	p
Familiarité	Score global	0,136 [0,068 ; 0,202]	0,001	26,59	< 0,001
Familiarité	Capacité d'application	0,179 [0,112 ; 0,245]	< 0,001	48,62	< 0,001
Fréquence	Score global	0,049 [-0,019 ; 0,117]	0,359	39,13	< 0,001
Fréquence	Capacité d'application	0,038 [-0,030 ; 0,106]	0,493	74,18	< 0,001
Durée d'une session	Score global	-0,027 [-0,096 ; 0,041]	0,647	4,03	0,133
Durée d'une session	Capacité d'application	-0,124 [-0,190 ; -0,056]	0,002	16,19	< 0,001

Note. Valeurs p ajustées par la procédure de Benjamini-Hochberg sur la famille des dix-huit corrélations (trois indicateurs, six variables dépendantes).

La figure 3 donne la forme de ces relations, et elle est en U. Pour la fréquence d'utilisation, le score global vaut 2,654 chez les utilisateurs de faible fréquence (n = 243), chute à 2,373 au niveau intermédiaire (n = 255) et remonte à 2,656 au niveau élevé (n = 327). Le creux est encore plus prononcé sur la capacité d'application : 2,536, puis 1,995, puis 2,516. Les comparaisons de Games-Howell confirment que le niveau intermédiaire se distingue des deux autres (sur le score global, contre le niveau faible : d = 0,44 ; p < 0,001 ; contre le niveau élevé : d = -0,46 ; p < 0,001 ; sur la capacité d'application, d = 0,65 et d = -0,67, p < 0,001), alors que les niveaux faible et élevé sont indiscernables (d = -0,004 ; p = 0,999 sur le score global). Le terme quadratique est significatif (b = 0,282 ; p < 0,001 sur le score global, b = 0,531 ; p < 0,001 sur la capacité d'application) et la spécification saturée porte le R² de 0,000 à 0,041 pour le score global et à 0,085 pour la capacité d'application : la relation existe et la spécification linéaire la masque intégralement. Le même patron, moins marqué, s'observe pour la familiarité (terme quadratique b = 0,169 ; p < 0,001 sur le score global).

La compétence déclarée ne croît pas avec l'usage : elle chute au niveau intermédiaire

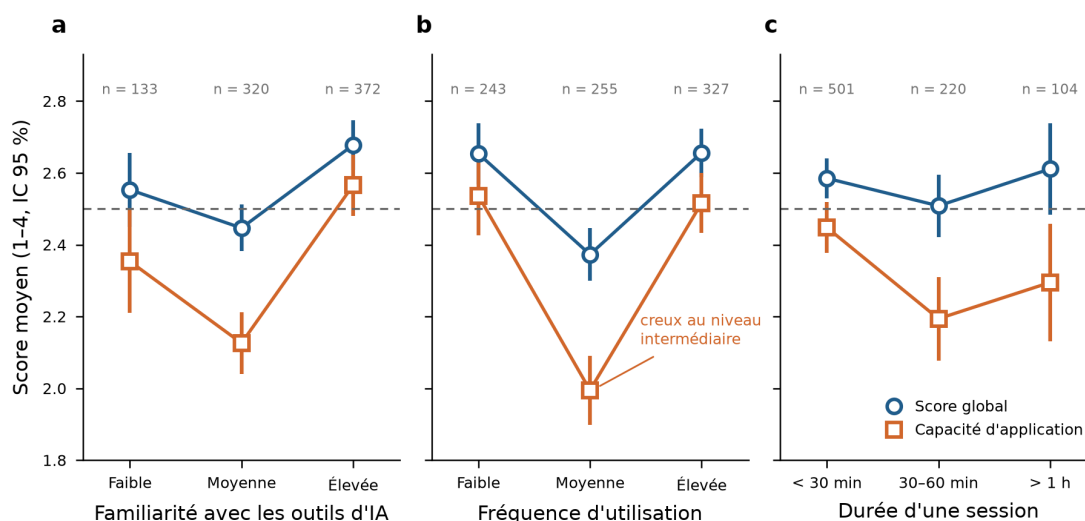


Figure 3. Score de littératie en IA selon les trois indicateurs d'usage (N = 825). Moyennes et intervalles de confiance à 95 %, pour le score global et pour la capacité d'application. La ligne horizontale marque le point milieu de l'échelle. Les effectifs par niveau sont indiqués en haut de chaque panneau.

Ce résultat résiste à deux mises à l'épreuve. Après contrôle du genre, du champ disciplinaire, de la promotion et de l'âge, l'écart du niveau intermédiaire demeure : $b = -0,200$ ($p < 0,001$) sur le score global et $b = -0,432$ ($p < 0,001$) sur la capacité d'application, le contraste entre niveau intermédiaire et niveau élevé restant significatif ($p < 0,001$). Après exclusion des trente-neuf participants signalés par les indicateurs de qualité de réponse, les coefficients augmentent légèrement (respectivement $-0,233$ et $-0,477$), ce qui écarte l'explication par un style de réponse central chez des répondants peu engagés. L'hypothèse H5 est soutenue.

4.5. Modèles multivariés

Le tableau 7 et la figure 4 présentent les régressions par blocs. Les indices d'inflation de la variance restent inférieurs à 1,43, écartant tout problème de colinéarité.

Tableau 7. Régressions hiérarchiques prédisant la littératie en IA et la capacité d'application (N = 825, erreurs types HC3)

Variable dépendante	Modèle	R ²	R ² ajusté	Variation de R ²	F de Wald du bloc	p
Score global	M1 : sociodémographie	0,074	0,066			
Score global	M2 : + indicateurs d'usage	0,084	0,072	0,009	8,70	0,034
Score global	M3 : + auto-efficacité académique	0,338	0,329	0,254	295,39	< 0,001
Capacité d'application	M1 : sociodémographie	0,076	0,068			
Capacité d'application	M2 : + indicateurs d'usage	0,098	0,087	0,022	19,11	< 0,001
Capacité d'application	M3 : + auto-efficacité académique	0,221	0,210	0,123	134,70	< 0,001

Trois enseignements se dégagent. Premièrement, les variables sociodémographiques expliquent 7,4 % de la variance du score global, essentiellement par le champ disciplinaire (sciences de la santé : $b = 0,184$ [0,072 ; 0,297] ; sciences et technologie : $b = 0,224$ [0,123 ; 0,325], en référence aux sciences sociales et humaines). Deuxièmement, l'ajout des trois indicateurs d'usage n'apporte que 0,9 point de pourcentage (F de Wald = 8,70 ; $p = 0,034$), et la familiarité est le seul indicateur dont le coefficient reste significatif ($b = 0,072$ [0,021 ; 0,123] ; $p = 0,006$). Sur la capacité d'application, en revanche, ce bloc apporte 2,2 points ($p < 0,001$), et la durée d'une session y contribue négativement ($b = -0,107$ [-0,181 ; -0,033] ; $p = 0,005$). Troisièmement, l'ajout de l'auto-efficacité académique générale fait passer le R² de 0,084 à 0,338 : à lui seul, ce prédicteur explique 25,4 % de la variance du

score de littératie (beta standardisé = 0,516 ; $b = 0,512$ [0,453 ; 0,570]), soit près de vingt-sept fois l'apport de l'ensemble des indicateurs d'usage. Son apport est nettement plus faible sur la capacité d'application (12,3 points ; $\beta = 0,359$).

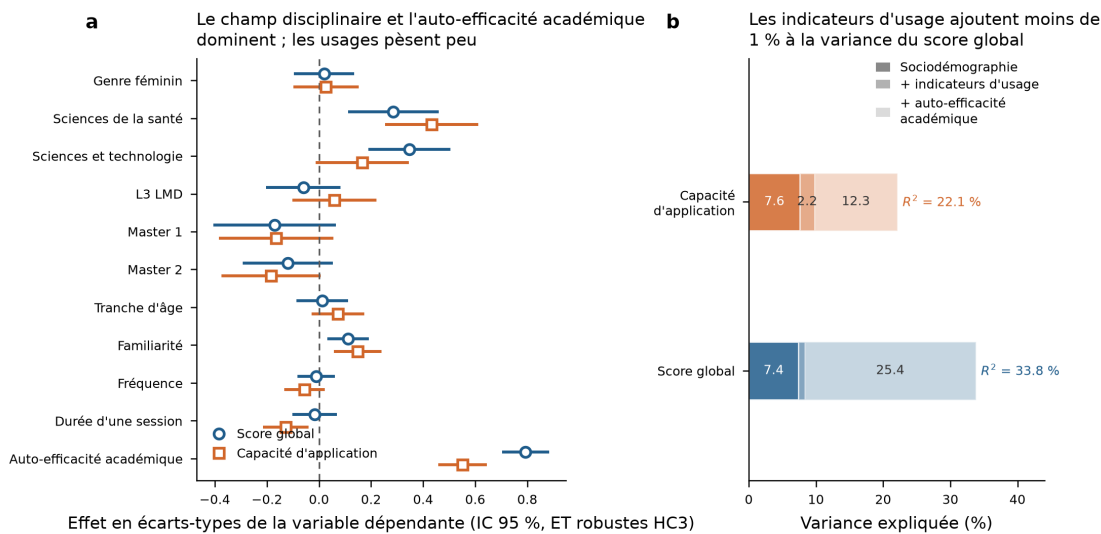


Figure 4. Prédicteurs de la littératie en IA et de la capacité d'application ($N = 825$). (a) Coefficients du modèle complet M3, exprimés en écarts-types de la variable dépendante, avec intervalles de confiance à 95 % calculés sur erreurs types robustes HC3 ; les modalités de référence sont le genre masculin, les sciences sociales et humaines et la promotion L2 LMD. (b) Décomposition de la variance expliquée par blocs successifs de prédicteurs.

4.6. Robustesse

Trois vérifications encadrent les résultats précédents. L'exclusion des trente-neuf participants signalés par les indicateurs de qualité de réponse ($N = 786$) laisse inchangées toutes les conclusions : l'effet du champ disciplinaire reste significatif avec une ampleur voisine ($\omega^2 = 0,060$ contre $0,064$ sur le score global ; $0,050$ contre $0,061$ sur la capacité d'application), et l'absence d'effet du genre est confirmée ($d = -0,087$; $p = 0,222$). Les tests non paramétriques conduisent aux mêmes décisions que les tests paramétriques pour l'ensemble des comparaisons rapportées.

5. DISCUSSION

5.1. Une compétence déclarée plutôt qu'une compétence démontrée

Le premier apport de cette étude est de délimiter ce que mesure l'instrument. Trois résultats convergent pour indiquer que le score obtenu relève davantage de la compétence perçue que de la compétence effective.

D'abord, l'auto-efficacité académique générale, mesurée par une échelle sans rapport de contenu avec l'IA, explique à elle seule 25,4 % de la variance du score de littératie, soit environ trois fois l'apport conjoint du champ

disciplinaire, de la promotion, du genre, de l'âge et des trois indicateurs d'usage. Une mesure de compétence spécifique dont le principal prédicteur est une disposition auto-évaluative générale partage une part importante de sa variance avec la manière dont l'étudiant se juge, indépendamment du domaine. Ce constat est cohérent avec ce que la littérature sur les biais de méthode commune prévoit lorsque prédicteur et critère sont recueillis par le même mode de réponse (Podsakoff et al., 2003), et il ne peut être écarté par les données disponibles.

Ensuite, la structure du profil est difficile à interpréter comme un profil de compétences. Les quatre dimensions relevant du jugement, à savoir la moralité, l'esprit critique, la capacité cognitive et l'auto-efficacité, sont toutes situées au-dessus du point milieu et s'échelonnent sur un intervalle de 0,10 point, tandis que la seule dimension engageant des tâches concrètes se détache par le bas de 0,22 à 0,34 écart-type intra-sujet. Cette dichotomie correspond à la structure à deux facteurs que l'analyse parallèle extrait des onze items. Autrement dit, ce que l'échelle distingue empiriquement n'est pas cinq compétences, mais une adhésion générale à des énoncés de principe d'un côté, et une capacité d'application déclarée de l'autre.

À cette observation s'ajoute un argument plus direct, l'échelle de littératie ne se sépare pas de l'échelle de perception de l'IA, composée d'items de facilité et d'utilité perçues : le rapport hétérotrait-monotrait atteint 0,901 et la corrélation entre facteurs latents 0,905. Deux instruments dont l'un interroge des capacités et l'autre des croyances d'acceptation devraient se distinguer ; ils ne le font pas. Ce résultat conforte la lecture proposée ici, celle d'une disposition évaluative commune, et il fournit la raison technique de ne pas incorporer ces croyances au score de littératie ou de les considérer comme une variable externe : le score est indissociable d'un jugement global favorable aux outils.

Enfin, l'écart intra-individuel entre confiance et application, positif chez 56,1 % des étudiants, est de nature à la fois substantielle et méthodologique. Substantielle, parce qu'il documente une confiance qui excède la capacité déclarée d'agir, précisément le patron que la recherche sur la calibration des jugements de compétence observe chez les personnes peu compétentes (Kruger et Dunning, 1999 ; Dunning, 2011). Méthodologique, parce qu'il montre que l'échelle capte cet écart : les items d'application discriminent là où les items de jugement saturent. Cette lecture reste néanmoins bornée par le fait que les deux termes de l'écart sont auto-rapportés ; seule une épreuve de performance permettrait de trancher entre une surestimation de la confiance et une sous-estimation de la capacité (Hornberger et al., 2023 ; Wang et al., 2022).

La conclusion qui s'impose est que le niveau global de littératie en IA des étudiants de l'UNIKIN ne peut être qualifié de satisfaisant. Le score global excède le point milieu de 0,07 point, près d'un étudiant sur deux se situe en dessous, et le seul indicateur qui engage l'action concrète est majoritairement en désaccord.

5.2. Le déficit d'application et sa lecture disciplinaire

Le déficit d'application n'est pas uniformément réparti. Il est le plus marqué en sciences sociales et humaines, où l'écart avec les sciences de la santé atteint 0,72 écart-type, et le champ disciplinaire explique plus de variance sur cette dimension que sur toute autre. Ce résultat est cohérent avec l'idée que les formations scientifiques et de santé multiplient les occasions de mobiliser des outils numériques sur des tâches disciplinaires identifiables : analyse de données, recherche documentaire spécialisée, résolution de problèmes formalisés (Zawacki-Richter et al., 2019 ; Chan et Hu, 2023).

Deux précisions s'imposent. La première est que le patron observé n'est pas une gradation : les sciences de la santé et les sciences et technologie ne se distinguent pas l'une de l'autre, et c'est le champ des sciences sociales et humaines, qui rassemble 55,8 % de l'échantillon, qui se situe en retrait. La formulation d'un classement des trois champs serait donc infondée. La seconde est que l'invariance configurale de l'échelle selon le champ n'est pas atteinte : la structure factorielle n'est pas la même d'un champ à l'autre, et une partie de l'écart pourrait refléter une compréhension différenciée de certains items selon la filière plutôt qu'une différence de niveau latent. Cette réserve n'annule pas le résultat, dont l'ampleur est notable et qui se reproduit sur les cinq dimensions, mais elle interdit de l'énoncer comme un écart de compétence latente et impose de le formuler comme un écart de score observé. Un instrument dont les items fonctionnent de la même manière dans les trois champs serait nécessaire pour trancher.

L'absence de progression avec l'avancement dans le cursus complète ce tableau. Les étudiants de L2 déclarent une littératie supérieure à celle des étudiants de L3, de Master 1 et de Master 2, ce qui contredit l'hypothèse d'un développement cumulatif par l'expérience (Bandura, 1997) et rejoint la thèse selon laquelle la littératie en IA dépend d'occasions d'apprentissage spécifiques plutôt que de l'ancienneté académique (Long et Magerko, 2020). En l'absence de formation structurée à l'IA dans les curricula de l'UNIKIN, il n'existe aucun mécanisme institutionnel par lequel le niveau progresserait d'une promotion à l'autre. Une lecture alternative doit être mentionnée : les étudiants avancés, mieux à même d'évaluer l'exigence des tâches académiques, pourraient se juger plus sévèrement, ce qui produirait le même patron sans différence de compétence effective. Les données transversales ne permettent pas de départager ces deux explications.

5.3. L'usage n'est pas un levier linéaire

Le résultat le plus inattendu de cette étude est la forme de la relation entre usage et littératie déclarée. Une lecture linéaire conduirait à conclure que la fréquence d'utilisation est faiblement et positivement associée à la littératie, ce que suggère un coefficient de corrélation de 0,05. Cette lecture est erronée : la relation est en U, et la spécification linéaire l'annule. Les utilisateurs de fréquence intermédiaire déclarent une compétence d'application inférieure d'environ deux tiers d'écart-type à celle des utilisateurs occasionnels comme des utilisateurs intensifs, dont les niveaux sont par ailleurs indiscernables ; sur le score global, l'écart est de l'ordre de 0,45 écart-type.

Deux mécanismes peuvent produire ce patron, et ils ne sont pas exclusifs. Le premier est un mécanisme de calibration. Les utilisateurs occasionnels disposent de peu d'éléments pour évaluer leur propre maîtrise et s'attribuent un niveau proche de la norme sociale, tandis que les utilisateurs intermédiaires ont rencontré assez d'échecs et de résultats douteux pour percevoir l'écart entre ce qu'ils attendent des outils et ce qu'ils en obtiennent, sans avoir encore acquis les routines qui le comblent. Les utilisateurs intensifs auraient franchi ce seuil. Ce mécanisme est celui que documente la littérature sur l'auto-évaluation des compétences, où la précision du jugement croît avec l'expérience mais où la confiance ne décroît pas dans les premiers stades (Zell et Krizan, 2014). Le second mécanisme est substantiel : la pratique délibérée, faite de contrôle et de correction, se distingue de l'exposition simple (Ericsson et al., 1993), et le passage d'un usage occasionnel à un usage régulier peut n'être qu'un accroissement d'exposition, sans effet formateur.

L'association négative entre durée des sessions et capacité d'application, robuste au contrôle des covariables, converge avec cette seconde lecture. Une session longue n'atteste pas d'un travail approfondi ; elle peut signaler une difficulté à formuler une demande productive et à en exploiter le résultat. Les recommandations qui font dépendre la littératie de la qualité de l'encadrement plutôt que du volume d'utilisation (UNESCO, 2024) trouvent ici un appui empirique direct.

Ces résultats invitent enfin à abandonner la pratique consistant à agréger familiarité, fréquence et durée en un indice d'« usage de l'IA ». La cohérence interne d'un tel indice est nulle dans ces données, et sa composante de durée est négativement liée aux deux autres. Un score moyen de ces trois indicateurs ne mesure aucun construit identifiable ; les analyses qui reposent sur un tel indice sont difficilement interprétables (Bollen et Lennox, 1991).

5.4. Ce que l'absence d'effet du genre autorise à dire

L'absence de différence significative entre hommes et femmes rejoint plusieurs travaux récents sur la réduction des écarts de compétences

numériques dans l'enseignement supérieur (Laupichler et al., 2022 ; Chiu et al., 2023). Nos données invitent cependant à la prudence dans la formulation. Le test d'équivalence n'établit pas l'absence d'un effet de l'ordre de 0,20 écart-type ($p = 0,097$), et l'intervalle de confiance du d de Cohen, $[-0,24 ; 0,02]$, reste compatible avec un léger avantage féminin. Conclure que le genre « ne constitue pas un facteur déterminant » excède donc ce que les données permettent. La formulation défendable est plus modeste : aucune différence de genre n'est détectée, et si un écart existe, il est petit.

5.5. Implications pour l'enseignement supérieur congolais

Trois orientations se dégagent, et elles diffèrent de celles qu'un simple constat de niveau global suggérerait.

La première est de cibler l'application plutôt que la sensibilisation. Les dimensions de jugement, éthique, critique et cognitive, recueillent déjà une adhésion majoritaire ; les dispositifs de sensibilisation aux enjeux de l'IA agiraient donc sur ce qui est le moins déficitaire. Le déficit porte sur la capacité à mobiliser ces outils dans des situations réelles, ce qui appelle des activités adossées à des tâches disciplinaires évaluables, et non des modules généraux d'initiation.

La deuxième est de traiter en priorité les sciences sociales et humaines, où le déficit d'application est le plus marqué et qui rassemblent la majorité des effectifs. Les applications disciplinaires y existent, de l'analyse documentaire à la constitution et au traitement de corpus, mais elles demandent d'être explicitement construites, ce que les formations scientifiques réalisent plus spontanément.

La troisième est d'accompagner les usages intermédiaires. Le creux observé au niveau intermédiaire d'utilisation désigne un moment du parcours où l'étudiant a rencontré les limites des outils sans disposer des moyens de les dépasser. C'est à ce point que l'encadrement, sous la forme de cellules d'appui aux usages pédagogiques de l'IA, est susceptible de produire le plus d'effet, alors qu'un dispositif visant uniformément l'ensemble des étudiants diluerait son action.

5.6. Limites

Cinq limites bornent la portée de ces résultats. La première, la plus importante, est que la littératie est mesurée par auto-déclaration : l'étude porte sur la compétence perçue, et la section 5.1 a montré que celle-ci partage une part substantielle de sa variance avec une disposition auto-évaluative générale. Une épreuve de performance à réponses objectivement corrigées (Hornberger et al., 2023) constitue le complément nécessaire, et permettrait de tester directement l'hypothèse de calibration avancée à la section 5.3.

La deuxième est le nombre d'items par dimension : trois pour la capacité d'application et deux pour les quatre autres, ce qui place quatre coefficients de fiabilité sous le seuil conventionnel et, pour l'esprit critique, à un niveau (0,360) qui rend cette dimension non interprétable isolément. Les scores dimensionnels ont été utilisés à des fins descriptives et de comparaison intra-individuelle, non comme variables de modèles inférentiels, mais toute lecture dimension par dimension doit tenir compte de cette imprécision, qui atténue les associations observées. Une version de l'instrument comportant au moins quatre items par dimension serait nécessaire pour analyser les dimensions séparément.

La troisième est que les indicateurs d'usage sont des questions uniques à trois modalités. Elles ne permettent de distinguer ni les types de tâches confiées aux outils, ni la nature du contrôle exercé sur leurs productions, distinction pourtant centrale dans l'interprétation proposée à la section 5.3.

La quatrième est l'échantillonnage non probabiliste sur une seule université, qui limite la généralisation, et la surreprésentation des sciences sociales et humaines comme des promotions de licence dans l'échantillon.

La cinquième est le caractère transversal du dispositif, qui interdit toute inférence causale. À quoi s'ajoute que l'analyse exploratoire et l'analyse confirmatoire de la structure portent sur le même échantillon, de sorte que les indices d'ajustement rapportés à la section 4.1 comparent des modèles concurrents sans constituer une validation sur données indépendantes. La relation en U entre fréquence d'usage et compétence déclarée est compatible avec un effet de l'usage sur la calibration comme avec une sélection des étudiants dans les niveaux d'usage. Un suivi longitudinal, ou un dispositif expérimental portant sur des activités d'apprentissage structurées, serait requis pour trancher.

CONCLUSION

Cette étude visait à évaluer la littératie en IA des étudiants de l'Université de Kinshasa et à en analyser les facteurs associés, en prenant au sérieux la question de ce que mesure une échelle auto-rapportée de compétence.

Sur le plan de la mesure, l'échelle à onze items fournit un score global fiable, mais sa structure empirique compte deux facteurs et non cinq, les scores dimensionnels à deux items sont trop peu fiables pour porter des analyses inférentielles séparées, et l'invariance configurale n'est pas atteinte selon le champ disciplinaire. La facilité d'utilisation perçue et l'utilité perçue, qui relèvent du modèle d'acceptation des technologies, ne font pas partie du construit et n'ont pas à figurer dans un profil de littératie ; les données montrent de surcroît qu'elles ne s'en distinguent pas empiriquement, ce qui rendrait un tel profil inintelligible.

Sur le plan du niveau, le résultat central n'est pas la valeur du score global, à peine supérieure au point milieu de l'échelle, mais la forme du profil : la capacité d'application est la seule dimension déficitaire, elle se détache nettement des quatre autres, resserrées sur un intervalle de 0,10 point, et un étudiant sur deux se déclare plus assuré qu'il ne se déclare capable.

Sur le plan des facteurs, le champ disciplinaire est le seul déterminant sociodémographique d'ampleur appréciable, avec un retrait des sciences sociales et humaines maximal sur la capacité d'application. La littératie déclarée est plus élevée à l'entrée dans le cursus qu'aux niveaux supérieurs, aucune différence de genre n'est détectée, et la relation avec l'usage n'est pas linéaire : les utilisateurs intermédiaires se déclarent les moins compétents. Les indicateurs d'usage n'ajoutent que 0,9 % à la variance du score global, contre 25,4 % pour l'auto-efficacité académique générale.

Ces résultats plaident pour une politique d'intégration de l'IA qui porte sur la pratique disciplinaire encadrée plutôt que sur l'exposition aux outils : des activités d'application évaluables inscrites dans les curricula, un effort ciblé sur les filières de sciences sociales et humaines, et un accompagnement des étudiants au moment où ils rencontrent les limites des outils. Ils plaident aussi pour une évolution des instruments : aussi longtemps que la littératie en IA sera mesurée par des énoncés d'accord, l'écart entre ce que les étudiants déclarent savoir faire et ce qu'ils savent faire restera hors de portée de l'observation.

BIBLIOGRAPHIE

- African Union. (2024). *Continental artificial intelligence strategy*. African Union Commission.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289-300.
- Bofula Lobebe, T., Enguta Mwenzi, J., & Kyamundu Katungu, S. (2023). Usages des numériques et rendement académique des étudiants de l'Université de Kinshasa. *Mouvements et Enjeux Sociaux*, (126), 219-226.
- Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), 305-314.

- Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILES: Meta AI literacy scale. Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change and meta-competencies. *Computers in Human Behavior: Artificial Humans*, 1, 100014.
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1), 43.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.
- Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M.-T. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4, 100118.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.
- Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38(2), 269-277.
- Dunn, T. J., Baguley, T., & Brunsdon, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399-412.
- Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. *Advances in Experimental Social Psychology*, 44, 247-296.
- Eisinga, R., Grotenhuis, M. te, & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, 58(4), 637-642.
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363-406.
- Games, P. A., & Howell, J. F. (1976). Pairwise multiple comparison procedures with unequal n's and/or variances: A Monte Carlo study. *Journal of Educational Statistics*, 1(2), 113-125.
- Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But... *Communication Methods and Measures*, 14(1), 1-24.

- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115-135.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185.
- Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about artificial intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence*, 5, 100165.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134.
- Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355-362.
- Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, 3, 100101.
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (p. 1-16). ACM.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum.
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041.
- Nong, Y., Cui, J., He, Y., & Zhang, P. (2024). Development and validation of an AI literacy scale. *Journal of Artificial Intelligence Research*, 1, 17-26.
- OECD. (2021). *OECD digital education outlook 2021: Pushing the frontiers with artificial intelligence, blockchain and robots*. OECD Publishing.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903.
- République démocratique du Congo. (2021). *Politique nationale du numérique*. Gouvernement de la République démocratique du Congo.
- Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright, & M. Johnston (dir.), *Measures in health psychology: A user's portfolio* (p. 35-37). NFER-Nelson.

- Target Sarl. (2025, 19 mars). *Perception et usage de l'intelligence artificielle par les Congolais*.
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO.
- UNESCO. (2024). *AI competency framework for students*. UNESCO.
- Wang, B., Rau, P.-L. P., & Yuan, T. (2022). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, 42(9), 1324-1337.
- Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, 38(3-4), 330-336.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16, 39.
- Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, 9(2), 111-125.

Centre de recherche d'interface université-société

COMITÉ DE RÉDACTION ET ADMINISTRATION

- Directeur-Editeur : Sylvain Shomba Kinyamba, Professeur Ordinaire
- Rédacteur en Chef : Gaston Mwene Batende, Professeur Emérite
- Rédacteur en Chef adjoint : Patrick Develtere, Professeur Ordinaire
- Secrétaire de rédaction : Michel Kandolo Edungu, Chef de Travaux
- Relation avec les partenaires : Jean Pierre Mpiana, Professeur Ordinaire
- Archivage et logistique : Ingrid Mulamba Feza, Professeure Associée

CONSEIL SCIENTIFIQUE

Professeurs Baeke Viviane, Batamba Balembu Antoine ; Bongeli Yeikeo Ya Ato Emile ; Develtere Patrick ; Ekala Pierre ; Gretry Lucille ; Huib Huyse ; Inswan Bidum Pierre ; Kaminar Théo-Macaire ; Kapalayi Mutanga Roger-Vincent ; Labana Lasay'abar Jean Berck Mans ; Malemba N'sakila Gilbert ; Matabisi Nico ; Mbalanda Willy ; Mbela Hiza Login ; Mbuyi Musongela François ; Mokonda Bonza Floratin ; Mukoka Nsenda François ; Mulamba Feza Ingrid ; Muluma Munanga Albert ; Musa Alokpo Dieudonné ; Musao Kalombo Célestin ; Mutunda Mwembo Pierre ; Mpiana Tshitenge Jean Pierre ; Mwabila Malela Clément ; Mwene Batende Gaston ; Ndeke Zamba Casimir ; Ngoma Binda Elie ; Ngoy Bisongo Colby ; Nkuanzaka Inzanza Adélard ; Ongevela Van Jan, Osokonda Okenge Basile, Ponsar Paul, Reyhler Luc ; Sekimonyo Wa Magango Côme ; Shomba Kinyamba Sylvain ; Wingenga Jeannot ; Yandju Marie-Claire ; Yere Afobe Innocent Zimango Gama Romain.

Immeuble CDS n°3, Trafic, Route de l'Ouest,
Campus de l'Unikin, Tél. : 0999903563 /
shombakin@yahoo.fr,
www.mesrids.org
Dépôt légal : MR 3.02103.57117
N° ISSN : 2790-3095 (impr.)
N° ISSN : 2790-3109 (ligne)

