

Évaluer les apprentissages avec les outils numériques : plus-values, moins-values et perspectives

André Tricot



Centre national d'étude des systèmes scolaires

Plan

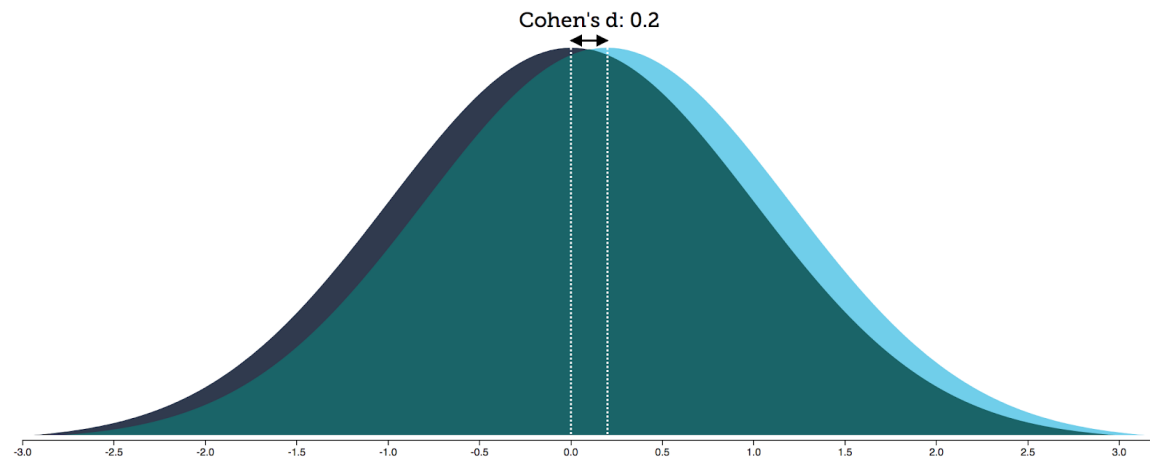
1. Pourquoi évaluer les apprentissages ?
2. Comment évaluer les apprentissages ?
3. Quels sont les plus-values et moins-values des outils numériques pour évaluer ?
4. Perspectives

Pourquoi évaluer ?

- Un pari rationnel
 - Étant donné un objectif à atteindre
 - Des moyens ont été mis en œuvre
 - Cet objectif a-t-il été atteint ?
 - Si non, comment améliorer les moyens ?

Pour comprendre la suite... les tailles d'effet

d de Cohen = différence entre les moyennes divisée par l'écart-type cumulé

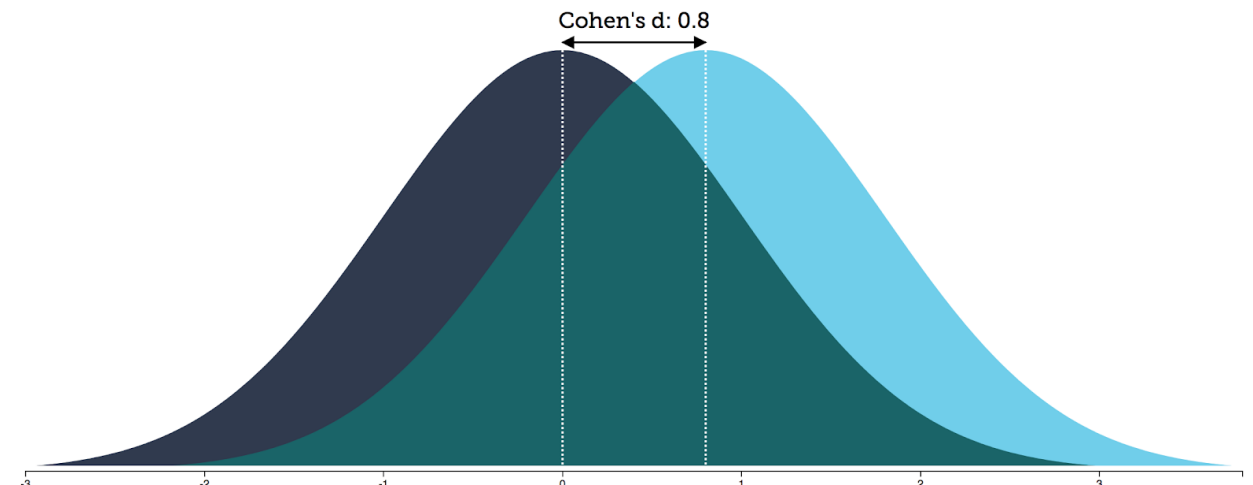
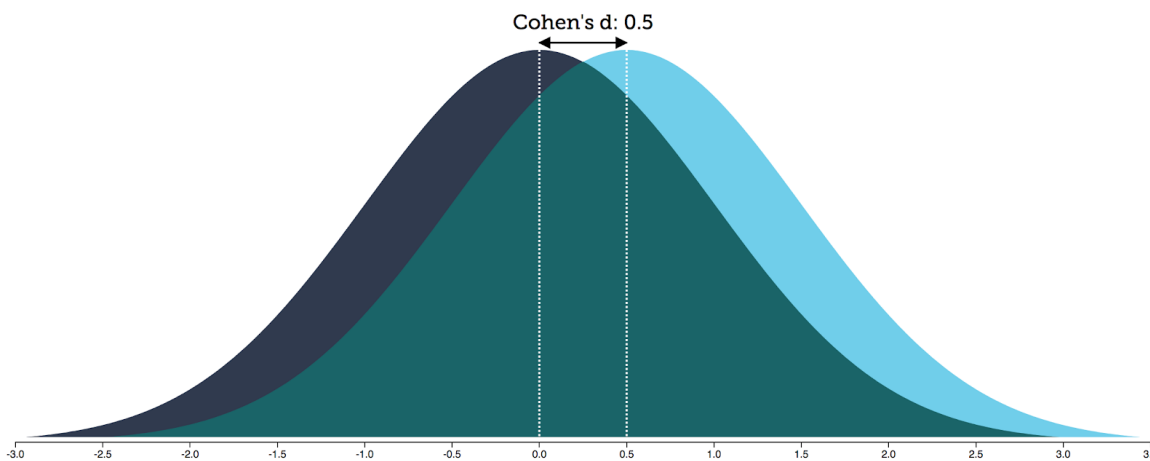


valeur d interprétation grossière

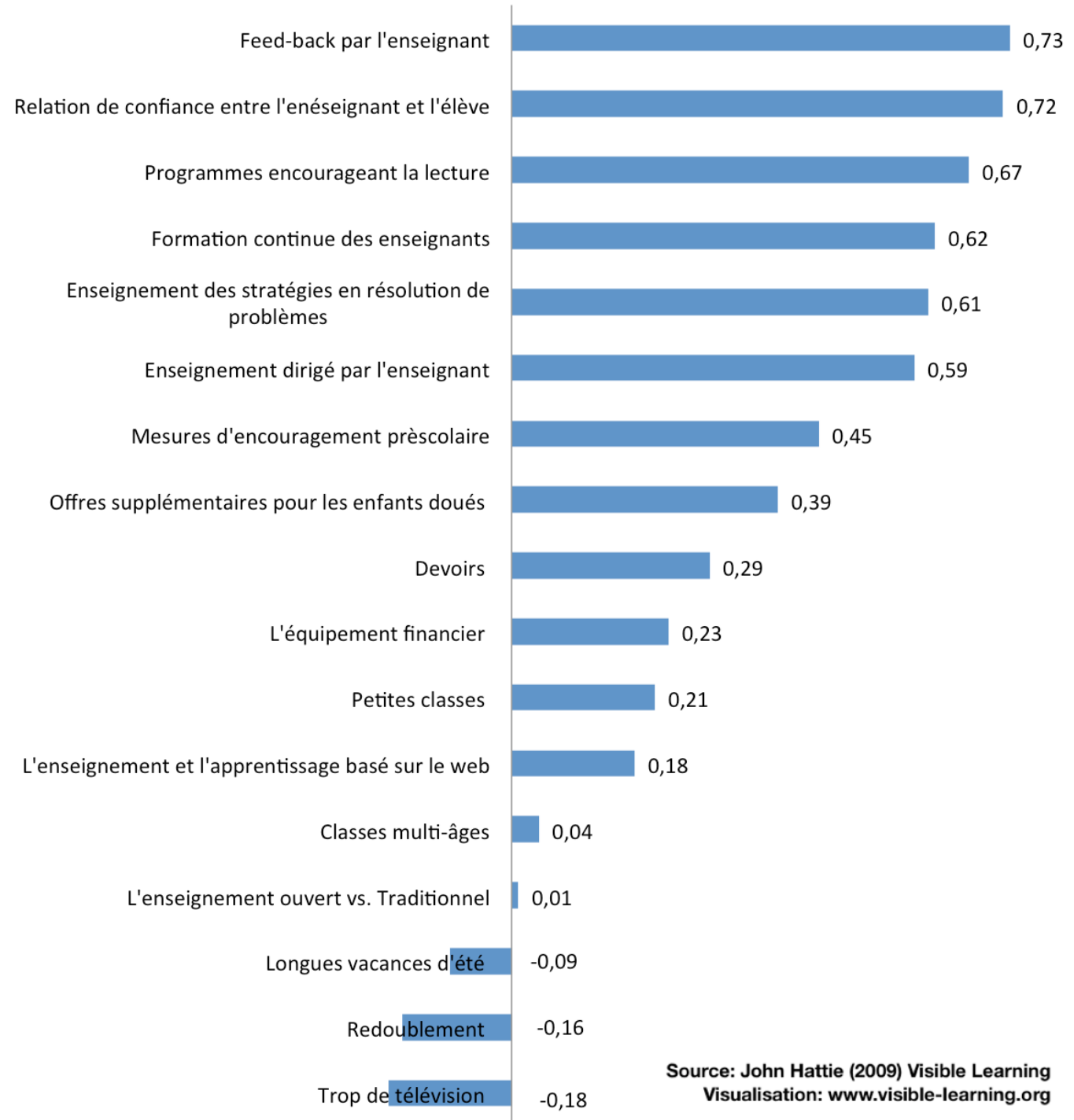
0,2 Petit effet

0,5 Effet modéré

0,8 Grand effet



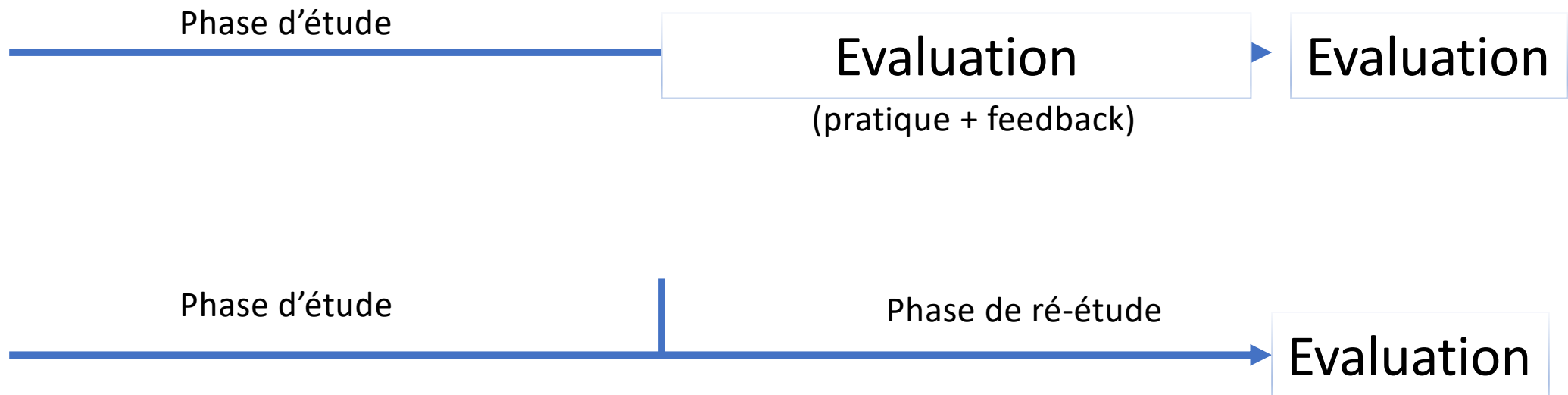
Les facteurs qui favorisent la réussite des étudiants (Hattie, 2017)



Source: John Hattie (2009) Visible Learning
Visualisation: www.visible-learning.org

L'effet test

- À temps total égal, quand les étudiant.e.s sont évalué.e.s pendant l'apprentissage elles et ils sont plus performant.e.s à l'évaluation en fin d'apprentissage
- Effet ancien (Abbott, 1909; Gates, 1917)



L'effet test : quelle efficacité ?

La méta-analyse de Rowland 2004

- 61 études, publiées ou rapportées entre 1975 et 2013
- g de Hedges moyen = 0,50
- Effet positif obtenu dans 81% des études recensées
- $g = 0,69$ quand l'intervalle entre les sessions > 1 jour
- Avec feedback $g = 0,73$; sans feedback $g = 0,39$
- Porte surtout sur des tâches de mémorisation littérale
- 77% en laboratoire

L'effet test : quelle efficacité ?

La méta-analyse de Adesope et al. (2017)

- 118 études publiées
- $g = 0,61$
- Quelle pratique lors de l'évaluation est la plus efficace ?
 - Rappel libre $g = 0,62$
 - Rappel indicé 0,58
 - Choix multiple 0,70
 - Réponse courte 0,48
 - Format mixte 0,80
 - Autre 0,57
 - Non renseigné 0,78

L'effet test : quelle efficacité ?

La méta-analyse Schwieren et al. (2017)

- Porte sur des cours de psycho, en classe
- 19 études publiées entre 1984 et 2015 (74 manipes)
- d de Cohen moyen = 0,56
- Effet positif obtenu pour 77% des effets analysés
- Avec feedback $d = 0,60$; sans feedback $d = 0,47$

L'effet test : quelle efficacité ?

La méta-analyse Yang et al. (2021)

- 222 études publiées
- Études dans des classes
- d de Cohen moyen = 0,50
- Quelle pratique lors de l'évaluation est la plus efficace ?
 - Mise en correspondance $d = 0,91$
 - Remplissage (phrases à trous) 0,77
 - Réponse courte 0,64
 - Choix multiples 0,57
 - Rappel indicé 0,32
 - Rappel libre 0,24
 - Mixte 0,34
 - Autres 0,25
 - Non renseigné 0,46

L'effet test : quelles explications ?

Roediger, Putnam & Smith (2011)

- L'évaluation permet d'identifier les lacunes en termes de connaissances,
- conduit les étudiant.e.s à apprendre plus la fois suivante,
- améliore l'organisation des connaissances,
- améliore le transfert des connaissances,
- permet de mobiliser des connaissances qui n'ont pas été préalablement évaluées,
- améliore le contrôle métacognitif,
- prévient l'interférence avec les contenus préalables quand on aborde un nouveau contenu,
- fournit un retour aux enseignant.e.s,
- fréquence encourage les étudiant.e.s à apprendre.

Résumé

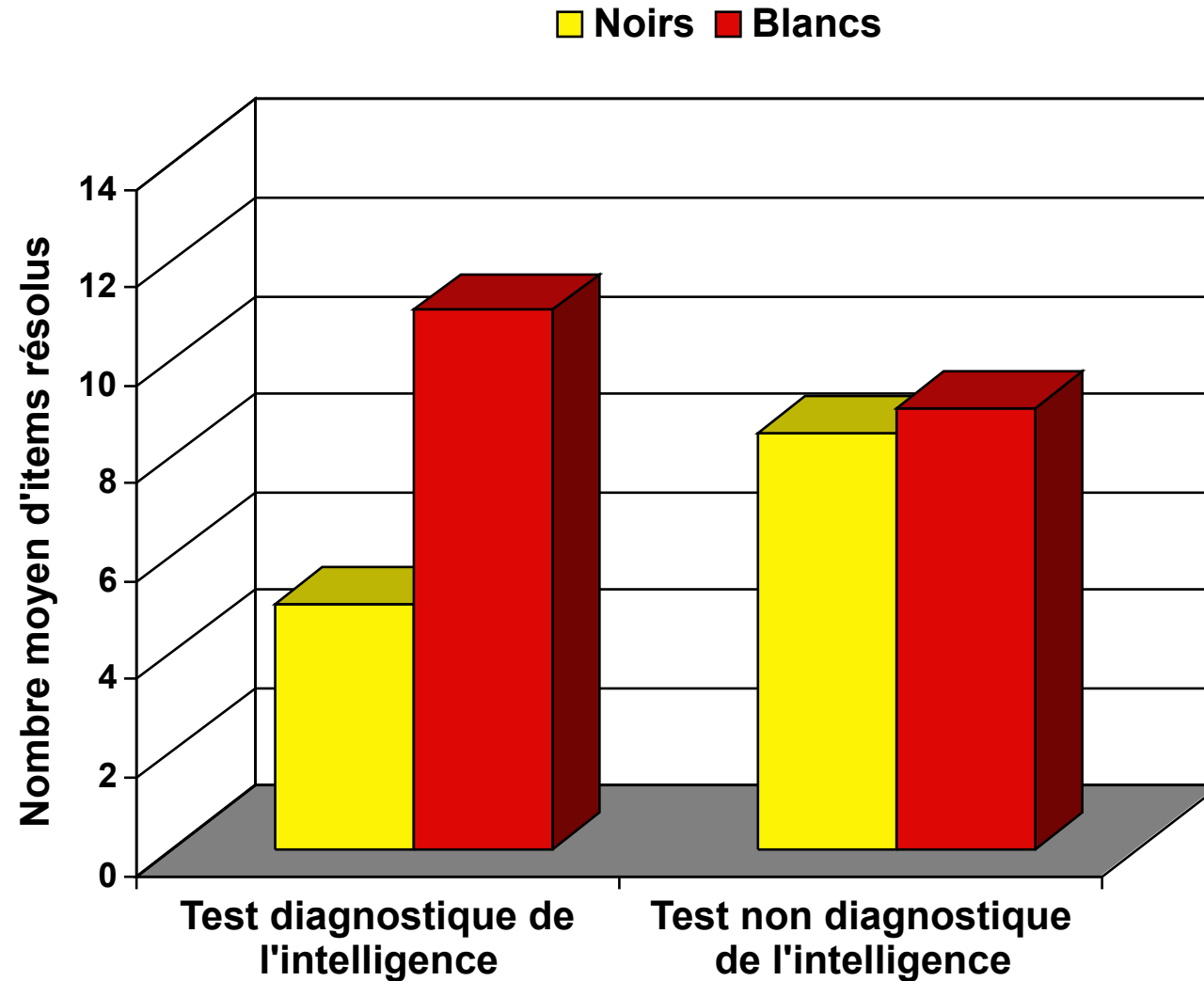
- L'évaluation pendant l'apprentissage est efficace
- Surtout avec feedback
- Elle agit principalement sur la métacognition

Plan

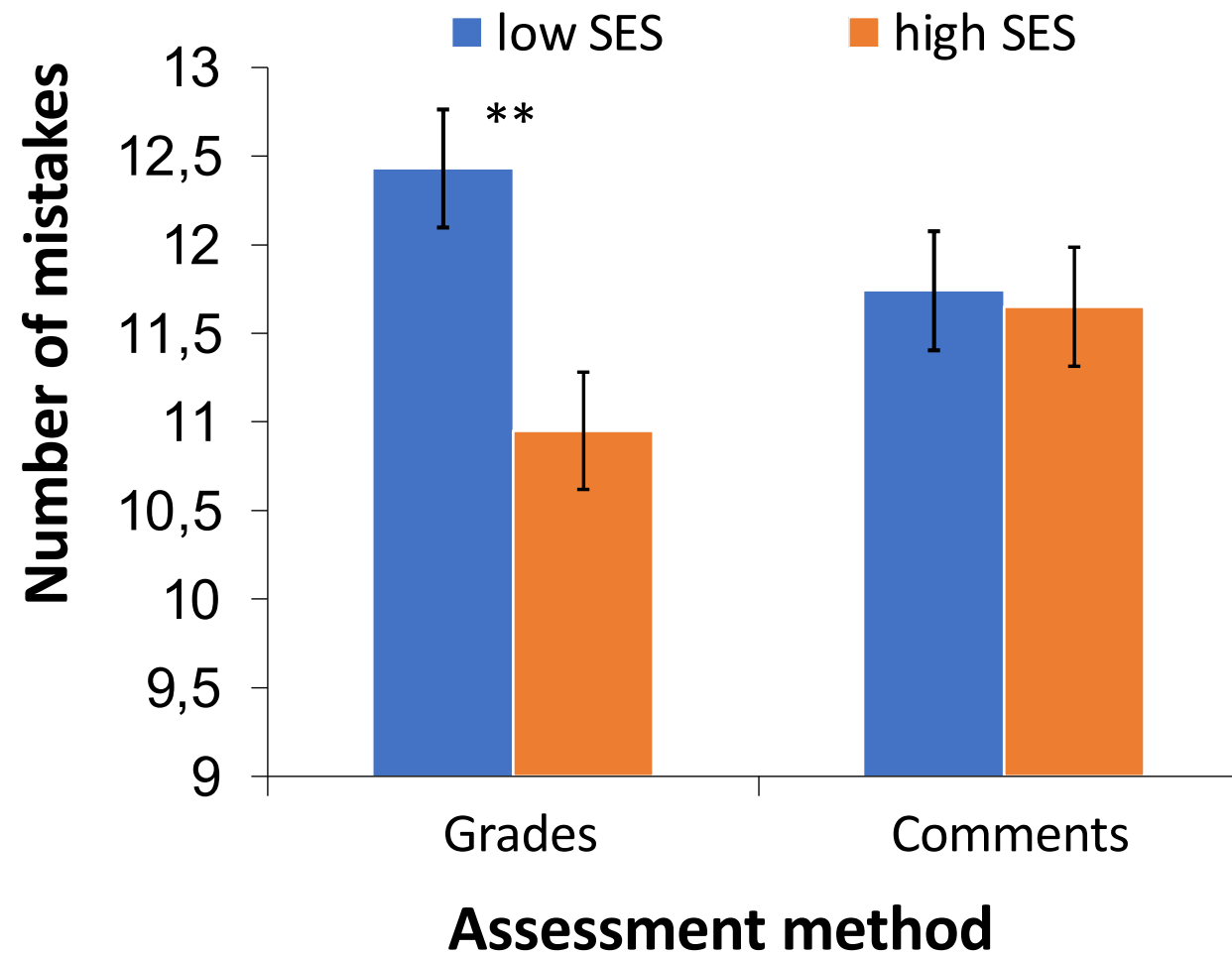
1. Pourquoi évaluer les apprentissages ?
2. Comment évaluer les apprentissages ?
3. Quels sont les plus-values et moins-values des outils numériques pour évaluer ?
4. Perspectives

La menace du stéréotype

(Steele & Aronson, 1995)



Évaluations normative et formative, et correction de dictée



Autin, F., Batruch, A., & Butera, F. (2019). The function of selection of assessment leads evaluators to artificially create the social class achievement gap. *Journal of Educational Psychology, 111*, 717–735 .

Évaluation : modèle de Hattie et Timperley (2007)

- Une évaluation efficace répond à trois questions
 - Quel est le but à atteindre ?
 - Où en suis-je ?
 - Que faire maintenant ?

Évaluation : modèle de Hattie et Timperley (2007)

- Une évaluation efficace réponds à trois questions
 - Quel est le but à atteindre ?
 - Où en suis-je ?
 - Que faire maintenant ?
- Elle porte sur quatre niveaux
 - La tâche à réaliser
 - Les processus à mettre en œuvre
 - La régulation à mettre en œuvre
 - Le soi

3 types et 4 niveaux de feedback

	Niveau de la tâche	Niveau du processus	Niveau de l'autorégulation	Niveau du soi
À propos du but	Tu as fini / pas fini, tu as atteint / pas atteint le but	Tu as bien compris, tu as bien analysé	Est-ce que tu as fini ? Penses-tu que tu as atteint le but ? Que tu as bien compris ?	

3 types et 4 niveaux de feedback

	Niveau de la tâche	Niveau du processus	Niveau de l'autorégulation	Niveau du soi
À propos du but	Tu as fini / pas fini, tu as atteint / pas atteint le but	Tu as bien compris, tu as bien analysé	Est-ce que tu as fini ? Penses-tu que tu as atteint le but ? Que tu as bien compris ?	
À propos de la manière	Tu t'y prends bien, tu fais comme il faut	Tu as fait telle erreur au cours de telle étape	Penses-tu que tu t'y prends bien ? Que tu utilises la bonne stratégie? Pourrais tu en utiliser une autre ?	

3 types et 4 niveaux de feedback

	Niveau de la tâche	Niveau du processus	Niveau de l'autorégulation	Niveau du soi
À propos du but	Tu as fini / pas fini, tu as atteint / pas atteint le but	Tu as bien compris, tu as bien analysé	Est-ce que tu as fini ? Penses-tu que tu as atteint le but ? Que tu as bien compris ?	
À propos de la manière	Tu t'y prends bien, tu fais comme il faut	Tu as fait telle erreur au cours de telle étape	Penses-tu que tu t'y prends bien ? Que tu utilises la bonne stratégie? Pourrais tu en utiliser une autre ?	
À propos de la prochaine étape	La prochaine étape à réaliser c'est ...	La prochaine étape est de comprendre..., d'analyser ...	Que vas-tu faire à la prochaine étape? Pourquoi ?	

3 types et 4 niveaux de feedback

	Niveau de la tâche	Niveau du processus	Niveau de l'autorégulation	Niveau du soi
À propos du but	Tu as fini / pas fini, tu as atteint / pas atteint le but	Tu as bien compris, tu as bien analysé	Est-ce que tu as fini ? Penses-tu que tu as atteint le but ? Que tu as bien compris ?	Ce dernier niveau n'est pas spécialement efficace. Important de ne pas le confondre avec les autres. Exemples de feedback centrés sur le soi : tu es fort, tu es intelligent, tu es le meilleur
À propos de la manière	Tu t'y prends bien, tu fais comme il faut	Tu as fait telle erreur au cours de telle étape	Penses-tu que tu t'y prends bien ? Que tu utilises la bonne stratégie? Pourrais tu en utiliser une autre ?	
À propos de la prochaine étape	La prochaine étape à réaliser c'est ...	La prochaine étape est de comprendre..., d'analyser ...	Que vas-tu faire à la prochaine étape? Pourquoi ?	

Quelques diagnostics : la route est longue encore

- L'étudiant.e manque de connaissances pour traiter la situation (il faut alors identifier lesquelles, précisément)
- L'étudiant.e manque de connaissances « implicites » (non apprises à l'école) pour traiter la situation
- L'étudiant.e est en surcharge cognitive (pour des raisons cognitives, émotionnelles, attentionnelles, affectives, etc. mais aussi manque d'automatismes)
- L'étudiant.e ne réussit pas à utiliser une connaissance qu'il ou elle la maîtrise pourtant dans d'autres situations
- L'étudiant.e ne traite l'information au bon niveau : pragmatique, formel, linguistique, etc.

Quelques diagnostics : la route est longue encore

- L'étudiant.e ne parvient pas à gérer sa propre activité
 - Il ou elle ne comprend pas ce qui est attendu, ne sait pas comment faire, comment planifier
 - L'étudiant n'a pas vérifié, il ou elle ne s'est pas relu ; ou ne sait pas *comment* vérifier
 - L'étudiant a arrêté trop tôt, il ou elle a abandonné, ne sait pas réguler sa propre activité, ne pas réussir à changer de façon de faire
 - L'étudiant utilise des stratégies stéréotypées, peu dépendantes de la tâche
 - L'étudiant ne parvient pas à savoir s'il a réussi ou pas
- L'étudiant.e ne s'engage pas, il ou elle n'essaie pas
 - Pour lui ou elle, l'école et les savoirs scolaires, sont non pertinents, il ne comprend pas leur sens, à quoi ils servent
 - Il ou elle a le statut d'étudiant.e faible, en difficultés
 - Il ou elle est persuadé.e d'échouer
 - Il ou elle préfère ne rien faire plutôt que de risquer d'échouer
 - Il ou elle ne fait que ce qu'il est certain de réussir

Résumé

- L'évaluation est soumise à de nombreux biais
- Elle ne doit pas être perçue comme menaçante
- Elle peut porter sur le niveau de la tâche, du processus ou de l'autorégulation
- Elle doit être fondée sur un véritable diagnostic

Plan

1. Pourquoi évaluer les apprentissages ?
2. Comment évaluer les apprentissages ?
3. Quels sont les plus-values et moins-values des outils numériques pour évaluer ?
4. Perspectives

L'approche tuteurs intelligents et ses limites

- *Big challenge* : réguler l'apprentissage, *i.e.* inférer un état de connaissances de l'individu, pour comprendre l'erreur qu'il a produite et lui proposer un feedback utile à l'apprentissage
 - Modèle des connaissances du domaine à enseigner
 - Modèle de la tâche et des erreurs
 - Modèle de l'élève (notamment de ses connaissances)
 - Modèle du feedback pertinent
- Apports du *machine learning, learning analytics, educational data mining*?
 - Détection automatique des étudiants en difficultés (depuis 15-20 ans)
 - Patterns de réussite, d'échec, de difficultés, d'abandon, d'engagement
- Pour l'instant : forte plus-value des feedbacks automatisés, par exemple avec les exercices (mais restreints aux domaines bien définis)

Plus-values des outils numériques

- Sur ordinateur l'élève bénéficie d'un retour immédiat, personnalisé, neutre, etc. l'enseignant de bénéficier de ce retour même quand il n'est pas en train de travailler spécifiquement avec cet élève.
- Le feedback fourni par un ordinateur perçu comme non menaçant, mais premières études peu répliquées (Resnik & Lammers, 1985; Beckmann et al., 2009 ; Mishra, 2006 ; Mumm & Mutlu, 2011 ; Nikou & Economides, 2016)

Plus-values des outils numériques

- L'ordinateur permet de fournir un feedback immédiat
 - Méta-analyse de Van der Kleij, Feskens et Eggen (2015)
 - quand l'évaluation est « élaborée », fournit une explication : $d = 0,49$
 - feedback concernant l'exactitude de la réponse a un effet nul : $d = 0,05$
 - celui qui indique la bonne réponse a un effet assez faible : $d = 0,32$
 - Méta-analyse d'Azevedo et Bernard (1995)
 - fort effet du feedback immédiat fourni par ordinateur : $d = 0,80$
 - effet faible du feedback différé : $d = 0,35$
 - Littérature importante (Erhel & Jamet, 2013 ; Narciss & Huth, 2004 ; Narciss, 2008).
 - sur le contenu des feedbacks élaborés (« comment faire », « pourquoi faire comme ça », « quelle connaissance permet de faire ça », et compréhension de la connaissance à mettre en œuvre
 - quantité de feedback fournie par l'outil numérique

Moins-values des outils numériques

- Domaines fermés, bien définis
- Difficulté à élaborer un modèle de l'élève.
- Difficulté à élaborer un diagnostic.

Résumé

- Les outils numériques apportent une plus-value à l'évaluation
- Permettent un feedback immédiat, explicatif
- Pour des étudiant.e.s en très grand nombre
- Peuvent être perçus comme moins menaçants
- Mais limite importante : domaines / tâches bien définis

Plan

1. Pourquoi évaluer les apprentissages ?
2. Comment évaluer les apprentissages ?
3. Quels sont les plus-values et moins-values des outils numériques pour évaluer ?
4. Perspectives

Domaines ou tâches ouverts

- Ex : évaluation de réactions (*automated essay evaluation*) : synthèses de Shermis & Burstein (2013) ou Dessus & Lemaire (2004)
- Logiciels comme e-rater® www.ets.org/erater, capables d'évaluer
 - contenu lexical, complexité lexicale, erreurs de grammaire, erreurs d'usage, erreurs mécaniques, style, organisation et développement des idées (thèmes), phraséologie idiomatique.
- Ces travaux ne prétendent pas évaluer aussi bien qu'un humain (Hoang & Kunnan, 2016 ; Moore & MacArthur, 2016)
- Peuvent être utilisés les étudiants comme premier retour avant de rendre leur rédaction

IA générative : exemple de l'étude de Stéphanie Roussel et Amanda Brown Peroy

- Fournir à chaque étudiant.e une version « réécrite » par ChatGPT de son propre texte
- L'étudiant.e doit analyser la version « réécrite » par ChatGPT
 - Immédiatement vs. une semaine après
 - Avec erreurs soulignées vs. erreurs soulignées + explications vs. sans
- Puis produire une nouvelle version, améliorée
- Comparaison avec un corrigé type produit par l'enseignante

Curiosité

- Développer la curiosité des enfants (9-11 ans), leur capacité à se poser des questions divergentes (pour évaluer et approfondir leurs connaissances), en les entraînant avec GPT-3 (avec des prompts)
- Abdelghani, R., Wang, Y. H., Yuan, X., Wang, T., Lucas, P., Sauzéron, H., & Oudeyer, P. Y. (2023). GPT-3-Driven Pedagogical Agents to Train Children's Curious Question-Asking Skills. *International Journal of Artificial Intelligence in Education*, 1-36.
- Abdelghani, R., Oudeyer, P. Y., Law, E., de Vulpillières, C., & Sauzéron, H. (2022). Conversational agents for fostering curiosity-driven learning in children. *International Journal of Human-Computer Studies*, 167, 102887.

Conclusion

- Les outils numériques apportent beaucoup à l'évaluation (par rapport à d'autres aspects de l'enseignement-apprentissage)
 - Feedback immédiat, explicatif, neutre
- Notamment pour les tâches / domaines bien définis
- Perspectives très encourageantes pour les tâches / domaines ouverts

Merci !

andre.tricot@univ-montp3.fr