



RODEZ

Évaluer c'est former ! Quel apport du numérique ?

Franck Silvestre - franck.silvestre@irit.fr
JARIUT - IUT Blagnac - 13 avril 2023

Franck Silvestre

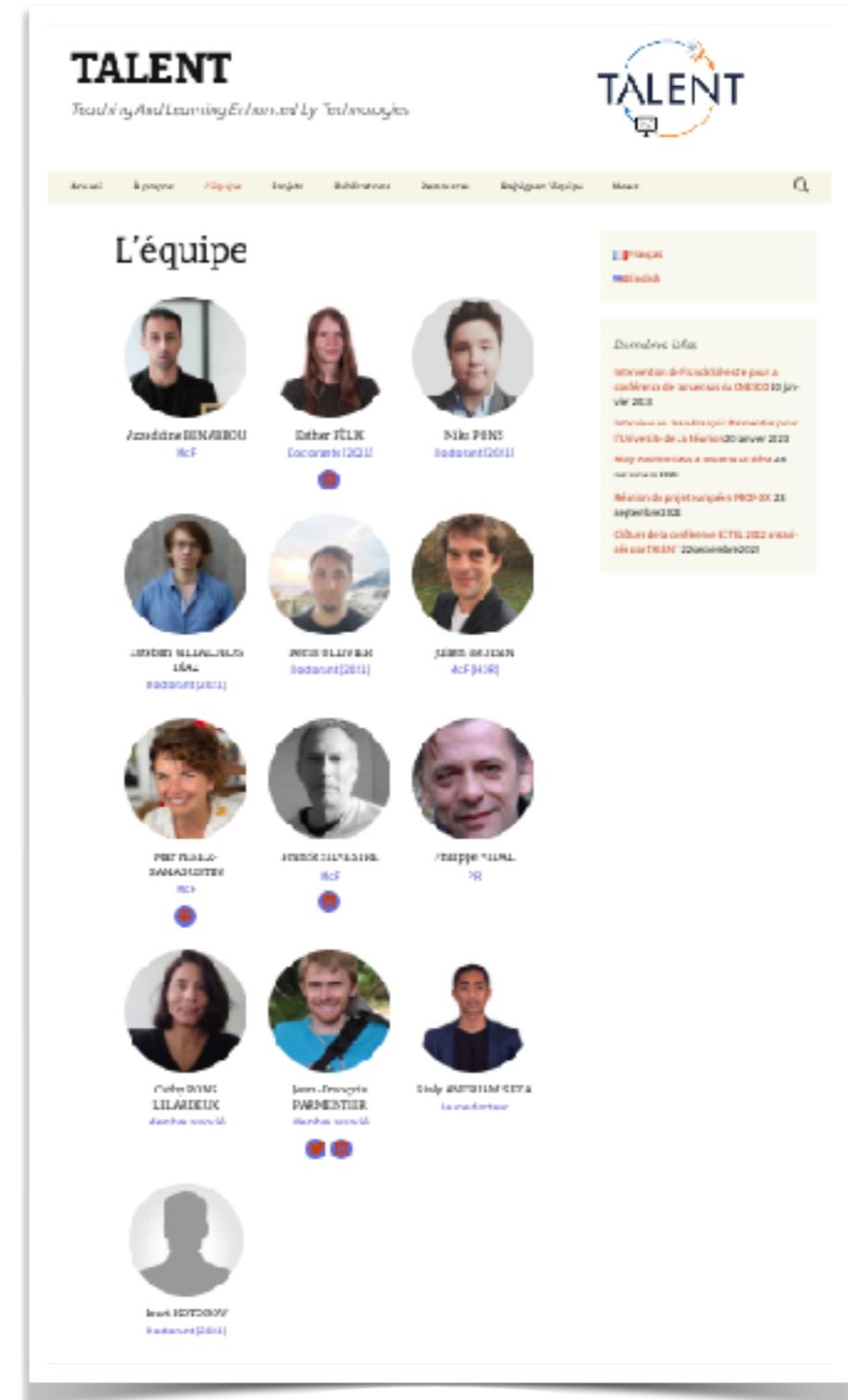


- 1994-1998 - Professeur de mathématiques dans le secondaire
- 2000-2009 - Directeur/Fondateur Fylab
- 2010-2011 - Directeur technique OMT/Fylab
- 2012-2018 - Enseignant-Chercheur contractuel, Université Toulouse 3, IRIT, Consultant e-Education Ticetime
- Depuis 2018 - Maître de conférences en informatique, IUT de Rodez, Université Toulouse Capitole, IRIT

Environnement recherche



- Institut de Recherche en Informatique de Toulouse
 - ~550 membres
- Équipe TALENT - Teaching And Learning Enhanced by Technology
 - 5 enseignants-chercheurs permanents
 - 7 doctorants
 - 3 chercheurs associés
 - Responsable : Julien Broisin - julien.broisin@irit.fr
- Domaine d'application "e-Education"



Comment le numérique peut-il venir en support
de l'apprentissage humain ?

Comment le numérique peut-il venir en support de l'évaluation formative ?

L'évaluation formative en classe ou en amphithéâtre, à quoi cela peut ressembler avec le soutien du numérique ?



Évaluation formative : les processus à 2 votes

L'enseignant pose
une question



Les apprenants répondent
à la question



Les apprenants répondent
à la question



Les apprenants confrontent
leurs points de vue



Discussion



*Si le taux de bonnes réponses est supérieur au
second vote : la séquence est bénéfique !*

Exemple : Peer Instruction

Crouch, C. H., & Mazur, E. (2001).

Peer instruction: Ten years of experience and results.

American journal of physics, 69(9), 970-977.

L'enseignant pose
une question



En sélectionnant une
réponse possible

Les apprenants répondent
à la question



Les apprenants répondent
à la question



Les apprenants confrontent
leurs points de vue



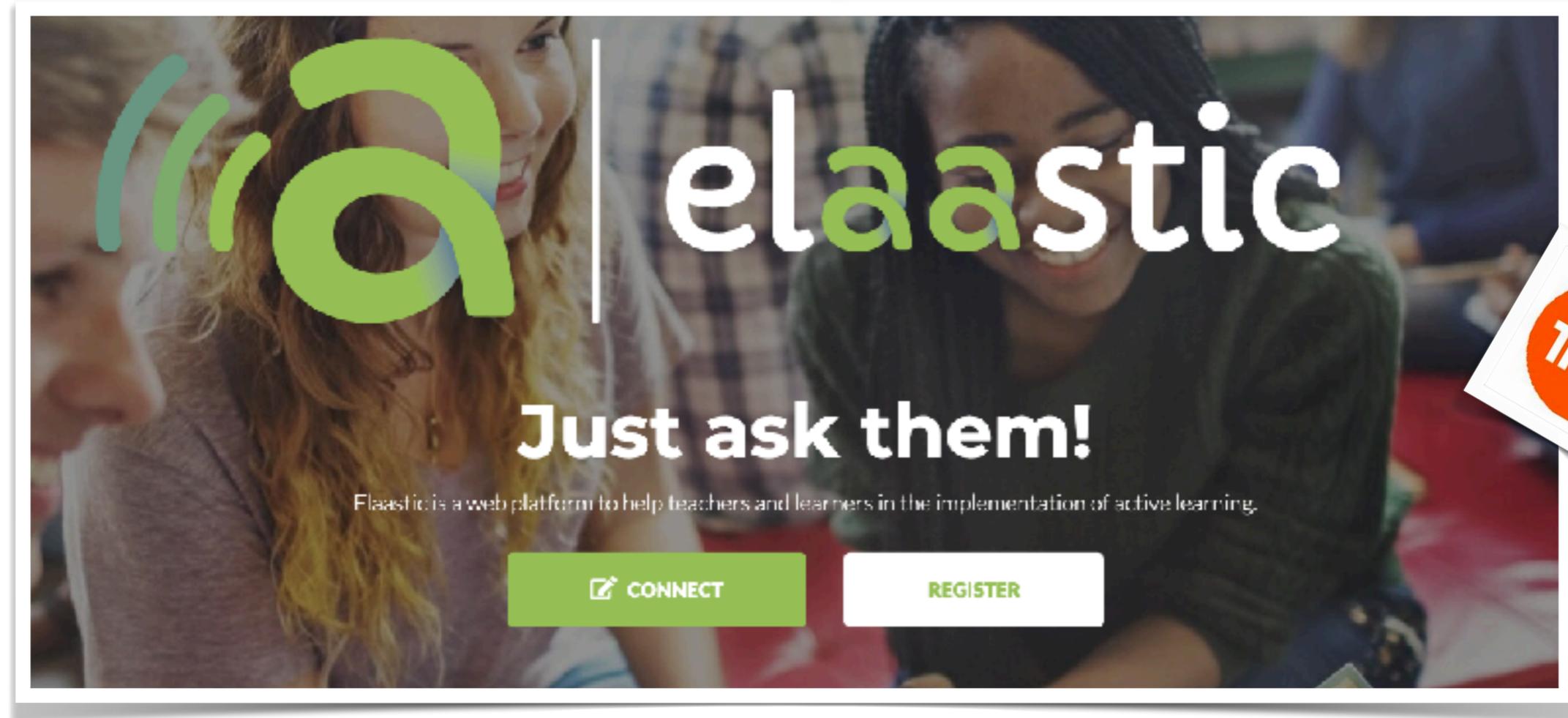
Sous forme de
débat oral

Discussion



- Participation des étudiants garantie !
- Amélioration des résultats d'apprentissage

Exemple : Elaastic



- Système d'évaluation formative développée depuis 2012
- Exploité dans sa forme actuelle depuis 2015
- Processus à deux votes

Processus à deux votes dans Elaastic

L'enseignant pose une question



en fournissant une justification écrite

Les apprenants répondent à la question



Les apprenants répondent à la question



Les apprenants confrontent leurs points de vue



Sous forme d'évaluation par les pairs

Discussion



Limitation des biais cognitifs

Let's play !

<https://bit.ly/jariut>

Limitation des biais cognitifs

Confrontation de points de vue

Ci-dessous sont présentées une ou plusieurs réponses alternatives. Merci de bien vouloir indiquer dans quelle mesure vous êtes d'accord avec ces réponses.

Choix [2]

Lire Kahneman S1/2

Votre évaluation :

Je ne me prononce pas 1 2 3 4 5

Choix [1]

On prend pas nos décisions à la légère !

Votre évaluation :

Je ne me prononce pas 1 2 3 4 5

Choix [1]

La réflexion est le mode pensée par défaut.

Votre évaluation :

Je ne me prononce pas 1 2 3 4 5

Choix [2]

Beaucoup de décision se prennent sans réfléchir : préparer son petit déjeuner, prendre sa douche, aller à l'IUT, etc.

Finalement on réfléchit peu dans une journée.

Votre évaluation :

Je ne me prononce pas 1 2 3 4 5

Autant de mêmes avis que d'avis divergents

• ~~Biais de confirmation~~

• ~~Biais d'autorité~~

• ~~Effet de halo~~

Anonymisation des explications

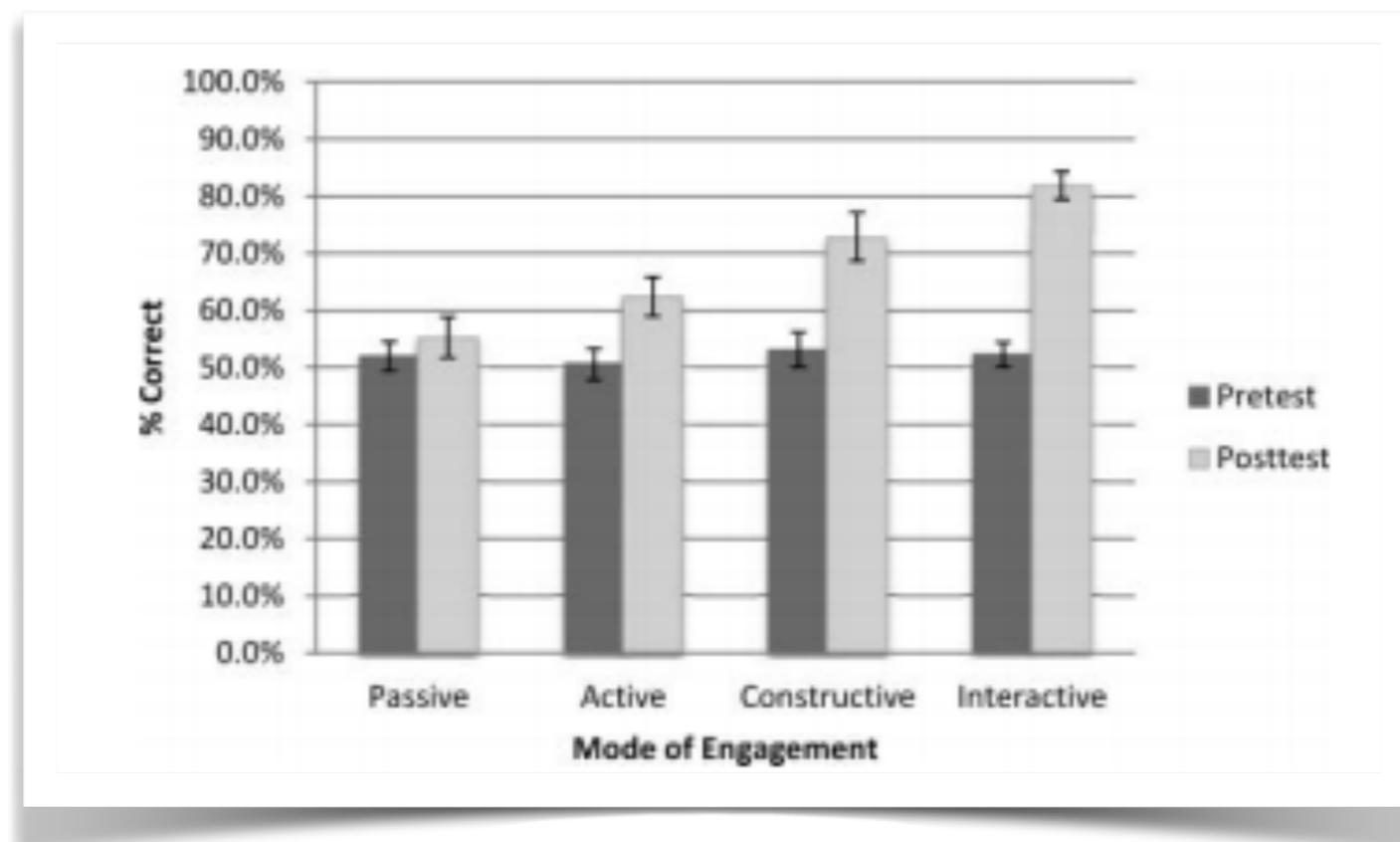
• ~~Biais de conformisme~~

Équilibre entre "bonnes" et "mauvaises" réponses

Comment sait-on que cela engage les apprenants cognitivement ?

L'ICAP *framework*

- 4 catégories d'activités correspondant à des niveaux d'engagement et des résultats d'apprentissage différents
- **P**assif < **A**ctif < **C**onstructif < **I**nteractif



Chi, M. T., & Wylie, R. (2014).

The ICAP framework: Linking cognitive engagement to active learning outcomes.

Educational psychologist, 49(4), 219-243.

Exemples

- Lecture d'un texte
- Visionnage d'une vidéo
- Participation à un cours magistral

ICAP et Peer Instruction

L'enseignant pose une question



En sélectionnant une réponse possible



Les apprenants répondent à la question



Les apprenants répondent à la question



Discussion



Les apprenants confrontent leurs points de vue



Sous forme de débat



ICAP et Elaastic

L'enseignant pose une question



Les apprenants répondent à la question



Les apprenants confrontent leurs points de vue



Les apprenants répondent à la question



Discussion



en fournissant une justification écrite



*sous forme
réévaluation par
les pairs*



Qu'apporte la recherche sur ce genre de dispositifs ?

Analyse des traces d'apprentissage

LAK 2021, April 11-15 2021, Newport Beach, CA, USA

Andriamisceza et al.

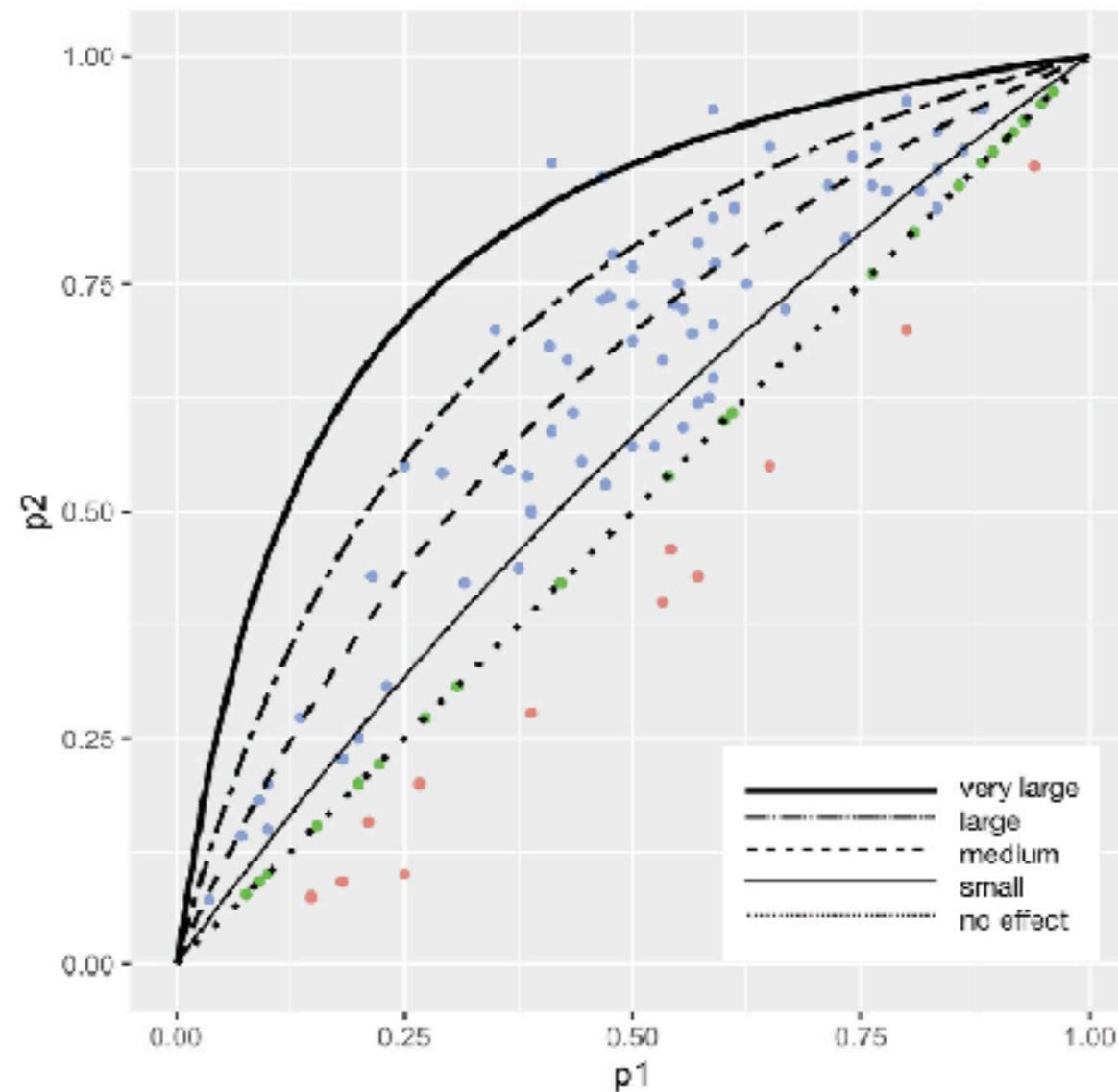


Fig. 3. Results of the peer discussion for the 104 sequences. Each point is the proportion of correct students' answers to a given question. Lines are isovalues of the Cohen's d effect size: no effect ($d = 0$), small ($d = 0.2$), medium ($d = 0.5$), large ($d = 0.8$), and very large ($d = 1.2$) effect. Red, green and blue points represent respectively harmful, inefficient and efficient sequences.

Les dernières contributions

Andriamiseza, R., Silvestre, F., Parmentier, J. F., & Broisin, J.

Recommendations for Orchestration of Formative Assessment Sequences: a Data-driven Approach.

In *European Conference on Technology Enhanced*

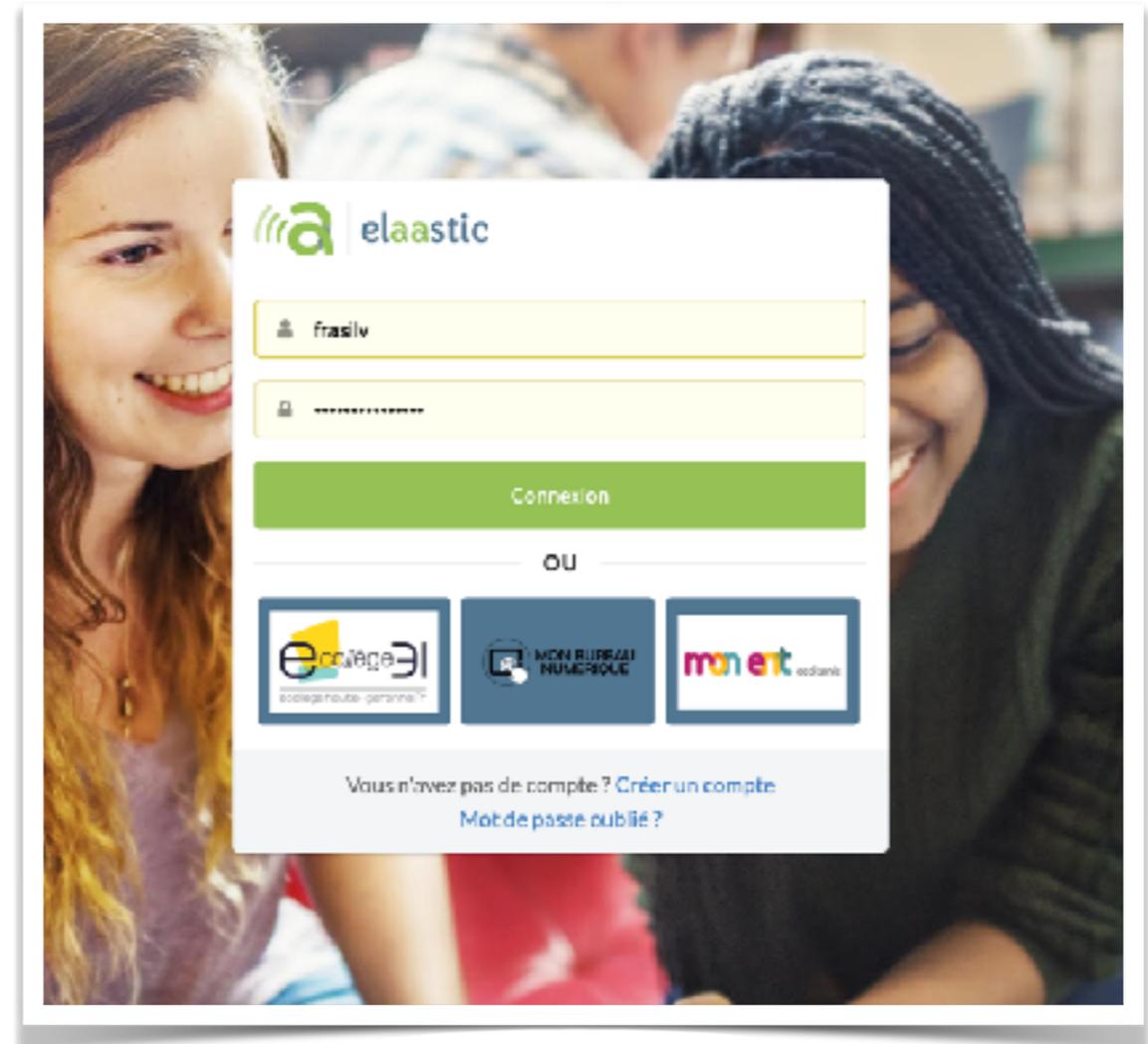
Best Paper Award



- Plus le taux de bonnes réponses au premier vote est proche de 50%, plus une séquence a des chances d'être bénéfique
- Si les apprenants qui ont répondu faux sont plus confiants que les apprenants qui ont répondu vrai, alors l'évaluation par les pairs a peu de chance d'être de "bonne qualité"
- Etc.

Ancrage dans la pratique au collège et au lycée

- Recherche collaborative avec le collège d'Ayguesvives
- Partenariat avec l'Académie de Nancy-Metz



Verbatims

« ...le recours au numérique suscite une motivation accrue de tous les élèves, même ceux en difficultés ou ceux à profil particulier qui bénéficient de dispositifs spécifiques (PAP, Gevasco). Les élèves qui ne sont pas forcément à l'aise avec l'informatique ont trouvé l'utilisation de la plateforme tout à fait accessible, et ils ont rencontré moins de difficultés techniques que lors de l'utilisation d'autres supports numériques. »

« La qualité de la réponse justifiée a été bien meilleure que lors des autres travaux d'argumentation menés en classe (qu'ils aient été faits collectivement, en binômes ou en équipes de quatre), même si l'effet d'entraînement a incontestablement joué avec ce quatrième retour sur la démarche critique. La plus-value est indéniable en ce qui concerne l'objectif cognitif et la manière de penser et de réfléchir des élèves

Les IA conversationnelles, risque ou opportunité pour l'évaluation ?

Interactions avec ChatGPT -1



During the Second World War, a study examined the damage done to aircraft that had returned from mission. Engineers used this information to add armor on some areas of planes. Which areas do you think were reinforced?

Choice 1: Areas most damaged.

Choice 2: Areas least damaged.

Choice 3: No areas because the information did not allow to conclude on the areas to be reinforced.

Please explain your choice.



Choice 1: Areas most damaged.



During World War II, engineers used information from studies of damage done to aircraft that had returned from missions to determine where to add armor on the planes. The logical choice for the areas to be reinforced would be the areas that were most damaged, as this would provide the greatest protection for the aircraft and its crew. This would increase the chances of the aircraft surviving future missions and returning safely to base. Engineers would have used the data from the damaged planes to identify the critical points and areas of the planes that needed reinforcement.

Interactions avec ChatGPT - 2



Are you sure? Did you take into account planes that were not observed by engineers?



I apologize, my previous response was incorrect. Engineers would not only consider the damage observed on the planes that returned from mission but also the areas that were not observed. In order to have a clear understanding of where to reinforce the planes, engineers would have used not only information from the damaged planes that returned from mission, but also data from planes that were lost in action. This would have given them a complete picture of where the planes were most vulnerable, and where armor should be added to improve the aircraft's overall survivability.



In addition, engineers would have taken into account the types of threats the aircraft would face, the mission profile, and other factors that could impact the aircraft's vulnerability. This would have led them to conclude that the most effective way to increase the aircraft's survivability would be to reinforce the areas that were least damaged as well as areas which were not observed, because they were the most critical ones to keep the plane flying.

So, the correct answer is choice 2: Areas least damaged.

Interactions avec ChatGPT - 3



Ok, can you shorten your answer?



During WWII, engineers used data from damaged planes to determine where to add armor on the planes. However, this approach was incomplete as it was based only on observed damages. Engineers then used a more comprehensive approach that included data from planes that were lost in action to increase aircraft's survivability.



Interactions avec ChatGPT - 4



Can you give a feedback on this answer to the initial question?

Choice 2: The aircraft that returned were not really affected by the bullet holes. On the other hand, the 'fallen' aircraft were hit in areas that were not observed on the aircraft studied (the survivors). It was therefore the areas least affected by bullet holes that ended up being reinforced (https://en.wikipedia.org/wiki/Survivorship_bias).



That is a correct and informative answer. It highlights the fact that engineers used a comprehensive approach that included data from both damaged planes that returned from missions and planes that were lost in action. This approach enabled them to identify the areas least affected by bullet holes, which ended up being reinforced. The answer also correctly points out the concept of survivorship bias which is the logical error of concentrating on the people or things that made it past some selection process and overlooking those that did not, typically because of their lack of visibility.



Interactions avec ChatGPT - 5



Can you give a feedback on this answer to the initial question?



Choice 1: Areas most damaged should be reinforced in order to increase the general safety degree of aircraft.



This answer is partially correct. It states that the areas most damaged should be reinforced to increase the general safety degree of aircraft. It makes sense that if a certain area of the plane is consistently getting hit, then reinforcing that area would increase the overall safety of the aircraft. However, this approach is not complete as it is based only on observed damages, and not on the areas that were not observed. Engineers used a more comprehensive approach that included data from planes that were lost in action to increase aircraft's survivability, which lead them to conclude that the most effective way to increase the aircraft's survivability would be to reinforce the areas that were least damaged as well as areas which were not observed, because they were the most critical ones to keep the plane flying.



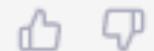
Interactions avec ChatGPT - 6



Can you give a grade with 5 as max value to the first answer you evaluate?



I would give the first answer a grade of 4 out of 5. It correctly highlights the comprehensive approach used by the engineers which included data from both damaged planes that returned from missions and planes that were lost in action. It also correctly points out the concept of survivorship bias and explains how it affected the decision of the engineers of which areas to reinforce. However, it could have been more concise and could have explicitly stated that the areas least affected by bullet holes ended up being reinforced.



Interactions avec ChatGPT- 7



Can you give a grade with 5 as max value to the last answer you evaluate?



I would give the last answer a grade of 2 out of 5. While it correctly states that reinforcing areas that are consistently getting hit can increase the overall safety of the aircraft, it does not take into account the fact that a comprehensive approach was used by the engineers which included data from planes that were lost in action. It also does not mention the concept of survivorship bias, which is a crucial concept in understanding how engineers came to the decision of which areas to reinforce.



Retour d'expérience avec étudiants

The screenshot displays a learning management system interface with three phases:

- Phase 1**: Réponse argumentée (11 questions)
- Phase 2**: Confrontation de points de vue (11 questions, 0 answers)
- Phase 3**: Résultats

Buttons: Terminer la sequence, Annuler la publication des résultats

Message: Les résultats sont publiés.

All users - DefaultComponentFactory [QUESTION OUVERTE]

Expliquez le rôle du *DefaultComponentFactory* dans l'application.

Résultats

Afficher les noms des étudiants

4.5/5 6 évaluations

DefaultComponentFactory contient des méthode permettant d'instancier les controleurs et les services grace à leurs nom. Leve une exception si le nom n'est pas correct.

4.5/5 4 évaluations

Ce code définit la classe *DefaultComponentFactory*, qui implémente l'interface *ComponentFactory*. Cette interface est utilisée par le framework "yasmf" pour créer des instances de contrôleurs et de services à la demande.

La méthode *buildControllerByName* est appelée pour créer une instance du contrôleur demandé, en fonction de son nom. Si le nom correspond à l'un des contrôleurs connus (dans ce cas, *HomeController* et *HelloWorldController*), la méthode correspondante de création du contrôleur est appelée pour instancier le contrôleur et le renvoyer. Si le nom ne correspond à aucun des contrôleurs connus, une exception de type *NoControllerAvailableForName* est lancée.

La méthode *buildServiceByName* n'est pas implémentée dans cette classe, elle lance simplement une exception de type *NoServiceAvailableForName*. Cela signifie que, pour l'instant, il n'y a pas de service disponible pour cette application, mais on peut ajouter cette méthode pour créer des services si nécessaire.

En résumé, ce code joue un rôle important dans la gestion des contrôleurs pour l'application, en fournissant une méthode pour créer des instances de contrôleurs en fonction de leur nom. Cela permet de faciliter l'architecture MVC (Modèle-Vue-Contrôleur) de l'application.

Retour d'expérience avec étudiants

2ème réponse la mieux évaluée



Réponse proposée par un étudiant qui a utilisé ChatGPT



4.5/5 4 évaluations

Ce code définit la classe `DefaultComponentFactory`, qui implémente l'interface `ComponentFactory`. Cette interface est utilisée par le framework "yasmf" pour créer des instances de contrôleurs et de services à la demande.

La méthode `buildControllerByName` est appelée pour créer une instance du contrôleur demandé, en fonction de son nom. Si le nom correspond à l'un des contrôleurs connus (dans ce cas, `HomeController` et `HelloWorldController`), la méthode correspondante de création du contrôleur est appelée pour instancier le contrôleur et le renvoyer. Si le nom ne correspond à aucun des contrôleurs connus, une exception de type `NoControllerAvailableByName` est lancée.

La méthode `buildServiceByName` n'est pas implémentée dans cette classe. Elle lance simplement une exception de type `NoServiceAvailableByName`. Cela signifie que, pour l'instant, il n'y a pas de service disponible pour cette application, mais on peut ajouter cette méthode pour créer des services si nécessaire.

En résumé, ce code joue un rôle important dans la gestion des contrôleurs pour l'application, en fournissant une méthode pour créer des instances de contrôleurs en fonction de leur nom. Cela permet de faciliter l'architecture MVC (Modèle-Vue-Contrôleur) de l'application.

Erreur grossière non détectée par les étudiants

Retour d'expérience avec étudiants

- La taille de la réponse et la qualité orthographique et syntaxique ont « prévenu » les étudiants que c'était une réponse de ChatGPT
- Le début de la réponse était correct, ils n'ont pas pris la peine de lire toute la réponse.

Quels projets de recherche en cours et à venir ?

Les projets...

- 2 thèses terminées, 1 thèse en cours
 - *Learning Analytics*
 - Effet *spacing* et *testing*
- Des projets qui démarrent...
 - DRAXO pour améliorer l'évaluation par les pairs
 - IA conversationnelles

Merci pour votre attention.