sugar camp paris #3

Data analysis from XO backup

Jonathan RAGOT, Kevin RAYMOND, Pierre VARLY

12-13 avril 2014

Outline

- Noky Komba project
- Backup procedures
- Data analysis :what for ?
- Data analysis
- Findings and next steps

OLPC deployment in Nosy Komba project

- 2009: 100 XO deployment
- 2010: + 60 XO, XS school server
- 2011: web and 1st malagasy content
- 2012: +50 XO (1.5) junior secondary school
- 2013: Inventory, maintenance, flash
- 2014 and next: +50 XO (1.5), internet practice, computer high school

2009:
100 XO deployment

2010:
+ 60 XO, XS school server


I
NOSY KOMBA
I
NOSY KOMBA
I
OSY KOMBA
I
OSY KOMBA

2011:
web and 1st malagasy content


I
NOSY KOMBA
I
NOSY KOMBA
I
NOSY KOMBA

2012: OLPC community @NK, +50 XO, secondary school, web,

2013: OLPC Fr @NK
Inventory, maintenance, flash


JOHN


7 + 2 =
6 + 6 =
14 + 2 =
8 - 3 =
5 - 5 =
milalao baolina i mora
sy ny namany.
ea amoran'ny dabo
no milalao ny mora

2014 and next: +50 XO (1.5) practice, computer high school

I ♥ NOSY KOMBA

Back up procedures /1

Sources https://git.sugarlabs.org/jparse
Backup the XO with dobackup.sh

Run this script from a USB dongle.

Saves datastore and GNOME files.

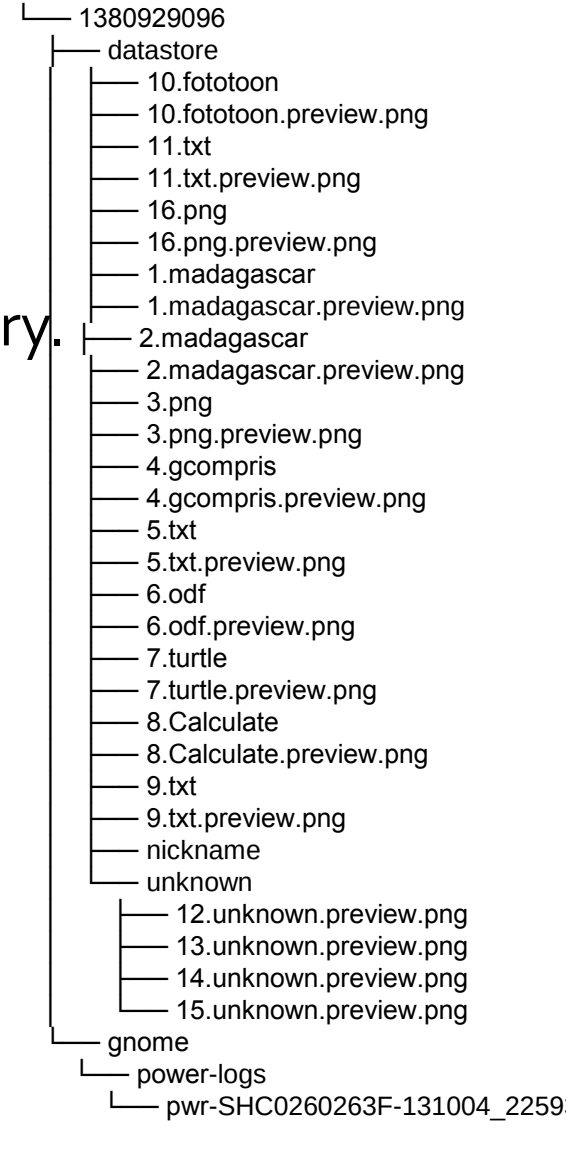
/home/olpc/.sugar/default/datastore
/home/olpc/{Desktop, Images…}

Back up procedures /2

Parse you backups with parse.sh

Run this script from you backup archives directory.

This extract all known and interesting file format per archive, on a specific directory. Set also the file extension.



Data analysis project

1. Request for data from OLPC members

1. Backup of XO by Kevin, Xavier (2010, 11, 12)

2. Data management by Abdallah ABARDA, statistician from Morocco (Varlyproject)

3. Research questions from Sandra and other OLPC France members

4. Data analysis by Adballah and Pierre

5. Presentation of findings in OLPC France meetings

6. Paper by Pierre (in French)+ Blog post to come (in english)

Data analysis : what for ?

Several deployments interested in collecting data from XO : Paraguay, Jamaïca, Nepal
http://www.olpcsf.org/node/204

What do the children do with the XOs ?

- Qualitative analyse : pupils' productions
- Quantitative analysis
- Comparative analysis : accross deployments

→ Learning

Guiding principles

TRANSPARENCY LEARNING FUN

Volume of data collected

Data	2010-2011	2011-2012	2012-2013
XO deployed	166	167	166
XO with back up	33	29	110
Activities deployed	64	66	67
Activities back up	31	31	14
Activities common (years)	8	8	8
Giga bit	na	1,54	11
Files	3,844	12,607	45,606

Nature of data collected

	Activities used by pupil	Number of time activity used by pupil	Time (but not duration)	Size files	Back up Gnome	Standards files use	Standard files readable	Internet (links)
2010-2011	X	X	X					
2011-2012	X	X		X		X		
2012-2013	X	inconsistent	X	X	X	X	X	X

Structure of data colected

One file = one activity = one line

File					Pupil		
File name	size	Date creation	location	Type file	Pupil ID	Gender	Grade
File 1	…	…	…	Jpeg	Pupil1	…	…
File 2	…	…	…	Memory	Pupil 1	….	…

Data limitations

1. Files are generated by the machine
2. Teachers and volunteers support and direct interventions
3. Children can share activities
4. Children can open multiple activities in the discovery of XO
5. Children can take many photos of the same thing …
6. Errors in manipulation…

Questions

Are children really using XOs ?

What activities are most used by children ?

How many times did they use each activity ?

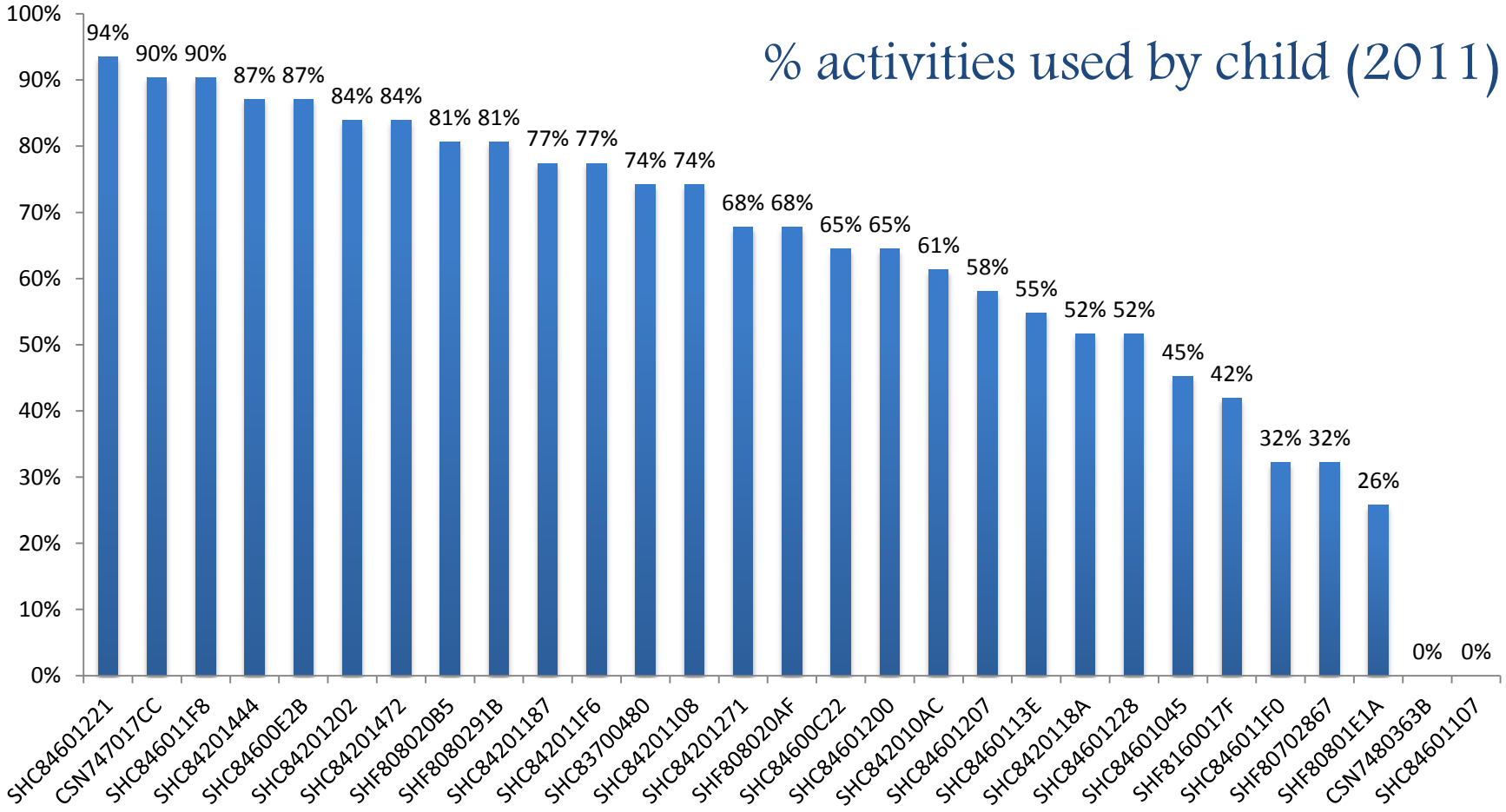
When and how often do they use activities ?

Are they using XO outside school ?

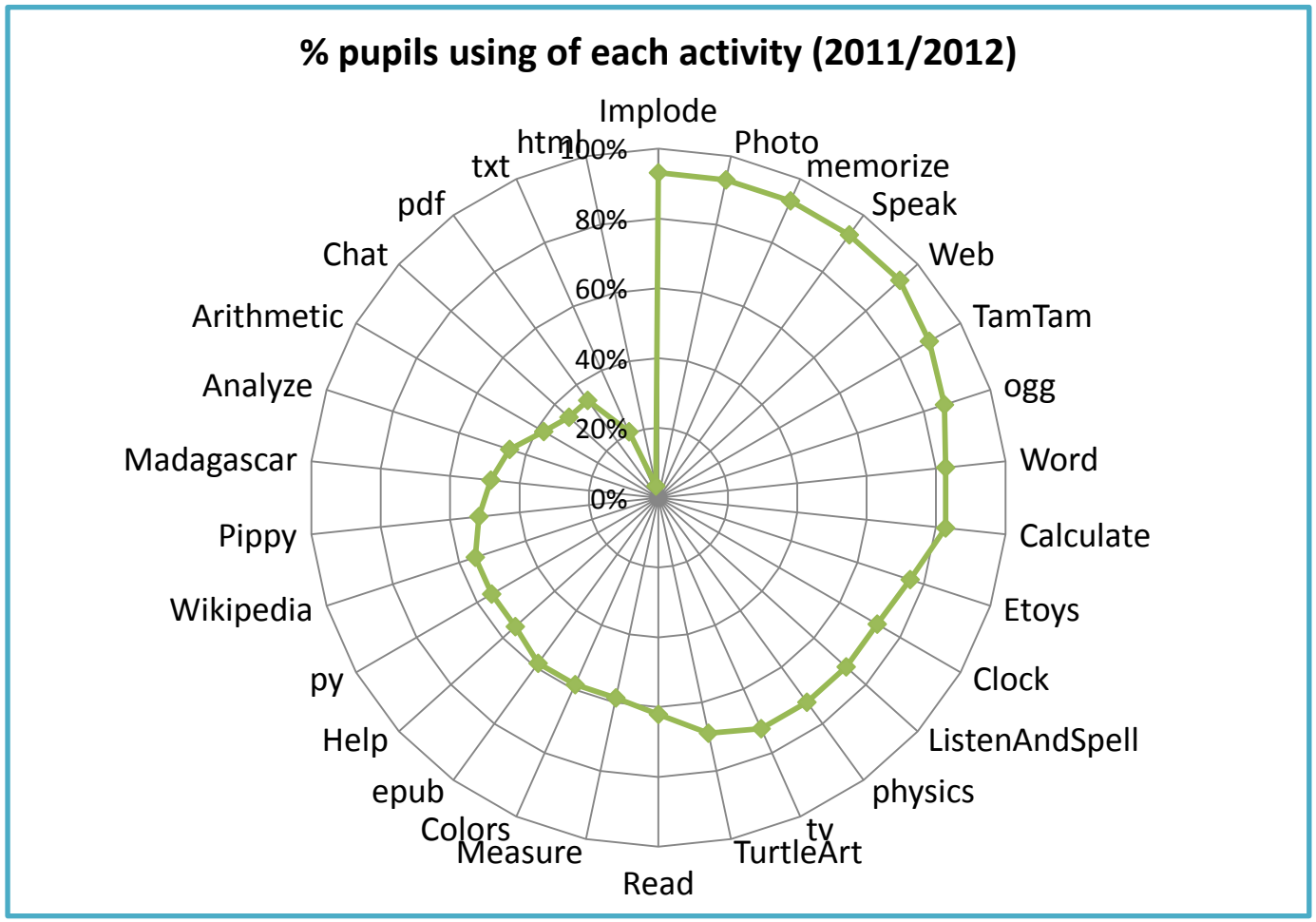
Are XO use differs according to grade and gender?

Can we define profiles of users ?

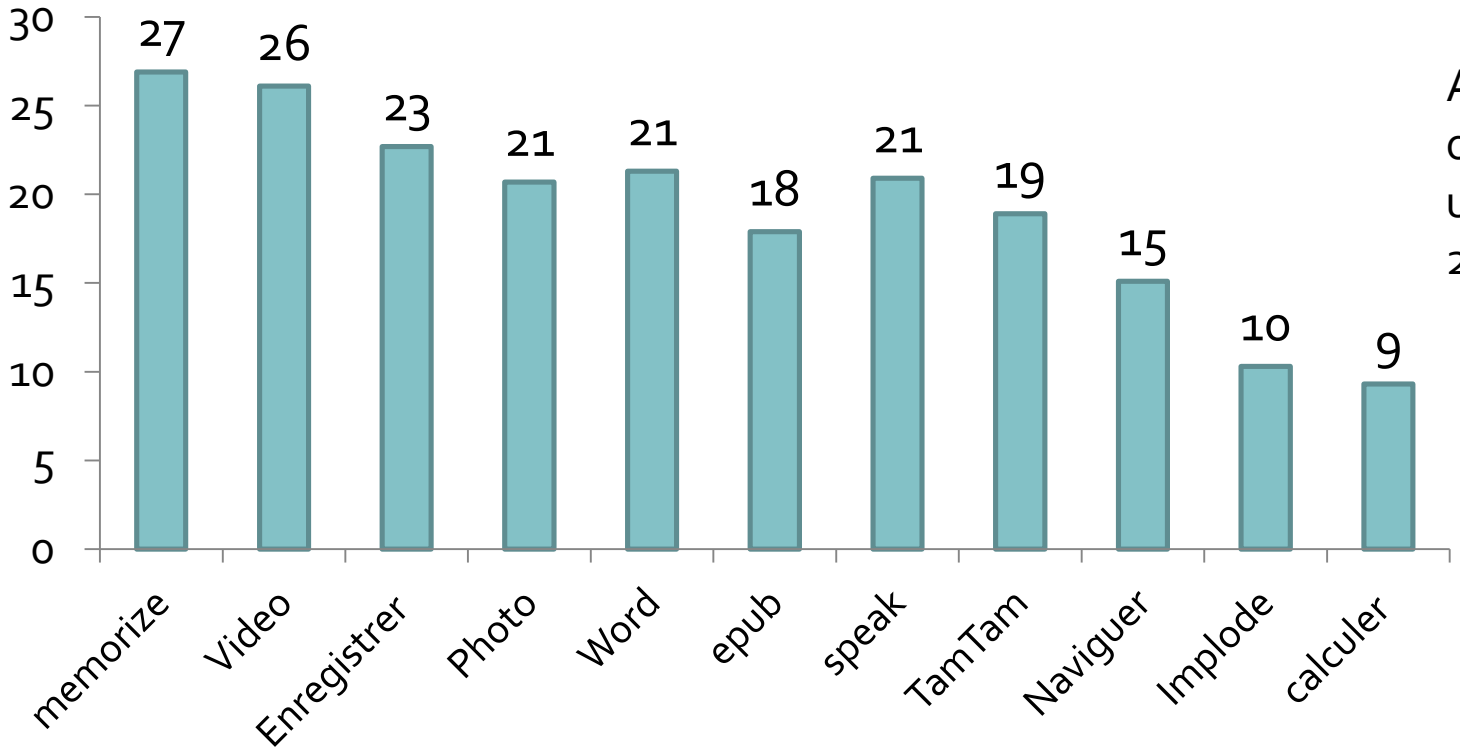
Are children using XO?



What activities are used the most ?



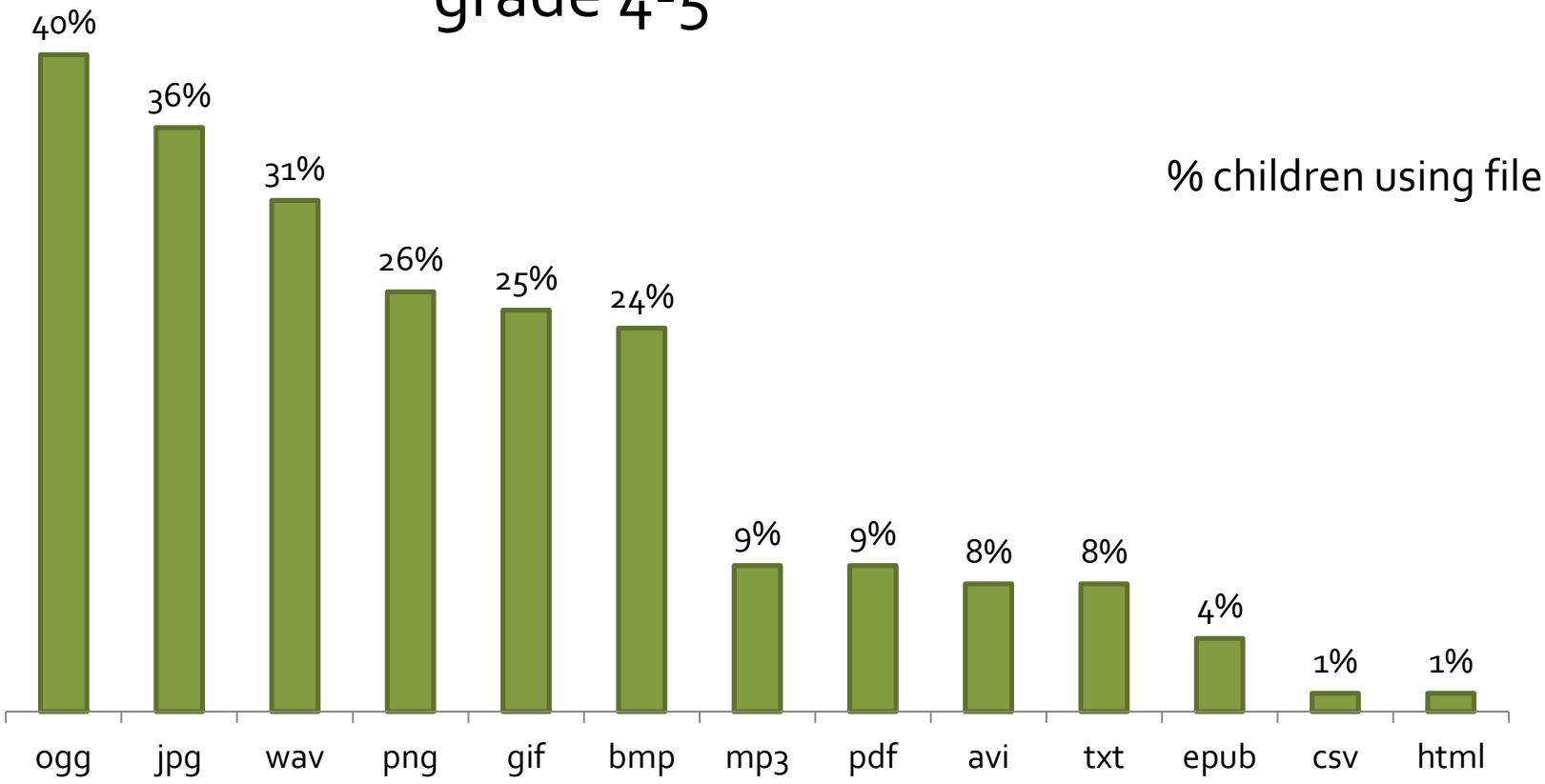
What activities are used the more ?



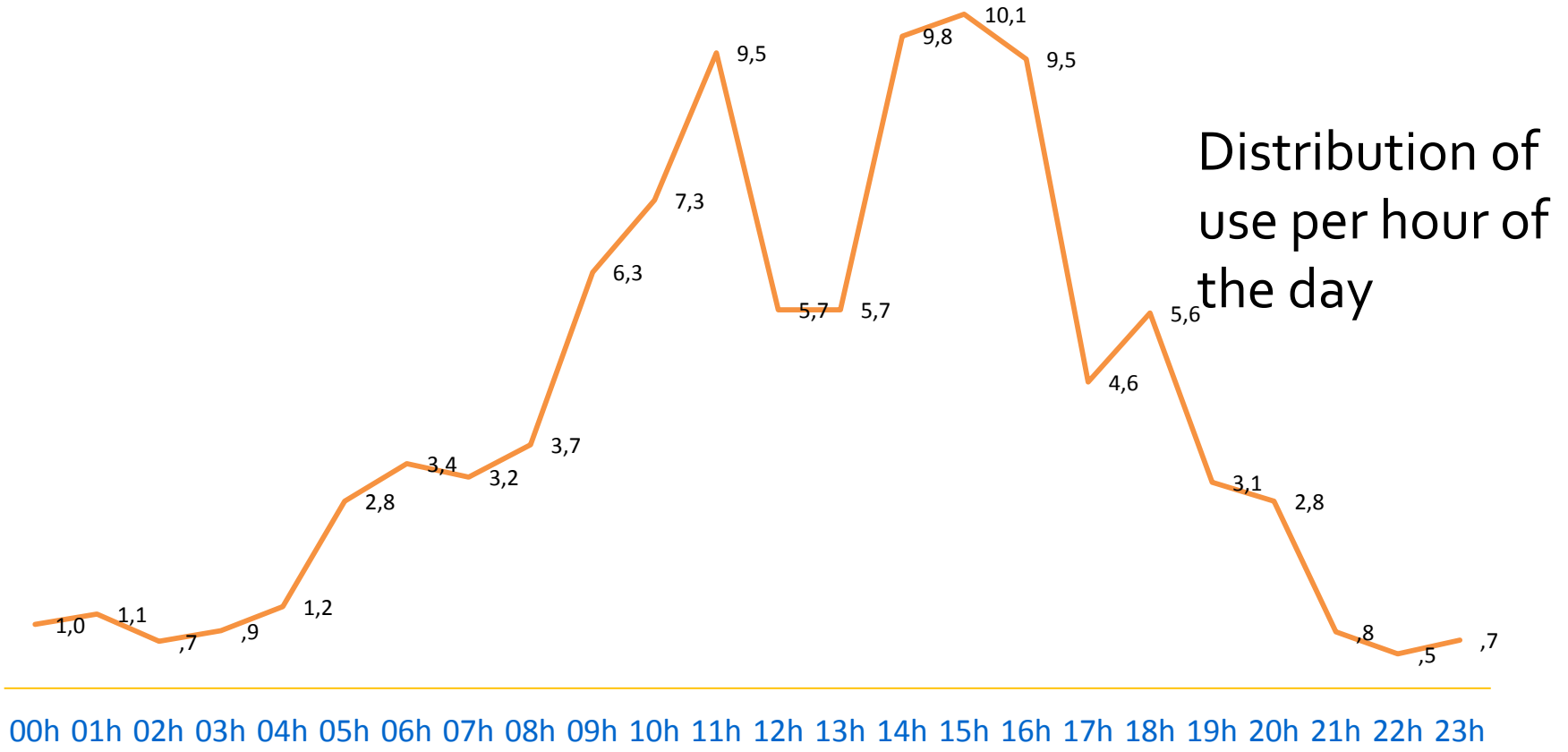
Average number of times each child use activity (2010-2011)

Gnome

Used by 81% of pupils, specially grade 4-5



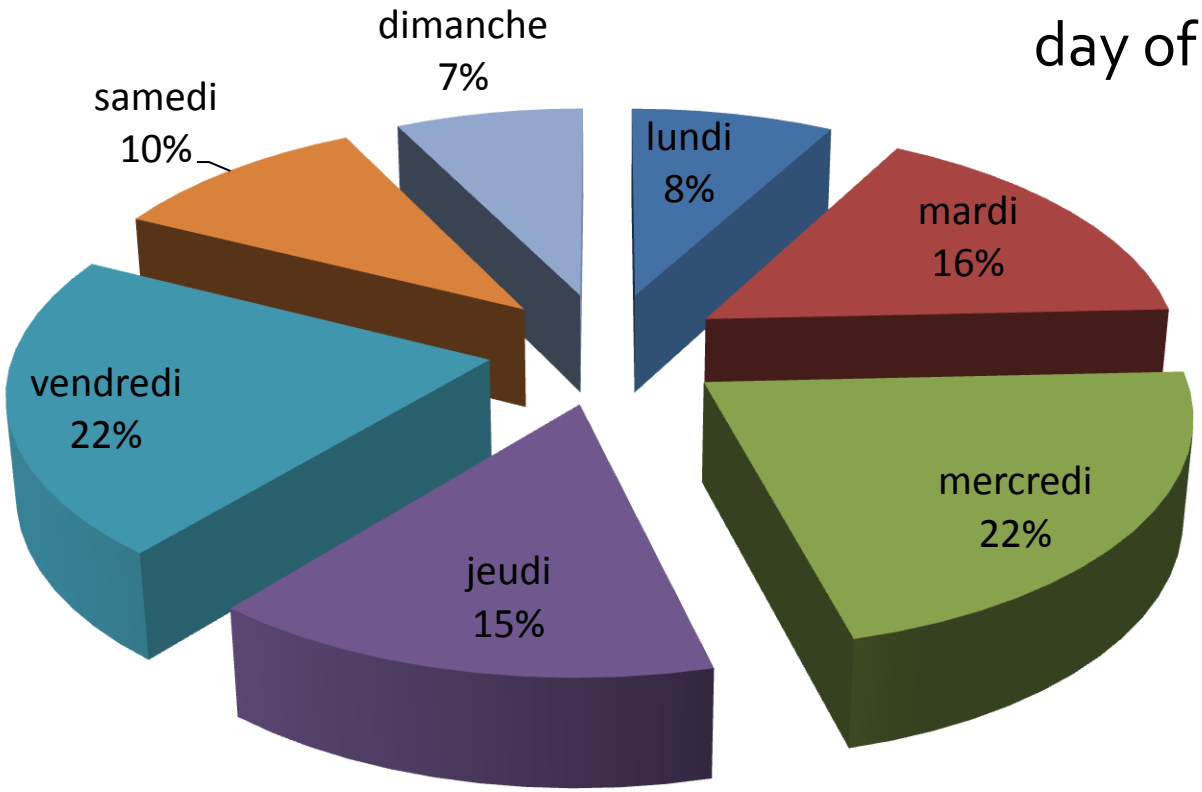
At what time are they using XO ?



Distribution of use per hour of the day

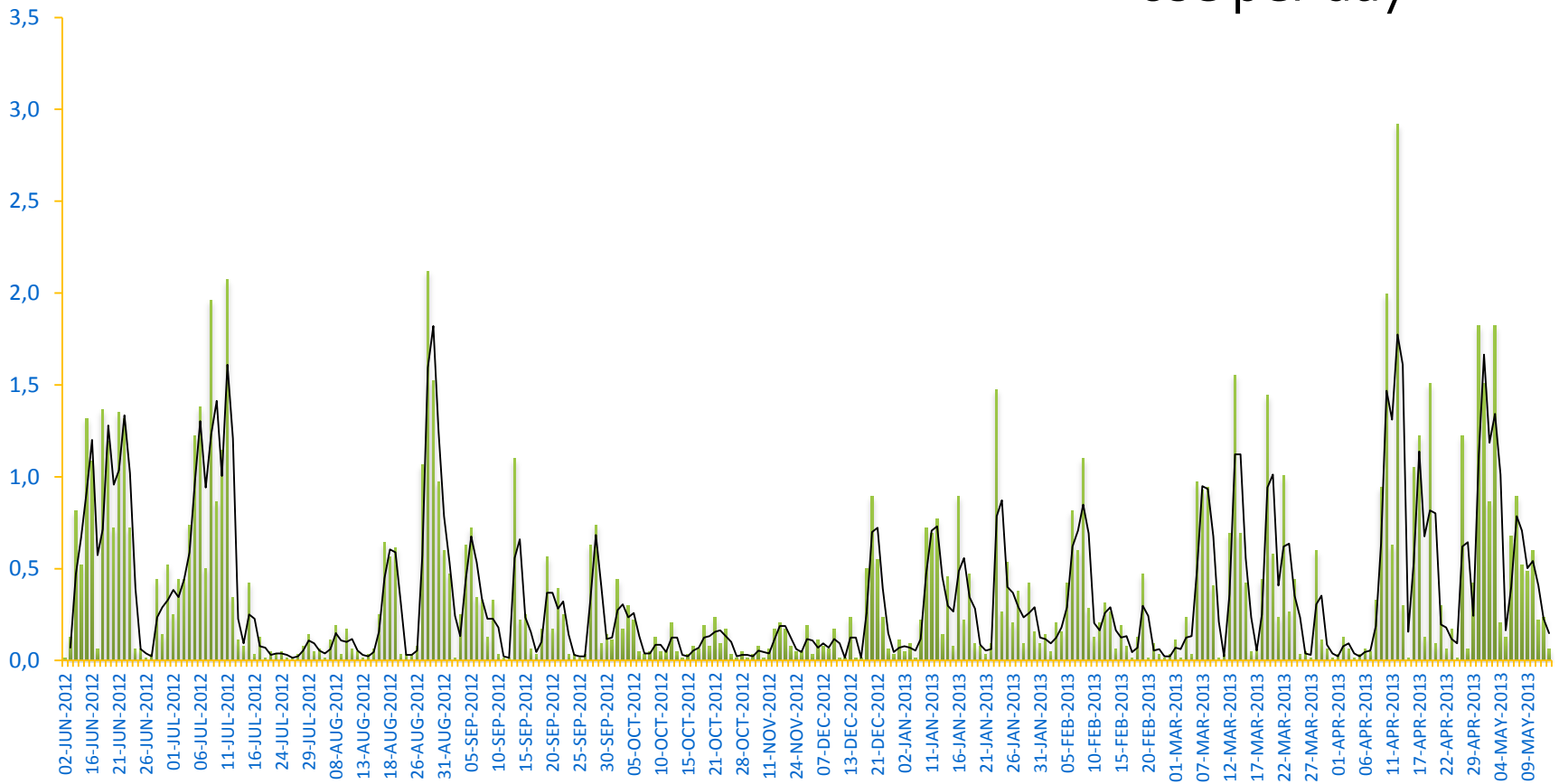
Which day are they using XO ?

Distribution of use per day of the week



At what period of the year?

Distribution of use per day



Does usage varie accross years ?

	2010-2011	2011-2012	2012-2013
calculer	85%	83%	76%
Chat	45%	34%	43%
Video	85%	86%	88%
Pdf	12%	34%	32%
Photo	88%	93%	95%
Rtf	6%	3%	23%
speak	100%	93%	93%
turtle	9%	69%	57%

% children using each activity by year

Does usage varie across grades ?

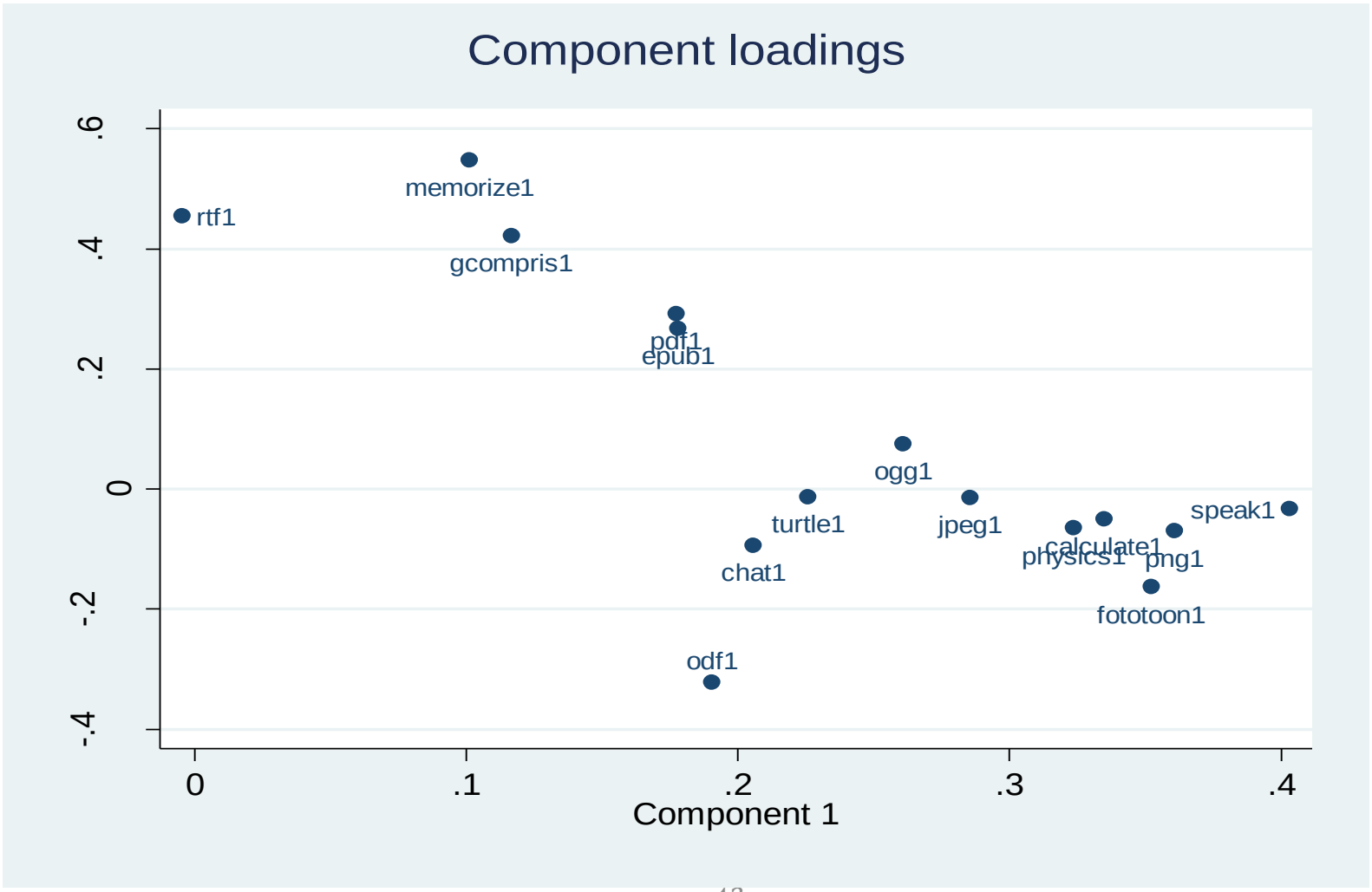
GRADE	JPEG	epub	PDF	OGG	mem	Gcom	speak	turtle	calc	chat	phys	fotot
1	100%	33%	0%	100%	33%	0%	100%	100%	100%	0%	100%	100%
2	88%	44%	28%	92%	56%	44%	96%	76%	88%	48%	92%	88%
3	97%	47%	34%	84%	53%	13%	91%	53%	69%	47%	75%	78%
4	100%	86%	43%	100%	29%	14%	100%	71%	86%	71%	100%	100%
5	94%	61%	39%	89%	50%	39%	94%	39%	72%	44%	89%	89%

Does usage varies by gender?

GENDER	JPEG	epub	PDF	OGG	Mem.	Gcom	speak	turtle	Calcul.	chat	Phys.	Fotot.	Mean
Girls	98%	53%	30%	93%	56%	30%	98%	65%	86%	53%	86%	93%	**70%**
Boys	91%	51%	37%	86%	47%	23%	91%	56%	70%	40%	86%	79%	**63%**
Average	94%	52%	34%	90%	51%	27%	94%	60%	78%	47%	86%	86%	**67%**

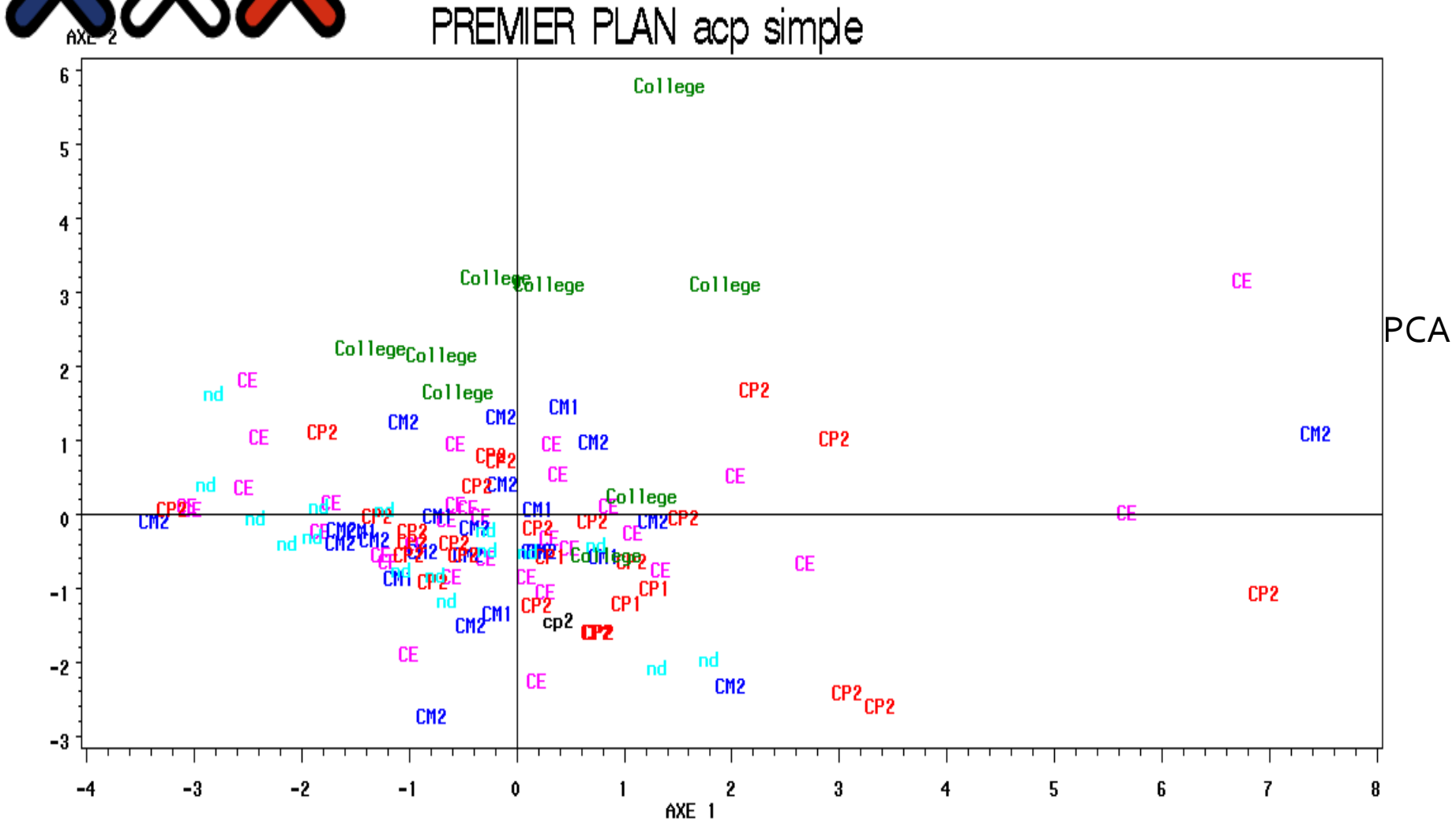
Correlations among use of activities ?

Cronbach alpha 0,74 > 0,7



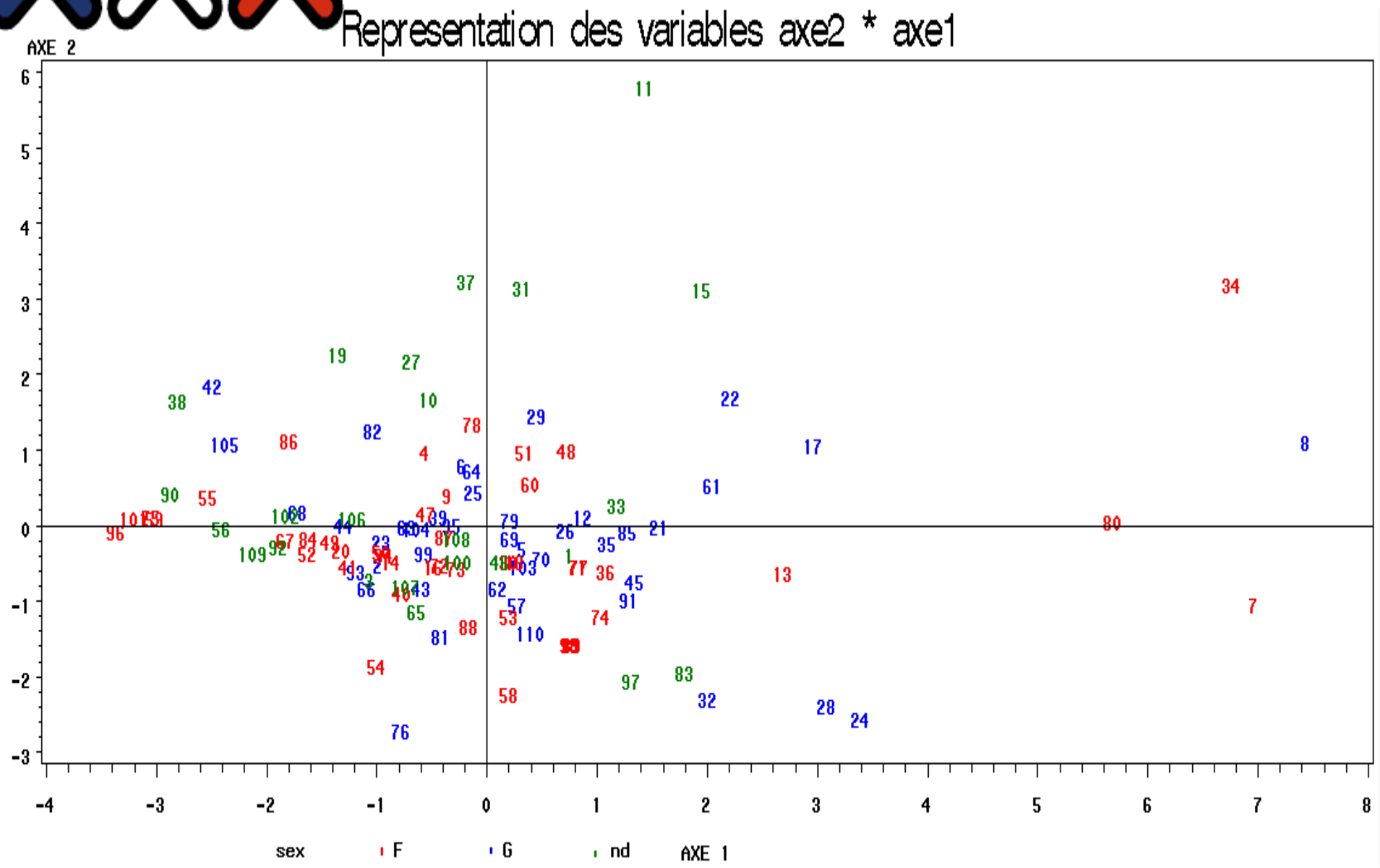
PCA

Use of activities driven by grade



PCA

Diversity of use by gender



PCA

Girls

Boys

Gender unknown

Findings

Data consistent

- Internaly
- With contextual variables
- With observations on field

Children use XO oustide school hours

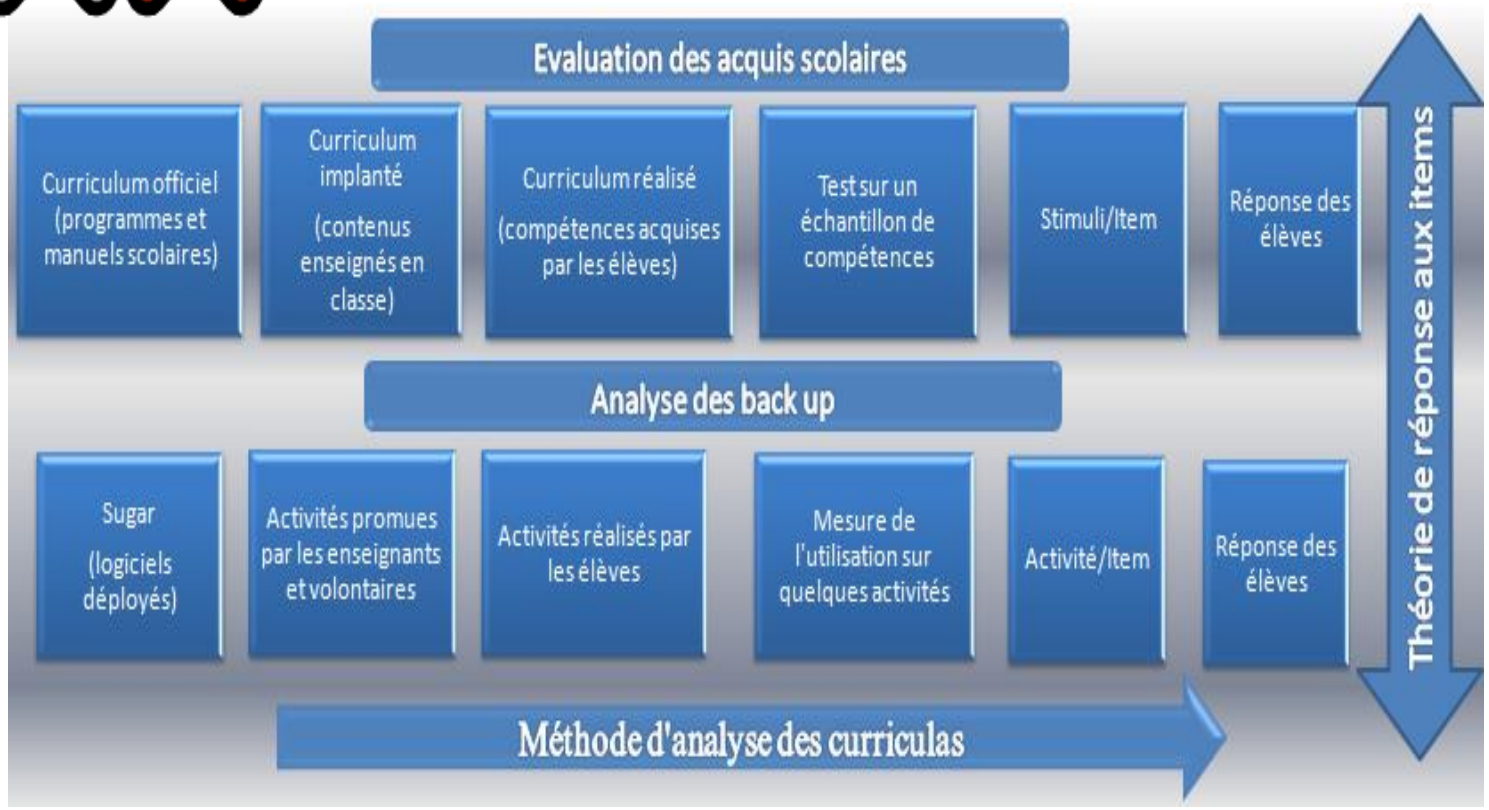
Children use XO along the year

Large diversity of usage among children

Lower secondary children use if much differently than in primary

Gender effect

Data analysis framework



Activity = stimuli → Children use : Yes /No

Item → Item response

Deontology

Testing specialits have developed their own norms (**APA**)

Simple norms should be set :

- Ask permission for data back up, to whom ?
- Data security
- Document data limitations to avoid misinterpretation
- **What date are for ? What data are not for ?**
- Data should not be used to judge teachers or pupils work
- Scientific purpose only

Simple things can be done

Collect information on pupils (using Quizz or other app.) :

Gender, grade, repetition, books, computer, electricity at home (and other goods), marks ?

Assess basic knowledge of pupils and teachers :

Academic and IT skills

Quick survey on activities that pupils/teachers like and use :

Demand driven versus experts driven

Cross check declaration with effective use

Complex things can be done

Define XO/Sugar learning metric (standardised indicators)

Write data analysis procedure with R (ongoing)

Develop better Journal/log activities

Share/standardise ? back up procedures

Set incentives for deployments to collect data

Share data

Compare use across deployments (anchor item-activities)

Great research potential

Gcompris

Speak

Abcdaire

Memory

Falabracam

Emergent Reader

Fluent Reader

What combination of applications increase pupils abilities ?

Merci

Thank you

Gracias

میرسی