

# Evaluation des environnements numériques pour l'apprentissage humain

Saison #01

---

Un MOOC proposé par  
Gaëtan Temperman, Bruno De Lièvre et Karim Boumazguida

soutenu par  
l'Université de Mons et l'Université Mohammed Ier d'Oujda

---

Prise de notes de Didier Dorsaz  
2021

# Table des matières

<b>1. Les caractéristiques d'un environnement numérique d'apprentissage.....</b>	<b>4</b>
1.1. Le numérique et l'innovation.....	4
1.2. Le profil initial des individus.....	4
1.3. La première étape en termes de scénarisation.....	5
1.3.1. Les tâches d'apprentissage.....	5
1.3.2. Outils et ressources.....	6
1.3.3. L'encadrement dans un environnement numérique.....	6
1.3.4. Gestion du temps et Organisation des groupes.....	6
1.4. Comment envisager l'évaluation d'un environnement numérique d'apprentissage ?	7
1.4.1. Quel lien entre recherche et innovation ?.....	7
1.4.2. Learning analytics.....	7
1.5. Quiz.....	9
<b>2. Quelles modalités d'évaluation de la performance dans un ENA ?.....</b>	<b>10</b>
2.1. Les indicateurs de l'évaluation des produits.....	10
2.1.1. La progression.....	10
2.1.2. Le niveau de maîtrise.....	11
2.1.3. Le transfert.....	11
2.1.4. L'équité.....	11
2.2. Création de l'outil d'évaluation.....	12
2.2.1. Varier les opérations mentales pour une même compétence.....	12
2.2.2. Intégrer les questions ouvertes et fermées.....	13
2.3. Validation de l'outil d'évaluation.....	14
2.3.1. Indice de difficulté des items ( $P_i$ ).....	14
2.3.2. Indice de discrimination.....	15
2.4. Quiz.....	16
<b>3. Que nous dit l'analyse du processus des apprenants ?.....</b>	<b>19</b>
3.1. La notion de trace.....	19
3.1.1. Les traces numériques.....	19
3.1.2. Les traces non-numériques.....	20
3.2. La nature des traces.....	20
3.2.1. Les réalisations.....	20
3.2.2. Les usages.....	20
3.2.3. Le temps.....	20
3.2.4. Les interactions.....	21
3.3. Utilisation pédagogique des traces.....	21
3.3.1. Niveau 1 : Miroir.....	24
3.3.2. Niveau 2 : Métacognitif.....	24
3.3.3. Niveau 3 : Adaptatif.....	24
3.4. Quiz.....	24

<b>4. Quels outils validés pour recueillir la perception ?</b>	<b>26</b>
4.1. Pourquoi mesurer les perceptions ?	26
4.2. Comment mesurer les perceptions ?	26
4.2.1. Mesurer l'utilité et l'utilisabilité	26
4.2.2. Mesurer les émotions	28
4.2.3. La charge cognitive	29
4.2.4. Une mesure globale de l'expérience utilisateur	30
4.2.5. Conclusion	30
4.3. Quiz	31
<b>5. Comment construire un outil pour analyser la perception ?</b>	<b>32</b>
5.1. Construire son propre questionnaire	32
5.2. Réaliser un entretien	33
5.3. Combiner plusieurs outils	34
5.4. Quiz	34
<b>6. Quel est l'intérêt de réaliser des analyses croisées ?</b>	<b>35</b>
6.1. Introduction	35
6.2. La comparaison	36
6.3. La corrélation	37
6.4. Les démarches prédictives	37
6.5. La catégorisation	38
6.6. Conclusion	39
6.7. Quiz	40
<b>7. Questionnaire final</b>	<b>42</b>

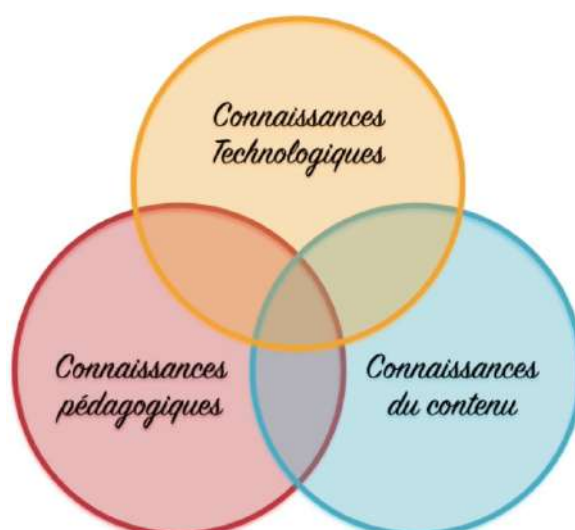
# 1. Les caractéristiques d'un environnement numérique d'apprentissage

## 1.1. Le numérique et l'innovation

Les tests PISA (2015) montrent qu'il n'y a pas vraiment de différence entre les pays qui ont massivement investi financièrement dans le numérique par rapport aux autres.

Souvent, l'outil technologique n'apporte pas d'innovation en soi ; ça ne permet pas forcément de faire mieux. Il permet surtout de nouveaux usages, lesquels pourront éventuellement apporter de l'innovation et des bénéfices.

L'utilisation d'un nouvel environnement d'apprentissage pertinent dépend de 3 dimensions (TPACK)



C'est un modèle qui donne des indications pour trois dimensions importantes. Tout l'enjeu réside dans la conception et la mise en place de tâches significatives en étant en équilibre sur ces trois dimensions.

Il y a énormément de contextes d'implémentations d'environnement d'apprentissages, il faut donc bien documenter le tout. Chaque enseignant a son style et chaque élève a son profil.

## 1.2. Le profil initial des individus

« Le facteur le plus important qui influence l'apprentissage est ce que l'élève sait déjà. » (Ausbuel, 1968). Ce qu'on **sait au départ** a un énorme impact sur ce qu'on va apprendre. Cette logique nous pousse donc à faire une évaluation au départ. Avoir ça en tête permet de mieux évaluer les progrès des participants.

Il faut aussi prendre en compte la **motivation** des apprenants. Elle peut être :

-intrinsèque : avec des buts de maîtrise / de compétence

-extrinsèque : avec des buts de performance

On peut les évaluer avec des questionnaires de buts motivationnels.

### 1.3. La première étape en termes de scénarisation

On peut d'abord se poser la question du « quoi » ? Quel est l'objet d'apprentissage ?

Il y a 5 variables-clés pour décrire les compétences dans un EA :



-**tâches** : les activités des apprenants (ce qui va être appris)

-**modalités d'encadrement** : rôle de l'enseignant

-**gestion du temps**

-**organisation du groupe** : dimension sociale

-**outils et ressources** (notamment numériques)

#### 1.3.1. Les tâches d'apprentissage

L'importance de la tâche renvoie l'idée que la connaissance est toujours de résultat d'une reconstruction par l'élève qui apprend. Il doit donc être proactif, il doit sélectionner les informations, les organiser et les intégrer dans les connaissances qu'il a déjà.

Il y a 5 types de tâches :

-**études** (pex : écoute d'un cours via une vidéo)

-**résolution de problème** : l'élève essaie d'atteindre un but ; il exerce une compétence (pex : questionnaire en ligne, un défi, une simulation, une carte conceptuelle)

-**recherche d'informations** : au lieu de chercher une solution par la réflexion (comme la résolution de problème), on va chercher des informations (pex : on cherche une solution sur le web)

-**dialogue** : travail à plusieurs (pex : collaboration, co-élaboration entraide)

-**jeux** : il permet de développer des compétences de manière active et interactive

Le mieux c'est de varier ces différents type de taches dans un scénario pour varier les processus.

### 1.3.2. Outils et ressources

On choisit un outil selon la fonction qu'on attend de ce dernier (on ne choisit pas le même outil pour évaluer que pour exposer un fait)

L'utilisation du numérique entraîne souvent la mobilisation du principe de double codage (qui est important). Il est important de mobiliser la mémoire auditive et visuelle ; le traitement parallèle des deux est plus efficace pour l'apprentissage.

### 1.3.3. L'encadrement dans un environnement numérique

C'est très important pour la réussite des élèves. Il y a plusieurs types de tutorats. Le tuteur peut être réactif (i.e. il réagit aux demandes des élèves), ou proactif (et intervenir de lui même dans le processus d'apprentissage).

L'encadrement peut être très importante pour la réussite des élèves (selon les outils qui sont à disposition). L'apprenant utilisera plus un outil si il est encouragé à le faire par le tuteur).

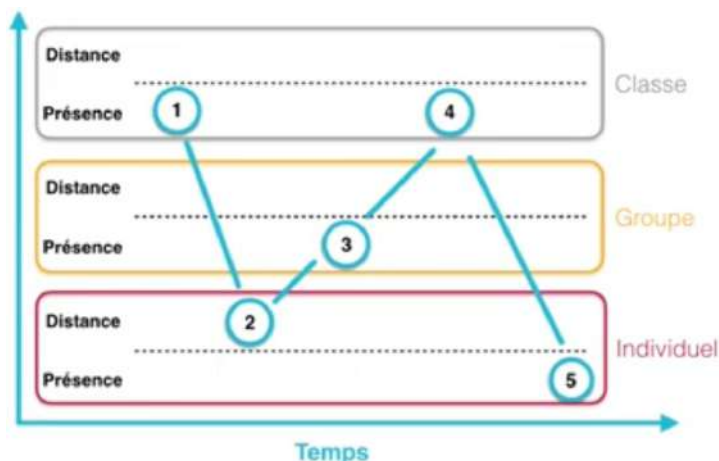
Le suivi peut être en lien avec la planification, le feedback, mais aussi le soutien socio-affectif.

### 1.3.4. Gestion du temps et Organisation des groupes

C'est crucial dans les apprentissages. La gestion concerne la durée et la manière de distribuer les apprentissages dans le temps.

Il y a plusieurs manières de distribuer les apprenants : individuels, petits groupes, plus grands groupes/classe.

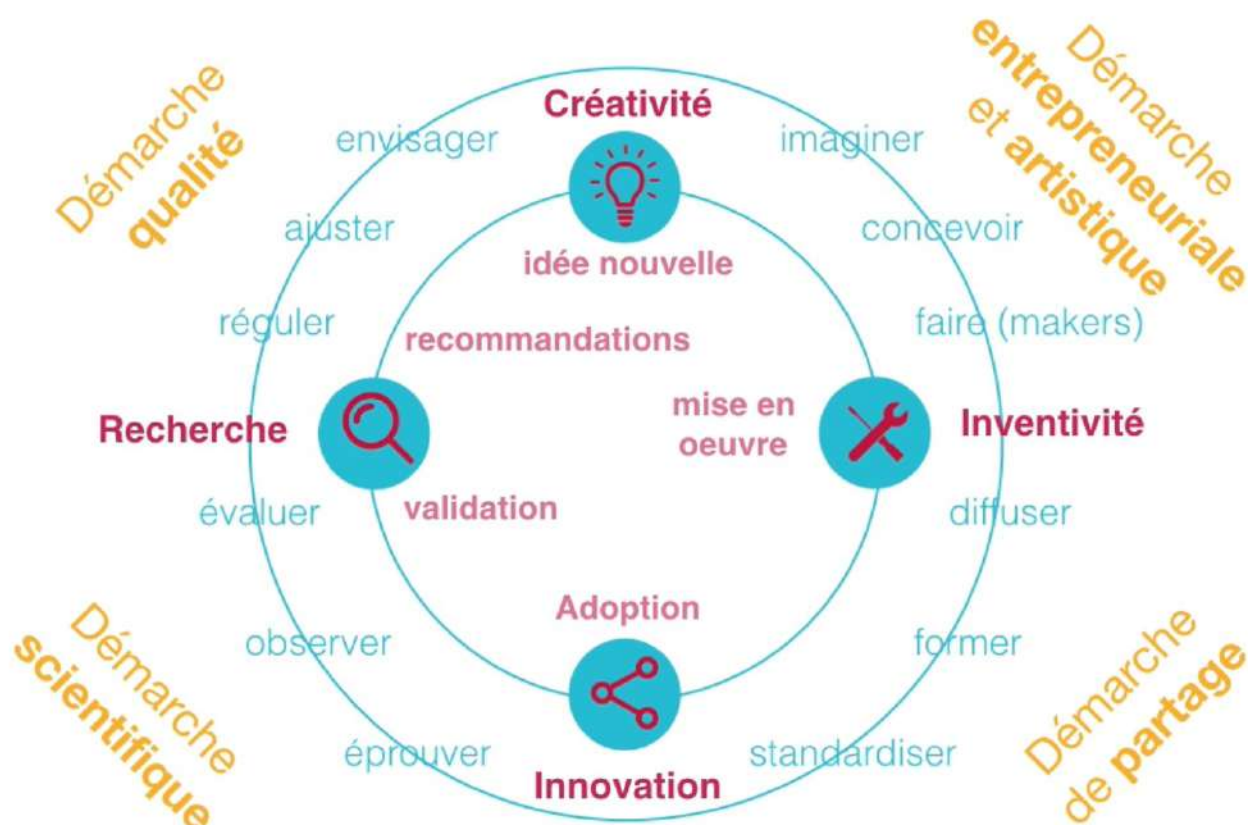
On peut croiser le temps et l'organisation des groupes. Le mieux c'est la modélisation de Dillenbourg (2016) pour représenter ça (comme ci-dessous).



Exemple de graphe d'orchestration de Dillenbourg

## 1.4. Comment envisager l'évaluation d'un environnement numérique d'apprentissage ?

### 1.4.1. Quel lien entre recherche et innovation ?



Quand on crée un EA, on est dans une démarche de créativité, d'inventivité, d'innovation et de recherche.

C'est important d'évaluer des EA pour voir si c'est réellement efficace.

### 1.4.2. Learning analytics

C'est « la mesure, la collection, l'analyse et l'interprétation des traces des apprenants et de leur contexte pour comprendre et optimiser leur apprentissage dans l'environnement dans lequel il se produit » (Siemens & Long, 2011)

Ça se fait sur 3 niveaux : micro (le cours lui-même), méso (une institution), macro (un système éducatif entier)

Au niveau micro, on peut regarder la performance des apprenants. On peut mesurer les produits des apprenants. Dans le MOOC on va se centrer sur 4 dimensions :

- la progression
- la maîtrise
- l'équité
- le transfert

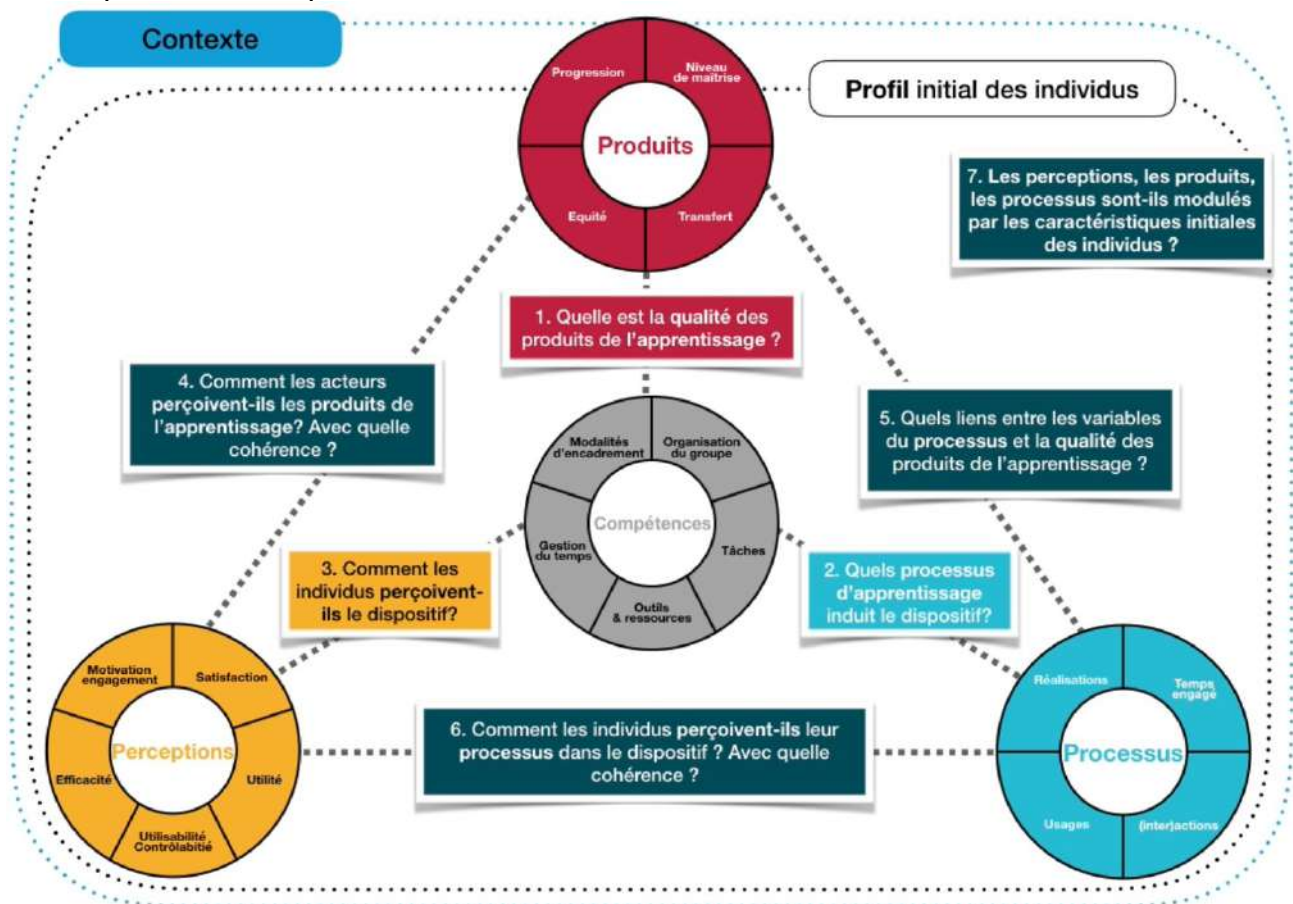
On peut aussi observer les processus mis en œuvre par les élèves et le formateur. Ça peut concerner :

- les réalisations
- les usages (des outils)
- les (inter)actions
- le temps engagé

Et finalement les perceptions :

- la motivation
- la satisfaction
- l'utilité
- l'utilisabilité, la contrôlabilité
- l'efficacité

En résumé, dans une démarche qualité, voici les différents éléments à prendre en compte et les questions à se poser.



Cette approche permet de générer des questions de recherches, et de guider les analyses.



## 1.5. Quiz

1. A quelle dimension peut être associée la notion d'instrumentalisation ?

les processus d'apprentissage

2. Pour évaluer les effets d'un environnement d'apprentissage, un chercheur peut s'intéresser à différentes dimensions complémentaires. Combien existe-t-il de sources d'observations distinctes à prendre en compte ?

Il y en a 4 : On peut s'appuyer sur le profil des acteurs, le produit, le processus et la perception.

3. Lequel des énoncés suivants est lié à la dimension « processus » ?

On s'intéresse au temps nécessaire à la réalisation de la tâche. L'analyse du temps est une variable importante dans toute analyse de situation d'apprentissage.

4. L'inventivité, c'est...

Faire des choix pour mettre en œuvre concrètement les choses.

5. A quoi correspond le niveau "micro" du champ de recherche "Learning analytics" ?

A l'activité de l'apprenant dans le contexte d'un environnement spécifique

6. Le champ de recherche du "Learning analytics" est constitué de :

3 niveaux: Micro- Meso - Macro

7. La prise en compte de la motivation au départ peut être associée à quelle dimension dans l'évaluation ?

profil initial

8. Selon Ausubel (1968), les *connaissances préalables* constituent le facteur qui influence le plus l'apprentissage.

9. La plus-value des technologies dans le modèle TPACK (Koehler, Mishra & Cain, 2013) réside dans l'interaction entre trois facteurs:

Un choix d'outils pertinents, des pratiques pédagogiques efficaces et l'analyse du contenu ciblé.

## 2. Quelles modalités d'évaluation de la performance dans un ENA ?

Objectifs spécifiques du module :

- Décrire les indicateurs relatifs aux produits de l'apprentissage
- Créer et valider un outil d'évaluation de la performance

### 2.1. Les indicateurs de l'évaluation des produits

Ce type d'évaluation se centre sur la dimension « produits » de l'évaluation. Elle est composée de 4 éléments :

- la progression
- l'équité
- le niveau de maîtrise
- le transfert

#### 2.1.1. La progression

Pex : on prend le niveau avant et après un cours pour voir le changement.

Pour mesurer cette progression on peut faire l'équation :

$$\text{si Post} > \text{pré} \Rightarrow \text{Gain relatif} = 100 \times \frac{\text{Posttest} - \text{Prétest}}{\text{ScoreMax} - \text{Prétest}}$$

=> c'est le rapport entre ce que l'élève a gagné et ce qu'il aurait pu gagner au maximum

$$\text{si Post} < \text{pré} \Rightarrow \text{Perte relative} = 100 \times \frac{\text{Posttest} - \text{Prétest}}{\text{Prétest}}$$

Exemple : Samuel fait 4/10 au premier examen, puis 7/10 au 2<sup>e</sup> => il y a un gain. Le gain relatif est de :  $100 \times (7-4)/(10-4) = 100 \times 3/6 = 50 \%$

Pour évaluer la valeur intrinsèque d'une intervention dans plusieurs établissements (dans une méta-analyse pex), on peut prendre le *d* de Cohen. Ainsi  $d = \frac{\text{Posttest} - \text{Prétest}}{\text{Ecart - type des données}}$

l'efficacité est :

- faible si  $d < 0.2$
- moyenne si  $d = 0.5$
- large si  $d = 0.8$
- très large si  $d = 1.3$

### 2.1.2. Le niveau de maîtrise

C'est le post-test. Le niveau de maîtrise doit être idéalement proche de 80 %. Il peut être évalué de manière individuel ou collectif (pour un apprentissage collaboratif).

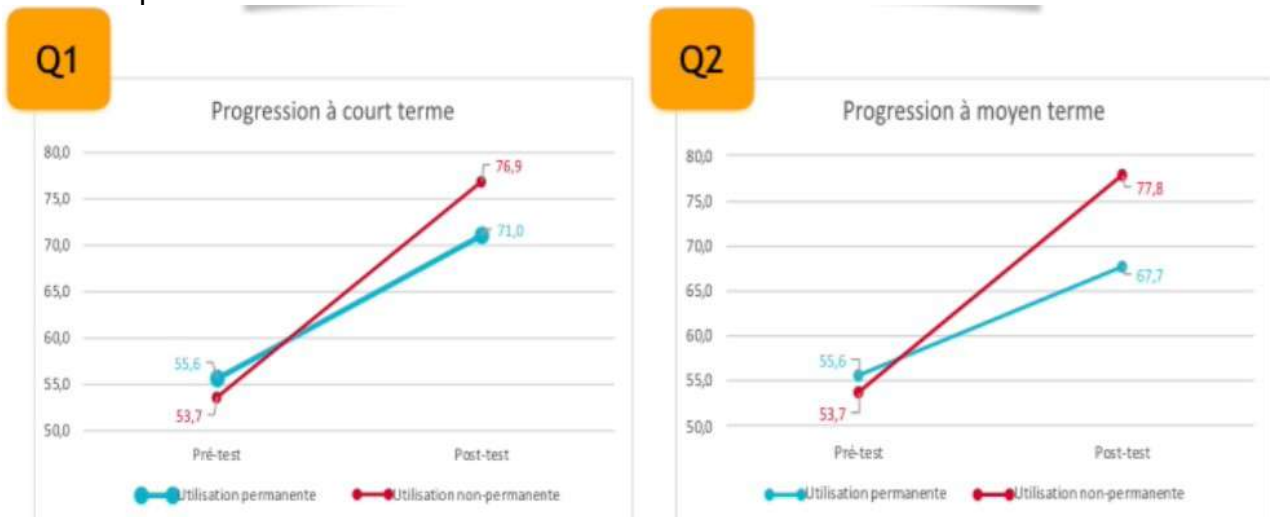
On peut le mesurer à la fin de l'intervention ou plus tard dans le temps (pour voir l'impact de l'oubli)

### 2.1.3. Le transfert

Il est en lien avec la logique de compétence. C'est la capacité à utiliser un acquis dans un contexte différent de celui dans lequel l'information a été apprise.

On peut le mesurer avec un posttest différé ou un posttest dans un autre contexte (ou les deux).

Par exemple :



On peut voir dans la recherche d'Housni (2019) que l'utilisation non-permanente d'une tablette aboutit à de meilleurs résultats à court et long terme.

### 2.1.4. L'équité

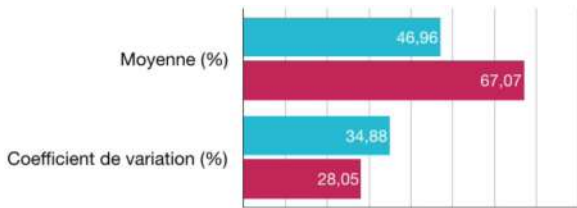
Elle permet de voir si les groupes deviennent plus homogène ou plus hétérogènes. Pour évaluer cette tendance, on va regarder le taux de variance (ou taux d'hétérogénéité) entre le pré-test et le post-test.

$$\text{Prétest : } CV_1 = \text{Taux d'hétérogénéité}_1 = \frac{\text{Ecart-type}_1}{\text{Moyenne}_1} \times 100 \quad (\text{CV} = \text{coefficient de variation})$$

$$\text{Posttest : } CV_2 = \text{Taux d'hétérogénéité}_2 = \frac{\text{Ecart-type}_2}{\text{Moyenne}_2} \times 100$$

Plus le nombre CV est faible, plus le groupe est homogène (i.e. les écarts entre individus sont faibles).

Si le taux d'homogénéité est plus faible à la fin du cours => il diminue les écarts entre les élèves



Pex : ce cours augmente l'équité entre ces élèves

On peut aussi voir l'efficacité d'une intervention en observant la corrélation entre le prétest et le gain relatif. Elle permet, de mettre en évidence pour quel type d'apprenant le dispositif a été le plus utile Si la corrélation est :

- positive => le cours a été profitable aux « meilleurs »
- négative => le cours a été profitable aux « moins bons »
- neutre => l'efficacité n'est pas liée au niveau de départ

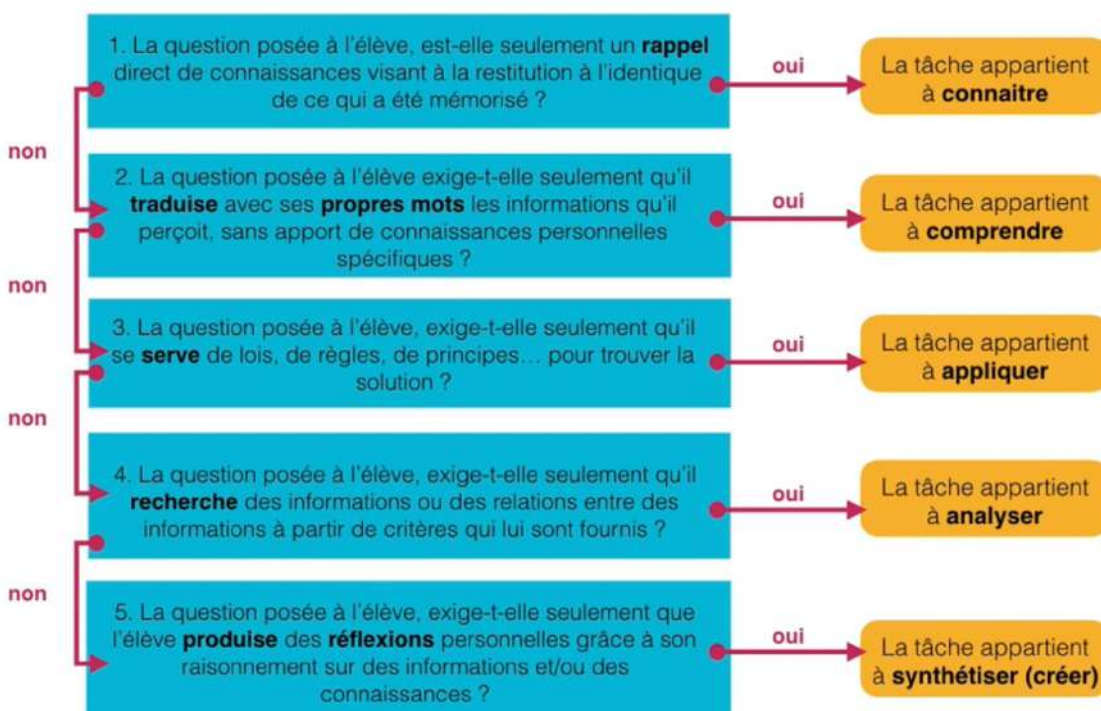
## 2.2. Création de l'outil d'évaluation

### 2.2.1. Varier les opérations mentales pour une même compétence

Les taxonomies aident bien à définir ses objectifs (et donc préparer l'instrument de mesure). Elles permettent de fixer la complexité d'une tâche, mais aussi de varier les opérations mentales.

C'est en lien avec la logique de compétence. L'idée est que les élèves travaillent sur les différents niveaux et atteignent les niveaux les plus élevés.

Exemple : la taxonomie de Bloom



## 2.2.2. Intégrer les questions ouvertes et fermées

Il est intéressant de mixer les questions ouvertes et fermées

### Les questions ouvertes sont des questions de « production »

Elles exigent que la personne élabore et rédige sa réponse. Le principal paramètre d'évaluation est l'étendue et la forme de la réponse attendue. Il y n'y a pas que des textes à faire, mais aussi des schémas à compléter, des textes à compléter et des phrases à compléter. Elles font appel à la mémoire d'évocation.

Pour évaluer des réponses longues, on va se baser sur une grille critériée

<b>C. SAVOIR-FAIRE REDACTIONNELS :</b>  se rapportent à l'exploitation et à l'analyse des résultats, à la rédaction d'un compte-rendu incluant la capacité de synthèse, la prise de recul...	C-1. Rédiger une introduction qui précise le contexte	Pas d'introduction	L'introduction reprend exactement le texte du TP	L'introduction est reformulée, mais elle reste centrée sur le déroulement de la séance	L'introduction est reformulée avec mention du contexte et des applications potentielles
	C-2. Établir un schéma du dispositif expérimental	Pas de schéma	Le schéma est ébauché de façon incomplète, ou reste inexploitable pour un lecteur ne connaissant pas le système	Le schéma est complet et exploitable mais manque de soin ou de rigueur	La schématisation est complète, exploitable, soignée et rigoureuse
	C-3. Présenter le principe et le protocole expérimental	Les résultats sont exposés directement sans que la manipulation ou les principes associés soient décrits	Le principe et le protocole sont copiés sans appropriation par l'étudiant	Le protocole expérimental est reformulé de façon claire et justifiée	Niveau CORRECT et au moins une proposition d'amélioration du protocole est formulée
	C-4. Tracer un graphique à partir de mesures	Le graphe est très brouillon, l'espace est mal utilisé (échelle), il manque le titre, le libellé des axes (grandeur représentée + unité), les graduations...	L'échelle est adaptée, les axes sont gradués régulièrement, mais mal libellés (absence de la grandeur représentée + unité), absence de titre	Les axes sont libellés, l'échelle est adaptée, le graphe comporte un titre et une légende, les mesures sont bien reportées avec une courbe de tendance	Niveau CORRECT et des barres d'erreur matérialisent les incertitudes ou des annotations permettent de faciliter l'interprétation des résultats
	C-5. Etablir une expression littérale (dont incertitudes)	Les expressions littérales préalables aux calculs numériques sont souvent absentes	Les expressions littérales sont établies, mais souvent erronées	Les expressions littérales sont généralement correctes et les notations introduites par l'étudiant sont explicitées	Les expressions littérales sont systématiquement correctes et les notations explicites
	C-6. Présenter un résultat finalisé	Les résultats sont présentés de façon brouillon et désordonnée	Les unités sont absentes ou les valeurs numériques sont fausses	Les résultats sont corrects, présentés avec les bonnes unités, et avec un nombre approprié de chiffres significatifs	Niveau CORRECT et une présentation originale fait ressortir les résultats importants
	C-7. Interpréter les résultats obtenus, en tirer des conclusions	Absence d'interprétation	L'étudiant discute les résultats obtenus de façon très superficielle (répétition du discours de l'enseignant, ...)	Les résultats sont comparés aux valeurs attendues, et les résultats aberrants sont signalés	Niveau CORRECT et les conclusions replacent les résultats obtenus dans un contexte plus général, en lien avec l'introduction

Ici, par exemple, il y a les critères en lignes et les échelles en colonne

### Les questions fermées sont des questions de « sélection »

Elles font appel à la mémoire de reconnaissance. Il faut opérer une action :

- choisir : pex les QCM ou les questions à choix dicotomique
- établir un ou plusieurs liens : pex les questions d'appariement ou d'ordonance

Pour varier les QCM, on peut ajouter des **solutions implicites** (par exemple, en ajoutant « toutes » « aucunes », « manque dans la donnée » ou encore « absurdité »).

On peut aussi ajouter un **degré de certitude** (en pourcentage par exemple, ça permet de sortir du pur « juste ou faux »). Par exemple : on peut utiliser ce schéma



1. Incompétence dangereuse
2. Incompétence reconnue
3. Compétence non-opérationnelle
4. Compétence opérationnelle

En gros, si un élève répond

- **faux** avec une **grande certitude**, c'est une **incompétence dangereuse**
- **faux** avec une **faible certitude**, c'est une **incompétence reconnue**
- **juste** avec une **faible certitude**, c'est une **compétence non-opérationnelle**
- **juste** avec une **forte certitude**, c'est une **compétence opérationnelle**

Basé là-dessus, on peut créer un **score de confiance** en additionnant les degrés de certitudes aux réponses justes, et un **score de prudence** en additionnant les degrés de certitudes aux réponses fausses.

Le choix des réponses varie selon la tâche : par exemple pour le livret, il est plus pertinent de demander le produit de 2 nombre plutôt que de donner 3 réponses QCM. Les réponses longues permettent d'évaluer des facultés d'analyse ou de synthèse.

## 2.3. Validation de l'outil d'évaluation

### 2.3.1. Indice de difficulté des items ( $P_i$ )

Il varie entre 0 et 1 : c'est la proportion d'élève qui ont répondu correctement à un item « i ».

$P_1 = 1 \Rightarrow 100\%$  des élèves ont fait juste l'item 1

$P_2 = 0.01 \Rightarrow 1\%$  des élèves ont fait juste l'item 2

Idéalement, il faudrait que la majorité des items soient autour de 5/10. Les **items à rejeter** sont soit **trop difficiles** ( $P_i = 0.15$ ) ou trop faciles ( $P_i = 0.85$ ). Les trop faciles discriminent les faibles et les trop difficiles discriminent les forts.

Trop de questions faciles font un **effet plafond** et trop de questions difficiles amènent à un **effet plancher**.

On peut aussi travailler avec une grille (voir ci-dessous)

Sujets n = 5	Items notés sur				
	/4	/20	/1	/7	/10
sujet 1	4	14	1	3	4
sujet 2	3	18	1	6	6
sujet 3	4	16	1	6	5
sujet 4	3	14	0	7	6
sujet 5	4	10	0	4	7
<b>Total</b>	18	72	3	26	28
<b>Moyenne</b>	18/5 = 3,6	72/5 = 14,4	3/5 = 0,6	26/5 = 5,2	28/5 = 5,6
<b>Indice de difficulté</b>	3,6/4 = <b>0,9</b>	14,4/20 = 0,72	0,6/1 = <b>0,6</b>	5,2/4 = 0,74	5,6/10 = 0,56

le plus facile

le plus difficile

Ici les colonnes correspondent aux items. Chaque carré est un ratio entre le nombre de points obtenus par le sujet pour l'item. Pex : pour l'item 1, il faut faire 4/4 pour avoir tous les points. L'item 1 est facile puisque les sujets ont fait en moyenne 3.6 sur 4

### 2.3.2. Indice de discrimination

Le but d'un questionnaire c'est de mesurer une compétence/un ensemble de connaissances. Il faut donc pouvoir discriminer les élèves sur ce critère.

L'indice de discrimination est une variante du coefficient de Pearson. Il est nommé  $r_{bis}$  (pour R biserial).

$$r_{bis} = \frac{\bar{X}_1 - \bar{X}_0}{S_y} \sqrt{\frac{p}{1-p}}$$

Il varie de -1 à 1. C'est la corrélation entre un item et le score total au test. On l'utilise quand les variables sont dichotomiques (i.e. quand les réponses peuvent être soit juste ou fausses).

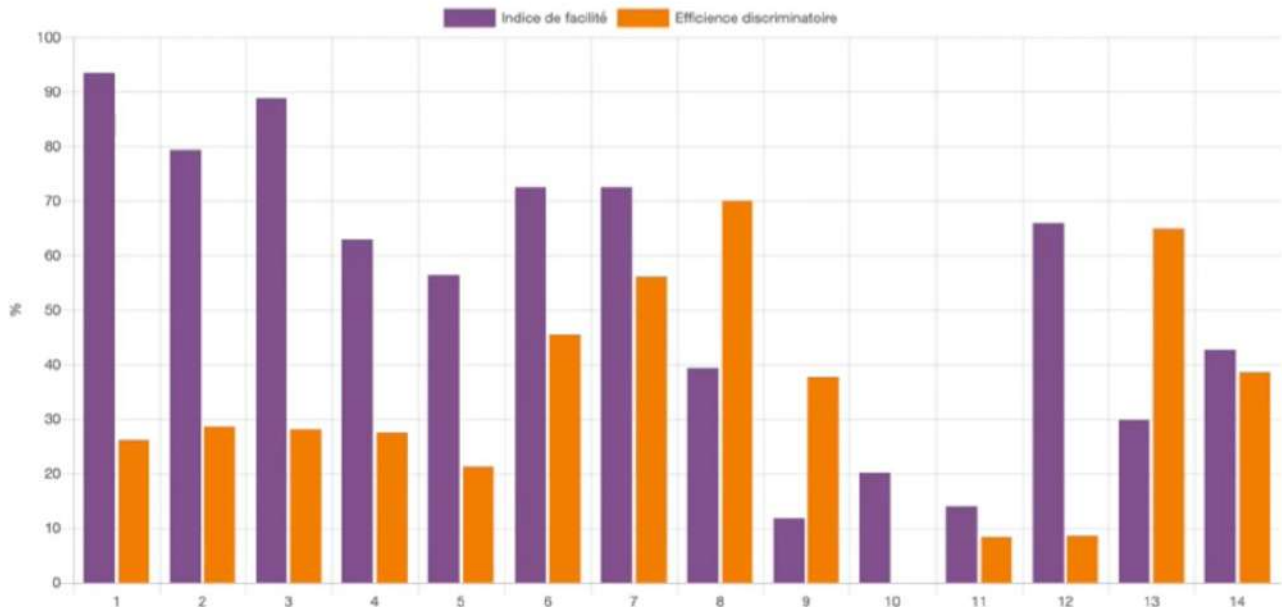
Si  $r_{pbi}$  est :

- proche de 0 (i.e.  $< |.4|$ ) alors la réussite de l'item est nulle ou très faible. On peut le rejeter
- positif et élevé => les élèves arrivent bien et il faut garder la question.

Si l'indice est faible (i.e. il y a une faible corrélation entre l'item et le test), alors ça veut dire que l'item mesure quelque chose d'autre que le test complet.

Un exemple de test à 14 questions

### Statistique des positions de question



En 6 et 7 : un item discriminant n'est pas forcément difficile

Un item à rejeter a presque toujours un indice de discrimination faible (pex : 1, 2 3)

Si le test a plusieurs dimensions (pex : plusieurs types de compétences) alors il faut faire l'indice entre l'item et la dimension.

Un test à une bonne **cohérence interne** si ses indices de discrimination élevés (i.e. proches de 1)

On peut utiliser **tacit 2.0** pour utiliser ces indices pendant l'apprentissage. Il permet d'ajuster la zone proximale de développement.

## 2.4. Quiz

### 1. Choisir les définitions correspondantes

Transfert	→	Capacité de l'apprenant à réutiliser les acquis de l'apprentissage dans une autre situation,
Équité	→	Effet du dispositif sur les écarts entre les élèves,
Niveau de maîtrise	→	Niveau atteint par l'élève à la fin de la formation,
Transfert	→	Post-test différé et/ou dans un autre contexte,
Équité	→	Coefficient de variation,
Efficacité	→	Gain relatif d'apprentissage,
Efficacité	→	d de Cohen



2. Pour concevoir un outil d'évaluation, il est intéressant de se tourner vers les taxonomies. Quels sont leurs avantages?

- Elles permettent de préciser le degré de complexité réel de la tâche. → Vrai,
- Elles permettent d'identifier les processus mis en place par les apprenants. → Faux
- Elle permettent de solliciter une réponse en faisant varier le niveau de certitude. → Faux
- Elles permettent de varier les opérations cognitives des tâches demandées. → Vrai

3. Lequel des énoncés suivants correspond à un élève faisant preuve de compétence non opérationnelle ?

L'élève donne la bonne réponse et coche un bas degré de certitude.

4. Un élève répond à un questionnaire comprenant des questions à choix multiples associées à des degrés de certitude. La situation la plus problématique est celle où l'élève répond incorrectement avec un degré de certitude élevé. Pourquoi?

La compétence n'est pas acquise et l'élève n'en est pas conscient.

5. Quels sont les indices qui sont utilisés pour valider l'outil d'évaluation?

Le degré de difficulté et l'indice de discrimination

6. Un élève a obtenu 6/10 au prétest et 8/10 au post-test. Cet élève a un gain de ... :  
50 %

7. Dans un questionnaire, qu'est-ce que l'indice  $P_i$ , variant entre 0 et 1?

C'est la proportion d'élèves ayant répondu correctement à un item.

8. A quoi correspond le rapport entre la progression observée et la progression possible ?

Le gain relatif

9. Les questions à choix multiples (QCM) comprenant des degrés de certitudes permettent à l'enseignant de mesurer de manière subtile le niveau opérationnel des répondants. Quel niveau d'opérationnalité peut être associé à un étudiant qui donne une réponse incorrecte avec un faible degré de certitude.

Incompétence reconnue

10. On s'intéresse à l'effet de différentes modalités d'intervention tutorale sur les performances d'étudiants engagés dans un travail collectif asynchrone. On a calculé la taille de l'effet pour chaque modalité et on a obtenu pour l'une de celle-ci 1,01. Que pouvez-vous en conclure ? Cette modalité se traduit par une taille d'effet ...

large

11. Il existe quatre variables complémentaires pour évaluer les résultats d'apprentissage : la progression, le degré de maîtrise, le transfert et l'équité.

Un enseignant réalise un prétest et un post-test et calcule le gain relatif de ses élèves en divisant la différence entre le post-test et le prétest par la différence entre le résultat maximum et le prétest. Quelle est la variable qui a été mesurée par l'enseignant ?

La progression

12. A l'issue d'une expérimentation, nous obtenons les résultats suivants pour le prétest : moyenne de 2,59 et écart-type de 1,78. Pour le post-test, la moyenne est de 4,83 et l'écart-type est de 1,93. Quel effet peut être mesuré en utilisant toutes ces données ?  
équité

13. Nous obtenons les valeurs suivantes pour le prétest : moyenne de 2,59 et écart-type de 1,78. Pour le post-test, la moyenne est de 4,83 et l'écart-type est de 1,93.  
En termes d'équité, le groupe est plus **homogène** à la fin de de l'apprentissage qu'au début de l'apprentissage.

14. Un enseignant souhaite observer l'efficacité d'un manuel avec de la réalité virtuelle, il souhaite pour cela calculer la progression de chacun de ses élèves. Concernant un de ses élèves, Pierre, il observe que celui-ci obtient pour son premier test 12/20. Après avoir suivi le cours dans l'environnement numérique mis en place par l'enseignant, Pierre repasse le même test et obtient à présent 9/20.  
La progression de Pierre est de **-25**

15. L'utilisation des degrés de certitude dans les QCM donne la possibilité de calculer un score de confiance et un score de prudence. Dans quelle situation l'apprenant a-t-il le meilleur niveau de maîtrise?

Quand l'apprenant a un score de confiance (a) et un score de prudence (b).  
**(a)** élevé - **(b)** faible

## 3. Que nous dit l'analyse du processus des apprenants ?

Objectif spécifique du module :

- Décrire les différents types de traces pour objectiver le processus d'apprentissage
- Utiliser les traces pour évaluer les effets d'un environnement d'apprentissage

### 3.1. La notion de trace

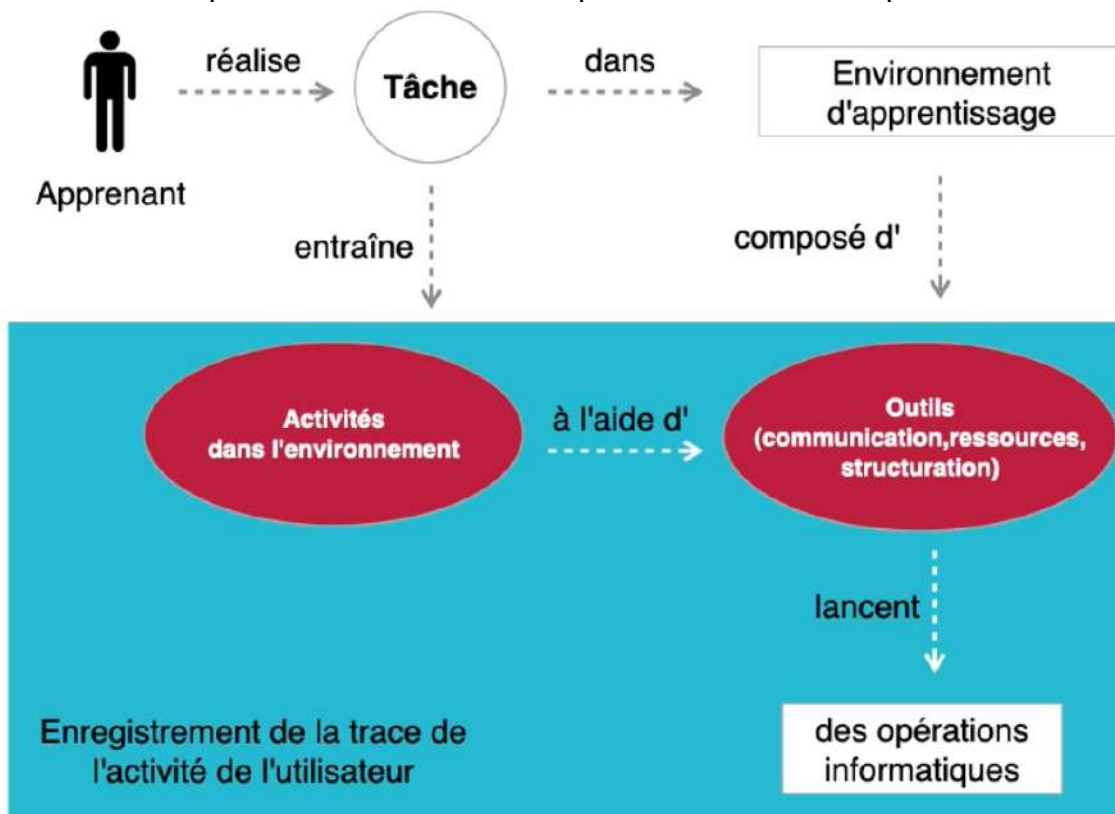
#### 3.1.1. Les traces numériques

Le processus d'apprentissage est lié au parcours des apprenants et aux données recueillies en cours de réalisation des activités

Pourquoi analyser les traces ?

Ça permet de suivre le parcours des apprenants. L'idée est que la compétence ne se voit qu'à travers la trace qu'elle laisse dans sa mise en œuvre. Elles permettent de mieux décrire le parcours des apprenants. La récupération de traces est plus facile dans un environnement numérique.

Au niveau informatique, voici comment la récupération de traces se passe :



On peut récupérer ces traces dans des bases de données. Par exemple, sur moodle on peut avoir le journal d'apprentissage de l'élève.

Il peut y avoir plusieurs types de traces :

- logs
- messages dans un chat
- éditions d'un wiki
- consultation de ressources en ligne
- scores à un questionnaire

### **3.1.2. Les traces non-numériques**

Les ENA n'ont pas toujours d'enregistrement de traces. On peut parfois avoir une approche plus « clinique » : on fait des observations directes.

Par exemple, dans le logiciel « oppia » il n'y a pas d'accès aux traces.

## **3.2. La nature des traces**

4 catégories de traces :

- complétions : les réalisations dans l'environnement
- interaction : les échanges entre apprenants ou avec le formateur
- usages des différentes fonctions ou des ressources
- le temps

### **3.2.1. Les réalisations**

On peut prendre toutes les tâches de production (pex : etherpad permet de sortir les résultats d'un brainstorming ou encore les quiz de moodle qui permet de comprendre les réussites d'une question ou la progression d'un participant).

### **3.2.2. Les usages**

Il y en a d'énormément de type, comme l'occulométrie (eye-tracking), le screencast, l'utilisation des aides, la lecture des vidéos, etc.

L'analyse de ces traces permet de déterminer différents profils d'élèves, styles d'apprentissage, ou encore de voir les difficultés ou les points d'intérêts des participants.

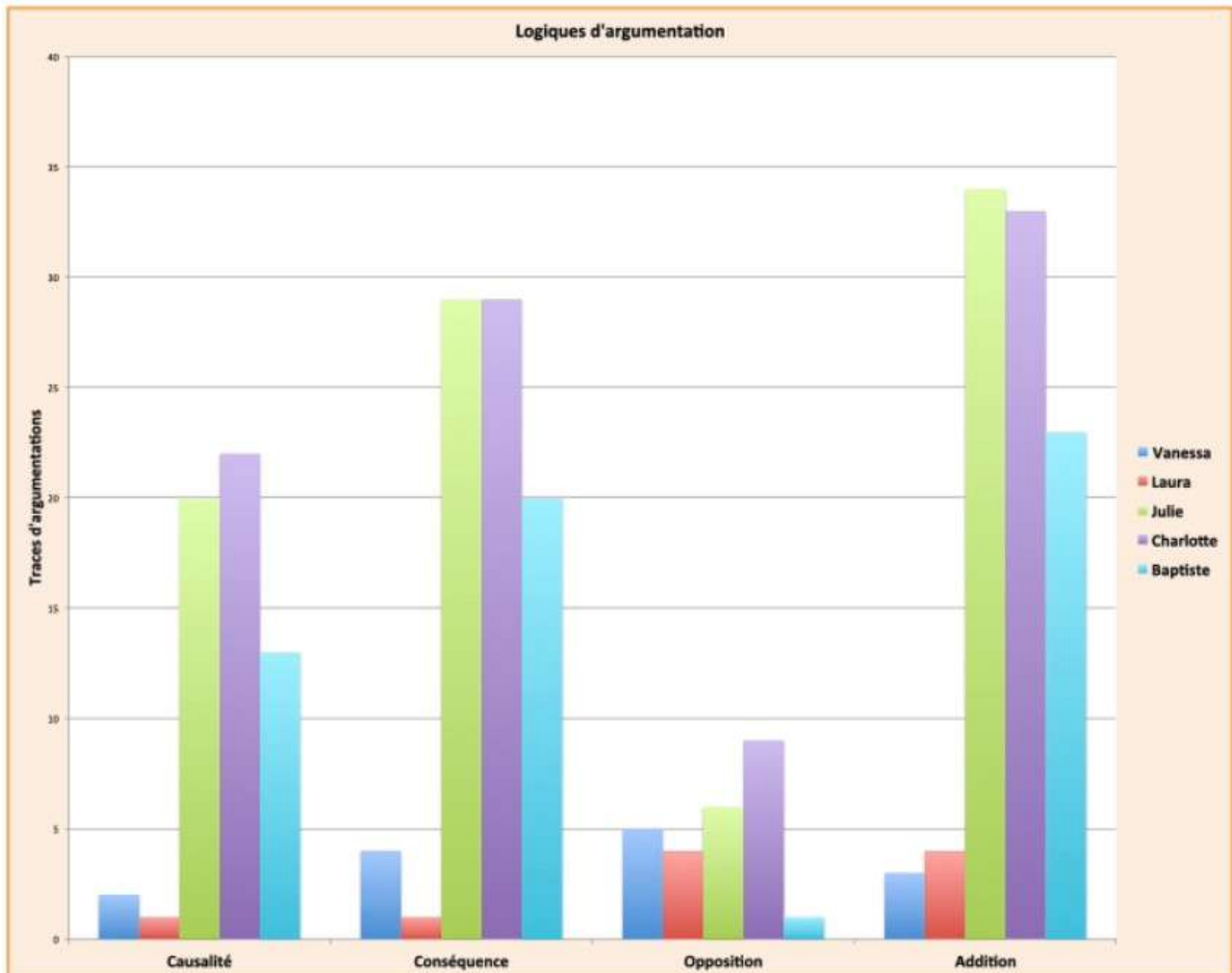
### **3.2.3. Le temps**

Il est extrêmement lié à la réussite. Il est très facile à observer dans un environnement numérique. Par exemple, sur le logiciel TACIT, on peut voir le temps pour répondre à chaque question de QCM. Selon Slavin, le temps, en particulier le temps effectif sur la tâche est un facteur explicatif de la qualité de l'apprentissage.

### 3.2.4. Les interactions

On va plutôt regarder les échanges entre apprenants. Ce qui est difficile, c'est que la masse d'information est très très importantes et donc l'analyse est chronophage et difficile.

On peut aussi regarder le contenu des échanges et décortiquer les discussions entre apprenants avec une analyse dite « lexicométrique ». On peut en déduire la « qualité ».

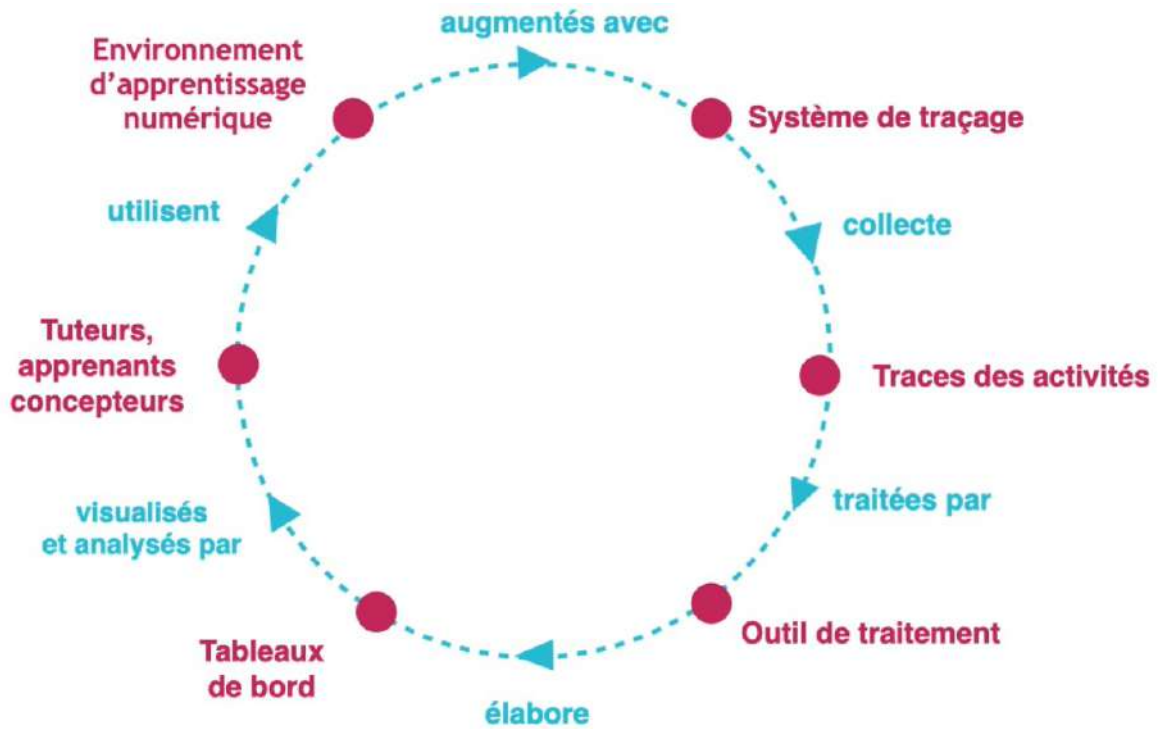


Exemple, d'analyse lexicométrique d'un échange argumentatif.

### 3.3. Utilisation pédagogique des traces

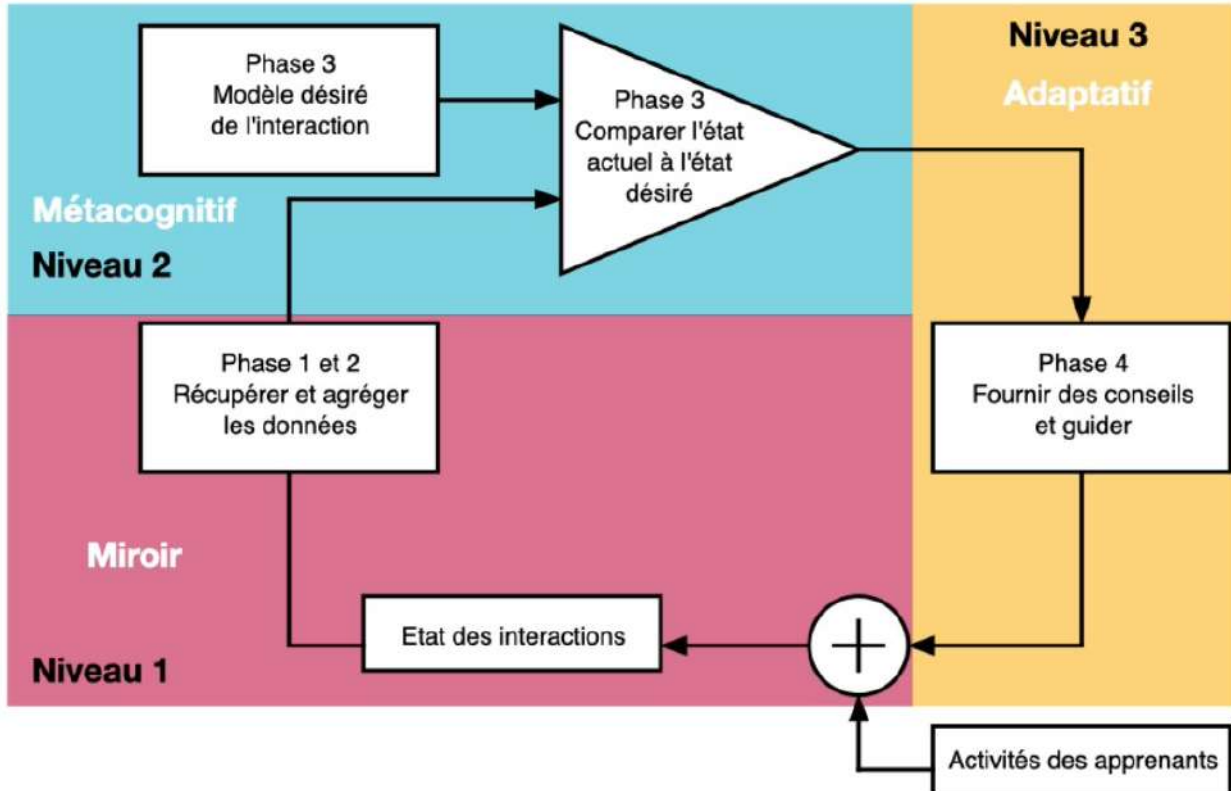
L'important est de synthétiser les traces via un tableau de bord et des indicateurs. Les traces peuvent être utiles pour les enseignants, mais aussi les apprenants (via l'auto-régulation). Ce n'est pas que l'activité qui fait progresser l'apprenant, mais aussi sa connaissance du résultat de celle-ci.

3. Que nous dit l'analyse du processus des apprenants ?



Modèle de May (2010)

Jermann et Dillenbourg proposent un modèle à trois niveaux



Modèle de Jermann & Dillenbourg (2008)

### **3.3.1. Niveau 1 : Miroir**

Il propose une simple description avec des indicateurs (sans objectif à atteindre). Dans Teams, un miroir montre l'ensemble de l'activité des apprenants (pex : échanges en réunions, réalisation des activités, consultation des ressources, etc.)

Jaillet propose de compléter le miroir avec 3 types d'informations :

- l'assiduité : fréquence de connexion à l'environnement
- disponibilité : durée de la connexion
- implication : nombre d'opérations effectuées.

### **3.3.2. Niveau 2 : Métacognitif**

Le tableau de bord va plutôt mettre une situation « idéale » à atteindre. Elle permet de stimuler l'auto-évaluation chez les apprenants.

Par exemple, on peut montrer le taux de prise de parole de chaque apprenant afin qu'ils dans une tâche de discussion. C'est aussi à ce niveau que fonctionne la smartwatch (on montre le nombre de calorie à brûler dans la journée, et ou on en est par rapport à cet objectif).

### **3.3.3. Niveau 3 : Adaptatif**

On peut proposer un parcours personnalisé sur la base des traces. Par exemple, duolingo adapte ses exercices en fonction des réussites des apprenants et propose des rappels du vocabulaire des apprenants en fonction de leur niveau de maîtrise.

## **3.4. Quiz**

*1. A quel type de trace "en cours de processus d'apprentissage" correspondent les énoncés suivants:*

*Un participant à un MOOC clique sur le tableau de bord de progression.*

→ Usages

*Deux élèves coordonnent les différentes tâches d'un travail via une messagerie instantanée.*

→ Interactions

*Un apprenant a effectué la séquence dans le livre numérique en 22 minutes.*

→ Temps

*Pour organiser ses connaissances, un élève construit une synthèse avec ses camarades dans un etherpad.*

→ Réalisations

*Un tuteur répond à une question posée sur le forum de discussion du cours en ligne.*

→ Interactions

*Lors de sa première tentative, un élève a obtenu la note de 8/10 au le test formatif qui concernant la matière travaillée durant le cours.*

→ Réalisations



2. *L'analyse des traces des apprenants permet de créer des indicateurs. Lesquels parmi les choix proposés sont des traces des apprenants ?*

Toutes (i.e. Les échanges entre le tuteur et l'apprenant, L'élaboration d'une carte conceptuelle dans le cadre de l'activité collaborative, Le temps de connexion d'un utilisateur, Le nombre de consultation d'une ressource)

3. *Les traces peuvent être mises à disposition des utilisateurs de manière didactique. Le développement de la visualisation s'opère en plusieurs étapes successives. Quelle est la première étape ?*

L'enregistrement de la trace

4. *Le tableau de bord de Tacit permet de visualiser quels types de traces ?*

Le temps

Les réalisations

5. *Selon Slavin, le temps **effectif sur la tâche** est un facteur explicatif de la qualité de l'apprentissage.*

6. *Pour Jaillet (2005), l'assiduité peut être associée à la durée de connexion.*

Faux

7. *L'analyse d'une synthèse réalisée dans un etherpad par une équipe restreinte offre une possibilité d'évaluer le degré de collaboration de celle-ci.*

Vrai

8. *Selon Jaillet (2005), l'implication correspond au nombre d'opérations réalisées dans l'environnement.*

Vrai

9. *Dans l'utilisation pédagogique des traces, un tableau de bord de type miroir fournit un objectif précis à l'apprenant.*

Faux

10. *L'oculométrie permet d'observer les fixations visuelles des apprenants sur l'écran.*

Vrai

## 4. Quels outils validés pour recueillir la perception ?

Objectif spécifique du module :

- Distinguer différents indicateurs relatifs à l'avis des acteurs
- Découvrir différentes démarches pour recueillir la perception des acteurs via des questionnaires standardisés

### 4.1. Pourquoi mesurer les perceptions ?

Les dimensions « Processus » et « produits » sont en lien avec ce que les participants réalisent effectivement ; elles se mesurent principalement via l'observation.

La dimension « Perception » est en lien avec ce que les apprenants expriment par rapport à leur expérience d'apprentissage. Parfois c'est la seule information disponible. C'est ce que les apprenants pensent du dispositif, mais aussi ce qu'en pensent les enseignants, concepteurs, tuteurs, responsables de formation, voire même les parents. C'est donc les croyances et opinions de toute personne liée au dispositif.

### 4.2. Comment mesurer les perceptions ?

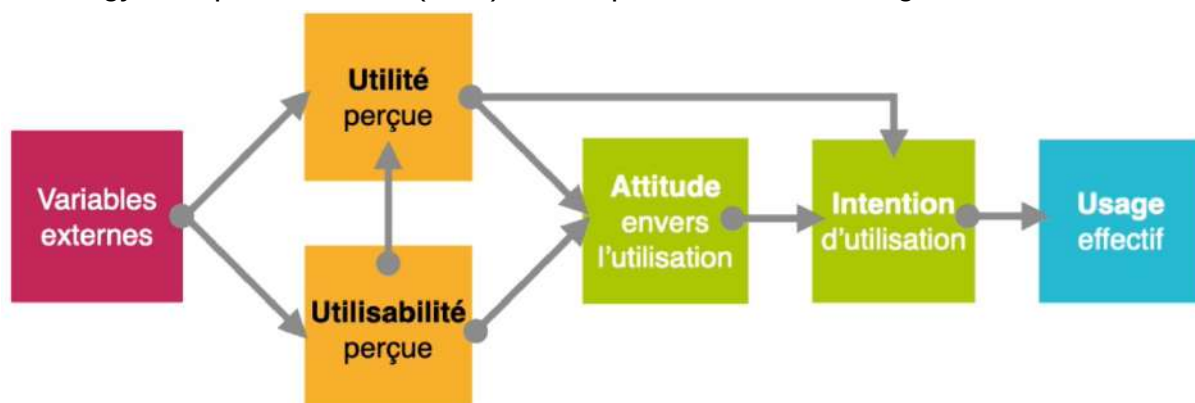
On peut le faire avec des questionnaires validés, des questionnaires à construire, ou via des entretiens.

L'approche par questionnaire d'opinion

C'est une bonne approche pour recueillir l'opinion d'un grand nombre de sujets. Voici quelques exemples de questionnaires validés

#### 4.2.1. Mesurer l'utilité et l'utilisabilité

Technology Acceptance Model (TAM) : l'acceptation des technologie.



Utilisabilité : c'est le degré selon lequel un produit/système peut être utilisé par l'utilisateur pour atteindre un but défini avec efficacité

Utilité perçue : c'est la possibilité d'atteindre des objectifs (ici d'apprentissage) en réalisant une tâche.

Ces deux variables sont importantes. Elles peuvent être mesurées avec le *questionnaire de Davis*, en 12 items, 6 par dimension.

		..... est un outil ...
<b>UTILISABILITE</b>	<b>Apprentissage</b>	pour lequel il m'a été facile d'apprendre le fonctionnement
	<b>Facilité</b>	qui m'a semblé facile à utiliser
	<b>Clarté et compréhension</b>	avec lequel l'interaction est claire et compréhensible
	<b>Contrôlabilité</b>	qui est facile à utiliser pour concevoir une carte conceptuelle
	<b>Habilité</b>	pour lequel m'a été facile de devenir compétent
	<b>Flexibilité</b>	avec lequel je peux interagir en souplesse
<b>UTILITE</b>	<b>Rapidité</b>	qui me permet de réaliser mes tâches plus rapidement
	<b>Performance</b>	qui améliore la qualité de mon travail
	<b>Productivité</b>	qui augmente ma productivité
	<b>Facilité</b>	qui facilite l'accomplissement de ma tâche
	<b>Efficacité</b>	qui améliore l'efficacité de mon apprentissage
	<b>Transfert</b>	qui me sera utile dans mon contexte professionnel

Le questionnaire *System usability scale* en 10 items, ne mesure que l'utilisabilité. Il est rapide à faire et est plutôt sensible.

1. Je pense que j'aimerais utiliser fréquemment ce système
2. J'ai trouvé ce système inutilement complexe
3. J'ai trouvé ce système facile à utiliser
4. Je pense que j'aurais besoin du support d'un spécialiste pour utiliser ce système
5. J'ai trouvé que les différentes fonctions de ce système étaient bien intégrées
6. J'ai trouvé ce système trop incohérent
7. Je pense que ce système sera facile à apprendre pour beaucoup de personnes
8. J'ai trouvé ce système très contraignant à utiliser
9. Je me suis senti(e) en confiance lorsque j'ai utilisé ce système
10. J'ai dû apprendre beaucoup de choses avant de me sentir familiarisé(e) avec ce système

Ce questionnaire produit un score qui va de 0 à 100. Un dispositif avec un score de 80 est considéré comme bon.

### 4.2.2. Mesurer les émotions

Les compétences émotionnelles jouent un rôle-clé dans l'apprentissage. Elles soutiennent l'attention, et sont importantes dans le traitement de l'information et la mémorisation.

Il y a plusieurs manières de mesurer les émotions (Redet et al., 2016) : ça peut se faire de manière objective (via la prise de données physiologiques) ou subjective (le ressenti des émotions).

Données	Outils de mesure	Moment	Intrusif	Durée installation/ consigne	Durée et difficulté d'analyse
OBJECTIVES	Électroencéphalogramme	Temps réel	●●●○	●●●●	●●●●
	Électrocardiogramme	Temps réel	●●○○	●●●○	●●●●
	Réactions électrodermales	Temps réel	●●○○	●●○○	●●●●
	Respiration	Temps réel	●●○○	●●●○	●●●●
	Expression faciale	Temps réel	○○○○	○○○○	●●○○
SUBJECTIVES	Questionnaires classiques	A posteriori	○○○○	●●○○	○○○○
	Questionnaires visuels	A posteriori	○○○○	○○○○	○○○○
	Entretien	A posteriori	○○○○	●○○○	●●●○
	Verbalisations	Temps réel	●●●○	○○○○	●●●○
	Auto-confrontation	Temps différé	○○○○	●●●○	●●○○

On peut voir que les questionnaires, bien que moins objectifs, sont bien plus simples à mettre en place et sont moins intrusifs.

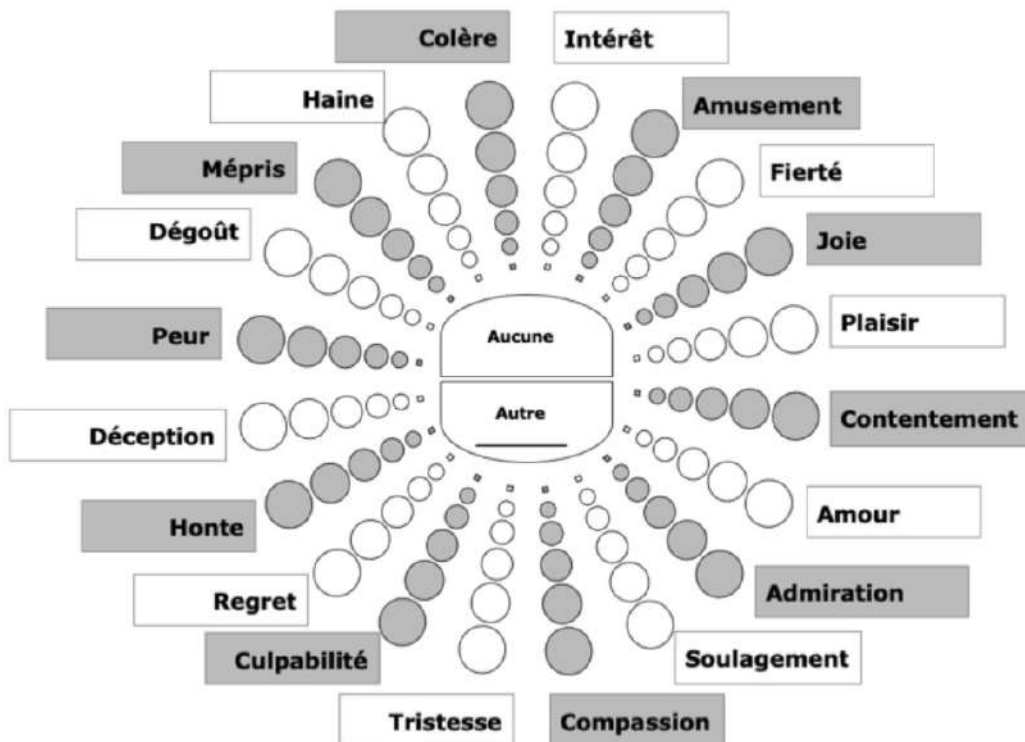
*Computer emotion scale (Kay & Loverock, 2008)*

C'est un questionnaire à 12 items qui mesure les émotions.

Emotions	1 jamais	2	3	4 tout le temps
1. Satisfait(e)				
2. Désespéré(e)				
3. Anxieux/se				
4. Irritable				
5. Excité(e)				
6. Découragé(e)				
7. Peu sûr(e) de moi				
8. Frustré(e)				
9. Curieux/se				
10. Impuissant(e)				
11. Nerveux/se				
12. En colère				

*La roue des émotions (Sacharin, Schlegel & Scherer, 2012)*

C'est une approche plus graphique pour mesurer les émotions



### 4.2.3. La charge cognitive

C'est la quantité de ressources mentales investies dans une tâche. Elle est présente dans le modèle SOI (Sélectionner, Organiser, Intégrer), lié au modèle d'apprentissage multimédia de Mayer. Il est important de l'évaluer, parce qu'elle impact l'apprentissage.

Elle dépend de la complexité de la tâche (le nombre d'éléments traité et à mettre en relation), les ressources de l'individu (i.e. ses connaissances de la tâche), et de la manière dont la tâche est présentée.

*Le questionnaire Nasa-TLX (Hart & Staveland, 1988)*

Il est composé de 6 sous-échelles

Contraintes imposées par la tâche	<b>Exigence mentale</b>	Quelle a été l'importance de l'activité mentale et intellectuelle requise ?
	<b>Exigence physique</b>	Quelle a été l'importance de l'activité physique requise ?
	<b>Exigence temporelle</b>	Quelle a été l'importance de la pression temporelle causée par la rapidité nécessitée pour l'accomplissement de la tâche ?
Interactions du sujet avec la tâche	<b>Performance</b>	Quelle réussite pensez-vous avoir eu dans l'accomplissement de votre tâche ?
	<b>Effort</b>	Quel degré d'effort avez-vous dû fournir pour exécuter la tâche demandée ?
	<b>Frustration</b>	Pendant l'exécution du travail vous êtes-vous senti découragé, irrité, stressé, sans assurance ?

### 4.2.4. Une mesure globale de l'expérience utilisateur

L'AttrakDiff (Hessenwahl et al., 2003)

C'est un questionnaire généraliste ; il se compose de 10 items sur 4 sous échelles :

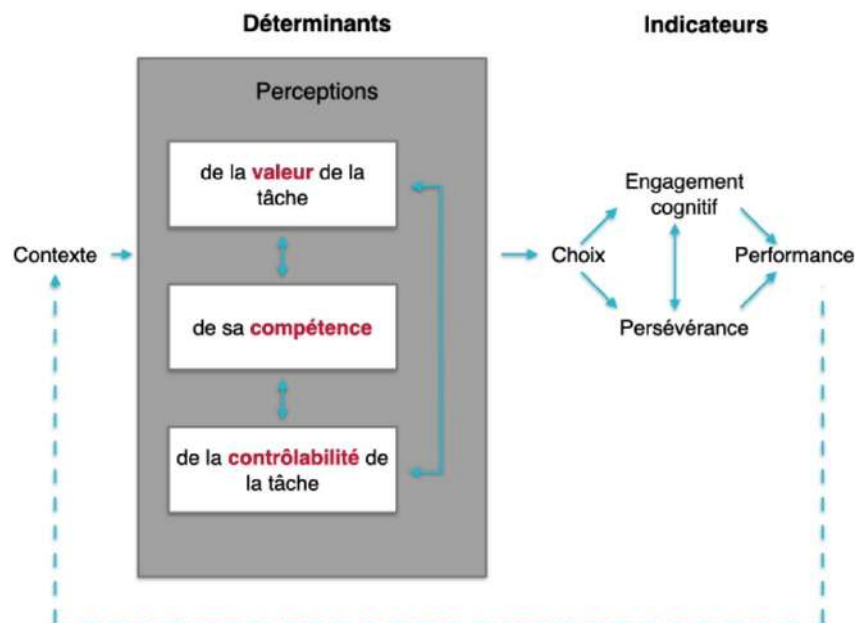
- 1. qualités pragmatiques : dans quelle mesure il permet à l'usr d'atteindre son but
- 2. qualités hédoniques – stimulation : dans quelle mesure l'individu est stimulé
- 3. qualités hédoniques – identité : dans quelle mesure l'usr s'identifie au produit
- 4. attractivité globale : valeur du produit basé sur la perception des qualités pragmatiques et hédoniques

Simple	○	○	○	○	○	○	○	Complicé
Laid	○	○	○	○	○	○	○	Beau
Pratique	○	○	○	○	○	○	○	Pas pratique
De bon goût	○	○	○	○	○	○	○	De mauvais goût
Prévisible	○	○	○	○	○	○	○	Imprévisible
Bas de gamme	○	○	○	○	○	○	○	Haut de gamme
Sans imagination	○	○	○	○	○	○	○	Créatif
Bon	○	○	○	○	○	○	○	Mauvais
Confus	○	○	○	○	○	○	○	Clair
Ennuyeux	○	○	○	○	○	○	○	Captivant

Il en existe aussi une version plus exhaustive de 28 items

### 4.2.5. Conclusion

Selon Viau, toutes ces mesures sont complémentaires et permettent d'appréhender la motivation des élèves.



La valeur de la tâche peut être associée à l'utilité et aux émotions. La perception de compétences, peut être associée à l'efficacité dans la réalisation de la tâche. La contrôlabilité peut être associée à l'utilisabilité et la charge cognitive.

### 4.3. Quiz

1. On peut classer les différents items du Nasa TLX en **deux** catégories principales.

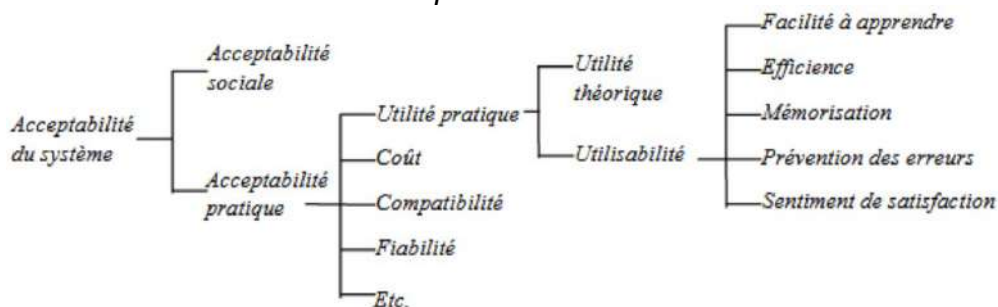
2. Le questionnaire de Davis (1989) permet de récupérer l'opinion d'un individu après une tâche. Pour chaque énoncé, identifiez s'il s'agit d'une opinion liée à l'utilité ou à l'utilisabilité des technologies par l'utilisateur.

- L'utilisation du wiki dans mon emploi augmenterait ma productivité. → **utilité**
- Mon interaction avec le wiki a été claire et compréhensible. → **utilisabilité**

3. Lors de l'utilisation d'outils pour évaluer les émotions, les données récoltées peuvent être d'ordre objectif (indicateur de l'activité émotionnelle), ainsi que subjectif (interprétation de l'utilisateur ou d'un observateur du ressenti des émotions). Ce sont les retours subjectifs des participants sur leurs émotions qui se révèlent bien plus exploitables que les données d'ordre objectif .

**Vrai**

5. Comme le modèle de Davis, Bagozzi et Warshaw, il en existe d'autres qui s'intéressent à l'utilité et l'utilisabilité d'un outil. En voici un exemple : le modèle de Nielsen. Laquelle des propositions ci-dessous est la description correcte de ce modèle ?



L'acceptabilité d'un système se subdivise en deux dimensions : l'acceptabilité sociale et l'acceptabilité pratique. L'utilité théorique et l'utilisabilité sont les facteurs déterminants l'utilité pratique de l'outil (c'est-à-dire le but que le système permet réellement d'atteindre). Cette dernière est une sous-dimension de l'acceptabilité pratique.

6. Dans le modèle de Viau, la valeur de la tâche peut être associée au principe de (d') **utilité**.

7. La **charge cognitive** impacte la qualité de l'apprentissage. Elle est liée à la mémoire de travail. Elle dépend de la complexité de la tâche , des connaissances de l'individu à propos de cette tâche et de la manière dont la tâche est présentée.

8. Le SUS permet de mesurer l'utilité de l'environnement numérique d'apprentissage. **Faux**

9. Selon le modèle de l'acceptation de la technologie, l'utilité perçue influence l'utilisabilité perçue. **Faux**

10. Lors de l'utilisation du questionnaire "Attrackdiff", il est possible de prendre en compte la mesure globale de l'expérience utilisateur.

**Vrai**

## 5. Comment construire un outil pour analyser la perception ?

Objectifs spécifiques du module :

- Elaborer un questionnaire d'opinions adapté à son contexte
- Distinguer les différentes démarches pour réaliser un entretien

### 5.1. Construire son propre questionnaire

Il y a plusieurs étapes pour concevoir un questionnaire :

1. Structurer son questionnaire pour en déterminer le fond
2. Conception du questionnaire
3. Passation et analyse des résultats

Pour structurer un questionnaire, il faut identifier les dimensions à explorer pour évaluer l'environnement (pex : l'encadrement, les tâches, les outils utilisés, etc.).

Pour chaque dimension, on peut associer un ou plusieurs items. Chaque item peut être associé à une variable. Pour chaque item doit être plutôt simple, avec des formulations courtes.

Il y a deux type d'items :

- les questions fermées
- les questions ouvertes

Les questions fermées

elles son pertinentes lorsqu'on interroge un grand nombre de personne. La passation et le traitement sont plutôt faciles.

Il y a plusieurs manières de récolter l'informations :

- échelle dicotomique
- échelle de likert
- échelle unidimensionnelle (une droite sur laquelle on se positionne)
- pictogrammes
- échelle bidimensionnelle (une échelle sur laquelle on se positionne plutôt vers un terme ou un autre. Pex : Simple    o    o    o    o    o    compliqué
- question à choix unique
- question à réponse multiple
- classement

Les échelles peuvent être paires ou impaires. Les échelles impaires permettent de mettre des réponses neutres

L'important est de chercher à maintenir la même échelle, histoire de faciliter le traitement statistique par la suite.



Avant d'analyser les données, il est important de mettre en place un codebook (un dictionnaire des variables mesurées et des items utilisés). On peut ensuite faire l'analyse de données dans des logiciels tels que [Jasp](#) (libre et gratuit).

Les questions ouvertes

Elles permettent d'avoir des réponses plus variées et plus riches, mais il est beaucoup plus long d'analyser les réponses. On peut accélérer l'analyse avec des méthodes telles que la lexicométrie.

Les items peuvent être des :

- complétions courtes (pex : nom, prénom)
- questions longues
- on peut aussi guider les réponses (pex : donner des catégories dans lesquelles répondre et proposer une valence)

pex :

Ce qui fonctionne bien ou ce qui doit être amélioré dans la continuité pédagogique (indiquez + si positif et - si négatif)

1 En tant que parents	2 En tant qu'enseignants	3 En tant qu'élèves - étudiants	4 En tant que responsables d'institution scolaire
-----------------------	--------------------------	---------------------------------	---

Dans un premier temps, il est possible de faire une analyse descriptive. On peut ensuite analyser les liens entre les réponses fournies.

Pour les analyses, on peut utiliser *Voyant tools*, un logiciel d'analyse de texte. Il permet par exemple, de dresser un réseau des termes les plus fréquents en fonction de leur proximité

Questionnaires avec items fermés et ouverts

Par exemple, on peut justifier une position sur un item fermé. On peut facilement tout combiner (likert, binaire, etc)

Les logiciels de questionnaires : plickers, wooclap, googleform, etc.

## 5.2. Réaliser un entretien

Un entretien va réunir moins de personnes qu'un questionnaire. C'est une discussion entre deux personnes dont le chercheur est neutre et dirige l'entretien. Il va être mené avec un guide, en fonction du type d'entretien.

Le guide comporte des thèmes et des questions à proposer lors de la rencontre. Ils permettent d'obtenir des informations plus approfondies qu'avec un questionnaire.

Le type d'entretien dépend du thème de la recherche et aux nombres de personnes simultanées à interviewer.

Dans les entretiens individuels, on a :

l'entretien libre, un discours continu dont aucune question n'est préparée à l'avance.

L'entretien semi dirigé, c'est un discours par thème, avec un ordre qui varie selon l'interviewé.

L'entretien dirigé, basé sur des questions précises dans un ordre déterminé.

Dans les entretiens collectifs (i.e. au moins deux personnes simultanément), on a le focus groupe, un type d'entretien semi-dirigé. L'objectif de ce procédé est de récupérer des données à partir des interactions intragroupe.

A partir de l'enregistrement, il faut retranscrire l'entretien. On peut ensuite analyser le corpus. Il y en a plusieurs :

Dans la démarche déductive, les catégories sont issues d'une revue de la littérature ; si une information du corpus n'entre pas dans une des catégories, on ne l'encode pas. Dans une démarche inductive, le chercheur construit ses catégories à partir du corpus. Dans l'approche itérative, on travaille de la même manière que la déductive, excepté que les catégories peuvent être modifiées au cours de l'analyse de données.

Attention, l'entretien et les analyses peuvent être parasités par les biais cognitifs. Ils peuvent beaucoup influencer les analyses. Les biais peuvent provenir du participant (par exemple, le biais de courtoisie : chercher à rester correct sur le plan social, ou le biais de négativité : se rappeler uniquement des éléments négatifs), ou du chercheur.

### 5.3. Combiner plusieurs outils

On peut combiner des questionnaires validés et y combiner des questionnaires construits (ouverts et fermés). On peut aussi combiner questionnaire (pour faire une sélection par exemple, pour dégager des profils), et des entretiens (pour chaque cluster).

### 5.4. Quiz

1. L'utilisation de questions ouvertes permet d'obtenir, auprès des répondants, des informations plus **riches**

2. L'utilisation d'un questionnaire à **choix unique** permet au chercheur de récolter l'opinion ou le choix plus "tranché" du répondant.

3. L'effet de courtoisie est un biais cognitif qui est lié au fait que le chercheur est influencé de manière inconsciente par les résultats qu'ils souhaiteraient obtenir.

Faux

4. L'idéal dans la conception d'un questionnaire à items fermés est de maintenir la même échelle tout au long du questionnaire afin de faciliter par la suite son traitement statistique et la cohérence de l'analyse.

Vrai

5. Pouvez-vous associer la bonne démarche à sa description ?

*Le chercheur va effectuer une revue de la littérature la plus exhaustive possible afin d'identifier des catégories hiérarchiques dans lesquelles il pourra classer les différents verbatims du corpus.*

→ Déductive

*Le chercheur va directement partir du corpus pour créer des catégories hiérarchisées.*

→ Inductive

*Le chercheur réalise une revue de la littérature afin de concevoir une catégorisation hiérarchisée a priori. Il la modifie à partir de l'analyse de corpus.*

→ Itérative

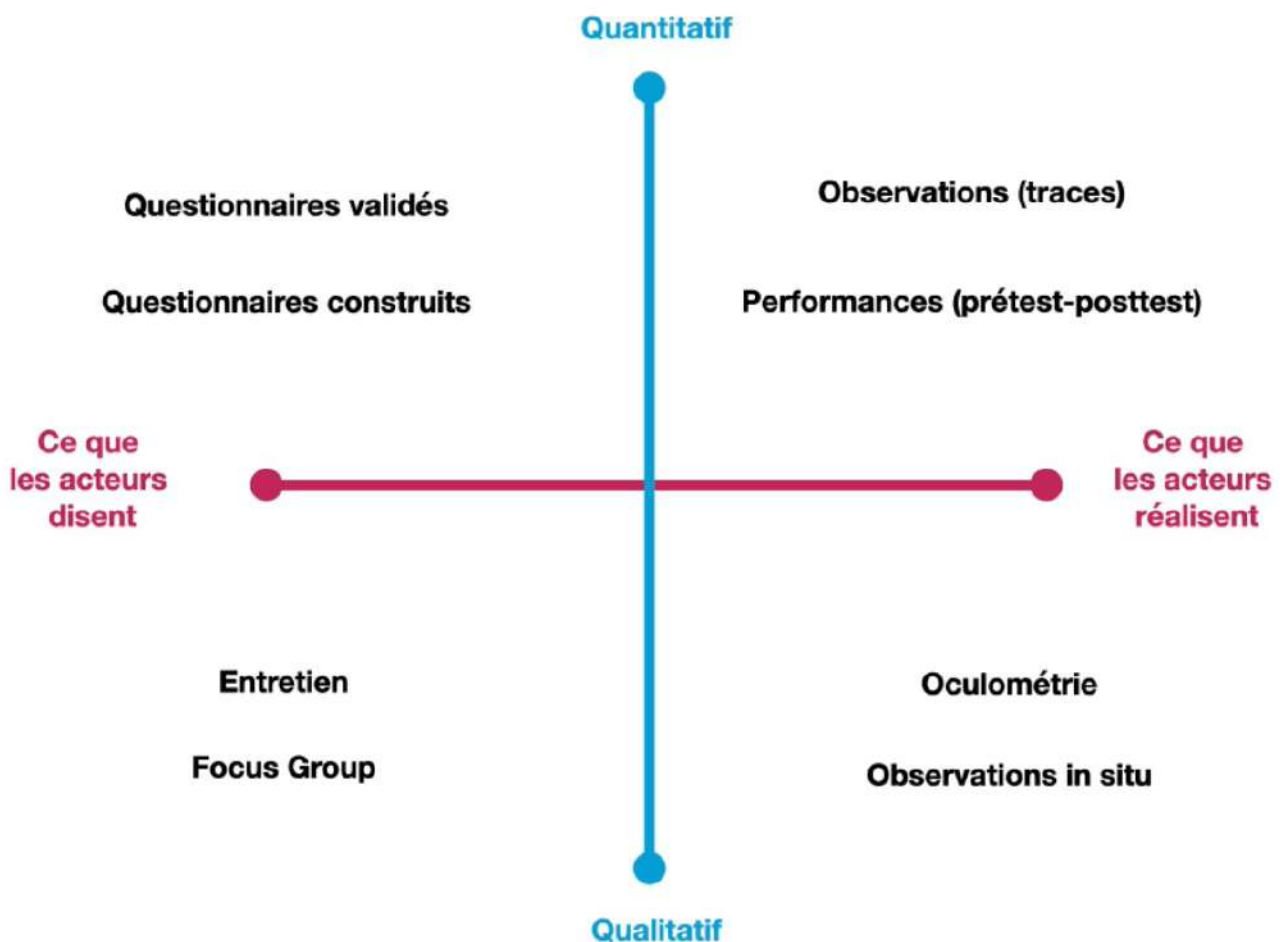
## 6. Quel est l'intérêt de réaliser des analyses croisées ?

Objectifs spécifiques du module :

- Distinguer les démarches de comparaison, de corrélation, de prédiction et de catégorisation
- Décrire ces démarches à l'aide d'exemples concrets

### 6.1. Introduction

Pour évaluer un ENA, on peut se baser sur ce que les acteurs disent et ce qu'ils réalisent ; les données peuvent être qualitatives (pour un nombre réduit de personnes), et quantitatives (pour un plus grand nombre de personnes).



Voici un tableau qui résume les différentes méthodes abordées dans le MOOC pour évaluer un dispositif.

Les méthodes croisées consistent à croiser les données issues de deux dimensions différentes (entre produits, processus, perception profils). Croiser deux dimensions permet de générer de nouvelles questions de recherche.

On peut procéder par comparaison, par corrélation (pour voir le lien entre variables), par prédiction ou par catégorisation.

## 6.2. La comparaison

Ce n'est pas comparer avec ou sans numérique. Il faut plutôt identifier les conditions ou les usages et bénéfiques sont positif.

Par exemple, on peut évaluer deux types d'usages de la tablette en classe (par exemple, une utilisation permanente de l'objet vs non-permanent). De cette manière on peut déduire un rapport de cause à effet.

Autre exemple : l'utilisation d'un chatbot pour l'apprentissage de la géométrie. Deux chatbots ont été développés en parallèle, l'un neutre et l'autre empathique (mais avec le même contenu en soit). Ils ont été utilisés dans 3 contextes : le secondaire, des étudiants et des étudiants en master. On a donc 2 effets principaux sur la perception des participants mesurée avec un questionnaire : le type de chatbot, le profil des utilisateurs et aussi le type d'interaction entre 2 les deux variables

Il est possible d'évaluer plus de 3 variables, mais ça devient plus compliqué. Par exemple, une recherche portant sur l'usage de l'animation pour apprendre a comparé la modalité de groupe (individuel/collaboratif), le type d'image (statique/animée) et la présence/absence de vignette sur l'écran. Pour analyser tout ça, il faut prendre en compte les différents effets principaux et les effets d'interaction.

On ne doit pas toujours comparer entre condition (logique inter), on peut aussi regarder l'évolution des sujets (logique intra) ; on parlera d'analyse diachronique. Par exemple, une recherche regardait l'évolution des aides en AR utilisé dans un livre augmenté. L'usage variait selon les tâche proposées.

Parfois il est intéressant de combiner des variables pour en créer de nouvelles. Par exemple, on peut combiner deux variables du processus : comme la réalisation d'une activité et le temps nécessaire pour la réaliser. Cette logique permet d'aboutir au principe d'efficience (mesuré en  $efficience = \frac{réalisation}{temps}$ ) qui permet d'évaluer le « coût-bénéfice » dans différentes conditions.

Pour objectiver les meilleures conditions pour apprendre et faire apprendre, on peut faire des traitement fréquentistes ou bayésiens sur des logiciels comme [Jasp](#).

### 6.3. La corrélation

La logique est un peu différente : on va regarder les liens entre deux variables.

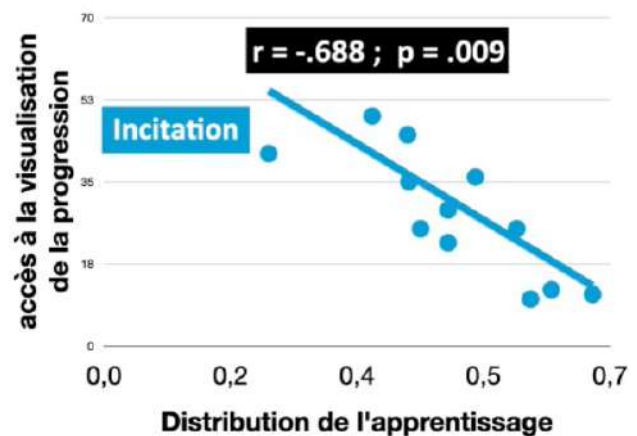
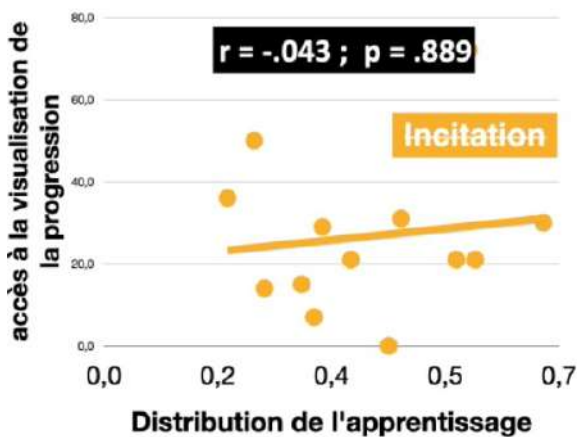
Si la corrélation = 0, alors il n'y a pas de lien entre les deux variables

Si la corrélation = 0.7, alors environ 50 % de la variable soit expliquée par l'autre. Il y a donc une forte relation positive entre les deux variables.

Si la corrélation < 0, il y a une relation négative entre deux variables.

Attention à ne pas confondre corrélation (les deux variables sont liées) et causalité (la variable 1 cause la variable 2). Aussi, la relation entre deux variables n'est pas forcément linéaire, elle peut suivre une courbe polynomiale.

On peut aussi mélanger les approches corrélationnelles et comparatives. Par exemple, une étude observait l'impact de l'utilisation d'un outil de visualisation de l'avancement d'une tâche sur la gestion du temps. Il y avait deux conditions : incitation à utiliser l'outil vs sans incitation.



On peut voir qu'il n'y a pas de lien entre l'usage de l'outil et la progression dans la condition où les apprenants ne sont pas incités à utiliser l'outil. Lorsqu'on les stimule, leur usage est corrélé négativement. En l'occurrence, la mesure de distribution se fait avec un coefficient de Gini qui va de 0 (la distribution est idéale) à 1 (la situation est massée dans le temps).

Sur Jasp, on peut facilement obtenir des matrices de corrélation et les obtenir en fonction de différentes conditions.

### 6.4. Les démarches prédictives

Dans une prédiction, on cherche une relation asymétrique (contrairement à la corrélation qui est symétrique). L'idée est qu'une ou plusieurs variables en expliquent une autre (souvent la performance des élèves).

L'équation est souvent représentée comme :  $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$

$B_0$  est la constante, les  $x$  sont les variables explicatives, les bêta sont les coefficients. Lorsqu'on standardise les variables, on peut déduire laquelle a le plus de poids dans l'explication. L'explication n'est jamais parfaite, c'est pourquoi on ajoute « epsilon », qui représente l'erreur aléatoire ; c'est la valeur résiduelle de la prédiction.

Par exemple, ils ont mesuré les variables qui ont un impact sur le niveau de maîtrise différé. Ils ont trouvé que le nombre de tentatives était le meilleur prédicteur ( $b=.302$ ), venaient ensuite les interactions ( $b=.181$ ) et l'intensité de lecture ( $b=.145$ ). En revanche, ils n'ont pas trouvé d'impact du temps sur la réussite.

Jasp permet de faire plusieurs types de régressions différentes.

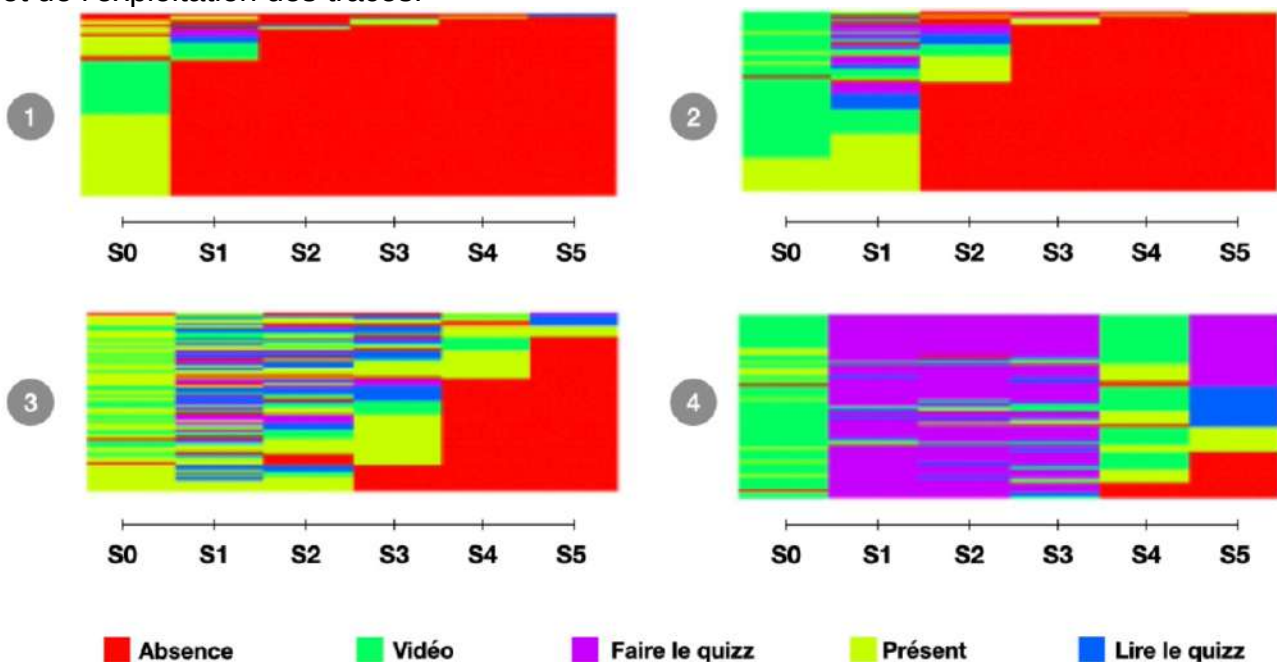
Il est possible de combiner plusieurs dimension pour faire une recherche (profils, produits et processus).

## 6.5. La catégorisation

C'est une procédure statistique qui vis à trouver une structure intrinsèque aux données en les organisant en groupes homogènes et distincts (clusters). Pour former un cluster, les éléments doivent être similaires entre eux et différents des objets regroupés dans d'autres clusters

C'est souvent utilisé en marketing pour faire des profils de consommateurs ou en médecine.

Pour un MOOC, on peut essayer de profiler les participants à partir de leur activité en ligne et de l'exploitation des traces.



Dans cette étude, les chercheurs ont pu mettre en évidence 4 profils sur la base de 5 variables. On voit que les profils 1 et 2 abandonnent très vite. Sur la base de ces

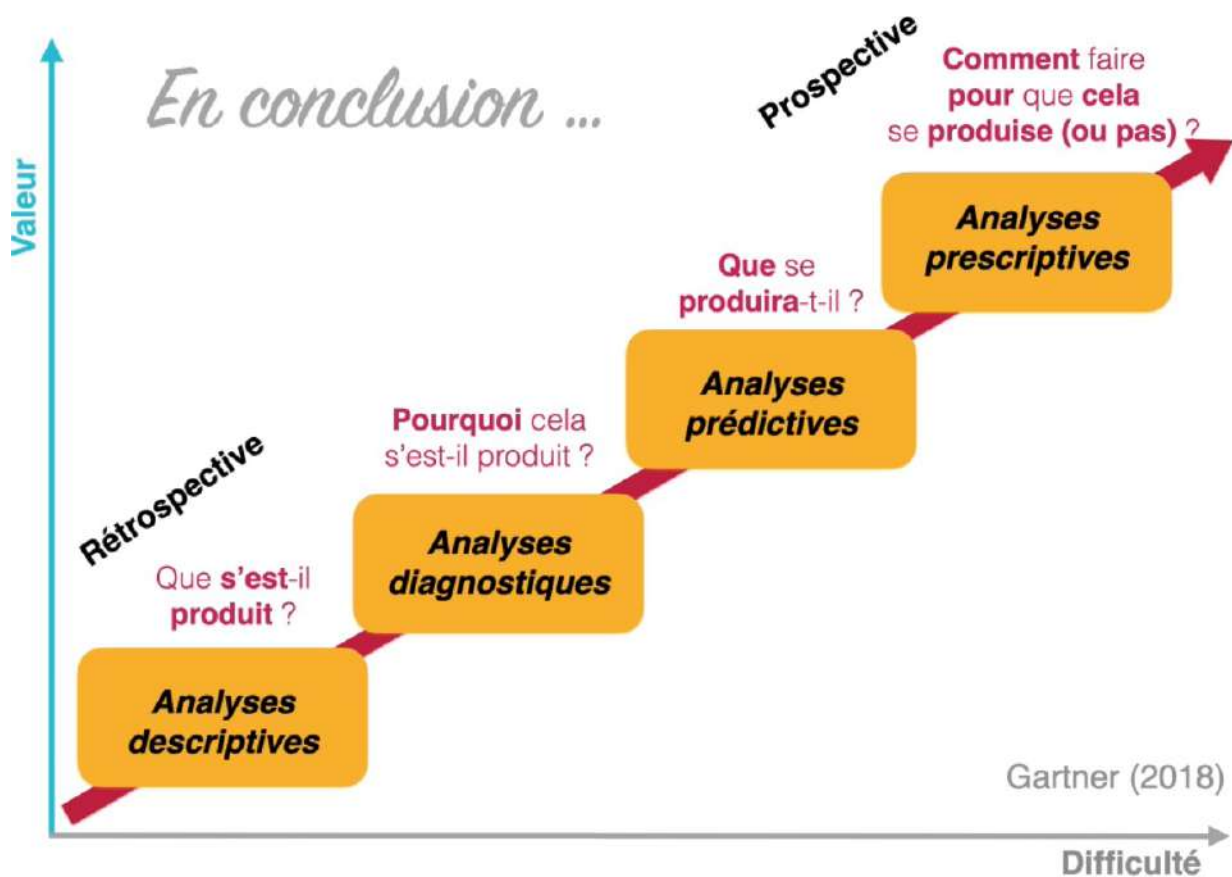
éléments, ils ont dégagé 3 profils : les actifs (qui font toutes les activités), les passifs (qui regardent uniquement la vidéo) et les inactifs.

Cette démarche permet de déduire des variables indépendantes à posteriori ; il permet aussi de synthétiser de l'information. Au niveau pédagogique, il est intéressant de dresser ces profils pour mieux aider les participants dans leur parcours. Ce sont des informations particulièrement utiles pour les tuteurs et les concepteurs.

Il est possible de faire ce type de catégorisations dans Jasp.

## 6.6. Conclusion

Sans croiser d'informations, on peut se poser plusieurs questions.



Il y a la question « que s'est-il produit » avec les analyses descriptives. Avec des analyses diagnostiques, on peut répondre à « pourquoi cela s'est passé » avec des analyses causales, et compléter le tout avec des analyses corrélationnelles. L'analyse par prédiction permet de savoir ce qui se passera dans la séquence d'apprentissage. Pex à partir de traces on peut chercher à déduire le degré de maîtrise final de l'apprenant. Les analyses prescriptives vont encore plus loin : ce sont les recommandations à donner aux élèves. Il y a une logique de différenciation simultanée : on va faire des relances différentes selon les profils identifiés à partir du comportement analysé en cours de processus. On passe d'une logique rétrospective à prospective.

A l'instar d'une taxonomie d'objectifs, le modèle multidimensionnel du MOOC permet de générer des questions de recherches auxquelles on n'aurait pas pensé au départ et ainsi envisager des analyses en fonction des faits observés. Ce tableau permet d'articuler des différentes questions possibles.

		Produits			
		Non		Oui	
Profil	Profil	Processus		Processus	
		Non	Oui	Non	Oui
Perceptions	Non	Quelles sont les habitudes technologiques des participants en dehors du dispositif ?	Quel est l'usage du tableau de bord dans l'environnement ?	Le type de tâche proposé aux participants différencie-t-il le niveau de maîtrise différé ?	Le nombre de tests réalisé est-il en lien avec le gain d'apprentissage ?
	Oui	Le niveau de motivation initial différencie-t-il l'utilité perçue ?	L'avis des apprenants par rapport au tableau de bord concorde-t-il avec leur usage de celui-ci ?	Le degré d'utilisabilité perçue est-il en lien avec le gain d'apprentissage ?	Quels types d'apprenants peut-on distinguer dans l'environnement en tenant compte de leur niveau de maîtrise, de leur perception et de leurs usages ?

## 6.7. Quiz

1. Pouvez-vous remettre ces différentes questions dans l'ordre en fonction du degré de difficulté dans la mise en oeuvre (1=très facile à 4 = très difficile) ?

Très facile → Que s'est-il produit

Facile → Pourquoi cela s'est-il produit

Difficile → Que se produira-t-il

Très difficile → Comment faire pour que cela se produise

2. Dans cet article, quelles sont les démarches utilisées ?

Veillez choisir au moins une réponse :

**Des comparaisons**

Des catégorisations

**Des corrélations**

Des prédictions

Aucune des réponses

Toutes les solutions sont correctes



3. L'idée d'évaluer le coût / bénéfice peut être associée à quel principe ?

Efficienc (ou rendement)

4. La catégorisation a pour objectif de trouver une structure intrinsèque aux données en les organisant en groupes hétérogènes et distincts, appelés « clusters ».

Dans cette explication, un mot est erroné. Lequel ?

Hétérogènes

5. La logique de différenciation simultanée va correspondre à quel niveau d'analyse :

Analyse prescriptive

6. Dans quelle zone du tableau (1 à 8), placez-vous cette question : Existe-t-il un lien entre l'usage du glossaire des concepts dans l'environnement et le gain d'apprentissage des participants ?

4

7. L'analyse diagnostique permet de savoir ce qu'il se passera pour l'apprenant dans la séquence d'apprentissage.

Faux

8. Dans une analyse prédictive, on peut combiner des données de profils, de produits et de processus.

Vrai

9. Dans une prédiction, on s'intéresse plutôt à l'analyse d'un lien asymétrique.

Vrai

10. Il est possible d'obtenir une valeur de corrélation de -2.

Faux

## 7. Questionnaire final

1. L'utilisation des degrés de certitude dans les QCM donne la possibilité de calculer un score de confiance et un score de prudence.

Pour ses réponses incorrectes, l'élève exprimer un degré de certitude faible.

2. Il existe plusieurs niveaux de degré d'étayage de la visualisation des traces. Pouvez-vous identifier la distinction entre le niveau miroir et le niveau métacognitif ?

Le niveau métacognitif sollicite la comparaison avec un état de référence.

3. Parmi les propositions suivantes, lesquelles correspondent à des traces ?

La consultation d'une ressource en ligne

Le temps de connexion en minutes d'un apprenant

Un échange sur un forum entre plusieurs apprenants

Les résultats à un test en ligne

4. Dans [cet article](#), quelle est la source d'informations qui n'est pas considérée dans l'évaluation du dispositif ?

Les perceptions

Les produits

**Les processus**

Les profils

5. Que vaut le prétest (sur 10) si le gain relatif est de 75 % et le posttest est de 8/10 ?  
le prétest vaut **2**

6. Voici un item d'un questionnaire d'opinion cherchant à récolter l'avis des utilisateurs :  
J'ai trouvé facile d'apprendre à réaliser une carte conceptuelle avec ce logiciel.

Quel aspect concerne cet item ?

Contrôlabilité (ou utilisabilité)

7. La mémoire visuelle fonctionne selon un traitement parallèle de l'information.

Vrai

8. Quand on calcule le rapport entre la qualité d'une production et le temps nécessaire pour la réaliser, nous cherchons à évaluer le, la, l' **efficience**.

9. Dans le modèle de Viau, la contrôlabilité de la tâche peut être associée à quel principe mesuré dans le questionnaire de Davis ?

utilisabilité

*10. Il existe différents types de questions fermées. Retrouvez la définition associée à ces différents types.*

*Le répondant se positionne selon l'intensité avec laquelle il est en accord ou non, à l'affirmation posée.*

→ Questionnaire à échelle unidimensionnelle

*Le répondant se positionne sur une échelle qui met en opposition deux idées ou deux qualificatifs.*

→ Questionnaire à échelle bidimensionnelle

*Le répondant n'a le choix qu'entre deux options.*

→ Questionnaire à échelle dichotomique

*Le répondant sélectionne plusieurs items sans trancher, et sans ordonnancement.*

→ Questions à réponses multiples

*Le répondant hiérarchise les items selon un niveau de référence.*

→ Classement