‘Evidencing Engagement and Performance’

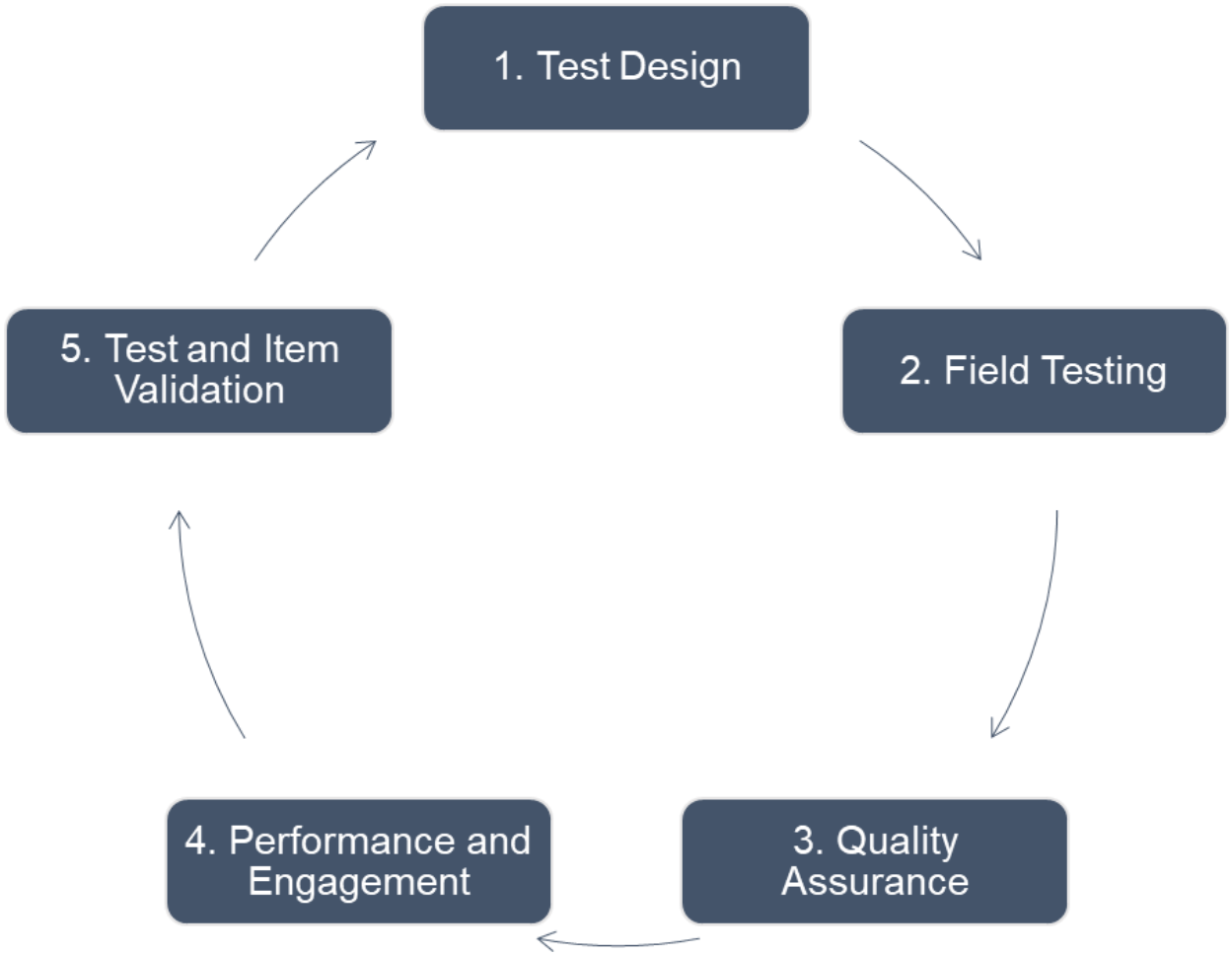
CEMO Lecture Series on *Process Data in Educational Assessment*

Lecture 3, January 2023.

Dr Bryan Maddox

UiO : University of Oslo

Assessment MicroAnalytics™



From Maddox, ‘The uses of process data in large-scale educational assessments, 2023 (forthcoming).

Skill and Will

*'An achievement test score can be viewed as a **joint function of skill and will**, of knowledge and motivation.* (Eklöf 2010, p345, emphasis mine).

Sources of data on student engagement:

Student motivation surveys (effort thermometer), 'Rapid Response' RT Data, 'Performance Decline' data (Wise and Kingsbury 2022), ethnographic observation.

Example 1

long puff out (35.15).

R: Can you tell me how far we are?
I: No.
R: Ah.
I: You still have some exercises. You are over the half, at two thirds. I’m speaking from experience
R: Yeah, yeah..
I: But I cannot influence the computer’s selection of exercises for you. So that … there can be a slight deviation.

I: *We have finished*.

R: *Oooooh,* ***finally***!

I: Was it a bit difficult?

R: **It was *killing* me**, this now … but …

I: OK, that’s it.

R: ***Damn***, I thought it would never end.

Example 2

The respondent makes a very loud sigh, shakes her head and rubs her eyes.

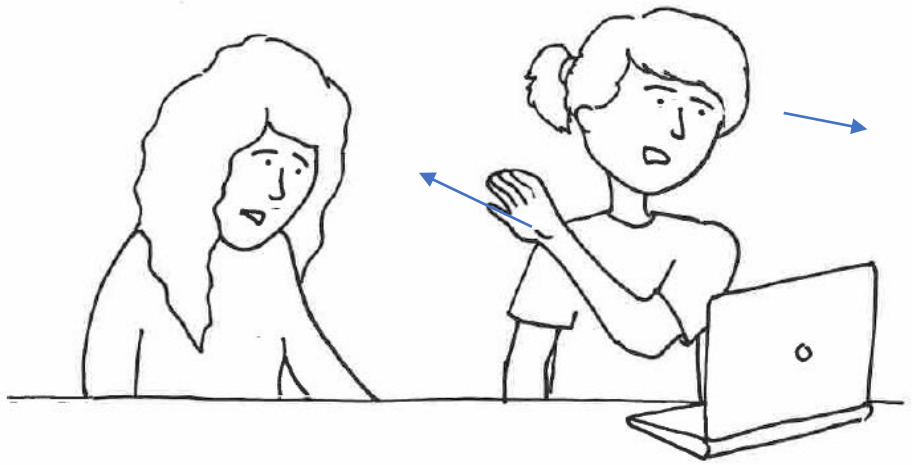
R: It won't work. I don't have any will, not to read or anything else. ((R sighs loudly, looks up to the clock and then looks directly at the test respondent))

I: Take a few minutes break.

R: **No!**, you know, because someone … someone is spurring me. The evening is almost over and I promised ...

I: I see, you told her you would *[content removed..]*?

R: Yeah, I promised her... And you're pestering me with this … I thought that I would be merely answering, not that it would be so tiresome!



'No!'

I: There are only few short exercises left.

R: No, I don't have any will left. When my morale is gone, it's gone totally.

I: Why?

R: I don't know.

I: Do you have a bad feeling?

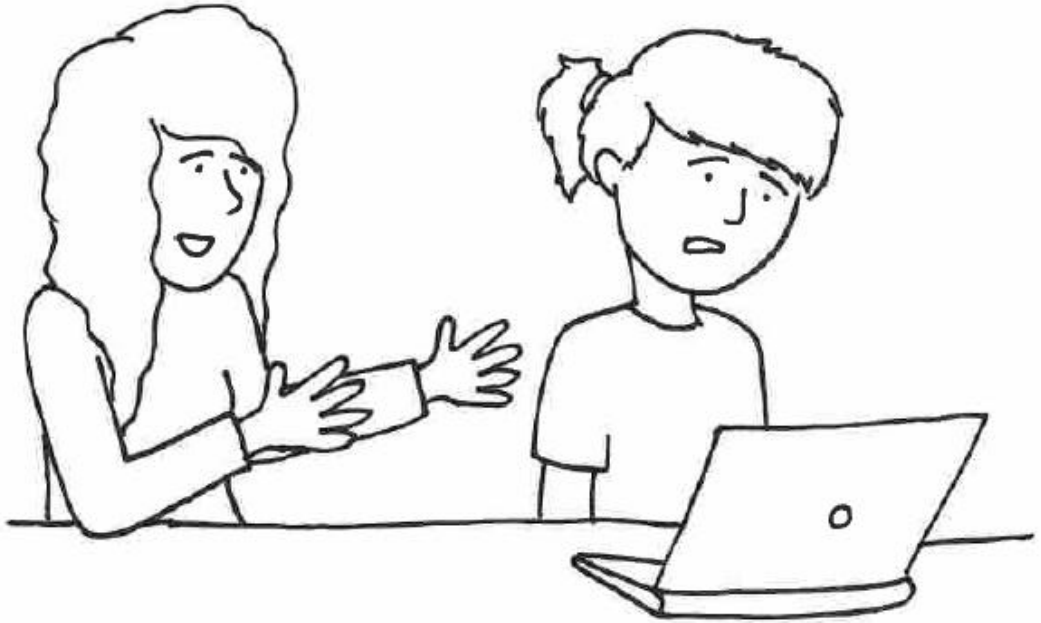
R: I have a bad feeling and I [

I: [Don’t feel bad about it

R: Ah, I don't feel like doing it at all!

I: Look at the next short exercises

R: When I don't feel like … at all … my mind doesn't … … work, really. I had completely different plans for today, really..



Example 3

I: It won't go?

R: No, this is not for me.

I: I see. If there are problems, you can freely move on, there's no …]

R: [It's not so difficult, but … I'm a bit …

((she smiles and gestures with her hand, indicating lack of concentration))

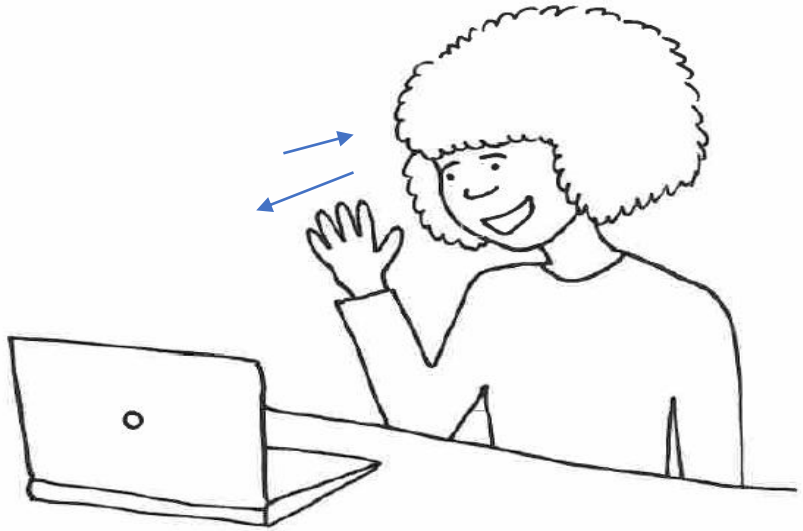
I: I see, it's long …]

R: [And this noise!

I: Yes, true! We have a bit … specific environment.

R: Yes!

((she laughs with a big smile, nods agreement and looks briefly at the test administrator and then back at the screen))



B: .. when there was too much text, did you ever give up and skip? Just pressed advance?

R: Um, I think here, in this one …

((Contact Employer ID: C304B711*))*

B: In that one, yeah.

R: In that one.

B: Any else?

R: Hm … I don't know. Maybe I just … yeah, here. Um

B: You skipped any others?

R: I don't know, maybe one, maybe one.

U304-ContactEmployer;C304B711;**StimulusAndQuestionLoaded**;10:50:41

U304-ContactEmployer;C304B711;**click**;10:50:41

U304-ContactEmployer;C304B711;highlightEvent;10:51:42

U304-ContactEmployer;C304B711;highlightEvent;10:52:21

U304-ContactEmployer;C304B711;highlightEvent;10:52:25

U304-ContactEmployer;C304B711;highlightEvent;10:52:26

U304-ContactEmployer;C304B711;highlightEvent;10:52:28

U304-ContactEmployer;C304B711;highlightEvent;10:52:31

U304-ContactEmployer;C304B711;highlightEvent;10:52:35

U304-ContactEmployer;C304B711;**click**;10:52:37

U304-ContactEmployer;C304B711;**onItemEnd**;10:52:39

B: … you had to concentrate, maybe it wasn’t so exciting, you still concentrated and …?

R: Yeah, I tried, but maybe in one or two cases I just … what’s the word for “označiti”?

B: Highlighted?

R: Yes, highlighted and I didn’t *care, if it’s true …*

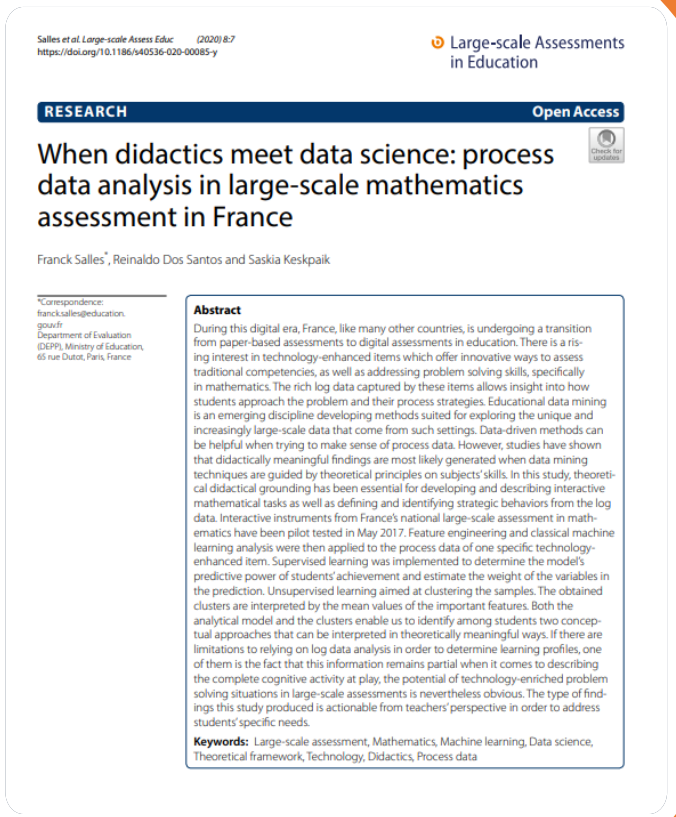
B: Oh, I see, so you just *finished*

R: Yeah, yeah.

B: … just highlighted a bit and then pressed advance.

R: Yeah.

Source: Maddox, 2017. ‘*Talk and Gesture as Process Data*’



Salles et al. Large-scale Assess Educ (2020) 8:7
https://doi.org/10.1186/s40536-020-00085-y

Large-scale Assessments in Education

RESEARCH Open Access

When didactics meet data science: process data analysis in large-scale mathematics assessment in France

Franck Salles*, Reinaldo Dos Santos and Saskia Keskpaik

*Correspondence: franck.salles@education.gouv.fr
Department of Evaluation (DEPP), Ministry of Education, 65 rue Dutot, Paris, France

Abstract

During this digital era, France, like many other countries, is undergoing a transition from paper-based assessments to digital assessments in education. There is a rising interest in technology-enhanced items which offer innovative ways to assess traditional competencies, as well as addressing problem solving skills, specifically in mathematics. The rich log data captured by these items allows insight into how students approach the problem and their process strategies. Educational data mining is an emerging discipline developing methods suited for exploring the unique and increasingly large-scale data that come from such settings. Data-driven methods can be helpful when trying to make sense of process data. However, studies have shown that didactically meaningful findings are most likely generated when data mining techniques are guided by theoretical principles on subjects' skills. In this study, theoretical didactical grounding has been essential for developing and describing interactive mathematical tasks as well as defining and identifying strategic behaviors from the log data. Interactive instruments from France's national large-scale assessment in mathematics have been pilot tested in May 2017. Feature engineering and classical machine learning analysis were then applied to the process data of one specific technology-enhanced item. Supervised learning was implemented to determine the model's predictive power of students' achievement and estimate the weight of the variables in the prediction. Unsupervised learning aimed at clustering the samples. The obtained clusters are interpreted by the mean values of the important features. Both the analytical model and the clusters enable us to identify among students two conceptual approaches that can be interpreted in theoretically meaningful ways. If there are limitations to relying on log data analysis in order to determine learning profiles, one of them is the fact that this information remains partial when it comes to describing the complete cognitive activity at play, the potential of technology-enriched problem solving situations in large-scale assessments is nevertheless obvious. The type of findings this study produced is actionable from teachers' perspective in order to address students' specific needs.

Keywords: Large-scale assessment, Mathematics, Machine learning, Data science, Theoretical framework, Technology, Didactics, Process data

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

depp Direction de l'évaluation, de la prospective et de la performance

CITO

Process Data and the Analysis of Performance. With thanks to Frank Salles at the DEPP, France and Eva de Schipper at the Cito Institute.

Student 1: Is it interesting? Very interesting this test? (*Filling in the feedback form)*
Student 2: Wait! (*doing an item*)
Student 1: We think it's interesting!
Student 2: There is no zero!
Student 1: No you do 2 minus 1.1
Student 2: Do I? wait, I'll look..
Student 2: That's it!
Pffff! It's actually completely stupid, but *no*, I am actually not stupid, so let's go!

"I am actually not stupid"

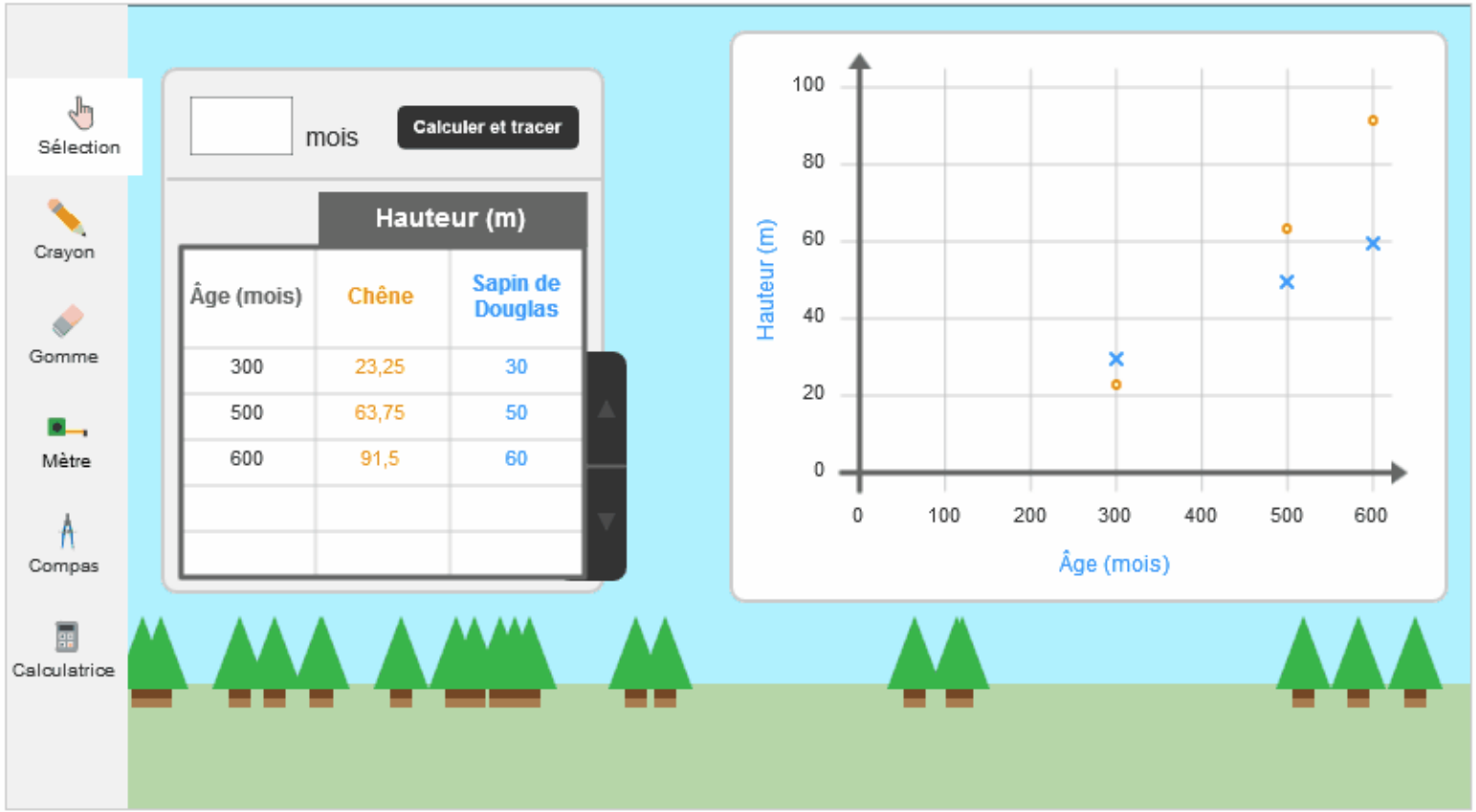
Deux graines d'arbres sont plantées au même moment : un chêne et un sapin de Douglas.

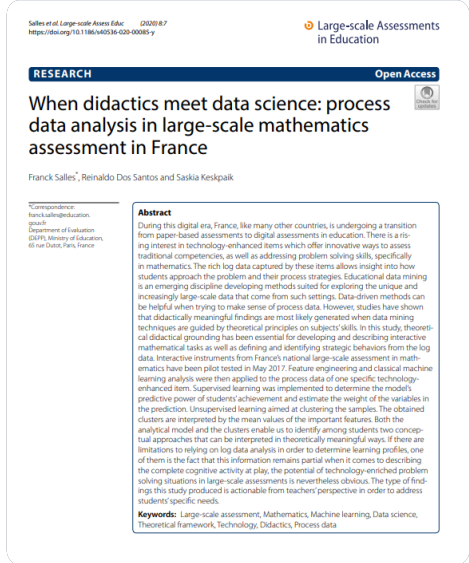
En entrant dans la première colonne, l'âge (en mois) des arbres, on obtient leur hauteur (en mètre) dans les deuxième et troisième colonnes.

Les points correspondants s'affichent sur le graphique : en orange le chêne, en bleu le sapin.

A quel âge (autre que 0 mois) ont-ils la même hauteur ?

L'âge est de _____ mois.




Salles et al. Large-scale Assess Educ (2020) 8:7
https://doi.org/10.1186/s40536-020-00085-y

Large-scale Assessments in Education

RESEARCH Open Access

When didactics meet data science: process data analysis in large-scale mathematics assessment in France

Franck Salles*, Reinaldo Dos Santos and Saskia Keskpaik

*Correspondence: franck.salles@education.gouv.fr
Department of Evaluation (DEPP), Ministry of Education, 65 rue Dutot, Paris, France

Abstract

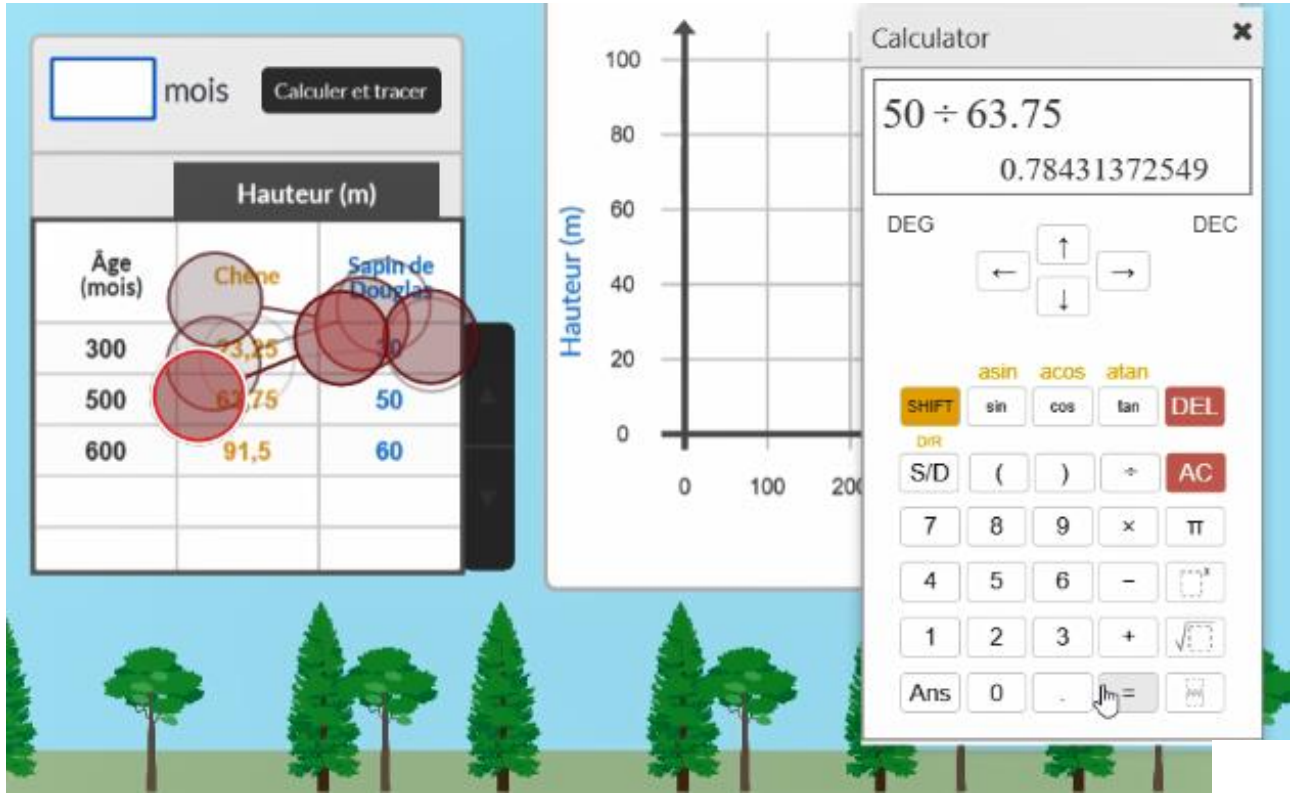
During this digital era, France, like many other countries, is undergoing a transition from paper-based assessments to digital assessments in education. There is a rising interest in technology-enhanced items which offer innovative ways to assess traditional competencies, as well as addressing problem solving skills, specifically in mathematics. The rich log data captured by these items allows insight into how students approach the problem and their process strategies. Educational data mining is an emerging discipline developing methods suited for exploring the unique and increasingly large-scale data that come from such settings. Data-driven methods can be helpful when trying to make sense of process data. However, studies have shown that didactically meaningful findings are most likely generated when data mining techniques are guided by theoretical principles on subjects' skills. In this study, theoretical didactical grounding has been essential for developing and describing interactive mathematical tasks as well as defining and identifying strategic behaviors from the log data. Interactive instruments from France's national large-scale assessment in mathematics have been pilot tested in May 2017. Feature engineering and classical machine learning analysis were then applied to the process data of one specific technology-enhanced item. Supervised learning was implemented to determine the model's predictive power of students' achievement and estimate the weight of the variables in the prediction. Unsupervised learning aimed at clustering the samples. The obtained clusters are interpreted by the mean values of the important features. Both the analytical model and the clusters enable us to identify among students two conceptual approaches that can be interpreted in theoretically meaningful ways. If there are limitations to relying on log data analysis in order to determine learning profiles, one of them is the fact that this information remains partial when it comes to describing the complete cognitive activity at play, the potential of technology-enriched problem solving situations in large-scale assessments is nevertheless obvious. The type of findings this study produced is actionable from teachers' perspective in order to address students' specific needs.

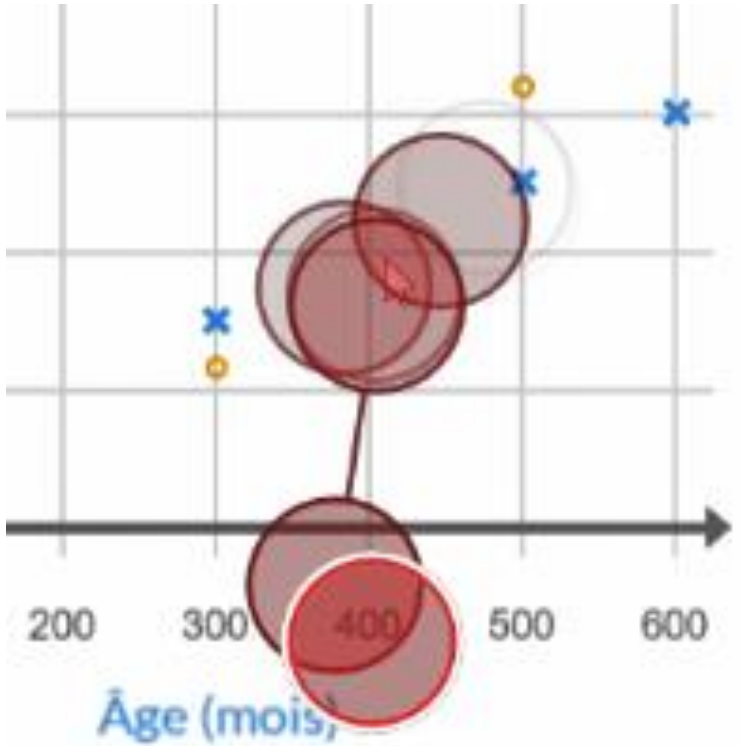
Keywords: Large-scale assessment, Mathematics, Machine learning, Data science, Theoretical framework, Technology, Didactics, Process data

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

depp Direction de l'évaluation, de la prospective et de la performance

Observing Digital Strategies



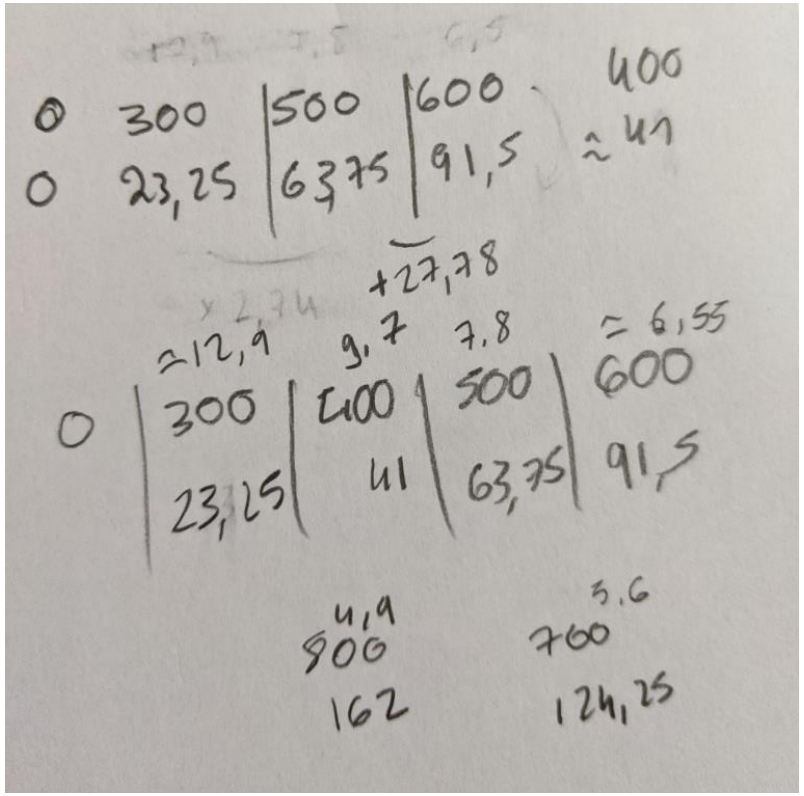


RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

depp Direction de l'évaluation, de la prospective et de la performance

AMA

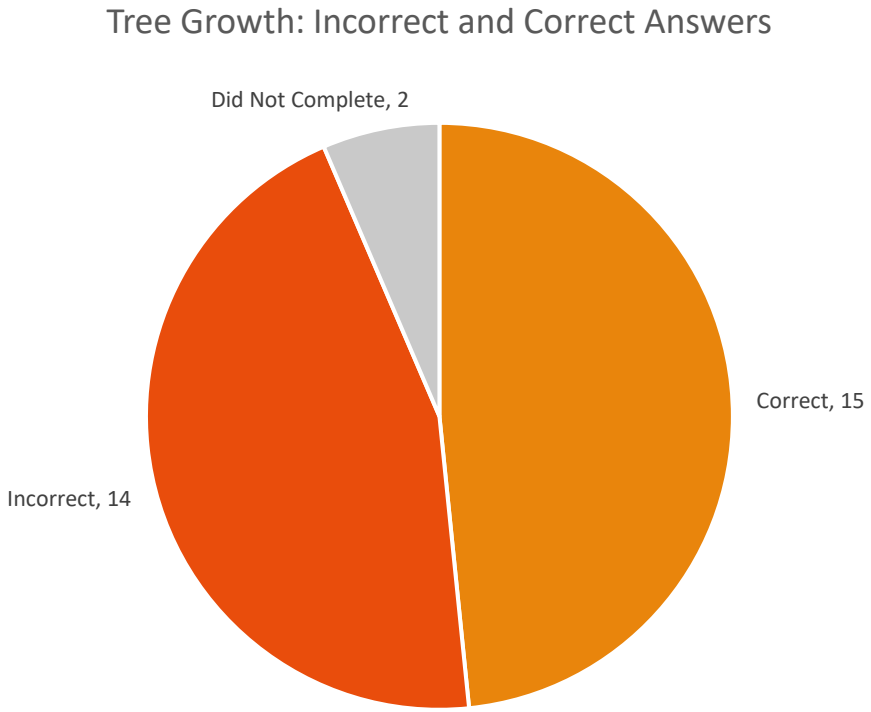
Between the clicks

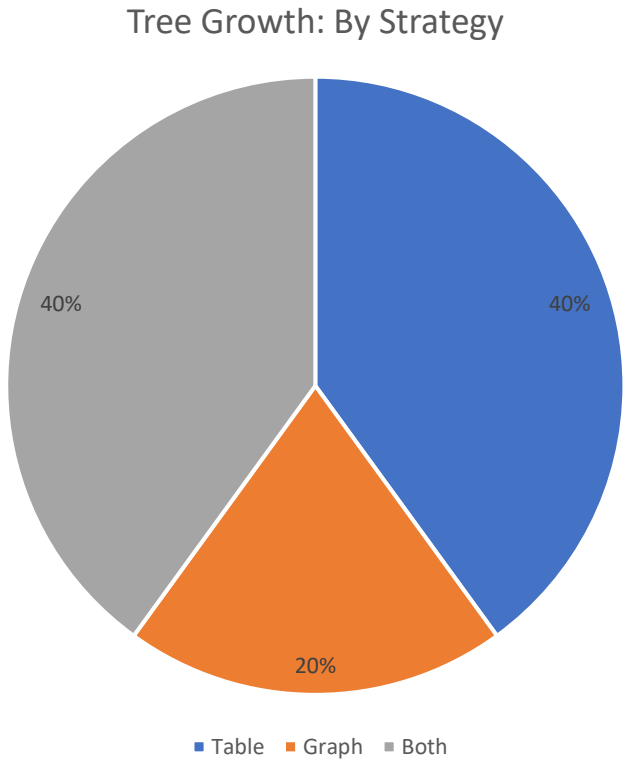


RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

depp Direction de l'évaluation, de la prospective et de la performance

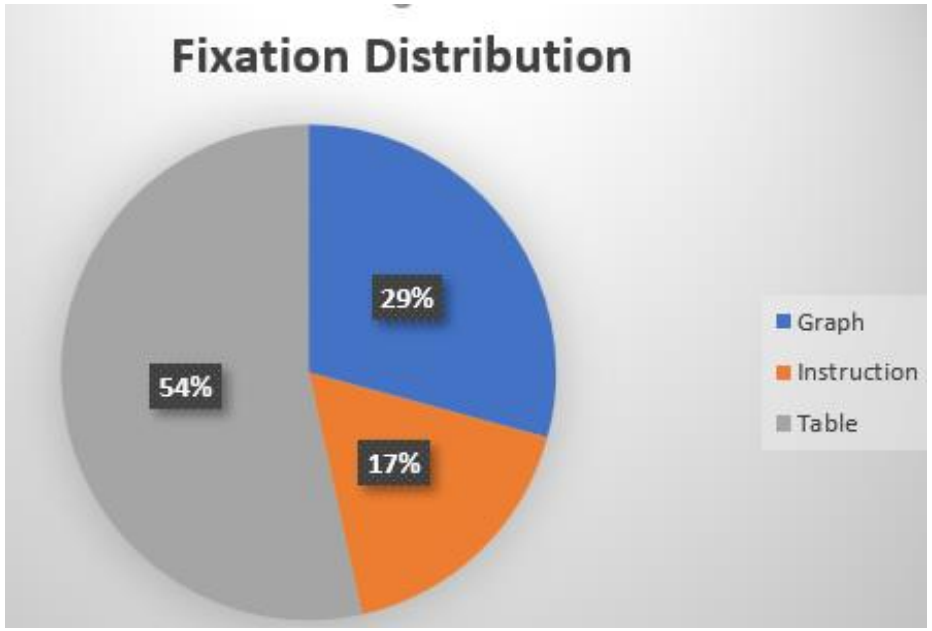
AMA

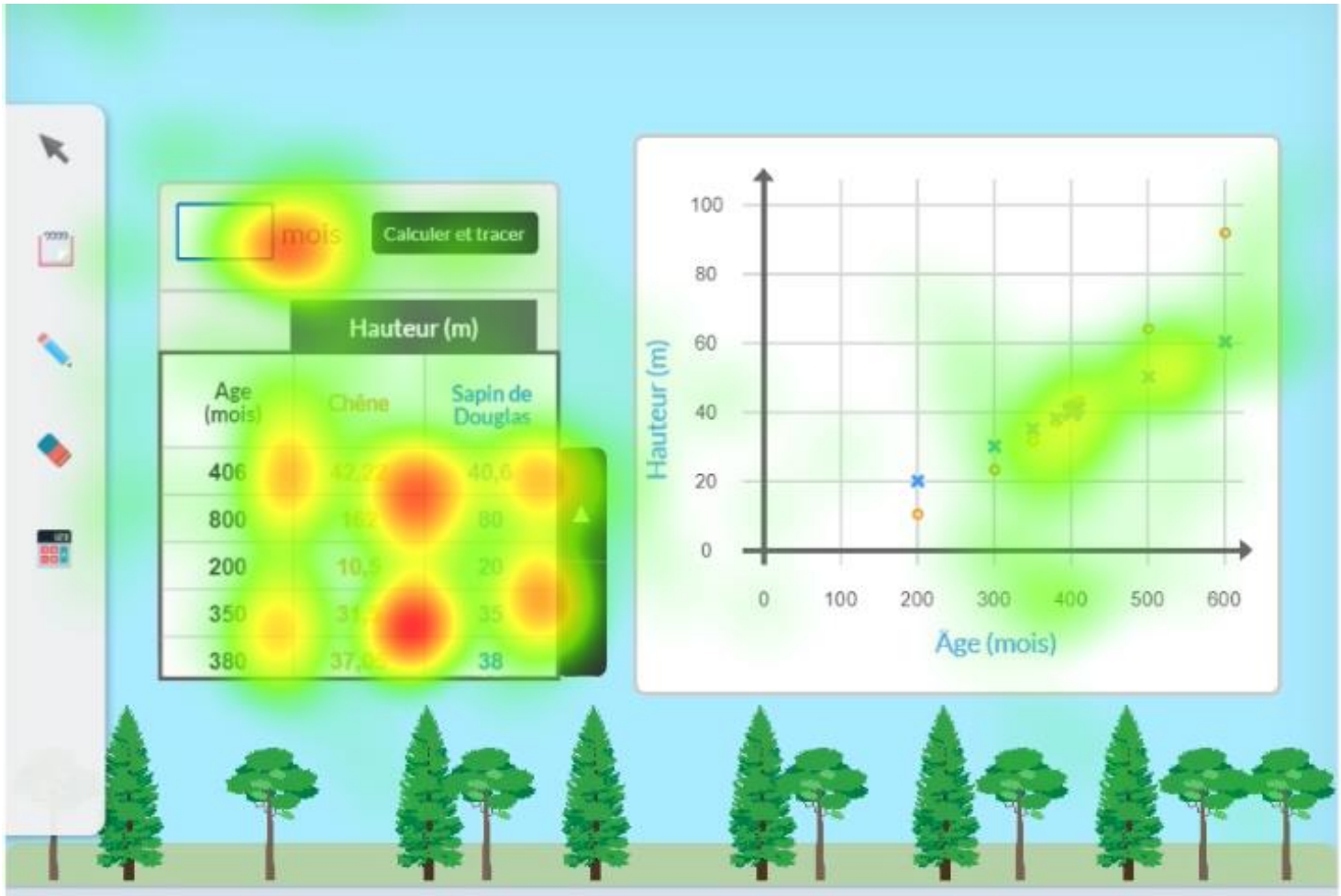

Tree Growth: Incorrect and Correct Answers
Did Not Complete, 2
Correct, 15
Incorrect, 14


Tree Growth: By Strategy
40%
40%
20%
Table
Graph
Both

MINISTÈRE
DE L'ÉDUCATION
NATIONALE ET
DE LA JEUNESSE

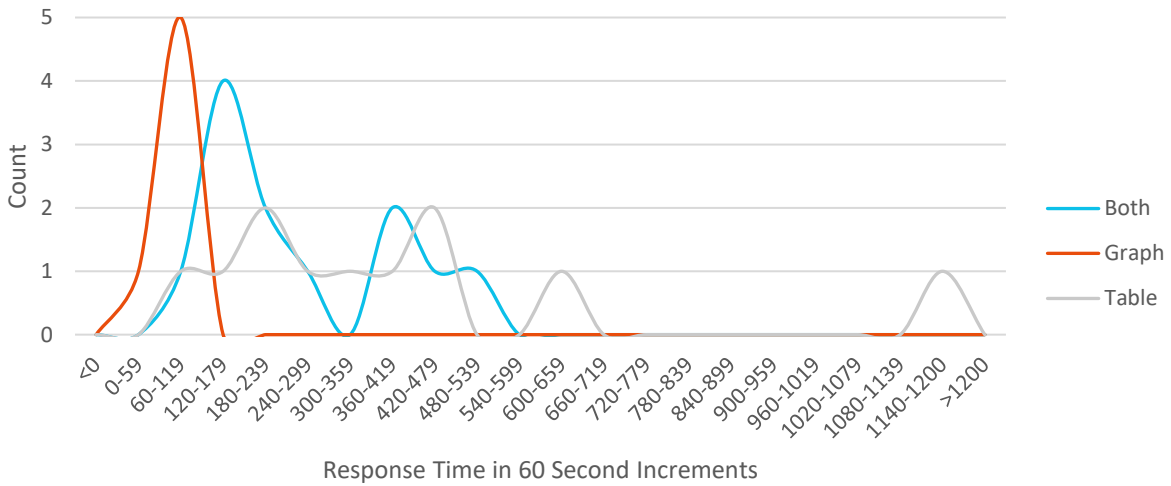
Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE



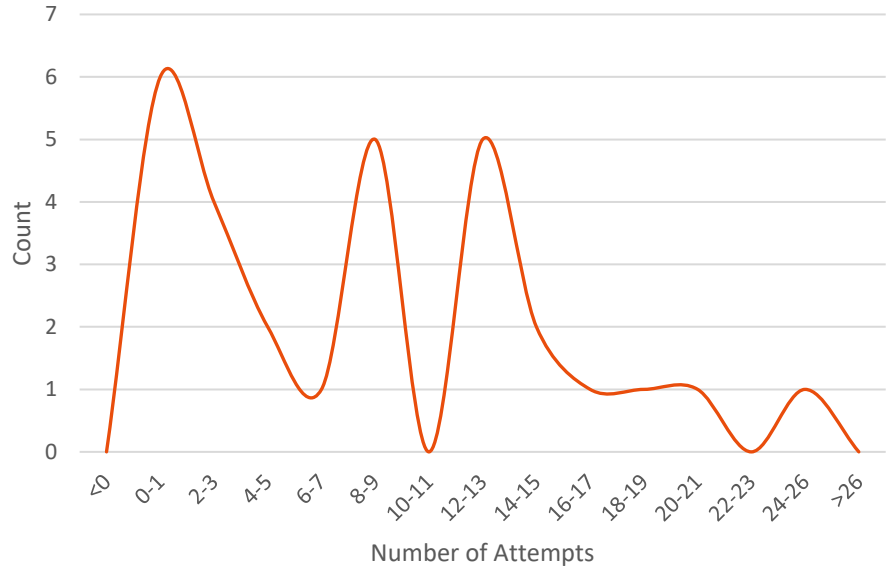


Tree Growth: Response Times

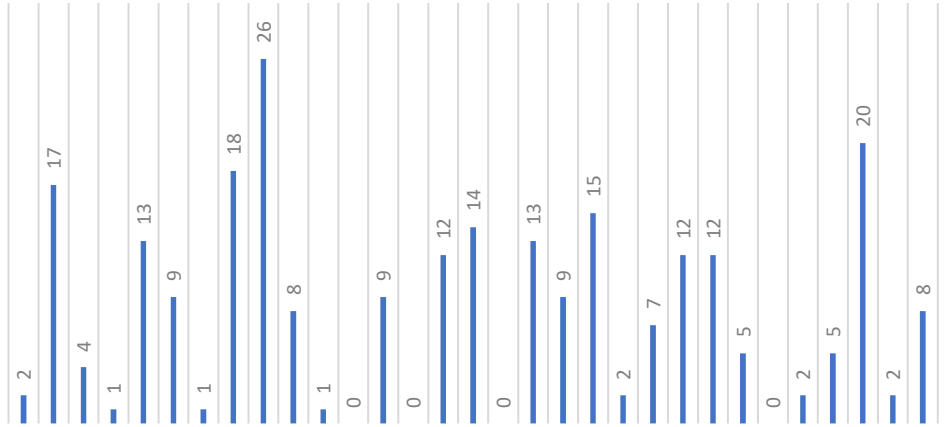
Tree Growth: Response Time by Strategy



Count - Number of Attempts

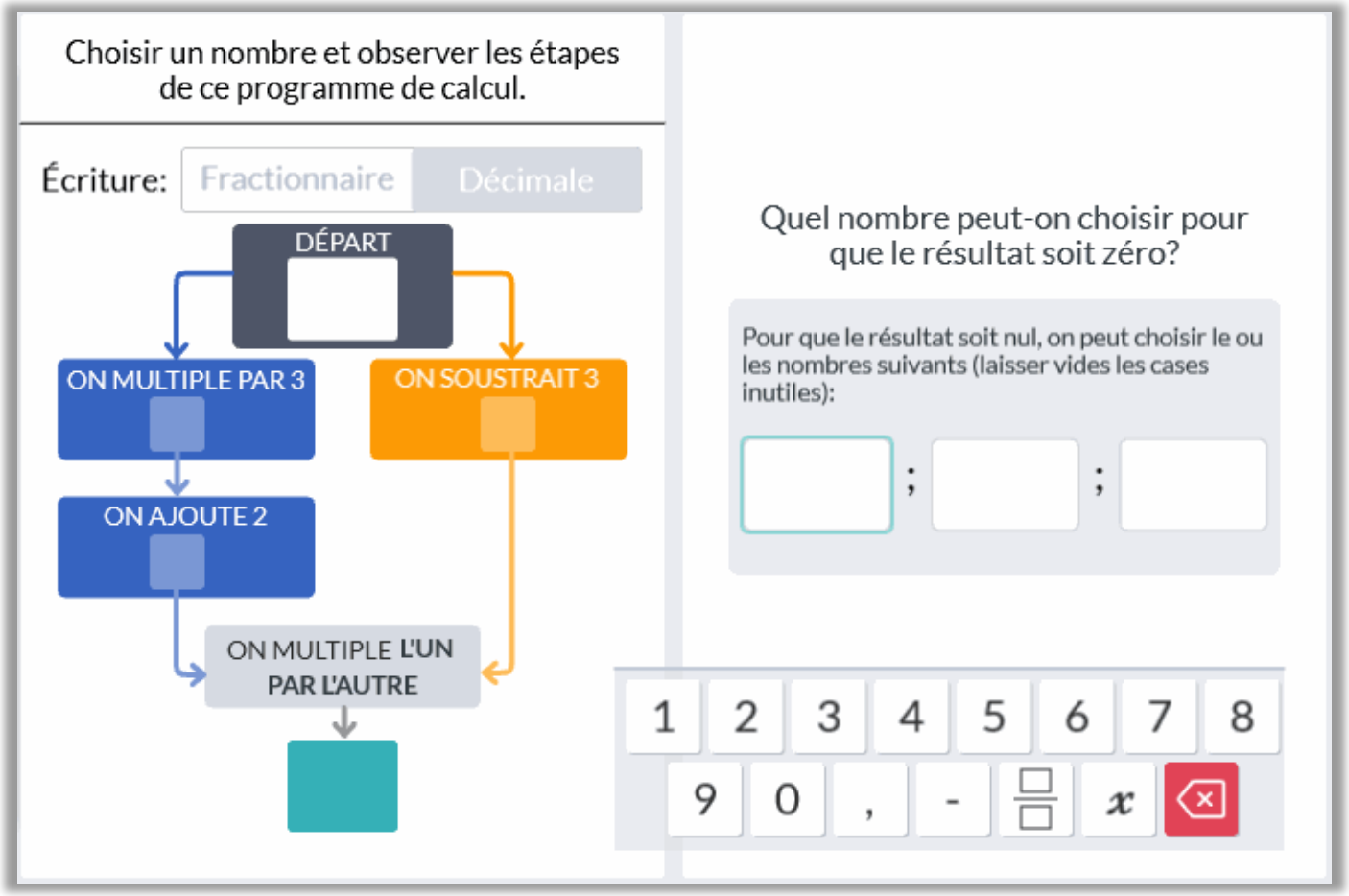


NUMBER OF ATTEMPTS (SEQUENCE)



RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

depp Direction de l'évaluation, de la prospective et de la performance



Numerical conception/approach

Equation = input/output relationship

Related strategy:

Trial and error iterations of

- input value,
- compare with target and previous result,
- decide on the next value

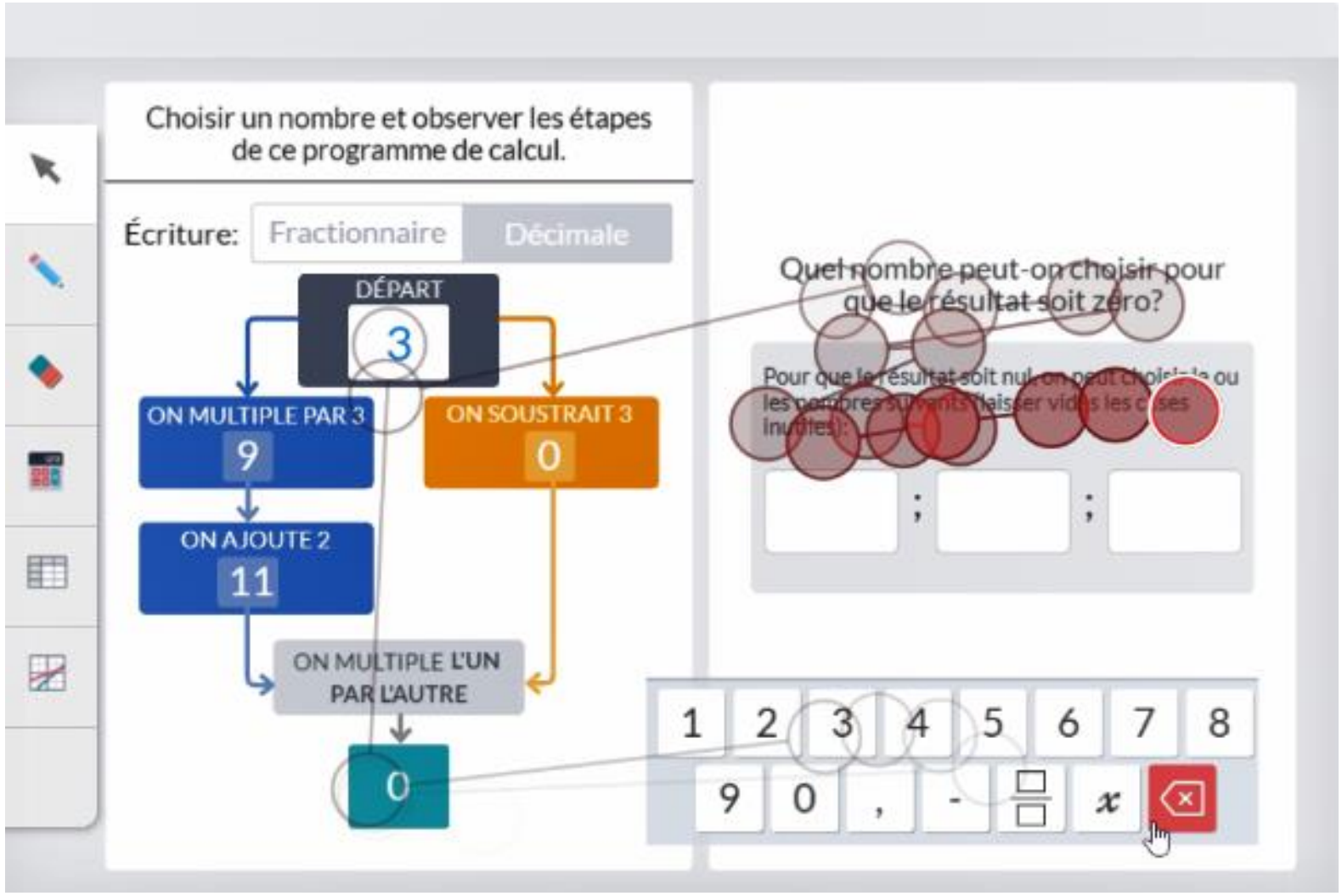
Algebraic conception

Equation = math object that can be transformed

Related strategy:

- Calculations (0-2/3; …),
- Using algebraic expression,
- Paper based solving

Eye Tracking Data: Reading the instructions

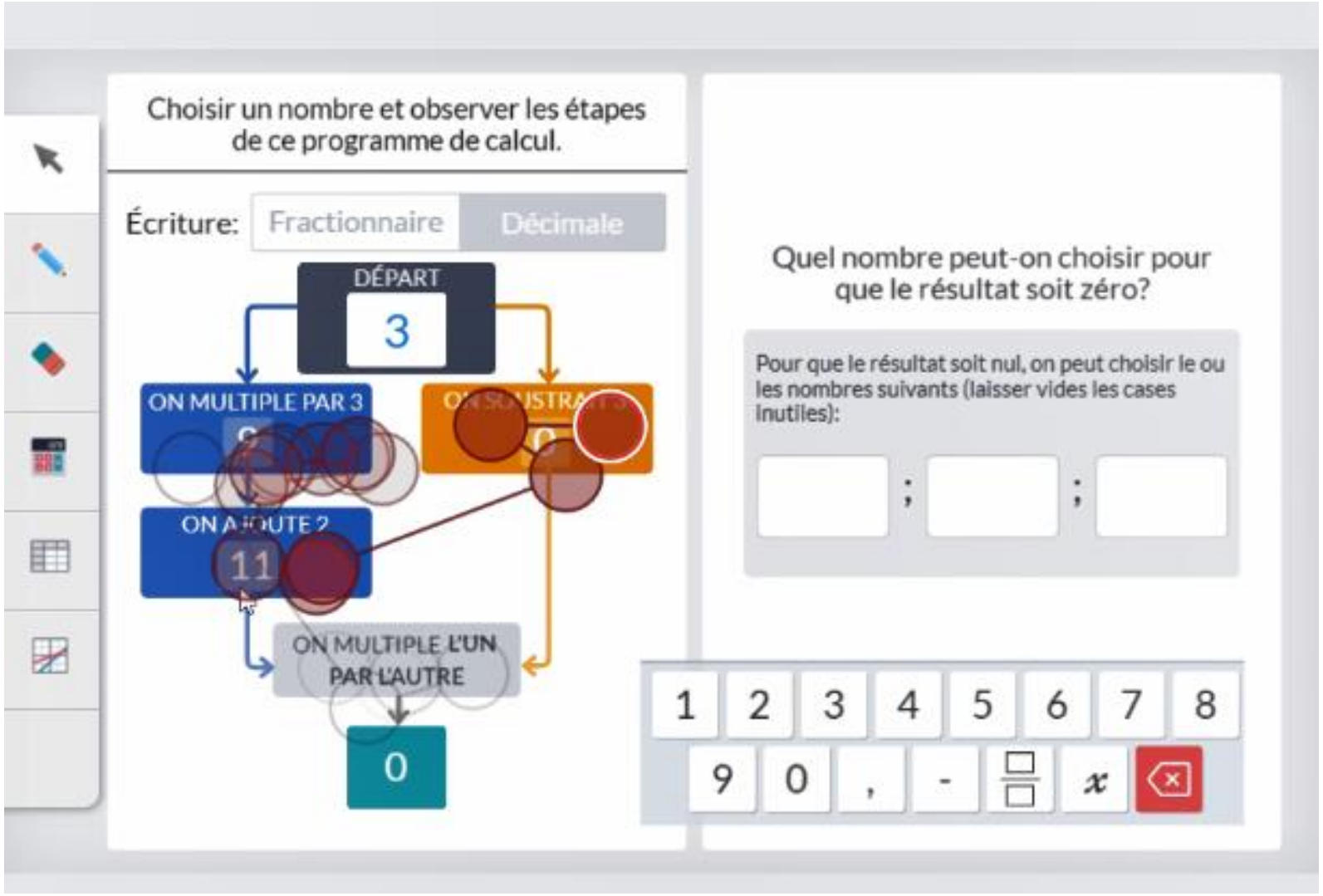


RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

depp Direction de l'évaluation, de la prospective et de la performance

AMA

Eye Tracking Data: Studying the diagram

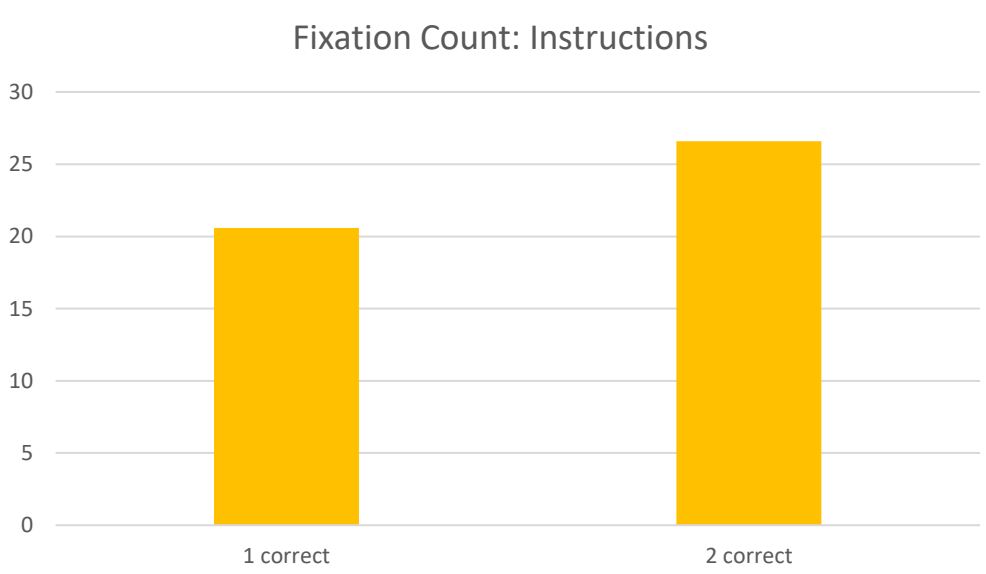


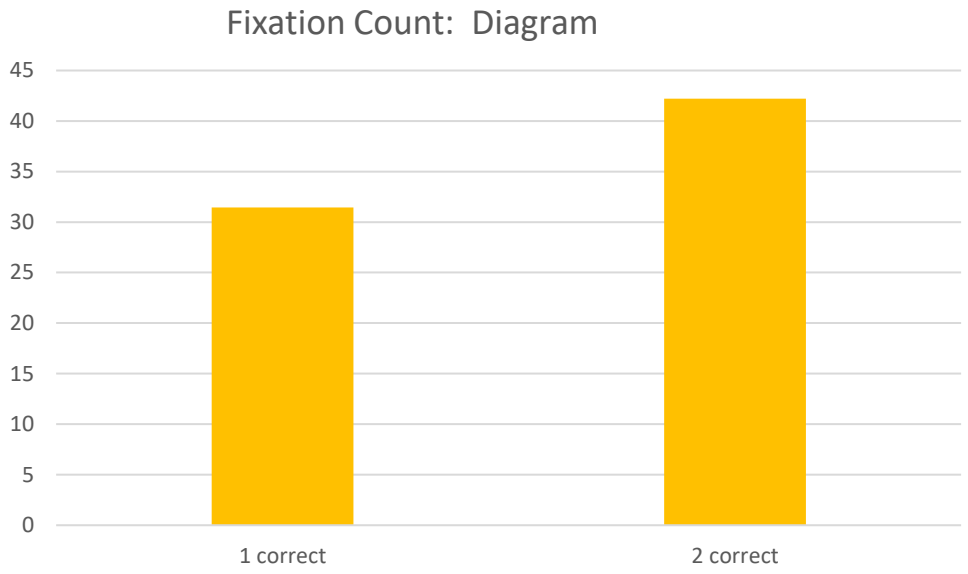
RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

depp Direction de l'évaluation, de la prospective et de la performance

AMA

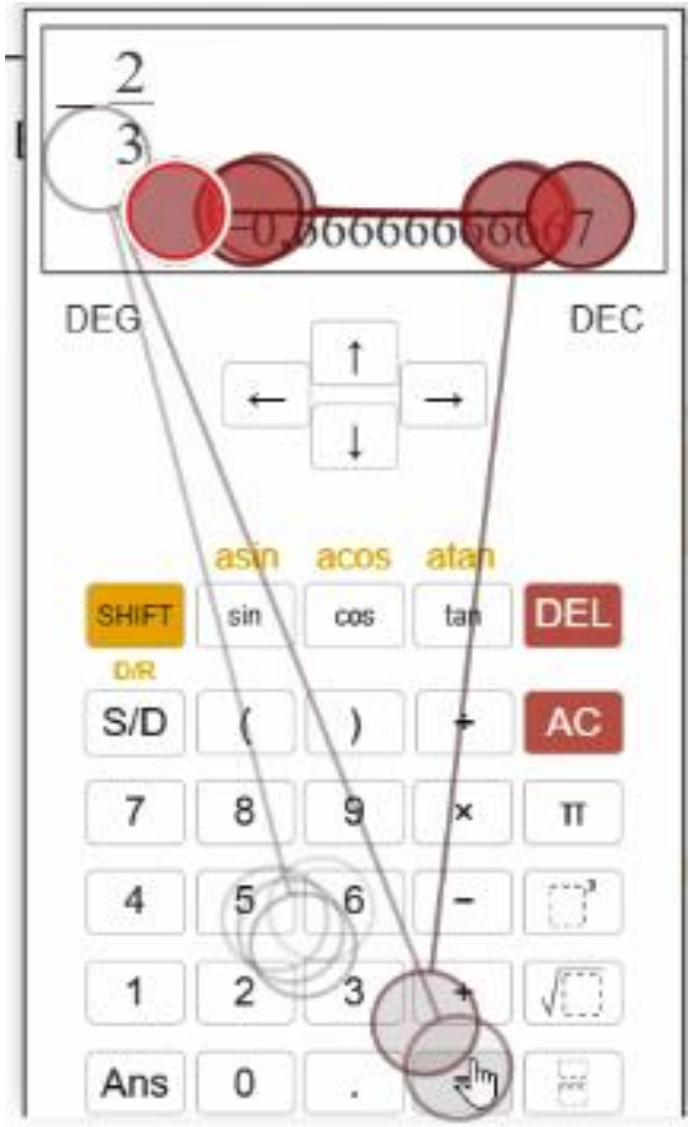
Eye Tracking: Average Fixation Counts *before first value tried*.

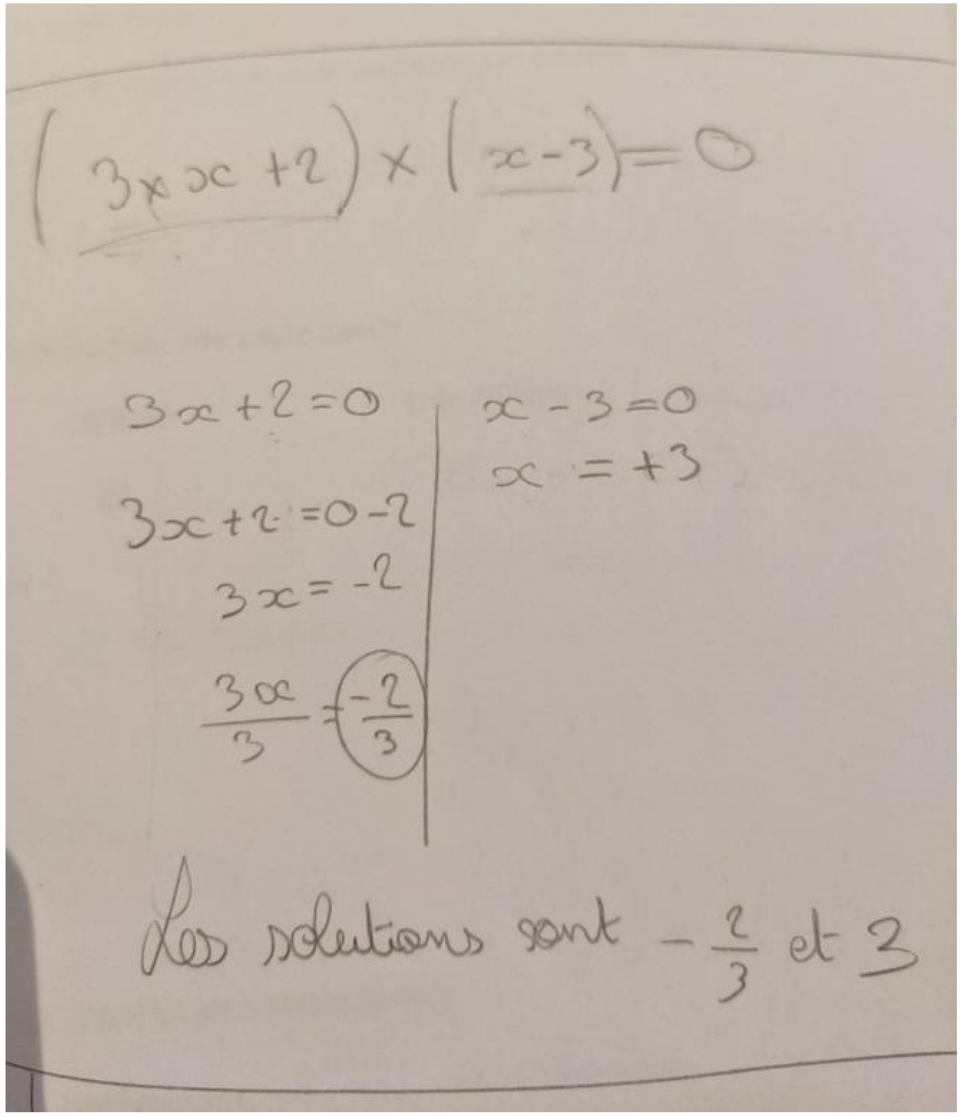




Students who got two correct answers had higher fixation counts on the item instructions (20.58/26.6), and the equation diagram (31.43/42.2). I.e., They studied the instructions and equation diagram in more detail *before trying an answer*.

Identifying Student Strategies



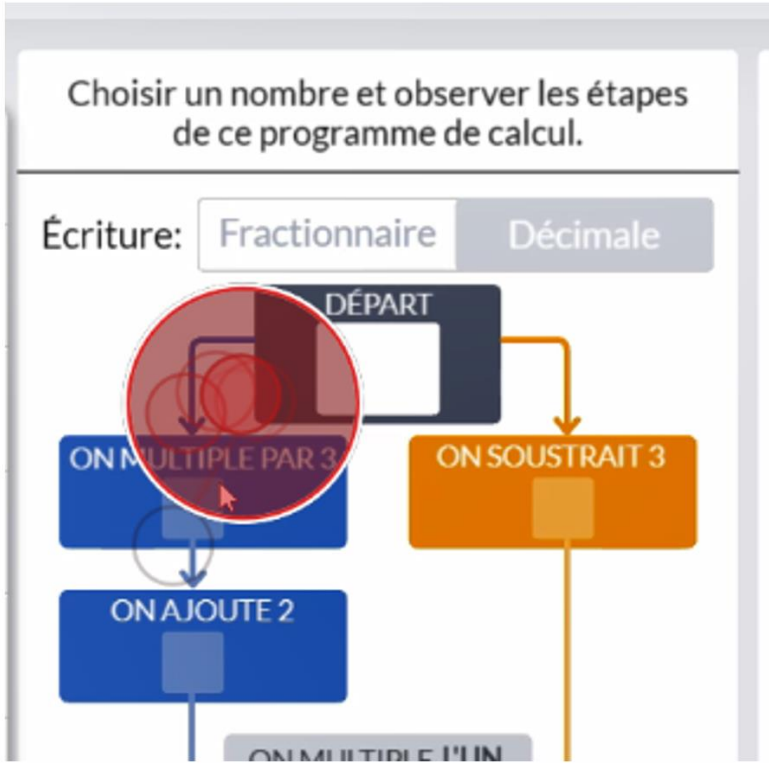


RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

depp Direction de l'évaluation, de la prospective et de la performance

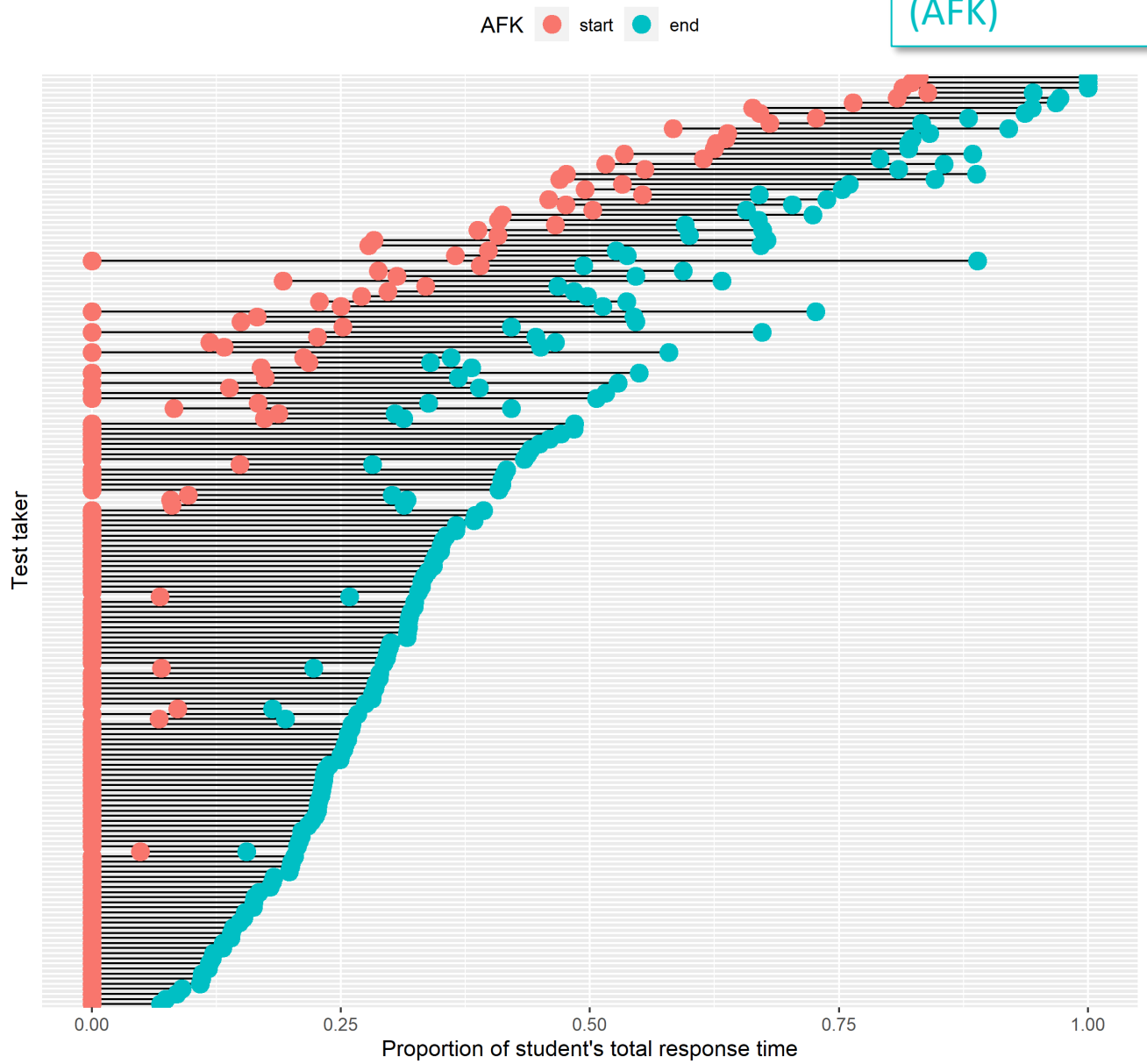
AMA

An answer without material resources

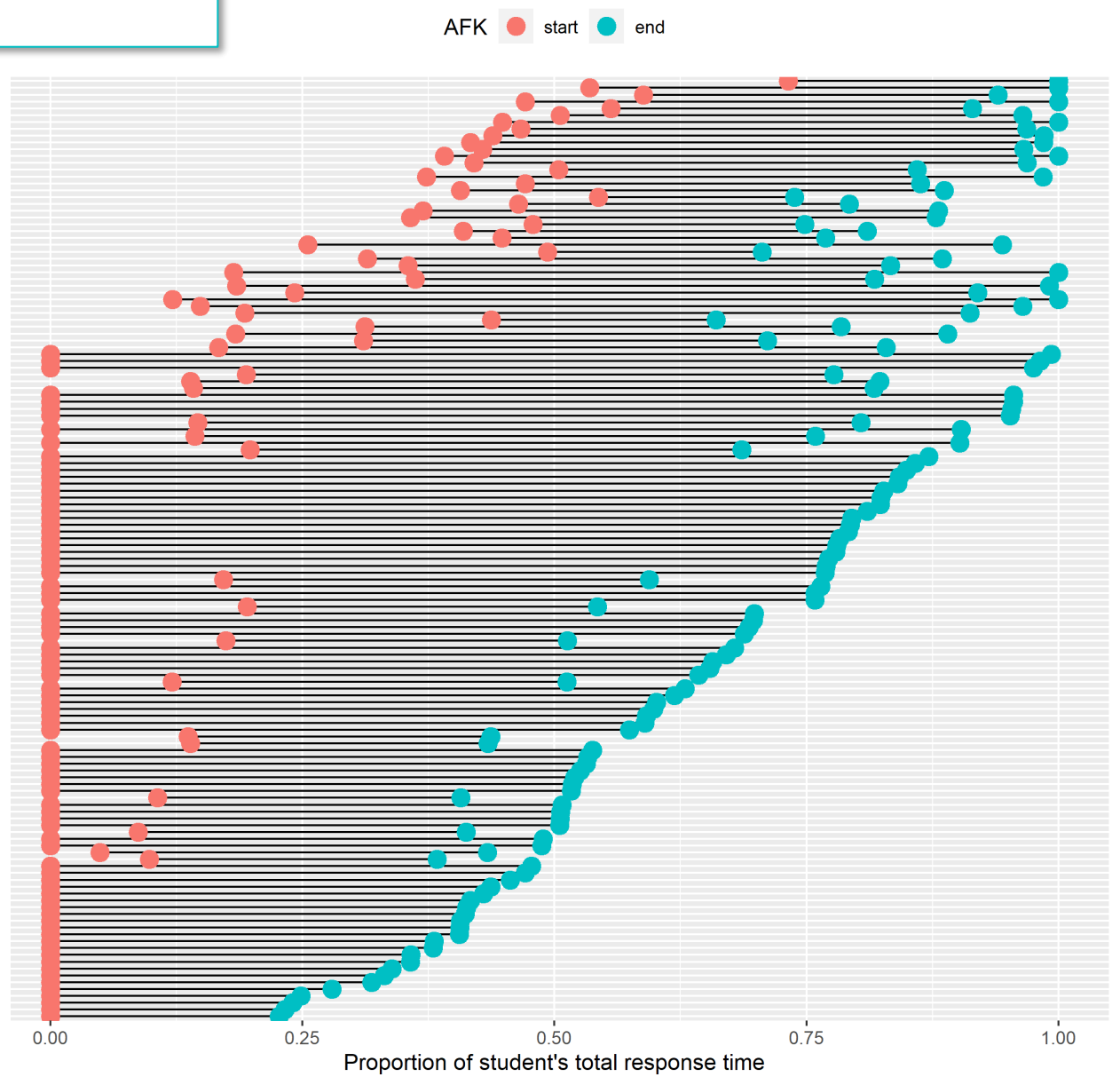


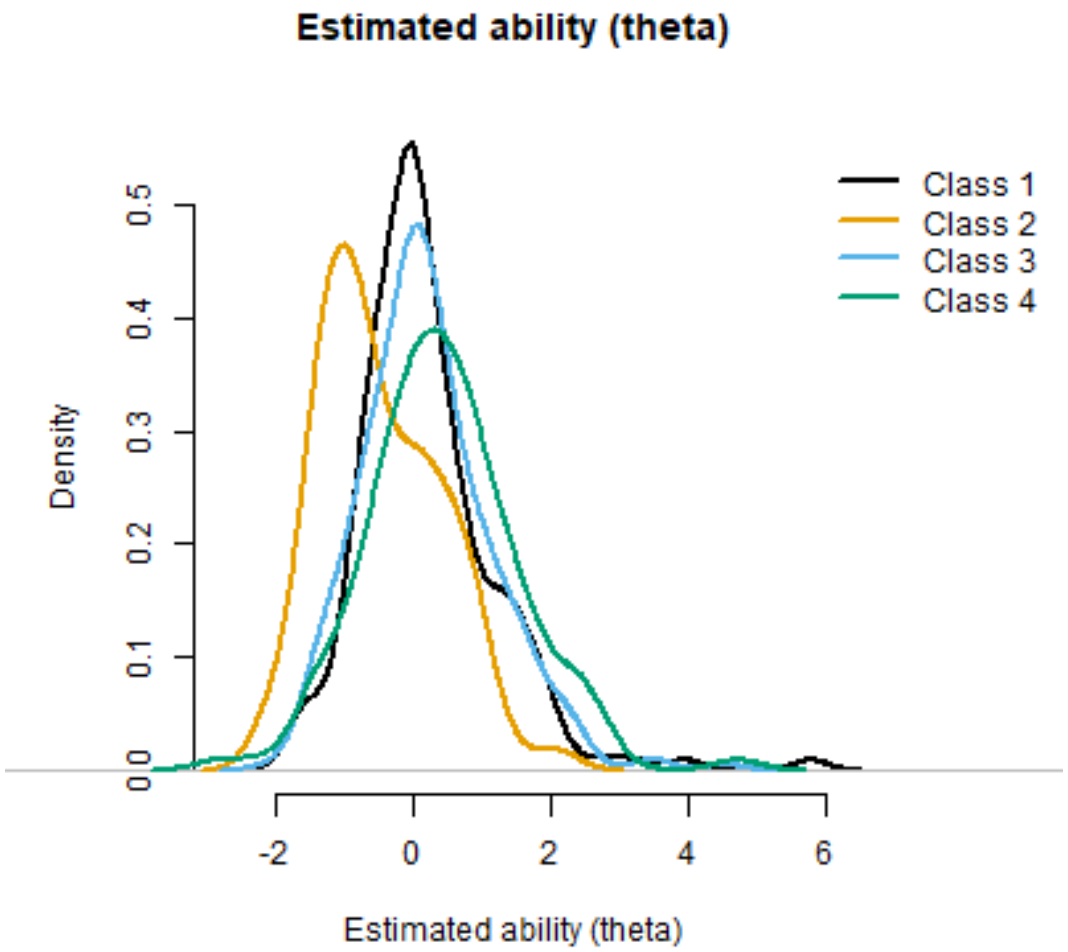
Duration of students' longest time "away from keyboard" (AFK)

Cluster 1: Trial & error?



Cluster 4: Algebraic?





Class 1: Often tried the values 3, x and -2/3.

- Some time away from the environment
- Roughly 10% manages to find the second answer to the item.
- Interact with the item the most.
- Wait relatively long before interacting with the item (but not as long as students in class 4).

Class 2: Hardly any attempt made

- Average estimated ability that is significantly lower ($p < 0.001$)
- Have not interacted a lot
- Spent the least amount of time on the item.
- Never find the answer '-2/3' and roughly 80% scores zero points on the item.

Class 3: Often tried the value 3, never the value -2/3.

- Most often succeeds finding 3. Never -2/3.
- Interact with the item a reasonable amount (not the most and not the least)
- The same goes for their response times

Class 4: Often tried the value 3 and have sometimes tried -2/3 or x.

- Wait the longest before interacting with the item.
- Spent the most time away from interacting
- Do not interacted that much (although more than students in class 2).
- These students find the second answer to the question more often.

References

Burstein, J., LaFlair, G. T., Kunnan, A. J., and von Davier, A. A. (2022). A Theoretical Assessment Ecosystem for a Digital-First Assessment – The Duolingo English Test. Available online at: https://duolingo-papers.s3.amazonaws.com/ other/det-assessment-ecosystem.pdf (accessed February 3, 2022).

Du Bois, J & Karkkainen, E. (2012). 'Staking a Stance on Emotion: Affect, Sequence and Intersubjectivity in Dialogic Interaction'. *Text and Talk*. 32-4. (433-451).

Eklöf, H. (2010). 'Skill and will: test-taking motivation and assessment quality'. *Assessment in Education: Principles, Policy & Practice*, 17:4, 345-356.

Fox, J. (2003). 'From products to process: An ecological approach to bias detection'. *International Journal of Testing*. 3 (1), 21-48.

Maddox B. (2017). 'Talk and Gesture as Process Data'. *Measurement: Interdisciplinary Research and Perspectives*. 15 (3-4). 113-127.

Salles, F., Dos Santos, R & Keskpaik, S. (2020). 'When didactics meet data science: process data in large-scale mathematics assessment in France'. *Large-Scale Assessments in Education*, 8 (7).

Wise, S. L. (2019) 'The Impact of Test-Taking Disengagement on Item Content Representation' Applied Measurement in Education, 33 (2), 83–94 https://doi.org/10.1080/08957347.2020.1732386

Steven L. Wise & G. Gage Kingsbury (2022) Performance Decline as an Indicator of Generalized Test-Taking Disengagement, Applied Measurement in Education, 35:4, 272-286, DOI: 10.1080/08957347.2022.2155651