







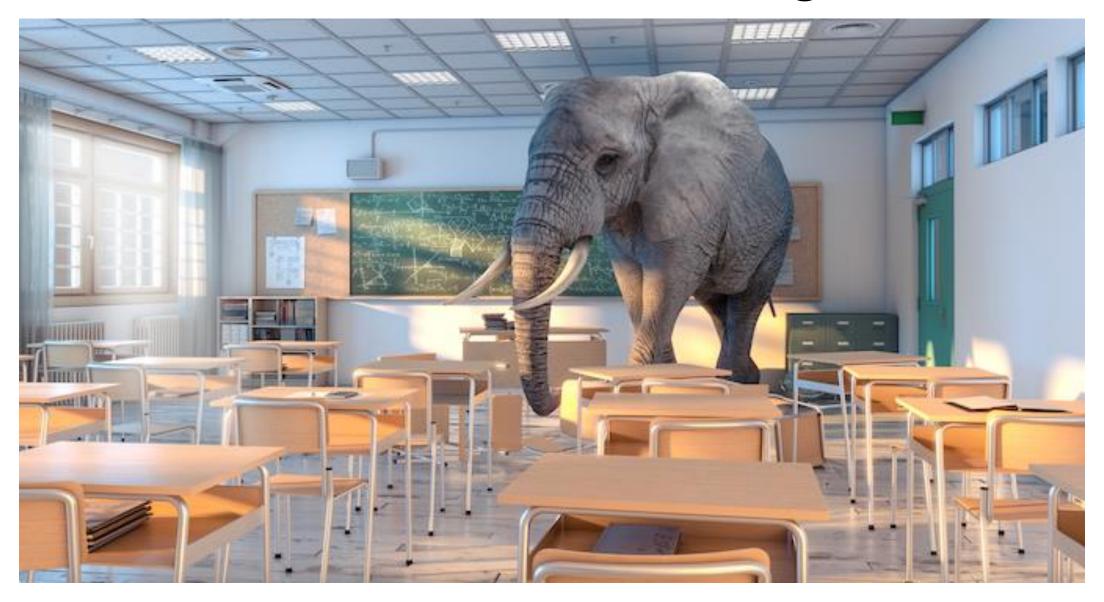
Reimagining Assessment in the Age of Artificial Intelligence

Dragan Gašević @dgasevic

> May 4th, 2024 CITERS 2024 University of Hong Kong



Generative Artificial Intelligence



https://blog.aare.edu.au/why-you-need-to-spot-the-invisible-elephant/



The Effects of ChatGPT in Schools and Why It's Getting Banned

Many schools are banning ChatGPT for plagiarism, accuracy and privacy concerns. However, the chatbot could help students and teachers with the right application.

By **April Miller**, Managing Editor of Consumer Technology at ReHack Magazine on June 13, 2023 in **Artificial Intelligence**

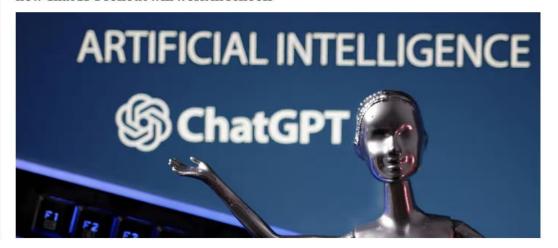
Policy decisions about the use of artificial intelligence in schools



Artificial intelligence (AI)

ChatGPT ban in Australia's public schools likely to be overturned

Government reveals a draft framework has been formulated for how ChatGPT rollout will work in schools





Artificial intelligence will not go away



Source: https://bit.ly/47NbA7D





Over 80% of jobs, especially writing and IT, are predicted have at least 19% exposure to generative AI



46% productivity increase in software engineers while maintaining code quality by using AI



Productivity in **problem solving support** increased by 35% for novice,
but not for experienced workers



Generative AI does not have inherent mechanisms to distinguish facts from falsehoods



≡

The Sydney Morning Herald

National Al

OPINION

Empowering learners for the age of artificial intelligence

Dragan Gasevic and George Siemens

December 10, 2020 - 7.30pm

Gašević, D. & Siemens, G. (2020). Empowering learners for the age of artificial intelligence, https://bit.ly/smh-empower-ai

Developing adaptive learners for the age of Al

Computers and Education: Artificial Intelligence 4 (2023) 100130



Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence



Empowering learners for the age of artificial intelligence



Gašević, D., Siemens, G., & Sadiq, S. (2023). Empowering learners for the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, *4*, 100130.

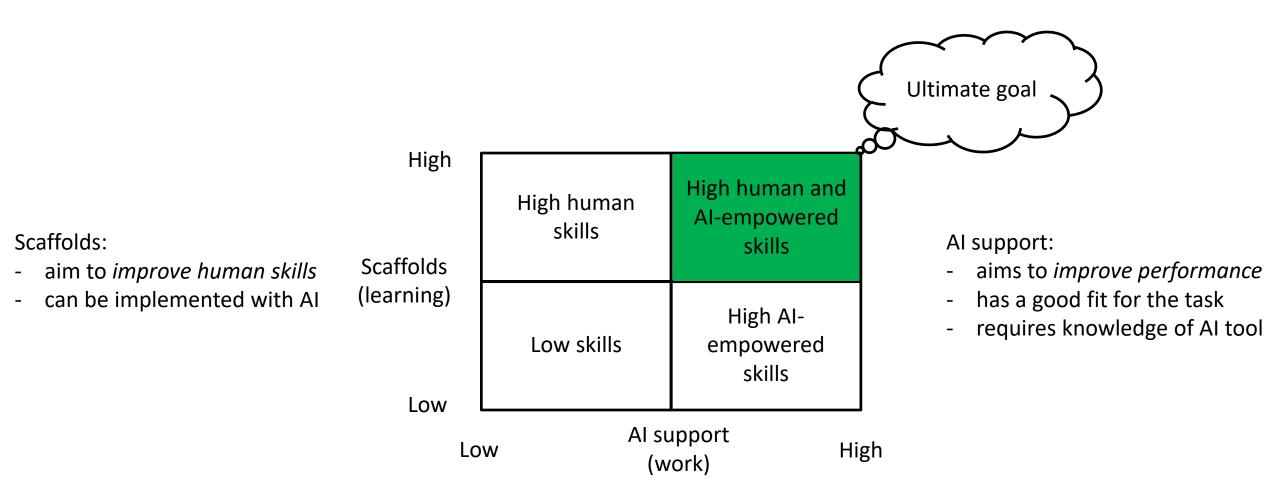


Ultimate goal

Developing human skills while benefiting from the power of Al



Towards high human and AI-empowered skills



Inspired by, but different conceptualization from

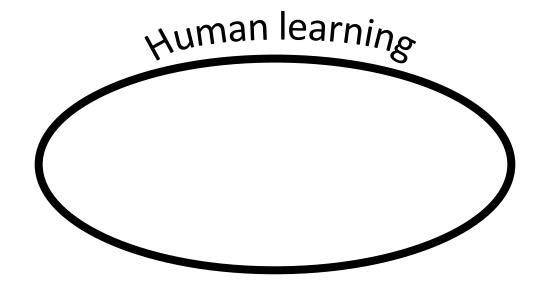
Cukurova, M. (2024). The interplay of learning, analytics, and artificial intelligence in education. *arXiv preprint* arXiv:2403.16081. Shneiderman, B. (2022). *Human-centered artificial intelligence*. Oxford University Press



Today's talk

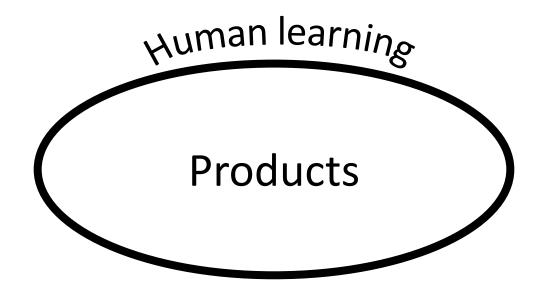
Reimaging the role of assessment in the age of Al





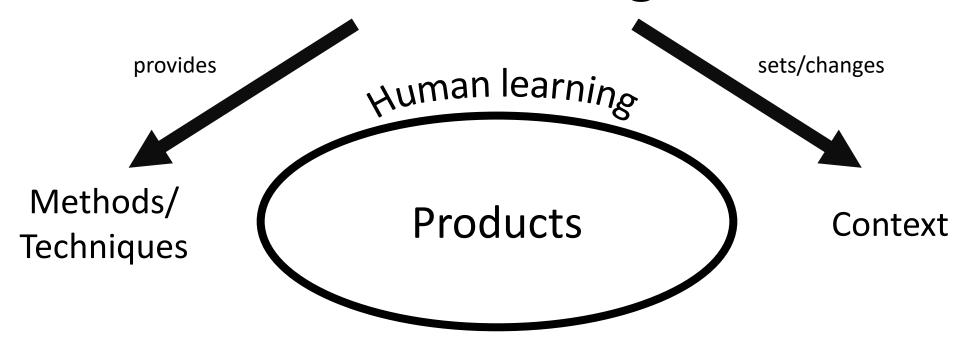
Assessment





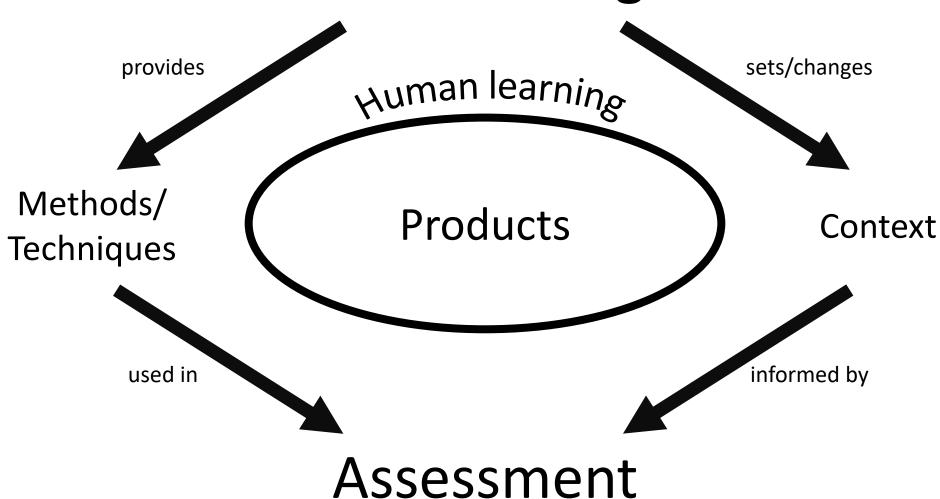
Assessment



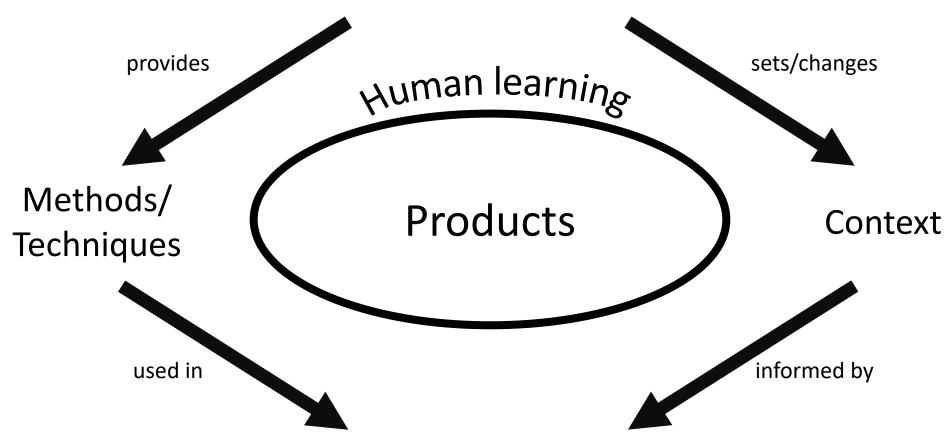


Assessment









Formative assessment



Key takeaway #1

Much promise of process assessment in authentic situations over time



Key takeaway #2

Limited understanding of combining human skills and Al support



Key takeaway #3

Metacognition and process sit at the core of assessment for the age of Al

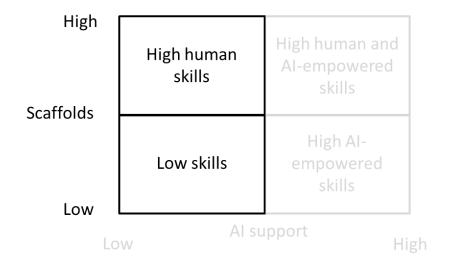


NEW METHODS – NEW CONTEXT– FINAL REMARKS



Reimaging assessment

Novel ways to enhance existing assessment practices





Opportunities for assessment provided by AI

onerous
discrete
uniform
inauthentic



feasible continuous adaptive authentic



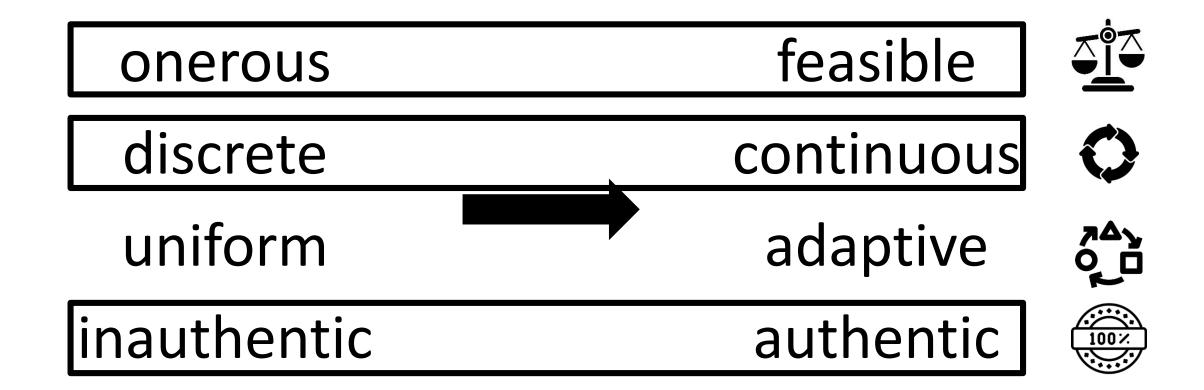








Opportunities for assessment provided by AI

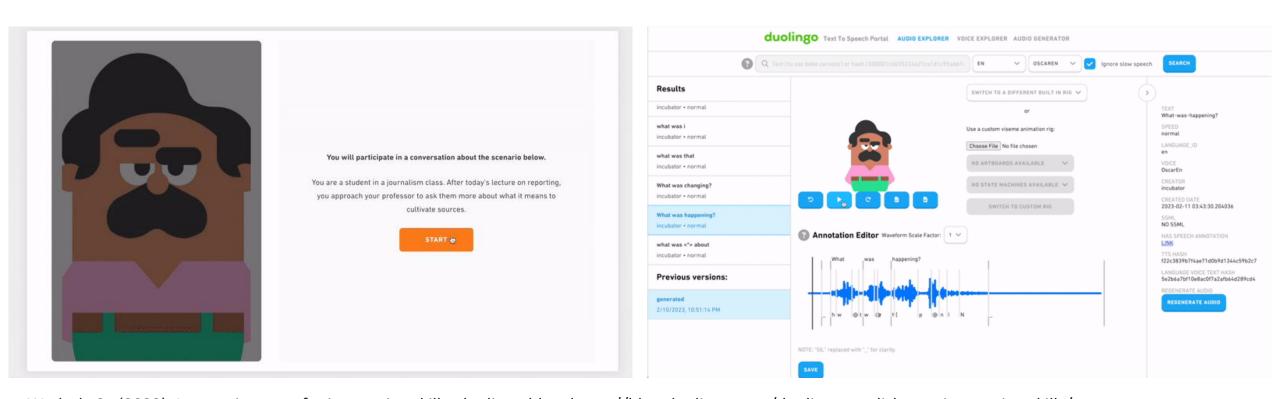






Feasible assessment with Al

Duolingo English Test - Al is part of every step



Wodzak, S., (2023). Interactive tests for interactive skills. duolingo blog, https://blog.duolingo.com/duolingo-english-test-interactive-skills/ Hao, J., von Davier, A. A., Yaneva, V., Lottridge, S., von Davier, M., & Harris, D. J. (2024). Transforming Assessment: The Impacts and Implications of Large Language Models and Generative AI. Educational Measurement: Issues and Practice, in press.





Feasible assessment with Al

Good reliability in automated scoring of essays and short answers

"Old fashioned" machine learning, deep learning, LLMs, IRT





Feasible assessment with Al

Good reliability in automated analysis of rhetorical structures

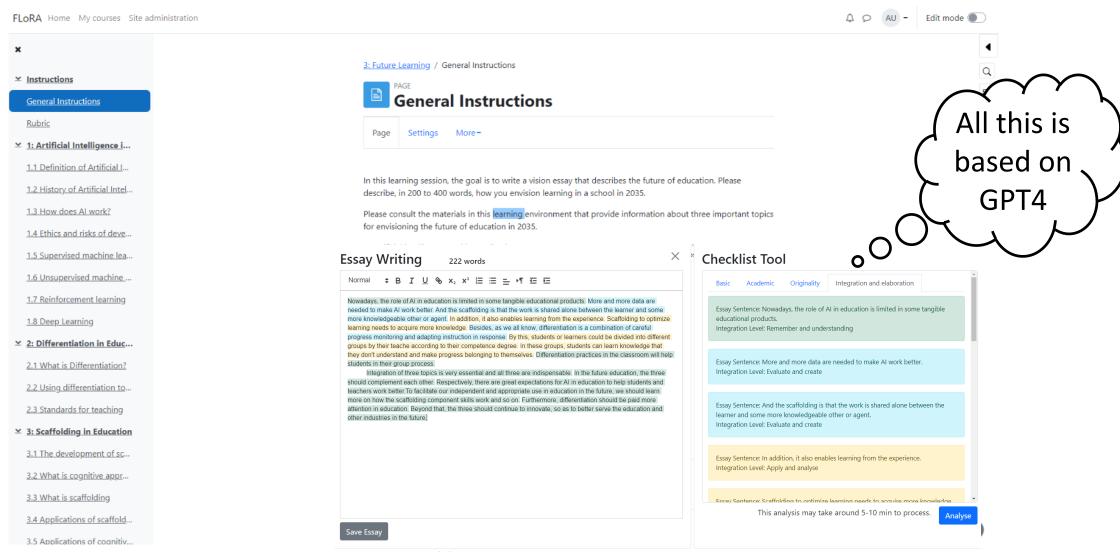
"Old fashioned" machine learning, deep learning, LLMs, IRT

Ferreira Mello, R., Fiorentino, G., Oliveira, H., Miranda, P., Rakovic, M., & Gasevic, D. (2022, March). Towards automated content analysis of rhetorical structure of written essays using sequential content-independent features in Portuguese. In *Proceedings of the 12th International Learning Analytics and Knowledge Conference* (pp. 404-414).





Feasible formative assessment with Al

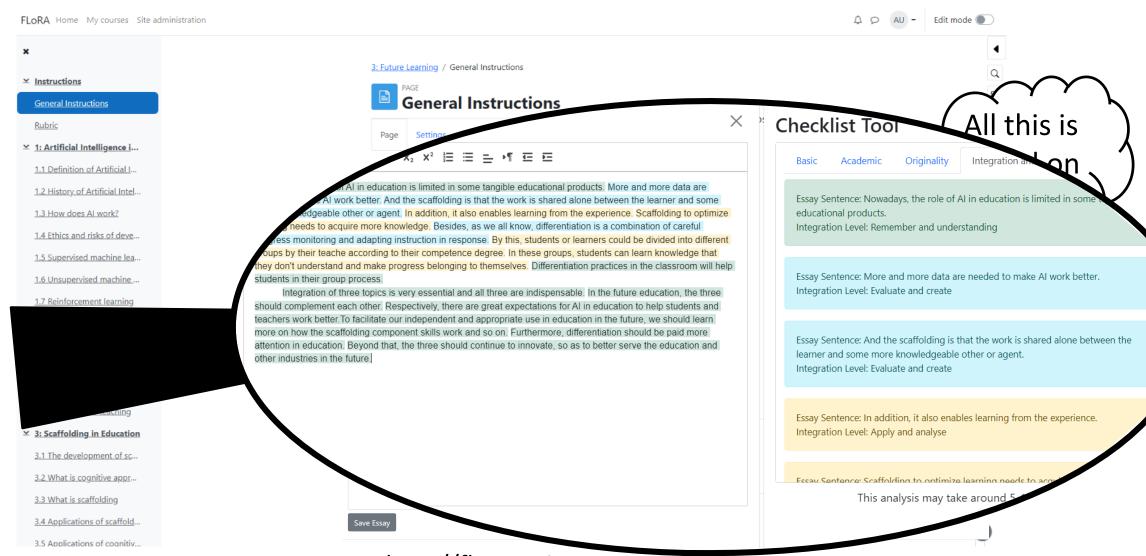


http://floraproject.org





Feasible formative assessment with Al



http://floraproject.org

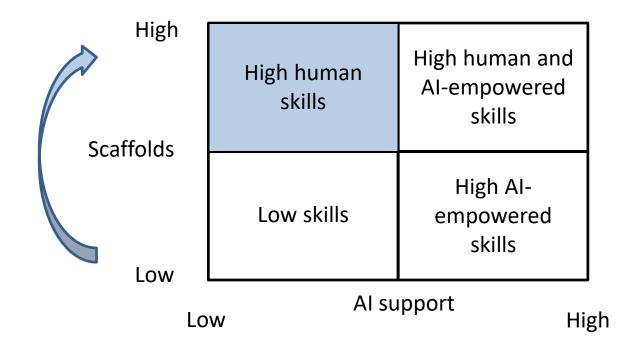


Effects of feasible formative assessment with Al

Increased judgement of learning with Al scaffolds for writing



Effects of formative assessment with Al







Continuous assessment

Process measurement is the *dark side* of assessment

Machine/deep learning and temporal/sequential modeling

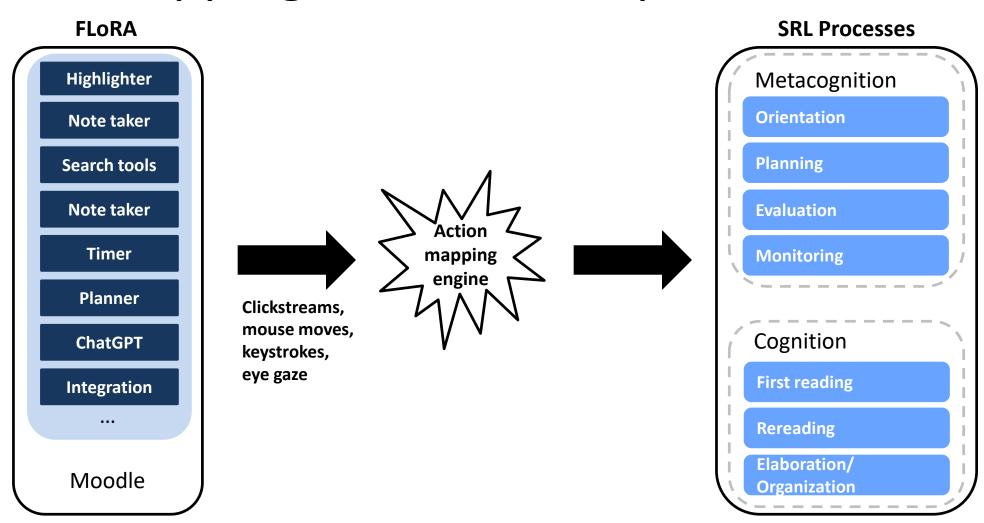


Far side of the Moon https://images.nasa.gov/details-art001e000268





Mapping trace data to processes

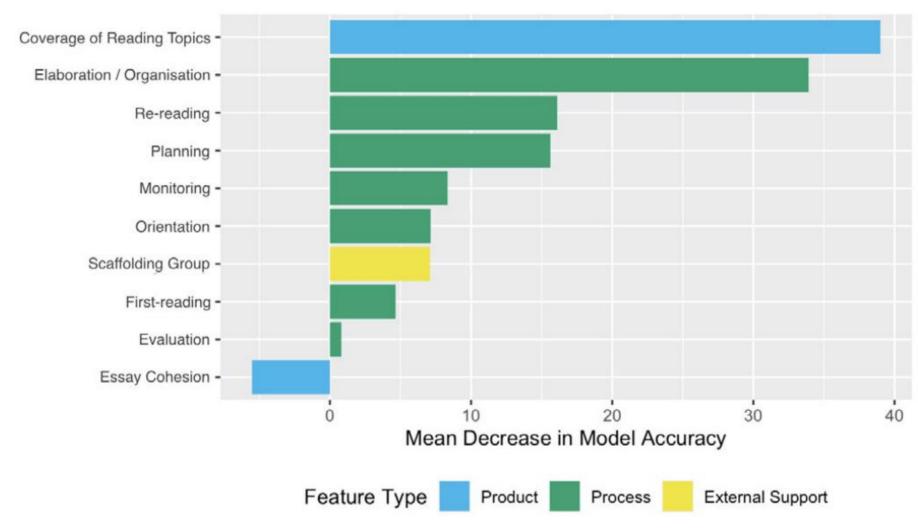


van der Graaf, J., Lim, L., Fan, Y., Kilgour, J., Moore, J., Bannert, M., ... & Molenaar, I. (2021). Do instrumentation tools capture self-regulated learning?. In *Proceedings of the 11th International Learning Analytics and Knowledge Conference* (pp. 438-448).





Process assessment with Al

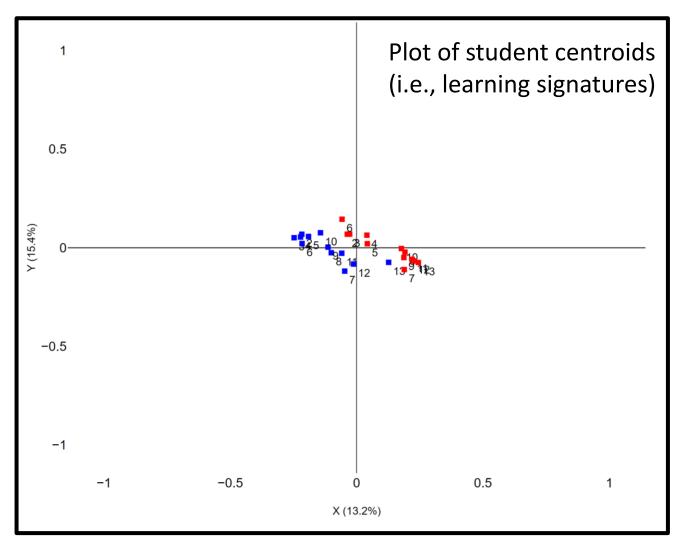


Raković, M., Iqbal, S., Li, T., Fan, Y., Singh, S., Surendrannair, S., ... & Gašević, D. (2023). Harnessing the potential of trace data and linguistic analysis to predict learner performance in a multi-text writing task. *Journal of Computer Assisted Learning*, 39, 3, 703-718.





Process assessment with Al

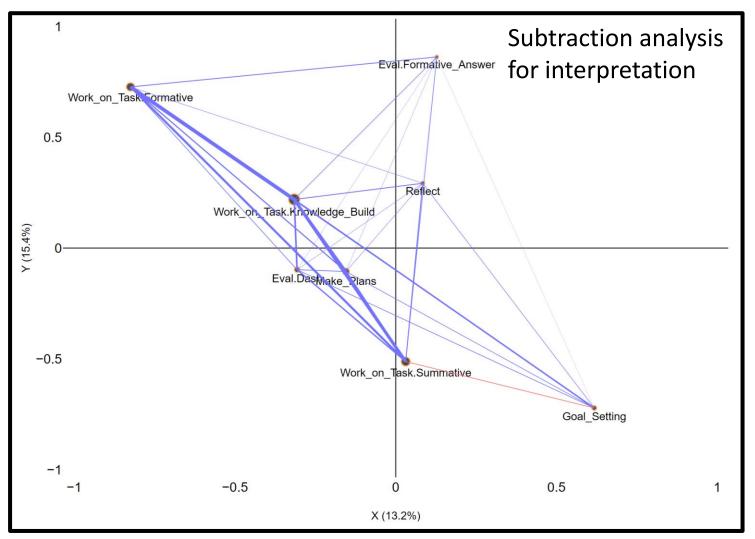


Saint, J., Gašević, D., Matcha, W., Uzir, N. A. A., & Pardo, A. (2020). Combining analytic methods to unlock sequential and temporal patterns of self-regulated learning. In *Proceedings of the tenth international conference on learning analytics & knowledge* (pp. 402-411).





Process assessment with Al

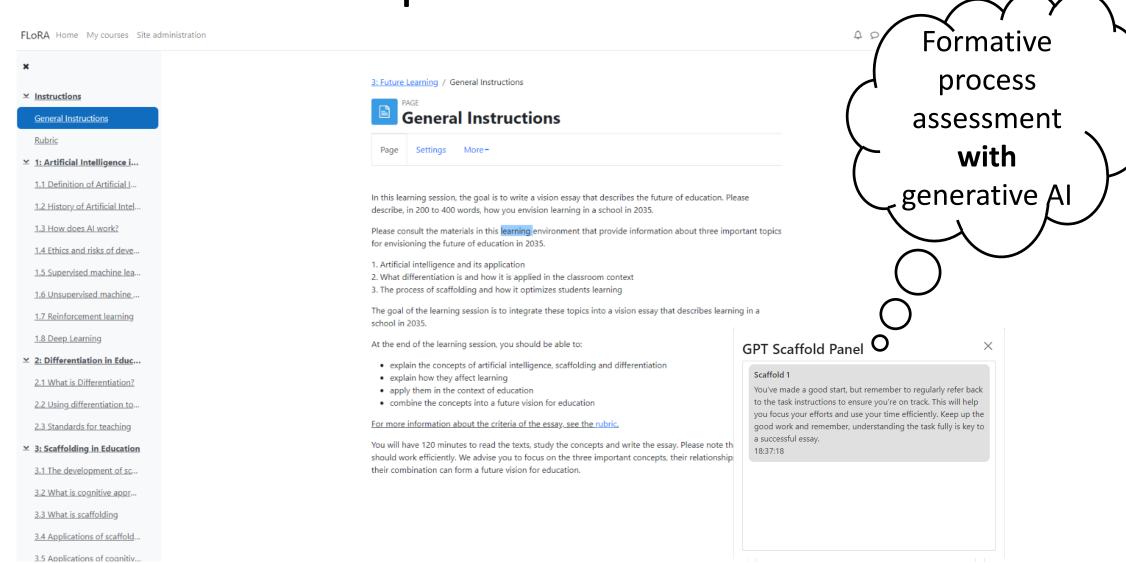


Saint, J., Gašević, D., Matcha, W., Uzir, N. A. A., & Pardo, A. (2020). Combining analytic methods to unlock sequential and temporal patterns of self-regulated learning. In *Proceedings of the tenth international conference on learning analytics & knowledge* (pp. 402-411).





Formative process assessment

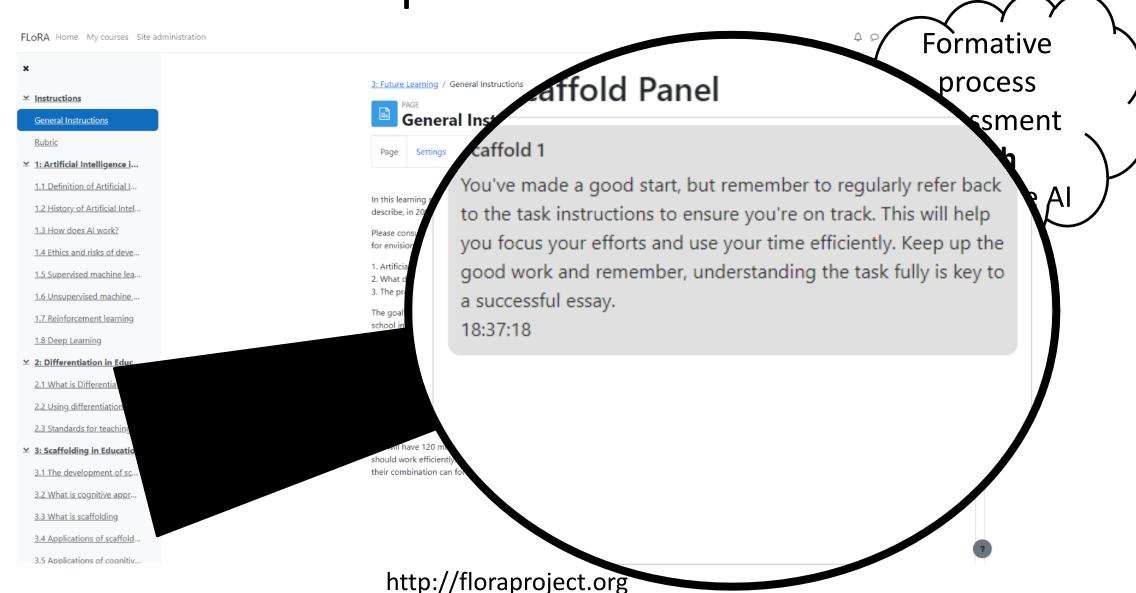


http://floraproject.org





Formative process assessment



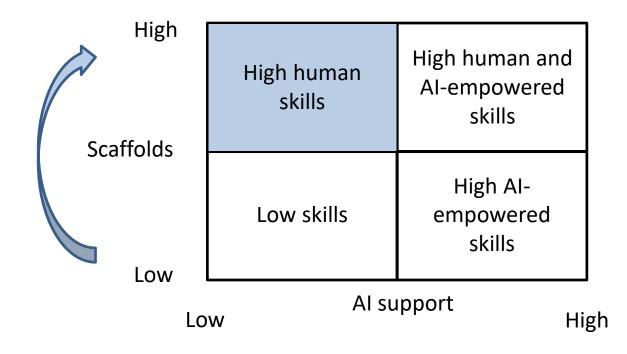


Effects of formative process assessment

Positive association with learning strategy improvement



Effects of formative assessment with Al







Authentic assessment

Measuring when and where learning happens





Authentic assessment



Assessment in physical space

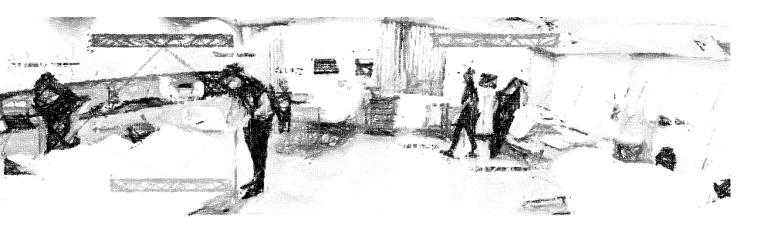
Assessment in digital space







Authentic assessment



Assessment in physical space

Assessment in digital space



















×

- ≥ 课程介绍

任务解释与说明

人工智能SP问诊

病史采集评分标准

- ン 阅读材料一:基础医学知识
 - 【解剖】胸壁结构与神经
 - 【解剖】胸膜及其神经
 - 【解剖】肺的血管和神经
 - 【解剖与生理】胸膜腔与...
 - 【解剖与生理】动脉结构...
 - 【生理】疼痛的产生机制
 - 【生理】胸痛的相关神经
 - 【生理】肺换气
 - 【病理生理】心包结构及...
 - 【病理生理】肺换气功能...
 - 【病理生理】冠脉供血与...

HTCR 3 / 人工智能SP问诊

■ 人工智能SP问诊

【场景】2023-06-13 15:09 门诊

【角色】你是一位内科门诊医生,现在来了一位病人。

【患者信息】许诺、男、25岁、未婚、住址:广东省汕头市金平区大学路21号。

【任务要求】

- 1. 请你与人工智能SP进行对话,完成完整的病史采集(系统回顾略)
- 2. 病史采集过程中,根据你自己对病史的掌握情况,在你认为合适的时候分别 诊断(1个)及鉴别诊断(至少2个)。可以参考阅读材料三中的相关疾病。
- 3. 若有知识盲区,可自行查阅阅读材料。

Formative authentic assessment with generative Al



问诊完成! 正在生成评分信

根据提供的对话及评分标准、以下是每项给出的 评分:

问诊内容评分(共80分):

1. 主诉(5分): 完全明确主要症状或体征及其发 病的具体时间或者持续时间,因此得分为 5 分。 2. 胸痛的特点 (25分): 覆盖了疼痛性质 (撕裂 样)和发病具体时间(2小时前,晚上八点)。覆 盖的信息点数量为3/14、约21%、四舍五入到 20%, 因此得分为 2 分。

2 Accopiated symptoms (20公): 对话由土明

土叫脉犬层

初步诊断: 主动脉夹层

鉴别诊断: 心肌梗塞, 肺栓塞

结束问诊

http://floraproject.org





Formative

authentic

sment

Authentic assessment with Al

× HTCR 3 / 人工智能SP问诊 ≃ 课程介绍 人工智能 任务解释与说明 【场景】2023 人工智能SP问诊 (系统回顾略); 病史采集评分标准 在你认为合适的时候分别 ≥ 阅读材料一:基础医学知识 设材料三中的相关疾病。 【解剖】胸壁结构与神经 【解剖】胸膜及其神经 【解剖】肺的血管和神经 【解剖与生理】胸膜的 【生 【病理 【病理生 【病理生理】冠脉供血与..

Medical Diagnosis

问诊完成! 正在生成评分信

10:27 DM

根据提供的对话及评分标准,以下是每项给出的评分:

问诊内容评分(共80分):

- 1. 主诉(5分): 完全明确主要症状或体征及其发病的具体时间或者持续时间,因此得分为 5 分。
- 2. 胸痛的特点 (25分): 覆盖了疼痛性质 (撕裂样) 和发病具体时间 (2小时前,晚上八点)。覆盖的信息点数量为3/14,约 21%,四舍五入到20%,因此得分为 2分。
- 2 Accopiated symptoms (20公)· 对于由土田

土叫脉犬层

初步诊断: 主动脉夹层

http://floraprojc 鉴别诊断: 心肌梗塞, 肺栓塞

结束问诊





× HTCR 3 / 人工智能SP问诊 ≃ 课程介绍 人工智能 任务解释与说明 【场景】2023 人工智能SP问诊 (系统回顾略); 病史采集评分标准 在你认为合适的时候分别 ≥ 阅读材料一:基础医学知识 讨对料三中的相关疾病。 【解剖】胸壁结构与神经 【解剖】胸膜及其神经 【解剖】肺的血管和神经 【解剖与生理】胸膜的 【生 【病理 【病理生 【病理生理】冠脉供血与..

Medical Diagnosis

2. 问诊过程(5分): 对话结构并不清晰, 条理也 不分明, 且未按照标准内容完整覆盖信息, 得分 为1分。

3. 问诊语言(5分):语言使用尚可,但对话未示 范解释医学术语或问题关注细节, 因此给分为 3

4. 宣教及对病史信息进行小总结(5分): 对话中 未进行宣教或总结,得分为1分。

总计分数:

问诊内容评分: 7/80 问诊技巧评分: 6/20

总计: 13 / 100

土动脉光层

初步诊断: 主动脉夹层

http://floraproje 鉴别诊断: 心肌梗塞, 肺栓塞

结束问诊

Formative authentic

sment



What's missing?

Assessment that has all four properties at the same time







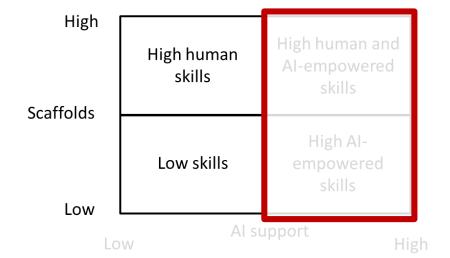




What's missing?



Modern assessment is more than measurement of Al literacy





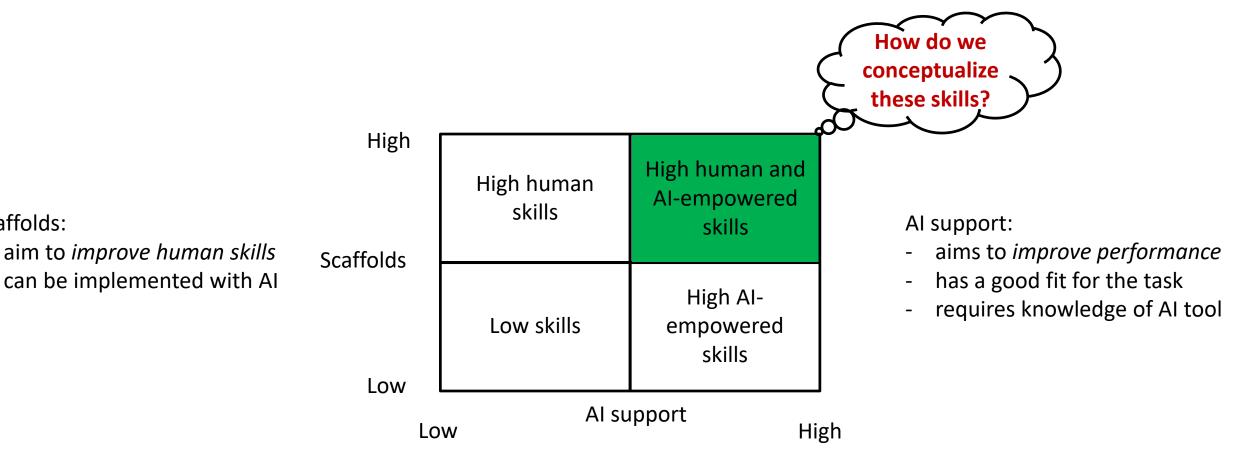
NEW METHODS – NEW CONTEXT – FINAL REMARKS

Key challenge

We need to know first what we are assessing



Towards high human and Al-empowered skills



Inspired by, but different conceptualization from

Scaffolds:

Cukurova, M. (2024). The interplay of learning, analytics, and artificial intelligence in education. arXiv preprint arXiv:2403.16081. Shneiderman, B. (2022). *Human-centered artificial intelligence*. Oxford University Press



Key challenge

Limited evidence of *combined* high human and Al-empowered skills



Key challenge

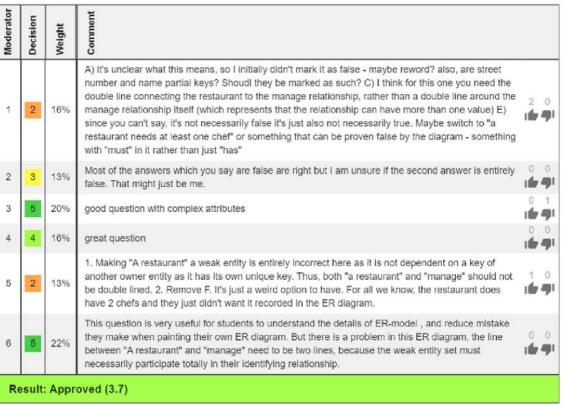
Limited evidence of *combined* high human and Al-empowered skills

Example study #1 – peer feedback



Al support and judgment of learning





Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., Siemens, G. (2024). Impact of Al Assistance on Student Agency. Computers & Education, 210, 104967.



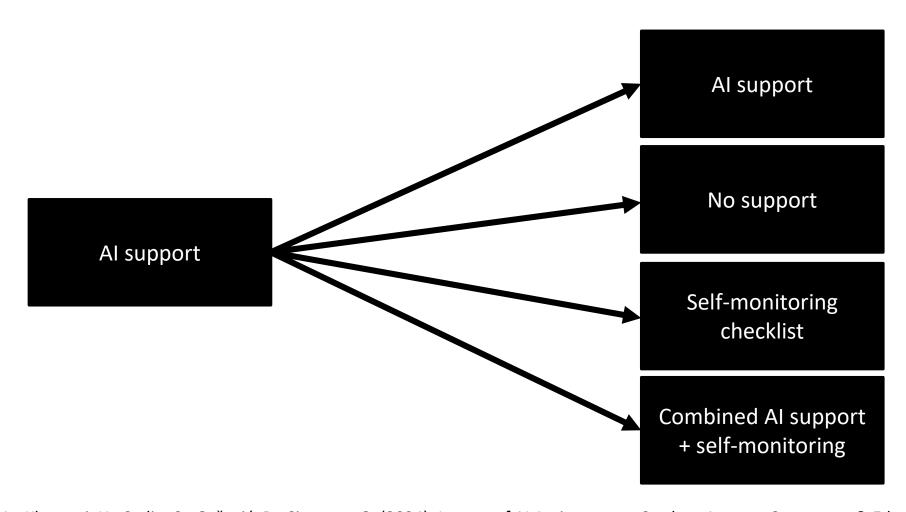
Al support and judgment of learning

Our algorithms have flagged that this comment may not be helpful.	Resource Feedback Please evaluate the resource based on the following criteria:				I don't want	to moderate this resource	
	Alignment with course content & objectives:	Poor	Needs Improvement	Satisfactory	Great	Outstanding	
 Please ensure that the comment: Is specific to this resoruce 	Correctness, clarity & ease of understanding:	Poor	Needs Improvement	Satisfactory	Great	Outstanding	
 Suggests meaningful improvements for the author 	Appropriateness of difficulty:	Poor	Needs Improvement	Satisfactory	Great	Outstanding	
 Aligns with the grades assigned in the rubric 	Encouragement of critical thinking and reasoning:	Poor	Needs Improvement	Satisfactory	Great	Outstanding	
MY FEEDBACK IS APPROPRIATE - SUBMIT ANYWAY EDIT COMMENT Justify your responses & provide feedback Please provide constructive feedback & justify your responses to the author so they can improve the resource. Align feedback with rubric Be detailed & specific Suggest improvements						①	
Our algorithms have flagged that this comment may not be helpful. • Your moderation comment is too similar to your previous comments The content of the question is good. However, please revise the answer. Distractor C is also correct in addition to the option A that has been selected. It would be better to change the question either by making it a multiple-answer or negating option C.							
Previous Comment Date	Decision Please rate the overall quality of this resource based on the criteria	Decision Please rate the overall quality of this resource based on the criteria above.					
Interesting question with well explanation. I could not add Sep 01, 2021	The overall quality of this resource is:	Poor	Needs Improvement	Satisfactory	Great	Outstanding	
more to it. Well done.	Rate your confidence in assessing this resource:	Very low	Low	Medium	High	Very High	
MY FEEDBACK IS APPROPRIATE - SUBMIT ANYWAY	COMMENT	SUBI	MIT				

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., Siemens, G. (2024). Impact of Al Assistance on Student Agency. Computers & Education, 210, 104967.



Longitudinal impact on judgment of learning



Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., Siemens, G. (2024). Impact of AI Assistance on Student Agency. Computers & Education, 210, 104967.

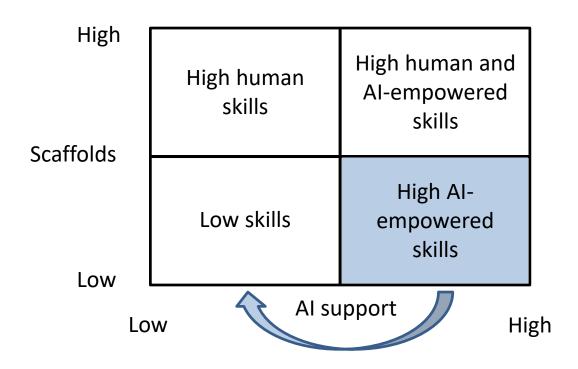


Al support and judgment of learning

Benefits deteriorate when Al support is removed



Long term effects of Al-support



Concern

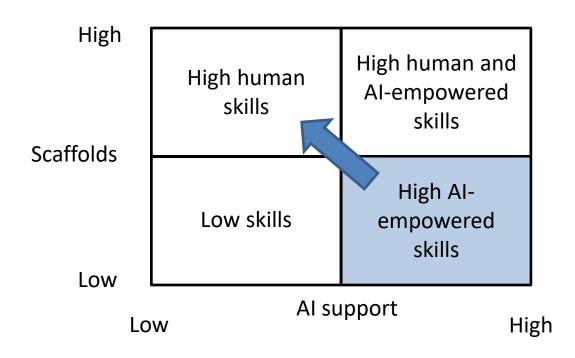


Al support and judgment of learning

Benefits remain when Al support replaced by scaffolds



Al-support after replaced with scaffolds



Positive, but not novel

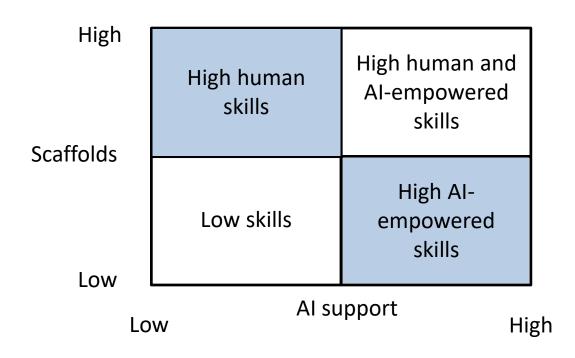


Al support and judgment of learning

Combining scaffolds and AI support does not produce additive effects



Towards high human and AI-empowered skills



No additive effect of scaffolds and Al support



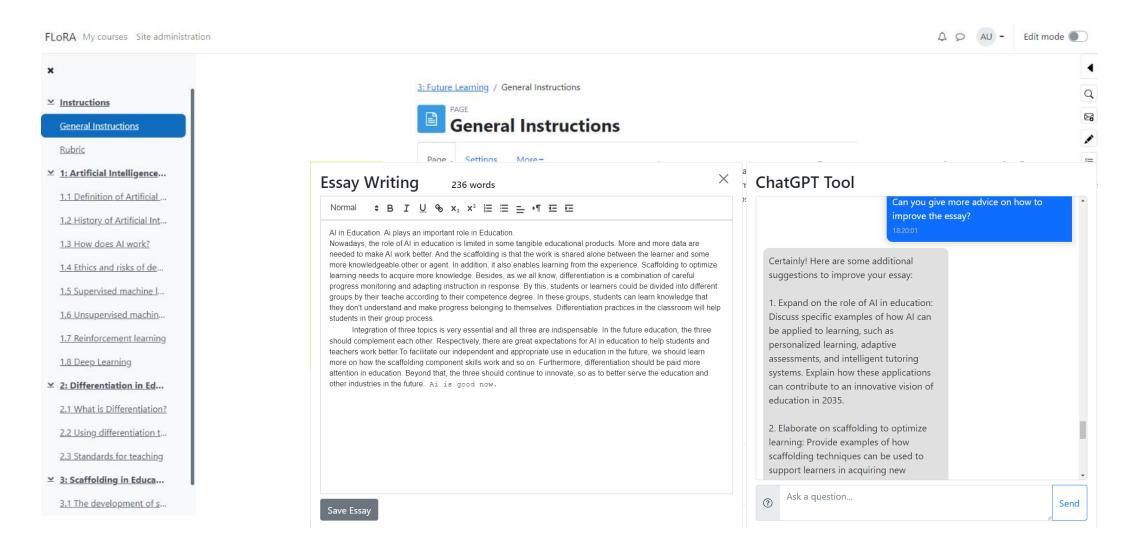
Key challenge

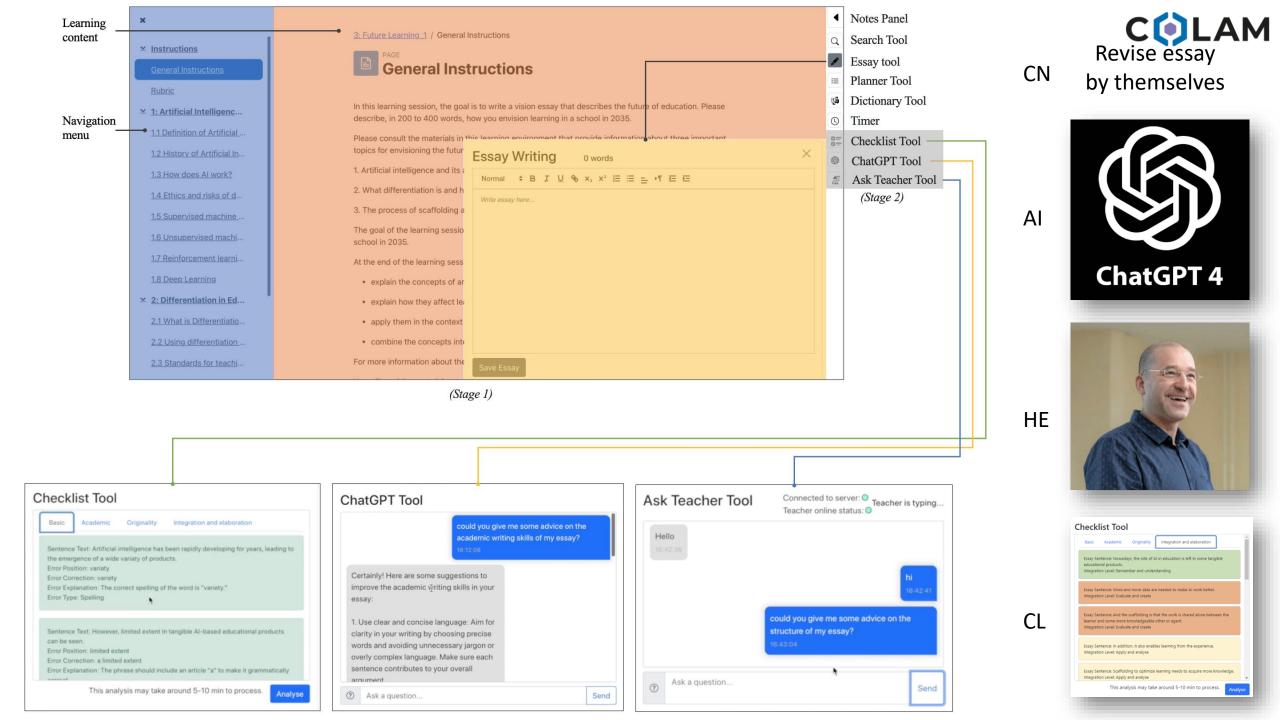
Limited evidence of *combined* high human and Al-empowered skills

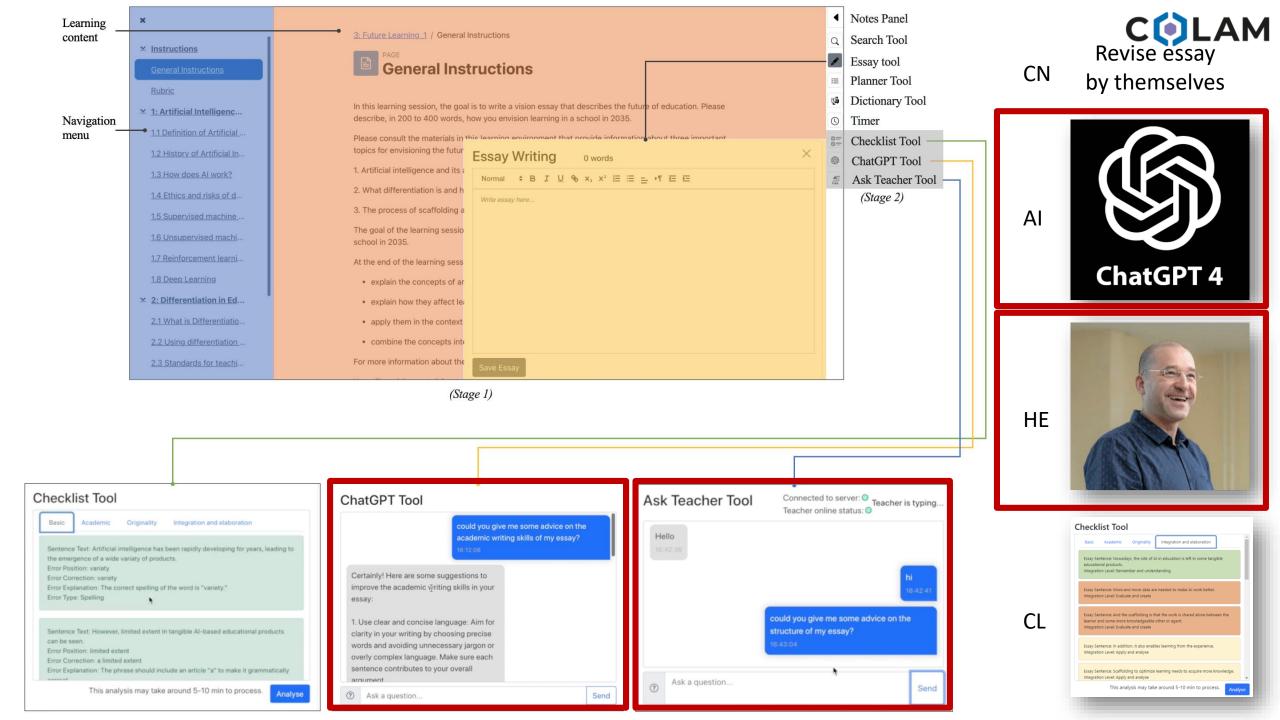
Example study #2 – information problem solving



ChatGPT in FLoRA









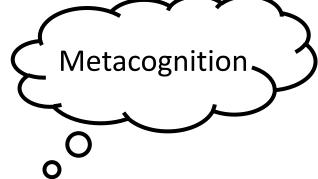
Writing performance

Unsurprisingly writing performance is improved when using ChatGPT

ChatGPT significantly higher essay scores than the other three



Metacognition in information problem solving

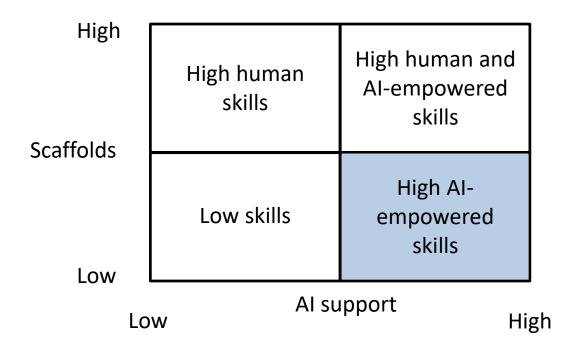


Limited evidence of judgment about information provided by generative Al



Towards high human and Al-empowered skills

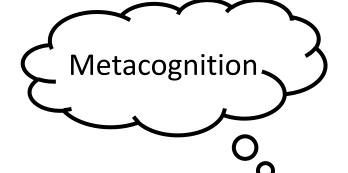
Ask ChatGPT



Chen, A., Xiang, M., Zhou, J., Jia, J., Shang, J., Li, X., Gašević, Fan, Y. (2024). Unpacking Help-Seeking Processes through Multimodal Learning Analytics: A Comparative Study of Learning Facilitated by ChatGPT and Human Expert, in preparation.



Metacognition in information problem solving

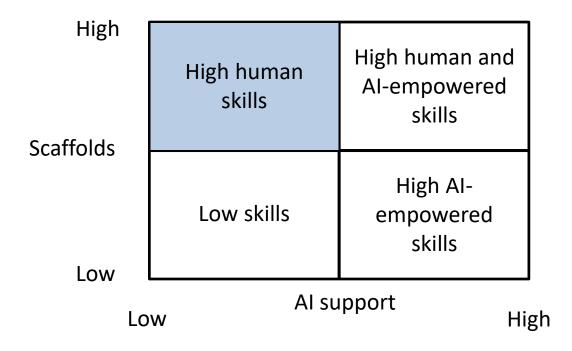


Always observed evidence of judgment about human provided information



Towards high human and Al-empowered skills

Ask Teacher





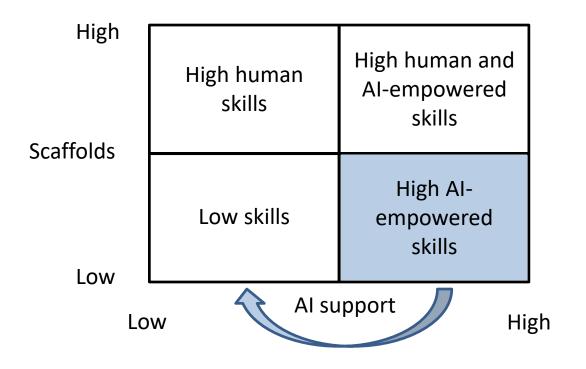
Metacognitive engagement

What are long term implications of limited metacognitive activity?

Unreliable information produced by GenAl



Long term effects of Al-support



Concern

High processing fluency influence selection of strategy



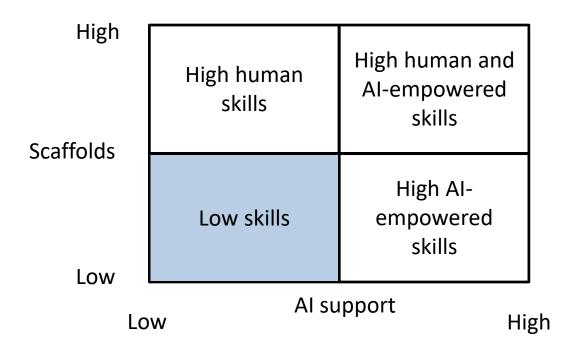
Inadvertent deception of ChatGPT

Inaccurate and verbose, but users prefer ChatGPT responses

Processing fluency – illusion of truth effect



Towards high human and AI-empowered skills



Al support:

- aims to improve performance
- has a fit for the task
- knowledge of AI tool



Corollary #1

Metacognition is a key dimension of high human and Al-empowered skills



Corollary #2

Longitudinal approach to assessment of high human and Al-empowered skills





Evidence-centered design



Example

Task

Evidence



High human and Al-empowered medical diagnosing



Task





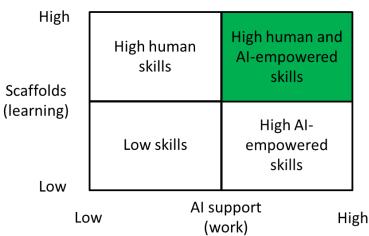
Tasks to be authentic





Task Evide





Tasks to include scaffolds, AI support, and metacognition

Task





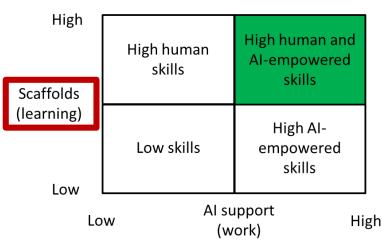
Evidence

Direction









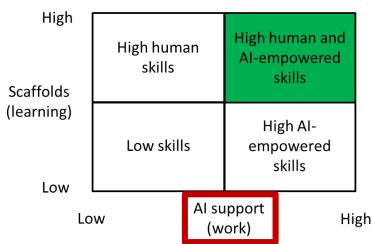
Tasks to include

scaffolds. Al support, and metacognition









Tasks to include scaffolds, Al support, and metacognition

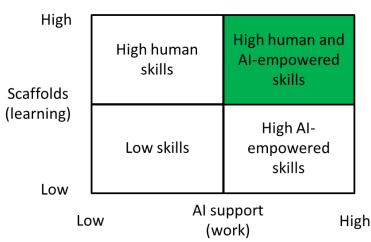


Dire

Task

Evidence





Tasks to include scaffolds, AI support, and metacognition



Task









Tasks to be over extended period of time

Longitudinal assessment



Example

Task

ividence

High human and Al-empowered medical diagnostics



Human skills





Scaffolding metacognition





Another role: COLAM
Instructing doctors
as scaffolding

Scaffolding metacognition as scaffolding









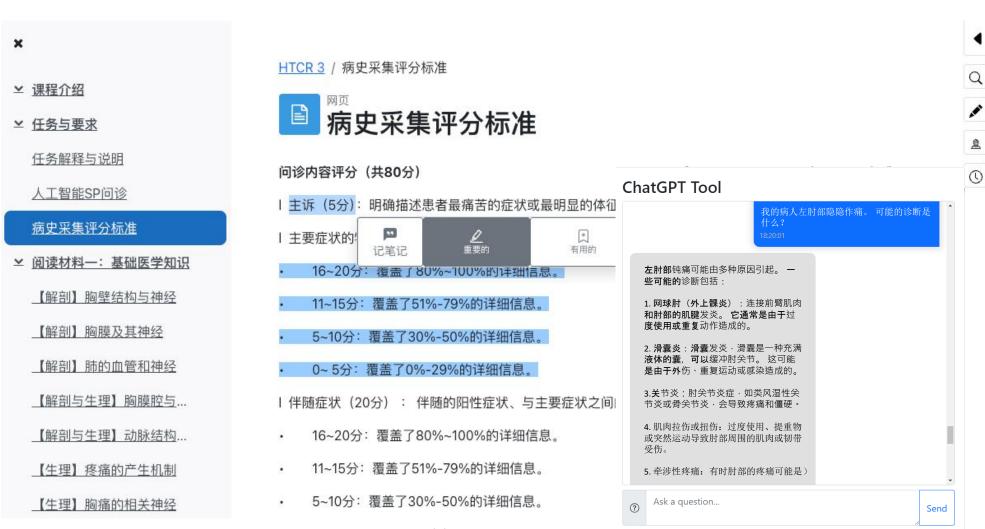
Al-empowered skills



http://floraproject.org



Metacognition



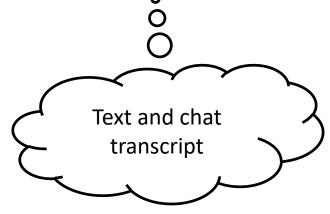
http://floraproject.org



Task Evidence

Assessment to include data about process and product

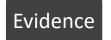
Clicks, mouse/eye
movements,
keystrokes

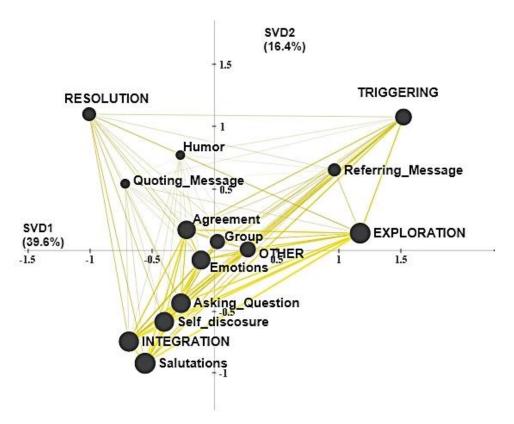




Assessment of modern skills

Task





Combining large language models with epistemic network analysis

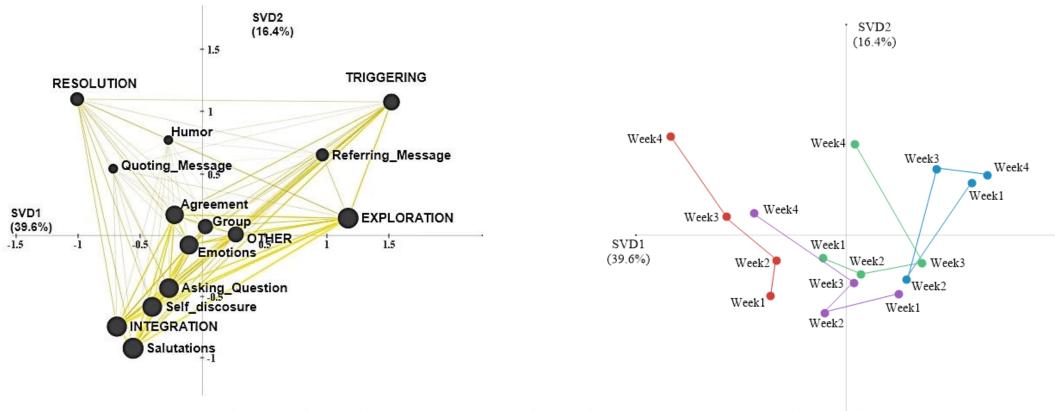
Rolim, V., Ferreira, R., Lins, R. D., & Gašević, D. (2019). A network-based analytic approach to uncovering the relationship between social and cognitive presences in communities of inquiry. *The Internet and Higher Education*, 42, 53–65.



Assessment of modern skills

Task





Combining large language models with epistemic network analysis

Rolim, V., Ferreira, R., Lins, R. D., & Gašević, D. (2019). A network-based analytic approach to uncovering the relationship between social and cognitive presences in communities of inquiry. *The Internet and Higher Education*, 42, 53–65.



















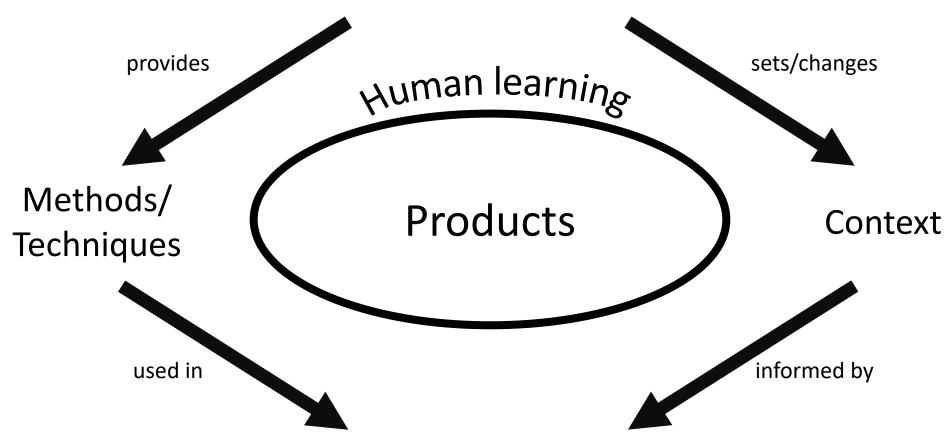
Assessment is feasible and scalable due to the use of Al



NEW METHODS – NEW CONTEXT – FINAL REMARKS



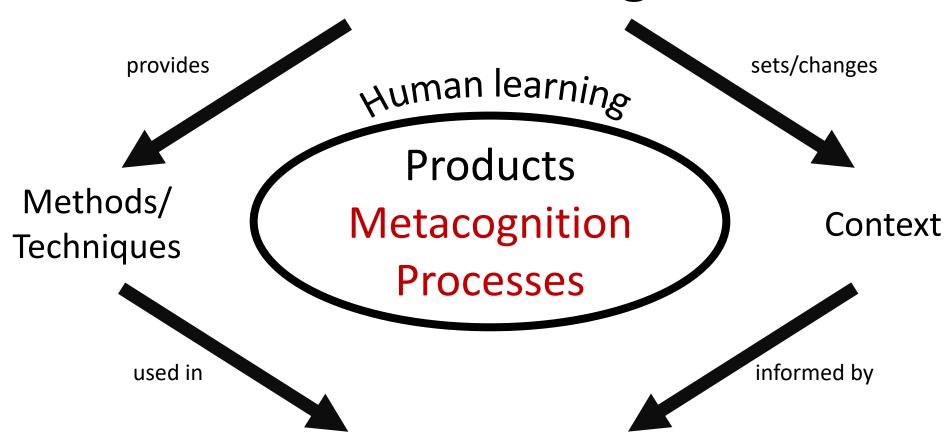
Artificial intelligence



Formative assessment



Artificial intelligence



Formative assessment



Prerequisite for assessment in the age of Al

Urgent need to understand *combined* high human and Al-empowered skills



Al and human learning



Embracing the opportunity

Image: Flaticon.com









Reimagining Assessment in the Age of Artificial Intelligence

Dragan Gašević @dgasevic







