

# Vocab@Leuven 2019

The Third-K slump:  
What Lexical Tutors can  
learn from Lextutor users

**Tom Cobb**

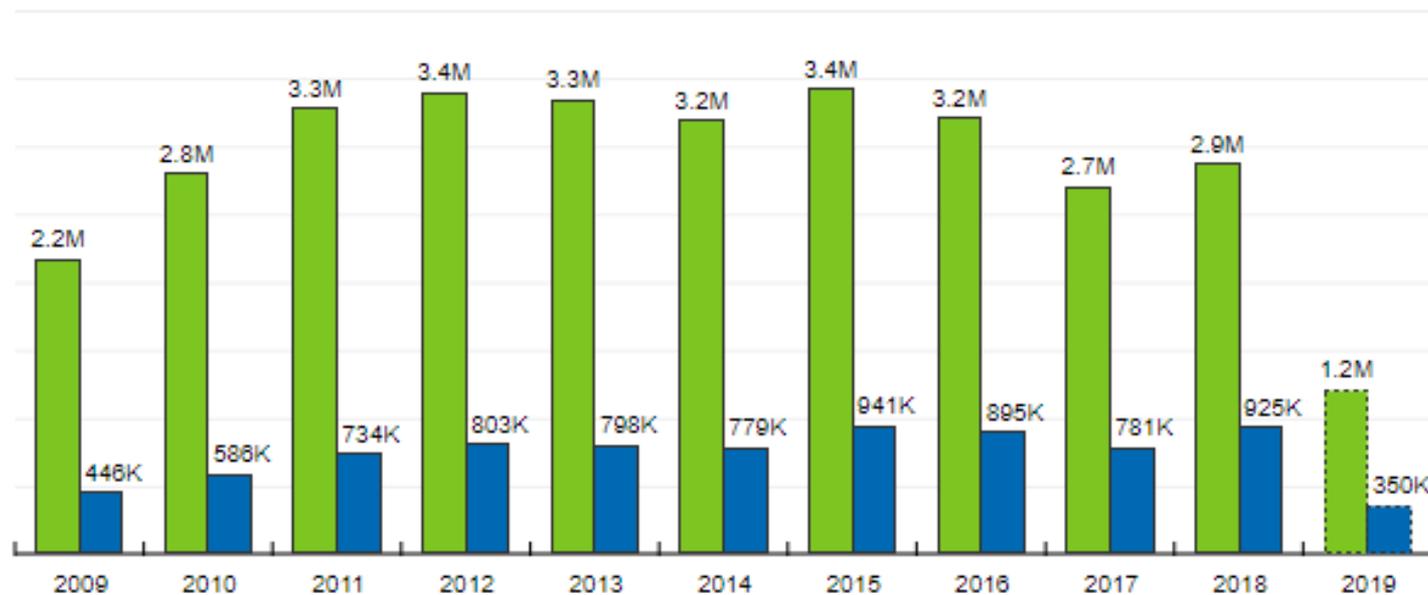
**Université du Québec à Montréal**

**UNESCO Curriculum Reform Program - Geneva**

Get this PPT at <https://www.lex tutor.ca/vocab-leuven-tom.pptx>

Lextutor 2009-2019 : Who are these users?  
>10 million users  
>35 million pages

Who are these users?  
What do they do?  
What does it mean?



# In this talk, I meant to show you some quirky usage features

- But inevitably a ‘theme’ emerged that could not be ignored
- A theme related to an old paper of mine
  - “Computing the vocabulary demands of L2 reading” (LLT, 2007)
- That looked for the reason that ESL learners typically know so few words
  - Even in the most elementary meaning of ‘know’

“So few words” →

## Typical VLT results in my old research

- Cobb (1997)
  - “Mean pre-test score for the experimental group on the 2000-level was **670** words, or 33.5% (SD 6.5), and their post-test score was **1100** words, or 55% (SD 10.5).”
- Cobb Horst Meara (1998)
  - “Average knowledge of the 2000 most frequent words was was estimated at **1203** words (sd = 348) and of the 5000 most frequent English words **2071** words (sd = 560).”

# This finding is quite general (or *was* in 1990s)

Table 1: English vocabulary size of foreign learners

Country	Vocab Size	Hours of Instruction	Reference (re-
Japan (EFL Univ.)	2000 2300	800-1200	Shillaw 1995 Barrow et al.
China (Eng. majors)	4000	1800-2400	Laufer in press
Indonesia (EFL Univ.)	1220	900	Nurweni and Read1999
Oman (EFL Univ.)	2000	1350+	Horst et al 1998
Israel (high school grad.)	3500	1500	Laufer 1998
France High school	1000	400	Arnaud et al. 1985
Greece "Age 15, high school "	1680	660	Milton & Meara 1998
Germany "Age 15, high school"	1200	400	Milton & Meara 1998

Laufer, Batia.  
*Task effect on instructed vocabulary learning: The hypothesis of 'involvement'.*  
AILA 1999, plenary address

# My 2007 paper argued...

- The reasons for these low figures are quite simple
- 1. Obviously, too few hours
  - Hours x size  **$r=.92$**
- 2. Less obviously, with no systematic vocabulary instruction ~
  - Learners are trying to learn from input
  - And the conditions for learning from input run dry at c. 2500 word families

	A	B	
1	<b>SIZE</b>	<b>HOURS</b>	
2	2000	1000	
3	2300	1000	
4	4000	2100	
5	1220	900	
6	2000	1350	
7	3500	1500	
8	1000	400	
9	1680	660	
10	1200	400	
11			
12	<b>CORRELATION</b>		
13		<b>0.92</b>	
14			

# In any text or corpus you look at,

- Somewhere in the lower 3k zone,
- Frequency drops below any known L2 learnability criteria
  - Number of occurrences is low
  - Time between occurrences is high

# Take e.g. “The Great Gatsby”

- A “lexically rich” text
- Here are the 2<sup>nd</sup> and 3<sup>rd</sup> thousand family lists with 6 occurrences highlighted →



# Learnability conditions at 3k

- There are 311 word families @ 3k in The Great Gatsby
- But the majority are just 1-5 occurrences

[1] = one instance, etc.

- [1] 184
- [2] 66
- [3] 33
- [4] 9
- [5] 6
- [6] 5
- [7] 7
- [8] 2
- [9] 1
- [10] 1
- [11] 1
- [12] 1
- [15] 1

So if 6 occurrences are *typically* needed to form a permanent representation...

Then, here there are just  $5+7+2+1+1+1+1+1$   
= **19** words up for learning

Appearing in a 92% known-word context  
(Assuming 1k+2k are known)  
(~1 word unknown in 10)

And this assumes different family members are recognized as re-occurrences

# My conclusion was that ~

- **An L2 reading lexicon is difficult to build from reading alone**
- In L1, reading can “do the whole job”
  - Timeframe is 15+ years, with a gradual increase of lexical complexity
    - in self-selected texts
- In L2, it cannot (in normal circumstances)
  - Timeframe is 1-2 years
  - Reading rate is slow
  - Lexical complexity comes all at once
    - in assigned texts

# A paper that embroiled me in a long debate with ~

- Krashen & McQuillan
  - Cobb, T. (2008). [What the reading rate research does not show: Response to McQuillan & Krashen](#). *Language Learning & Technology* 12(1), 109-114.
- Nation & McQuillan
  - Cobb, T. (2016) [Numbers or Numerology? A response to Nation \(2014\) and McQuillan \(2016\)](#). *Reading in a Foreign Language*, 28 (2), 299-304.
- And brought down the wrath of the extensive reading establishment

# Now, new data from Lextutor users

Sheds further light on this phenomenon

- For two RQs:

**RQ1:** Is low vocab still true?

- After 30 years of ‘the vocab revolution’
  - since Laufer’s survey in 1999

**RQ2:** What do learners themselves do about it?

- Since no one else seems willing to help them
  - Except to say ‘You should read more’

To answer these  
2 questions we  
look at  
Lextutor's most  
used routines



Page Views	Page Title
964	VP 1-WORD (HOME)
865	Compleat Lexical Tutor
853	CORPUS CONCORDANCE OUTPUT
806	VOCABPROFILE COMPLEAT - INPUT
789	VOCABPROFILE COMPLEAT - OUTPUT
748	VP 1-WORD (EDIT-TO-PROFILE)
484	CORPUS CONCORDANCE OUTPUT
417	VOCABPROFILE ENGLISH
338	VST (computer/test )
328	CORPUS CONC ENG
318	VP HOME
292	VOCABPROFILE ENGLISH OUTPUT
255	TESTS HOME

964	VP 1-WORD (HOME)	←
865	Compleat Lexical Tutor	
853	CORPUS CONCORDANCE OUTPUT	←
806	VOCABPROFILE COMPLEAT - INPUT	←
789	VOCABPROFILE COMPLEAT - OUTPUT	←
748	VP 1-WORD (EDIT-TO-PROFILE)	←
484	CORPUS CONCORDANCE OUTPUT	←
417	VOCABPROFILE ENGLISH	←
338	VST (computer/test )	←
328	CORPUS CONC ENG	←
318	VP HOME	←
292	VOCABPROFILE ENGLISH OUTPUT	←
255	TESTS HOME	←
247	CORPUS CONC ENG	←
185	ONLINE CONCORDANCERS	←
112	Compleat Lexical Tutor	←
103	LARGER CONTEXT	←

For a typical day

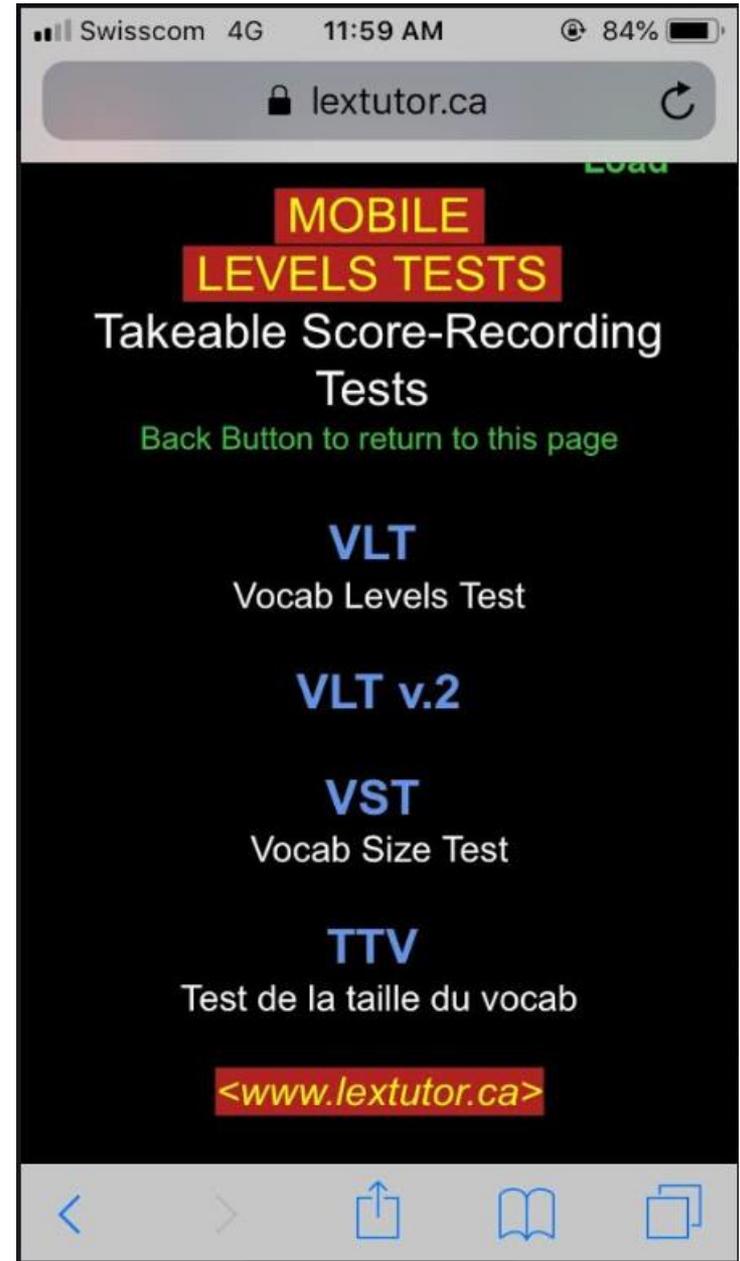
ROUTINE	HITS
<b>VP</b>	4,334
<b>Conc</b>	2,294
<b>Tests</b>	593
<b>Group Lex</b>	200

# Levels Tests on Lextutor

Four multi-platform, dual-mode, frequency-level-based tests

Why am I collecting results?

Teacher requests for recorded results



# Levels Tests on Lextutor

- Mobile + computer
- Score + practice versions
- Only scored version is recorded

TEST	# SCORES RECORDED THUS FAR IN 2019
VST	<b>2983</b> scores
VLT	<b>2641</b> scores
VLT2	<b>1757</b> scores
TTV	<b>36</b> scores
	<b>7,417</b> total scores

**Vocab Size Test**  
Mode = test

1 3 4 5 6 7 8 9 10 11 12 13 14

- First 1000 [Go 2]
- SEE: They **saw** it.
    - cut
    - waited for
    - looked at
    - started
  - TIME: They have a lot of **time**.
    - money
    - food
    - hours
    - friends
  - PERIOD: It was a difficult **period**.
    - question
    - time
    - thing to do
    - book
  - FIGURE: Is this the right **figure**?
    - answer
    - place
    - time
    - number
  - POOR: We are **poor**.
    - have no money
    - feel happy
    - are very interested
    - do not like to work hard
  - DRIVE: He **drives** fast.
    - swims
    - learns
    - throws balls
    - uses a car
  - JUMP: She tried to **jump**.
    - lie on top of the water
    - get off the ground suddenly
    - stop the car at the edge of the road
    - move very fast
  - SHOE: Where is your **shoe**?
    - the person who looks after you
    - the thing you keep your money in
    - the thing you use for writing
    - the thing you wear on your foot
  - STANDARD: Her **standards** are very high.
    - the bits at the back under her shoes
    - the marks she gets in school
    - the money she asks for
    - the levels she reaches in everything

- Second 1000 [Go 3]
- MAINTAIN: Can they **maintain** it?
    - keep it as it is
    - make it larger
    - get a better one than
    - get it
  - NIL: His mark for that question was **nil**.
    - very bad
    - nothing
    - very good
    - in the middle

VST - SCORE  
TIME (EST): 2019.06.22 / 08.14  
LOCATION: IP\_170.157.22.217  
NAME: Student

Level	Percent
1k:	0
2k:	0
3k:	0
4k:	0
5k:	0
6k:	0
7k:	0
8k:	0
9k:	0
10k:	0
11k:	0
12k:	0
13k:	0
14k:	0

Size: 0 wds  
Recorded **here**

**VST - SCORE**  
**TIME (EST):** 2019.06.22 / 08.14  
**LOCATION:** IP\_170.157.22.217  
**NAME:** Student

Level	Percent
1k:	100
2k:	100
3k:	60
4k:	0
5k:	0
6k:	0
7k:	0
8k:	0
9k:	0
10k:	0
11k:	0
12k:	0
13k:	0
14k:	0
Size:	2600 wds
Recorded	<b>here</b>



# Example data file

```
2019.06.04 / 06.17 IP_81.31.109.202 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.14 IP_140.123.47.186 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.14 IP_140.123.47.213 6466 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.14 IP_140.123.47.205 Haha 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.13 IP_140.123.47.199 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.11 IP_140.123.47.183 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.11 IP_140.123.47.183 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.11 IP_140.123.47.183 1k= 2k= 3k= 4k= 5k= 6k= 7k= 8k= 9k= 10k= 11k=
2019.06.04 / 00.11 IP_140.123.47.183 Fsxf 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.11 IP_140.123.47.183 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.10 IP_140.123.47.183 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.10 IP_140.123.47.183 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.10 IP_140.123.47.183 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.07 IP_140.123.47.192 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.06 IP_140.123.47.192 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.06 IP_140.123.47.192 1 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.06 IP_140.123.47.193 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.05 IP_140.123.47.193 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.05 IP_140.123.47.190 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.05 IP_140.123.47.213 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.05 IP_140.123.47.211 Gfhfghfghfgxcghcvcvbfxhdfsgretgwergrngdfgdfzgz 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0
size=0 level=k01 secs=33
2019.06.04 / 00.04 IP_140.123.47.204 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.04 IP_140.123.47.211 1k= 2k= 3k= 4k= 5k= 6k= 7k= 8k= 9k= 10k= 11k=
2019.06.04 / 00.04 IP_140.123.47.197 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.04 IP_140.123.47.207 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.04 IP_140.123.47.197 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.04 IP_140.123.47.182 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.04 IP_140.123.47.203 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.04 IP_140.123.47.196 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.04 IP_140.123.47.220 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.04 IP_140.123.47.200 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.04 IP_140.123.47.197 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.04 IP_140.123.47.219 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.04 IP_140.123.47.185 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.03 IP_140.123.47.211 Kjljkljkljklfvhjkljhh 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0
secs=20
2019.06.04 / 00.03 IP_140.123.47.230 Student 1k=60 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.03 IP_140.123.47.204 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.03 IP_140.123.47.189 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.03 IP_140.123.47.191 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.03 IP_140.123.47.188 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.03 IP_140.123.47.198 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.03 IP_140.123.47.217 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
2019.06.04 / 00.03 IP_140.123.47.218 Student 1k=0 2k=0 3k=0 4k=0 5k=0 6k=0 7k=0 8k=0 9k=0 10k=0 11k=0
```

# Structure of data file

2019.05.31 / 06.00	IP_178.197.234.100							
2019.05.31 / 04.09	IP_81.106.166.144							
2019.05.31 / 04.09	IP_81.106.166.144							
Tom	1k=0	2k=0	3k=0	4k=0	5k=0	6k=0	7k=0	8k=0
Student	1k=0	2k=0	3k=0	4k=0	5k=0	6k=0	7k=0	8k=0
Student	1k=100	2k=80	3k=90	4k=90	5k=90	6k=70	7k=40	8k=90
9k=0	10k=0	11k=0	12k=0	13k=0	14k=0	size=0	level=k0	secs=36
9k=0	10k=0	11k=0	12k=0	13k=0	14k=10	size=100	level=k0	secs=3619
9k=40	10k=40	11k=0	12k=0	13k=40	14k=30	size=8000	level=k0	secs=3591

# Data Filter

- Record is deleted if :
  - There is no score for lowest level
    - 1k for VST
    - 2k VLT and VLT2
  - Time for test is < 180 seconds
    - 3 minutes
    - Enough to get to 3k part of test

# Takers from where?

219.88.224.45		New Zealand
222.164.160.228		Singapore
82.219.32.93		UK
101.127.10.96		Singapore
42.60.240.245		Singapore
175.0.79.213		China
82.15.195.149		UK
222.165.107.5		Singapore
41.182.165.148		Namibia
203.118.152.232		New Zealand
203.118.152.232		New Zealand
197.233.142.102		Namibia
101.127.22.181		Singapore
222.153.252.40		New Zealand
113.210.57.145	<b>ETC</b>	Malaysia
113.210.57.145		
113.210.57.145		
113.210.57.145	<b>100+ COUNTRIES</b>	
113.210.57.145		
113.210.57.145		

# Test Results -->

(By number of scores meeting  
criterion)

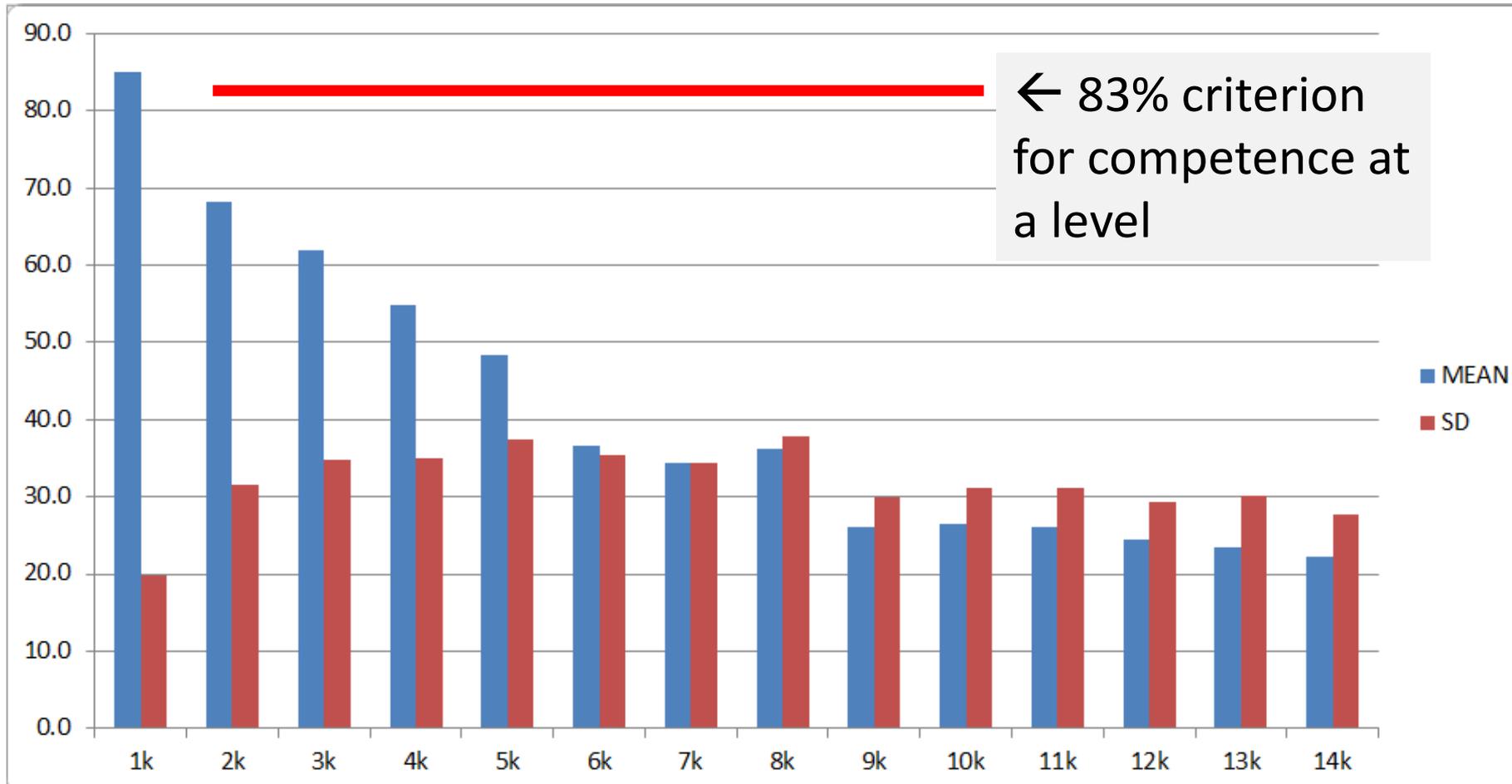
# Result VST

- Total scores requested : 2980
- After filter:  $K1 > 0 + \text{Time} > 180 \text{ secs} = 320$   
( $<11\%$  of requests)

VST	1k	2k	3k	4k	5k	6k	7k
MEAN	85.0	68.3	62.0	54.9	48.5	36.6	34.3
SD	19.8	31.6	34.9	35.0	37.4	35.4	34.4

8k	9k	10k	11k	12k	13k	14k
36.3	26.1	26.5	26.1	24.5	23.4	22.3
37.8	30.0	31.2	31.2	29.3	30.2	27.8

# Result VST

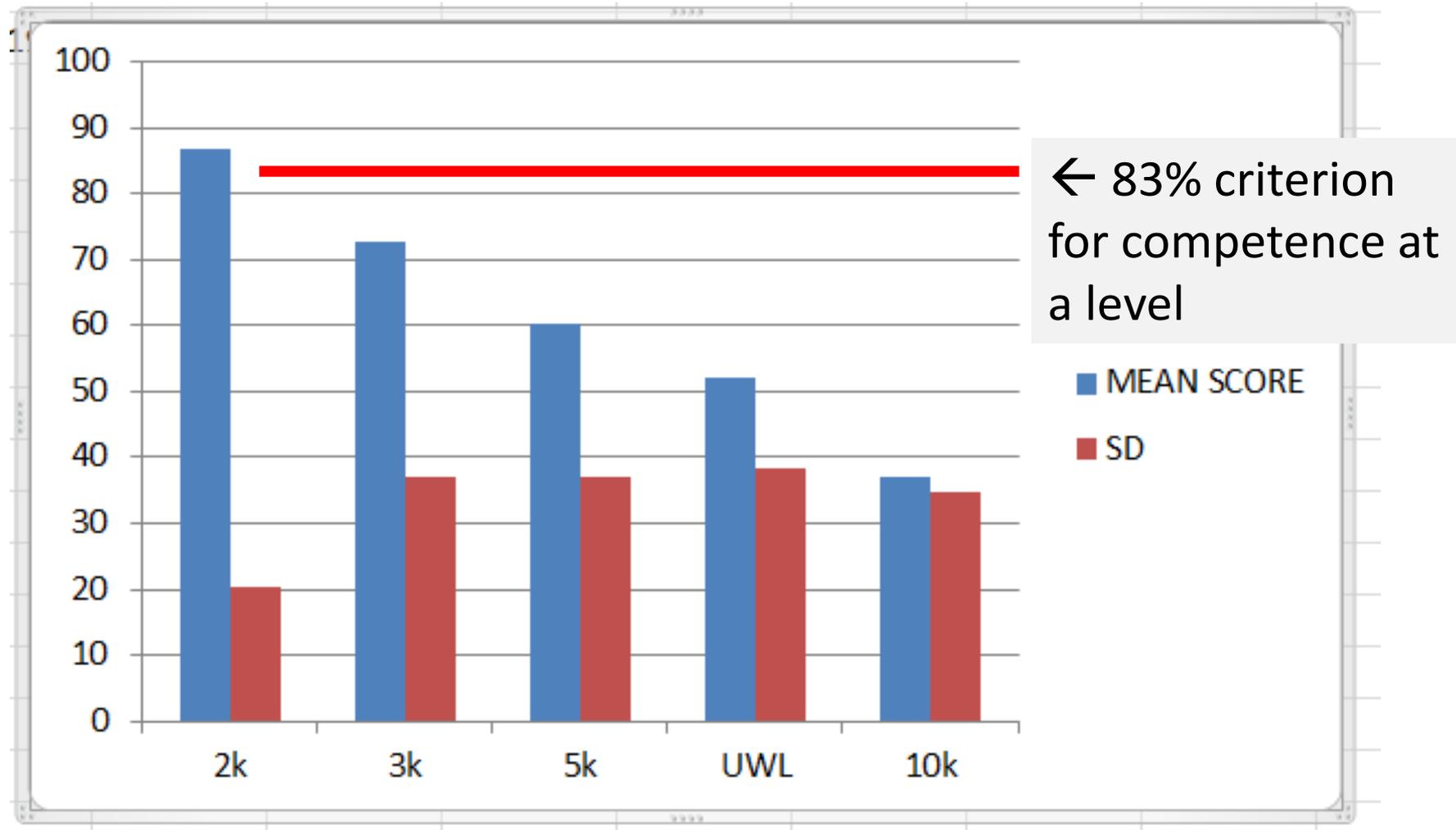


# Result VLT

- Total scores requested : 2646
- After filter:  $K2 > 0 + \text{Time} > 180 \text{ secs} = 1,769$  scores (67% of requests)
- Mean percentages by level

VLT	2k	3k	5k	AWL	10k
MEAN	86.6	72.8	60.4	52.0	37.2
SD	20.2	17.2	37.1	38.2	34.9

# Result VLT



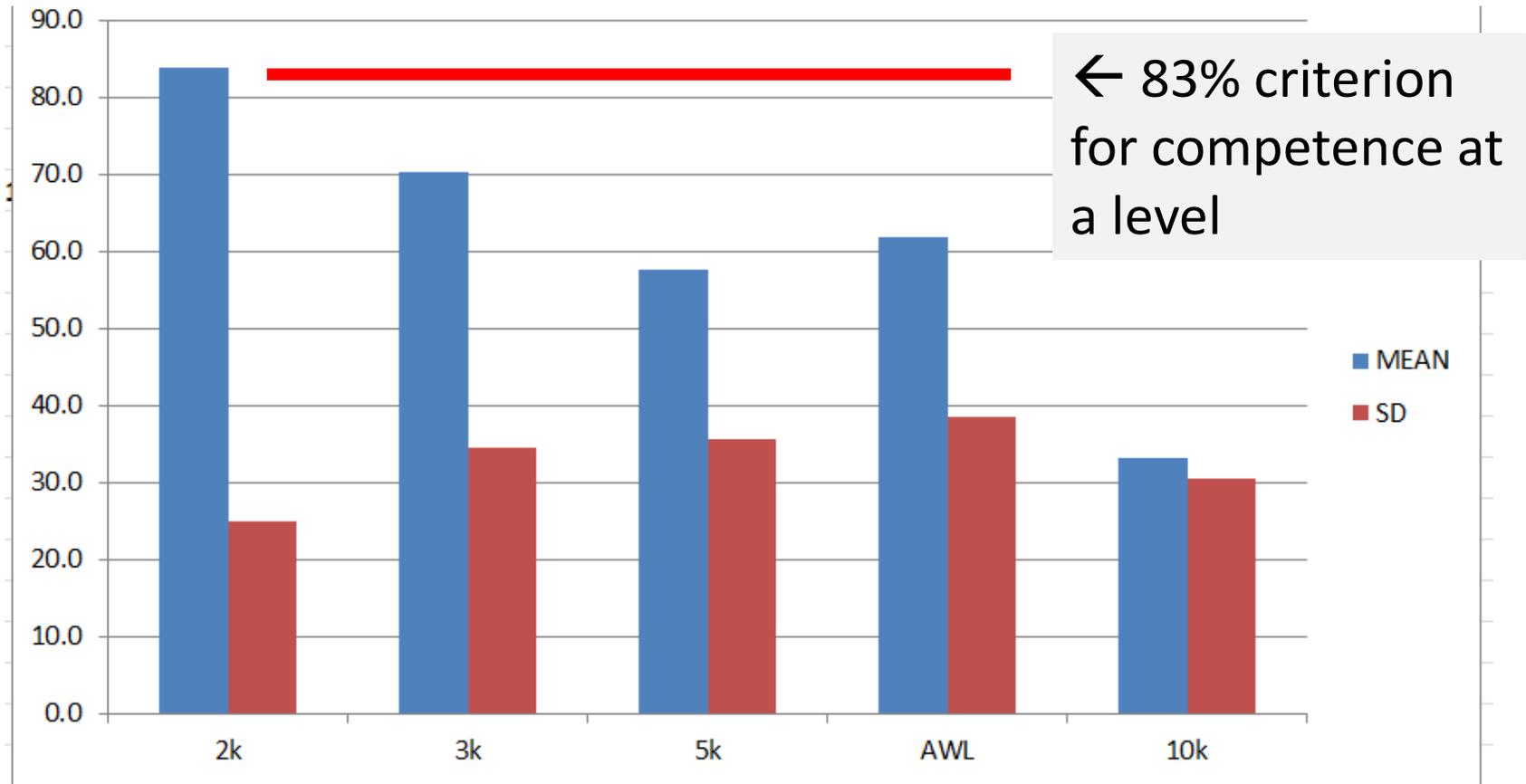
# Result VLT 2

(Schmitt, Schmitt & Clapham, 2001)

- Total scores requested : 1757
- Filter:  $K2 > 0$  + Time > 180 secs = 1,313 scores (75% of requests)
- Mean percentages by level

VLT2	2k	3k	5k	AWL	10k
MEAN	83.8	70.3	57.6	61.9	33.1
SD	24.9	34.6	35.6	38.4	30.6

# Result VLT 2



*(To be noted is that smaller tests seem to encourage more investment, VLT and especially VLT2)*

# These scores are probably too high

- An unknown number of test-takers have done these tests ‘as practice’ before doing ‘as test’
  - I.E. with answers

[Home](#) > Tests

## Vocabulary Tests

Mainly frequency-level based

In roughly chronological order

Name/ link	Language
VLT v.2 <a href="#">practice, test, results</a>	English
VST Vocab Size Tes <a href="#">practice, test, results</a>	English
<a href="#">PVST</a> Phrasal VST	English
<a href="#">CATSS</a>  Computer Adaptive Test of Size & Strength	English

# Conclusion RQ1

Three separate tests,

all properly validated in at least one setting  
with three sets of test items,  
all show identical results

There remains a general vocabulary weakness after 2k  
... despite 25 years of “the vocab revolution”

Consequences?

Knowing half of 3k = reading with about 85% coverage

- Or 3 words unknown per 20 in newspaper

# RQ2

## Are learners aware of this problem?

At some level, **Yes**

EVIDENCE FROM (1) 'GROUP LEX' ENTRIES AND (2) CONCORDANCE LOOK-UPS

- Group Lexical Database
- Ss enter word-example-meaning into a 'Group Data Base'
  - Significant investment of time
- Then sort and use in various ways
  - Make quizzes from each others' words, etc.
- This is the version that learners use **without** a teacher
  - 696 users registered in current version
  - Presently holding 2,226 records
    - Archived periodically

ALL ENTRIES:

Quiz Options>> [ Print | Interactive ] 1<sup>st</sup> 10 Any 20 None [+extras]

#	Qz	WORD	EXAMPLE	POS
1	<input type="checkbox"/>	glean	So what can we <b>glean</b> from our collection of titles	V Obtain (information) from sources, often by searching through them.
2	<input type="checkbox"/>	genre	Yet research can apply to all creative <b>genres</b> and so we should develop a more inclusive term.	N A style or category of literature.
3	<input type="checkbox"/>	fusion	All <b>fusions</b> were verified by DNA sequence analysis.	N The process of combining more things to form a new entity.
4	<input type="checkbox"/>	fragment	Besides ruins of wooden buildings, archeologists have found <b>fragments</b> of ceramic pottery.	N A small part of something.
5	<input type="checkbox"/>	focal	The <b>focal</b> theme this time was 'strategies for drought management and alternate cropping system in canal areas	ADJ Relating to the most important part of something.
6	<input type="checkbox"/>	fidelity	The result is maximum <b>fidelity</b> : a picture whose clarity, brilliance and color must be seen to be believed	N The degree of faithfulness or loyalty.
7	<input type="checkbox"/>	extrapolate	The computational model can <b>extrapolate</b> the morphogenetic movements of human organs such as the eye, heart, lung etc.	V Estimate or predict by extrapolating.
8	<input type="checkbox"/>	perennial	The debate on whether voice is teachable in L2 classroom is <b>perennial</b> .	ADJ long-lasting.
9	<input type="checkbox"/>	insight	Endurance could provide <b>insight</b> into the environmental history of the area, but there are risks.	N The capacity for deep understanding of something.
10	<input type="checkbox"/>	betterment	I believe politics and politicians are vital to the <b>betterment</b> of society.	N The improvement of something.
11	<input type="checkbox"/>	dovetail	Listening and communication skills must <b>dovetail</b> with other skills.	V Fit or cause to fit together conveniently.
12	<input type="checkbox"/>	cadre	- He trained <b>cadres</b> of engineers and built health clinics and schools in Iraq.	N A small group of people trained for a profession.
13	<input type="checkbox"/>	soluble	Not all problems are <b>soluble</b> , not all risks can be hedged at acceptable	ADJ (of a problem) able to be solved.

Group Lex M

**ADD ENTRY** X

Name

Group

New Word: Max 20

Example: Max 300

Word class

Meaning: Max 300

Add to Group Lex Now >

Arts	Muhamm...	2019.05.01 10:27
Arts	Muhamm...	2019.04.30 12:18
Arts	Muhamm...	2019.04.30 12:07
Arts	Muhamm...	2019.04.30 12:01
Arts	Muhamm...	2019.04.30 11:55
Arts	Muhamm...	2019.04.30 11:45
Arts	Muhamm...	2019.04.29 13:33
Arts	Muhamm...	2019.04.29 13:29
Arts	Muhamm...	2019.04.29 13:20
Arts	Muhamm...	2019.04.29 13:13

**truism** can be proclaimed in a manner that makes it startling. N obviously true and says nothing new or interesting.

QUIZ 1 - contexts from MUHAMMAD 27 Jul 19, 13:17 Quiz 2 -new contexts - after 100% on Quiz 1

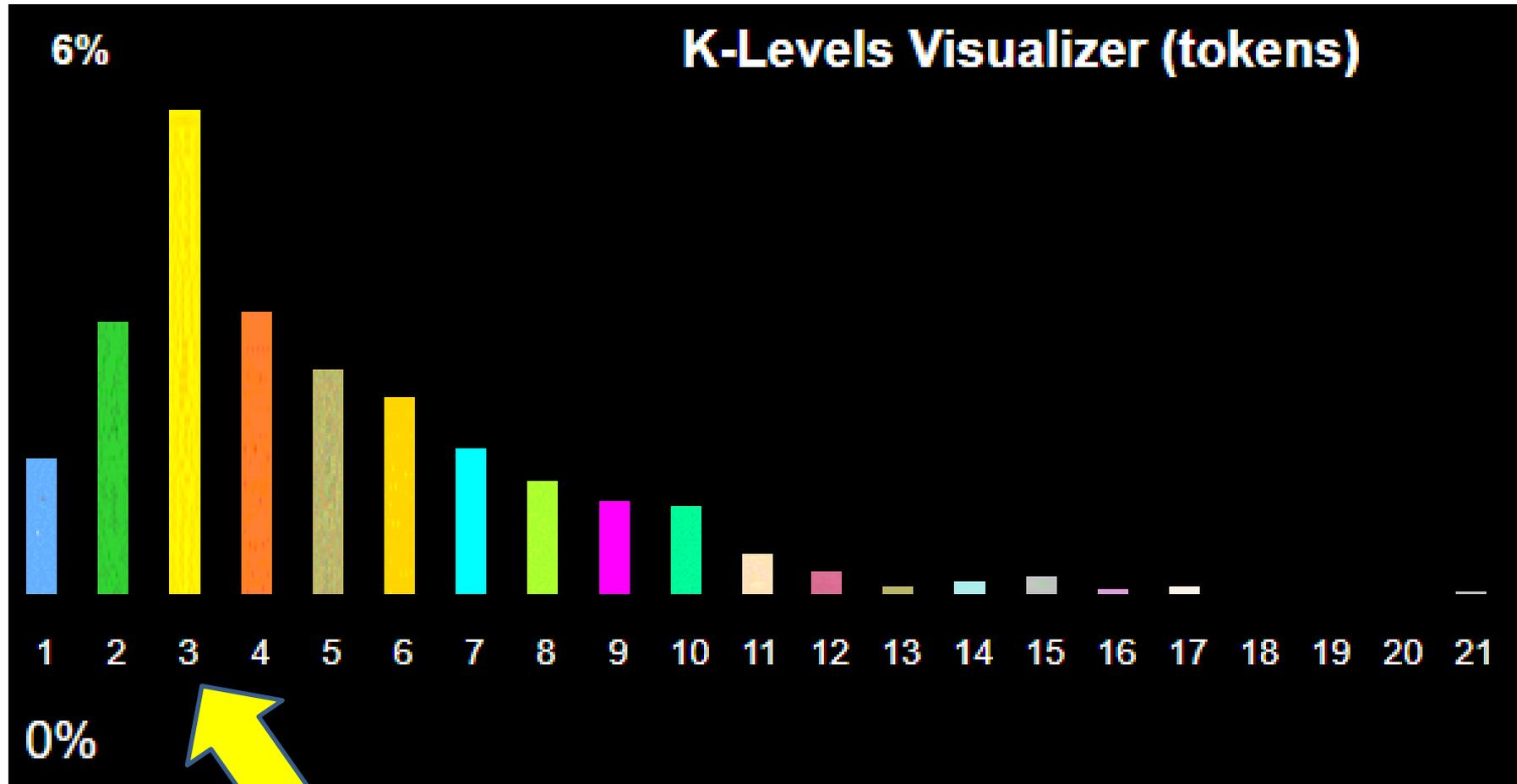
\*Check\* Questions: 12 Correct: 2 Tries: 2 Percent: 100

History >>

	NEW WORD	EXAMPLE	WORD CLASS	DEFINITION
1		Besides ruins of wooden buildings, archeologists have found _____s of ceramic pottery.	N	A small part broken off or separated from something.
2		So we _____ from our collection of titles	V	Obtain (information) from various sources, often with difficulty.
3		Yet not _____ apply to all creative _____s and so we should use an inclusive term.	N	A style or category of art, music, or literature.
4		The _____ model can _____ the morphogenetic movements of human organs such as the eye, heart, lung etc.	V	Estimate or conclude (something) by extrapolating.
5		- He _____ of engineers and built health clinics and schools in Iraq.	N	A small group of people specially trained for a particular purpose or profession.
6		Listening and communication skills must _____ with other skills.	V	Fit or cause to fit together easily and conveniently.
7		The _____ of the color was _____: a picture whose clarity, brilliance and depth were _____ to be believed	N	The degree of exactness with which something is copied or reproduced.
8		Not all _____ are _____, not all risks can be hedged at acceptable costs.	ADJ	(of a problem) able to be solved.
9		I believe that scientists and politicians are vital to the <b>betterment</b> of society.	N	The improvement of something.
10		Endemic diseases _____ provide <b>insight</b> into the environmental history of the area and the risks.	N	The capacity to gain an accurate and deep understanding of someone or something.
11		The _____ theme this time was 'strategies for drought management and alternate cropping system in canal areas	ADJ	Relating to the centre or most important part.

- betterment
- cadre
- dovetail
- extrapolate
- fidelity
- focal
- fragment
- fusion
- genre
- glean
- insight
- soluble

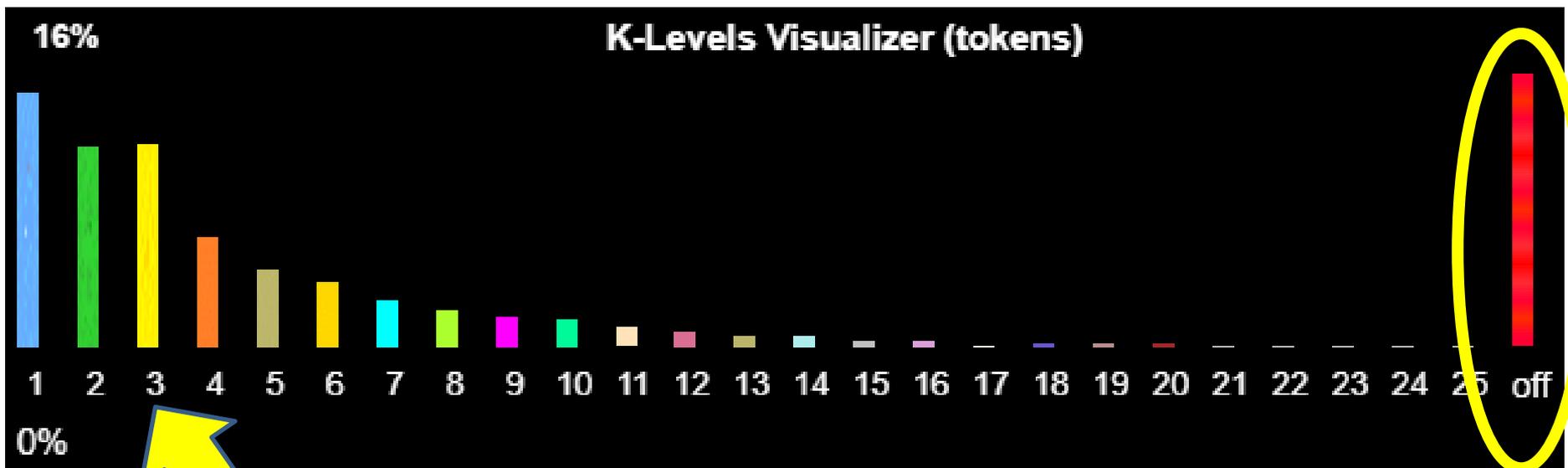
# So what words do Ss choose for Group Lex work?



# Similarly, Concordance look-ups

- Pattern = Group Lex
  - Except that entries are often phrases (include prepositions etc.)
  - And are more prone to spelling errors
    - Since Ss are not copying from an example as they are in Group Lex

For all corpora, 100 thousand look-ups



Again 3k gets double the interest of 4k

# Conclusion RQ2

- When learners get a chance, they focus directly on the 3k problem themselves
  - Working collaboratively with help of software
    - I.E. in data-driven learning

# Discussion

- Why do Ss focus on 3k?

Is it hugely more present in input than 4k?

- Not in “Great Gatsby,” whose VP →
  - 10.7% @ 3k
  - 9.1% @ 4k
- But overall, Yes, 3k items are way more frequent
  - And arguably constitute a ZPD

Freq. Level	Families (%)	Types (%)	Tokens (%)
<b>K-1 Words :</b>	1177 (40.5)	1968 (45.84)	22924 (86.6)
<b>K-2 Words :</b>	527 (18.1)	757 (17.63)	1359 (5.1)
<b>K-3 Words :</b>	311 (10.7)	390 (9.08)	607 (2.3)
<b>K-4 Words :</b>	266 (9.1)	337 (7.85)	503 (1.9)
			Coverage 95 
<b>K-5 Words :</b>	162 (5.6)	186 (4.33)	258 (1.0)
<b>K-6 Words :</b>	121 (4.2)	139 (3.24)	184 (0.7)
<b>K-7 Words :</b>	97 (3.3)	107 (2.49)	125 (0.5)
			Coverage 98
<b>K-8 Words :</b>	64 (2.2)	72 (1.68)	98 (0.4)
<b>K-9 Words :</b>	53 (1.8)	56 (1.30)	66 (0.2)
<b>K-10 Words :</b>	28 (1.0)	31 (0.72)	33 (0.1)
<b>K-11 Words :</b>	36 (1.2)	37 (0.86)	41 (0.2)
<b>K-12 Words :</b>	18 (0.6)	19 (0.44)	19 (0.1)
<b>K-13 Words :</b>	10 (0.3)	10 (0.23)	11 (0.0)
<b>K-14 Words :</b>	10 (0.3)	10 (0.23)	12 (0.0)
<b>K-15 Words :</b>	8 (0.3)	9 (0.21)	9 (0.0)

# 3k as a ZPD

Ss try to learn words they “sort of know”

Ss with 2k are likely to “sort of know” many 3k words

Much more frequent than 4k in “English at large”

>> In LOB corpus 3k is more than double 4k in coverage

(Nation, 2006)

TABLE 1

Tokens, types, and families at each of the 14 BNC word-family levels in the LOB corpus

Word list (1,000)	Token (%)	Types (%)	Families
1	78,944 (77.86)	4,487 (10.1)	998
2	23,477 (8.7)	4,131 (9.34)	998
3	37,511 (3.70)	3,239 (7.32)	998
4	18,198 (1.79)	2,683 (6.07)	998
5	10,495 (1.04)	2,226 (5.03)	969
6	7,080 (0.70)	1,789 (4.04)	928
7	6,633 (0.65)	1,542 (3.49)	887
8	4,096 (0.40)	1,382 (3.12)	836
9	3,217 (0.32)	1,118 (2.53)	734
10	3,228 (0.32)	1,025 (2.32)	719
11	1,609 (0.16)	753 (1.70)	587
12	1,434 (0.14)	646 (1.46)	498
13	1,211 (0.12)	529 (1.20)	441
14	973 (0.10)	339 (0.77)	288
15	18,519 (1.83)	2,878 (6.51)	2,878
Not in the lists	26,821 (2.65)	15,463 (34.96)	?????*
Total	1,013,9469	44,230	13,747

\* The RANGE program is not able to calculate families for words not in the lists.

# A name for it

- I propose this phenomenon be called the '3<sup>rd</sup>-K Slump'
- To mirror Chall & Jacobson's '4th-Grade Slump'
  - To which it is strongly akin

# *A name for it*

- **4th-Grade slump:**

- Gr 1-3 → “learning to read”

- I.E., to decode written language already understood in speech
    - Almost all kids get this

- Grade 4+ → “reading to learn”

- New vocab, vocab-from-vocab, new concepts
    - **Not** all kids get this
    - Teaching becomes crucial

- **3rd-K Slump :**

- 1k-2k is got for free, from input

- 3k+ → Teaching becomes crucial

# What would this 'teaching' look like?

Or 'provision for learning'?

Ideas for Post-2k FFI were once popular, though less now for some reason

- E.g., AWL is not updated
- Exception - interesting work on 'Mid-Frequency Vocabulary' by N & D Schmitt

How about these 2 ideas:

- 1. Flashcards for 3k divided in 100-fam. sets

Teamed with...

12:47 PM 72%

luxtutor.ca

LEXUTOR Down Load

MOBILE  
FLASHCARDS

ID

100-Family Sets x Freq.

K0-1	K1-2	K2-3	K3-4	K4-5
C1	C1	C1	C1	C1
C2	C2	C2	C2	C2
C3	C3	C3	C3	C3
C4	C4	C4	C4	C4
C5	C5	C5	C5	C5
C6	C6	C6	C6	C6
C7	C7	C7	C7	C7
C8	C8	C8	C8	C8
C9	C9	C9	C9	C9
C10	C10	C10	C10	C10

From BNC-COCA

[Go French](#)

[www.lexutor.ca](http://www.lexutor.ca)

# What would this 'teaching' look like?

Or 'provision for learning'?

- 1. Flashcards for 3k divided in 100-fam. sets

Teamed with...

- 2. Paul's 'Mid-Frequency Graded Readers'
  - E.g., "Alice in Wonderland" → (27,500 wds, similar Gatsby)
  - 97 K3 word-fams are met in 95% known-word contexts
    - (Assuming knowledge of K2)
    - One unknown word per 20
  - (But again only 6 fams > 5x Thus need to supplement 'input')

Freq. Level	Families (%)	Types (%)
K-1 Words :	878 (57.3)	1487 (60.50)
K-2 Words :	327 (21.3)	458 (18.63)
		Coverage 95
K-3 Words :	97 (6.3)	120 (4.88)
K-4 Words :	119 (7.8)	148 (6.02)
K-5 Words :	73 (4.8)	95 (3.86)
		Coverage 98
K-6 Words :	8 (0.5)	10 (0.41)
K-7 Words :	12 (0.8)	15 (0.61)
K-8 Words :	5 (0.3)	9 (0.37)
K-9 Words :		
K-10 Words :	2 (0.1)	3 (0.12)
K-11 Words :	3 (0.2)	4 (0.16)
K-12 Words :	2 (0.1)	2 (0.08)
K-13 Words :	1 (0.1)	1 (0.04)

# References

- Chall, J., Jacobs, V., Baldwin, L., (2009). *The reading crisis: Why poor children fall behind*. Harvard University Press.
- Cobb, T. (2007). [Computing the vocabulary demands of L2 reading](#). *Language Learning & Technology* 11(3), 38-63.
- Cobb, T. (2008). [What the reading rate research does not show: Response to McQuillan & Krashen](#). *Language Learning & Technology* 12(1), 109-114.
- Cobb, T. (2016) [Numbers or numerology? A response to Nation \(2014\) and McQuillan \(2016\)](#). *Reading in a Foreign Language*, 28 (2), 299-304.
- Laufer, B. (1999). *Task effect on instructed vocabulary learning: The hypothesis of 'involvement'*. AILA proceedings.
- Nation, P. (2006). How big a vocabulary is needed for reading and listening? *Canadian Modern Language Review* 63 (1), 59-82.
- Nation, P. (No date.) [Mid-Frequency Readers](#).