In the context of solely trace-based indicators, we address in this paper the following research questions:

RQ 1: What are the indicators (metrics) allowing to measure collaborative writing ?

RQ 2: What are the collaborative writing strategies that can be deduced from these indicators ?

RQ 3: To what extent these indicators and strategies are interpretable by teachers ? (How close is the relationship between these measures and their interpretation).

**3 Research Method**

….

3.2.2 Questionnaires

We used a questionnaire to measure to what extent the indicators are interpretable by teachers who are not expert users of Labnbook (RQ3). Our objective is to verify whether the definitions of the indicators make sense. In the LabNbook database, we selected 12 LabDoc co-written by 2 to 4 students in real conditions. The writing contributions of different students were highlighted.

These labdocs are chosen with various levels of collaboration (from low level of collaboration

to high level of collaboration). Teachers are asked to read the 12 LabDocs. After reading each of them, they indicate whether they perceive collaboration in the written documents (yes, no, perhaps) and argue their responses. In addition, for each Labdoc, we ask the teachers to indicate

(1) the writing strategy between 5 choices (see section 5.3)

(2) the level of each indicator (Low (L), Medium (M) or High (H));

(3) a numerical value for each indicator between 0 and 1.

The questionnaire ended with questions on teaching practices and student monitoring. The questionnaire results are interpreted in Section 5.

**4 Indicator construction**

*[déplacer de la section 5 à la section 4 la demo CW < BC]*

*[concernant la remarque 1 :*

*je propose de simplifier l’article en ne discutant pas le choix avec et sans racine carrée car, in fine, les données ne nous permettent pas de trancher. Cela allège les tableaux (et graphiques éventuels) et nous laisse plus d’espace pour discuter le reste des résultats ]*

*Anis :*

*Je suis tout à fait d’accord pour ne pas discuter ce choix, mais pourquoi tu as gardé les indicateurs sans la racine carrée (meilleurs resultats dans le questionnaire !?) ? Le fait de mettre la racine carrée. En fait en* raison de la mise au carré des contributions moyennes (*xi*,.,.) dans l’équation (2), l'[unité](https://fr.wikipedia.org/wiki/Unit%C3%A9_de_mesure) de la variance est le carré de ces contributions (ex. : si le caractère est en kg, sa moyenne est en kg, mais sa variance est en kg2),

**5 Résultats de l’évaluation des indicateurs auprès des enseignants**

**5.1 Choix des Labdocs soumis aux enseignants**

*[j’utilise MV = machine value à la place de TL]*

|  |  |  |
| --- | --- | --- |
| **LD** | **MV BC** | **MV CW** |
| 1 | 0,49 | 0,41 |
| 2 | 0,87 | 0,11 |
| 3 | 0,84 | 0,77 |
| 4 | 0,99 | 0,41 |
| 5 | 0,49 | 0,44 |
| 6 | 0,99 | 0,51 |
| 7 | 0,51 | 0,07 |
| 8 | 0,99 | 0,74 |
| 9 | 0,48 | 0,17 |
| 10 | 0,92 | 0,09 |
| 11 | 0,30 | 0,04 |
| 12 | 0,14 | 0,10 |

Nous avons choisi 12 combinaisons différentes des deux paramètres : BC varie entre 0.14 et 0.99 ; CW entre 0.04 et 0.77

*Anis :*

*On pourrait aussi mettre les stratégies machine dans ce tableau (redondant par rapport au tableau 5?!)*

*Fahima:*

*Mettre cette partie dans la data collection. Parler des critéres.*

*Mettre MV dans les tableaux 3 et 4.*

**5.2 Niveaux et valeurs numériques des deux indicateurs**

*[je propose de faire 3 sous-sections : 5.2.1 BC sur les aspects niveaux et valeur numérique, 5.2.2 CW sur les aspects niveaux et valeur numérique, 5.2.3 combinaison de BC et CW pour parler des stratégies]*

**5.2.1 Balance de contribution**

expliquer qu’on a d’abord cherché à savoir si les enseignants estiment le niveau des indicateurs en cohérence avec les indicateurs calculés

D’abord on demandant une estimation grossière en utilisant 3 niveaux : Low, medium, high (découpage de l’intervalle 0 - 1)

FAHIMA : Mettre ce paragraphe au debut dans data collection

résultat : dans le tableau

🡪 9/12 bien classés

ensuite on leur a demandé d’estimer plus précisément la valeur numérique

résultat : **moyennes**, rmsd (~ 0,3 ?) et distances = (moyenne (ens) - MV) dans le tableau

🡪 à commenter en fonction des résultats

insérer ici le tableau pour BC :

*[rmq : dans l’ancien tableau 3 colonne TLSR il y avait des erreurs sur la mise en gras:*

*LD 2 non gras ; LD3 gras ; LD5 non gras ; LD6 non gras -> au total aussi 3 gras comme pour TL ; donc cela ne permet pas de trancher entre TLSR et TL non plus…*

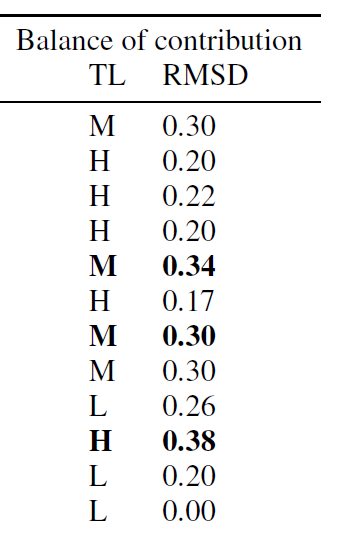
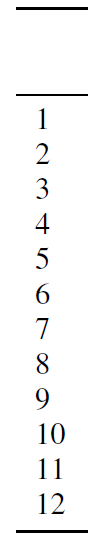
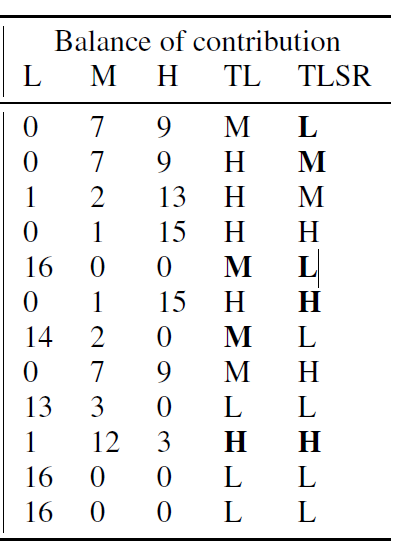
*-> d’où ma proposition de nous limiter aux MV (machine value - anciennement TL) sans la racine carrée]*

regrouper les résultats du table 3 (supprimer colonne TLSR) + table 4 : ajouter la moyenne des réponses enseignantes et les distances

Anis :

Quelles distances ? Ecart type par rapport à la moyenne des réponses ?

Si c’est le cas, je pense que c’est suffisant de mettre RMSD. Ce qu’on veut faire dans ce questionnaire c’est comparer les réponses des enseignant par rapport à la valeur machine. Pour ca, le RMSD est suffisant. Je suggére donc de garder cette table dans l’état.



**5.2.2 Co-écriture**

*la même chose que pour BC …*

estimation du niveau de l’indicateur CW par les enseignants entre L, M et H

résultat : dans le tableau

🡪 10/12 bien classés

estimation chiffrée

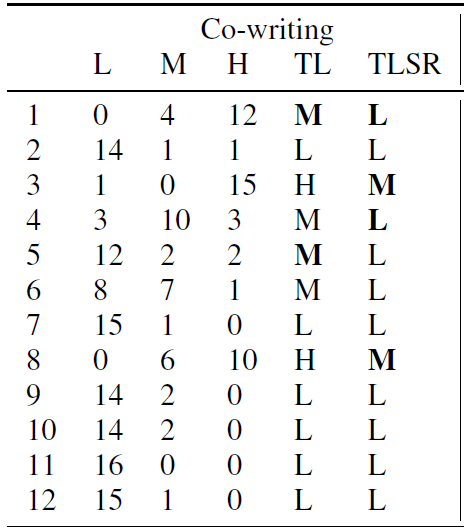
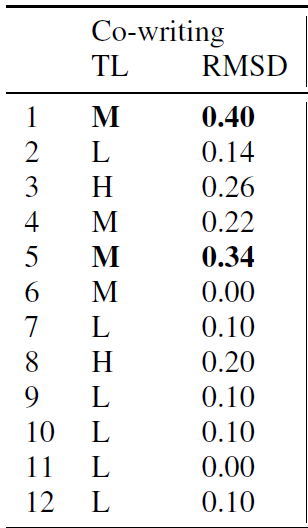
résultat : moyennes, rmsd *[ les 0.00 pour le LD6 et LD11 ne peuvent pas être corrects, Anis, à revoir ]* et distances -> dans le tableau

Anis :

En effet, il y’a une erreurs dans le RMSD de ces LD. Je vais les corriger.

🡪 commenter

insérer ici le tableau pour CW : regrouper table 3 (supprimer TLSR) et table 4 : ajouter la moyenne des réponses enseignantes et les distances



Anis :

Je pense que, comme la présentation de ces deux sous sections est symetrique, il serait peut etre préférable de mettre les deux tableaux dans un seul (à condition d’avoir de l’espace….). En plus je pense qu’on a pas assez de matiére pour faire deux sous-sections. Voila ce que je propose :

**5.2 Niveaux et valeurs numériques des deux indicateurs :**

**5.2.1 Niveux et valeurs numériques**

expliquer qu’on a d’abord cherché à savoir si les enseignants estiment le niveau des indicateurs en cohérence avec les indicateurs calculés

D’abord on demandant une estimation grossière en utilisant 3 niveaux : Low, medium, high (découpage de l’intervalle 0 - 1)

résultat : mettre un tableau pour les deux indicateurs.

***5.2.2* Stratégies d’écriture collaborative**

Suivre la proposition de Christian.

**5.2.3 Stratégies d’écriture collaborative**

Anis :

Que faire de la partie où on parle des stratégies dans le papier. On la garde ?

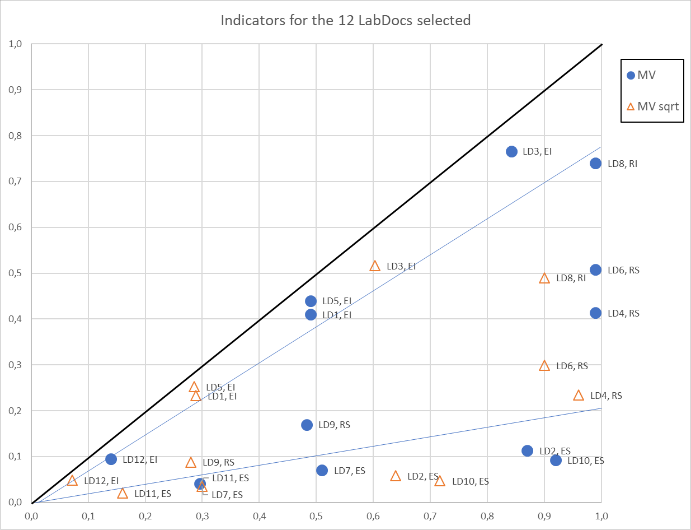
Fahima :

Elle propose de parler d’introduire les stratégies avant.

Introduire ici les deux stratégies summative et integrative en se basant sur la réf Onrubia& Engel ;

présenter ensuite quelles sont les stratégies utilisées dans les 12 Labdoc soumis aux enseignants, selon notre classification : 4 LD de la stratégie EI, 4 de la stratégie ES et 4 entre les deux (BSI) . Parmi les BSI : 3 avec tendance RS et 1 avec tendance RI. (tableau 5 mais commencer par les dernières colonnes : stratégie et tendance, cf. ci-dessous)

*[juste pour nous (pas à mettre dans l’article), voici comment les 12 LabDocs se repartissent dans le plan CW-BC]*



Exploitation de la partie du questionnaire qui concerne les stratégies au regard de la RQ3 :

Question 1 du questionnaire : (vise à répondre à la question si les enseignants classent les LabDocs dans les mêmes stratégies que les experts qui ont conçu les indicateurs)

on offre 5 choix : en gros I, S et BSI, plus deux choix intermédiaires (RI et RS) pour voir si une distinction plus fine est perçue.

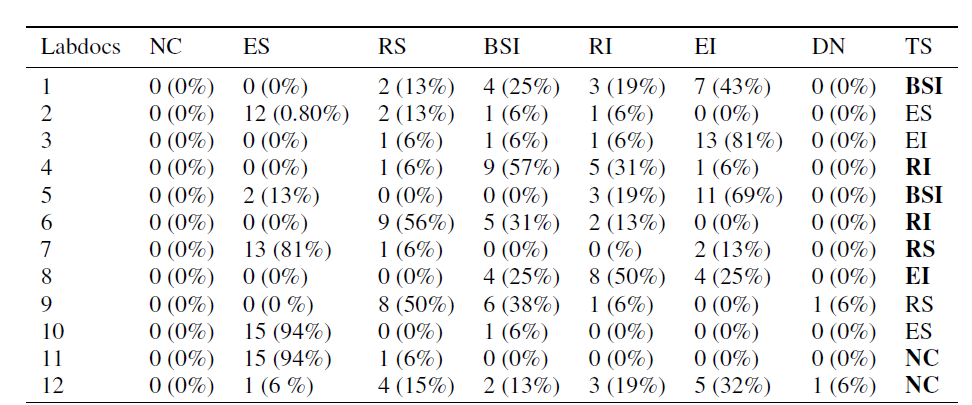
Résultat : pour tous les LabDocs, à deux exceptions près (LD1 et LD12), la majorité des enseignants classe les LabDocs dans la même catégorie que les experts. *à commenter davantage*

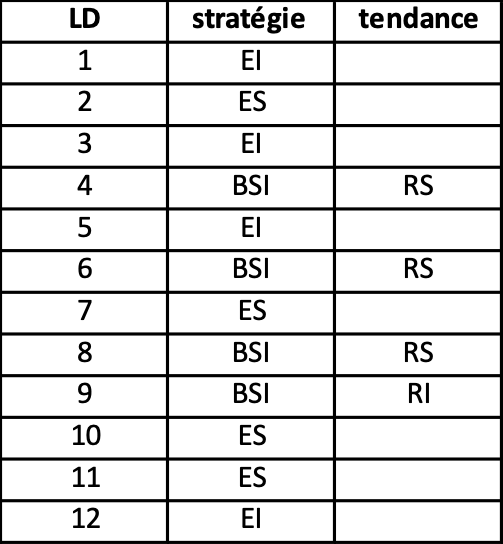
*Anis:*

*Les « vraies« stratégies dans le tableau sont bien celle qu’on a eu en combinant les deux indicateurs ?*

*Quelle est le role des experts ici ? Les tendances ?*

insérer ici le tableau 5 revu :



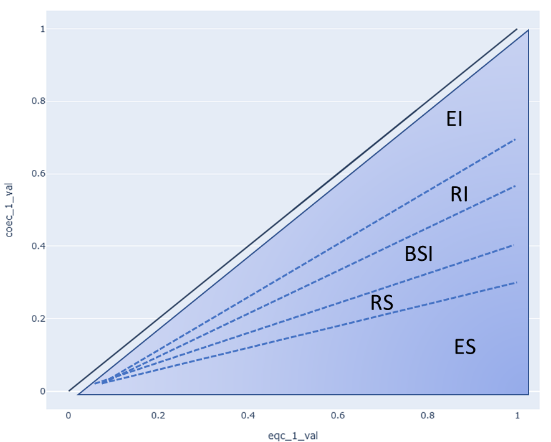


puis *à discuter* si on enchaine avec une comparaison de la position de leurs points (BC, CW) dans le plan BC-CW avec le point machine : est-ce qu’on insère un autre tableau avec les distances entre les barycentres (BC et CW moyennes des enseignants) et les points machine, plus les RMSD respectives ?

et/ou un graphique avec un ou deux exemples (cf. ci-dessous, les points dans le graphique sont des exemples et non pas basés sur les vraies valeurs) ?

**MV, LD1**

**écart-type des enseignants, LD1**



**Moyenne des enseignants, LD1**

Anis :

Je trouve que c’est pas une bonne idée de faire des moyennes car nous avons un petit échantillon. En revanche, j’aime bien l’idée de mettre un tel graphique car cela permet de montrer les stratégies.

**6. Discussion**

RQ1 : What are the indicators (metrics) allowing to measure collaborative writing ?

- les deux indicateurs séparément donnent des infos pertinentes pour caractériser la collaboration (référence aux concepts de ‘equality’ et ‘mutuality’)

BC semble assez bien compris,

estimations du niveau ok

donner des valeurs numériques plus difficiles, surtout si

- plus de 2 étudiants (LD5 et LD10)

- valeurs intermédiaires (LD7)

CW : idée également perçu par les enseignants

estimation du niveau OK, mais :

- limitation à BC (qui vient du fait qu’on tient également compte de l’équilibre de contribution, au niveau d’une phrase) n’est pas bien perçu (aussi dû au fait que ce n’était pas suffisamment expliqué dans le questionnaire) ; exemples LD 1, LD5

🡪 les enseignants ont pensé à deux indicateurs indépendants

Anis. :

Si on cite les LD pour on devrait les rendre accéssible dans le papier. Le lecteur doit les voir.

RQ 2: What are the collaborative writing strategies that can be deduced from these indicators ?

- nous avons exploré l’exemple de deux classes de stratégie importantes : summative et intégrative

-> faire le lien avec la différence entre coopération et collaboration

RQ 3: To what extent these indicators and strategies are interpretable by teachers ? (How close is the relationship between these measures and their interpretation).

- les enseignants reconnaissent les stratégies

* ils donnent des valeurs pour les indicateurs avec une dispersion assez importante : discuter les diverses raisons ici : contrainte BC < CW non-perçue ; écriture à plus de 2 auteurs plus difficile à interpréter ; confusion possible entre la fin de la phrase par le . et une nouvelle ligne dans nos LD exemples (est-ce que la phrase en intro du doc a été lue ?) ; etc.

Anis :

Pour moi, toutes les analyse faite sur les réponses du questionnaire permettent de répondre à la RQ3

**7. Conclusions**

quand ils s’agit de construire des indicateurs à fournir aux enseignants :

- des indicateurs descriptifs qui sont faciles à comprendre/interpréter (CW trop complexe ?)

- si on propose plusieurs, ils devraient être indépendants

🡪 doivent être « actionable » c’est-à-dire permettre d’engager des actions concrète dans l’enseignement, l’accompagnement des étudiants

quand il s’agit de créer des indicateurs pour un traitement automatique de traces, p.ex. analyser la qualité d’un travail collaboratif et renvoyer une information à ce sujet aux enseignants (alerte, curseur, etc.) afin qu’ils engage la discussion avec les étudiants :

- CW et BC sont de bons candidats, car ils caractérisent bien le degré et le type de collaboration

- ils pourraient être utilisé p.ex. comme paramètres d’entrée dans des modèles de machine learning (cf p.ex. papier Chejara, 2020)

-> à explorer dans le futur ?

Anis :

Je rajouterais:

la limite des deux indicateurs quant à la présence des formules.

Dire qu’une perspective possible est d’entrainer un model ML pour extraire les formules et les traiter séparémeent.