How and Why WordEngine Improves TOEIC and TOEFL Scores

JALT CALL

June 8, 2014

Presenter: Guy Cihi

info@lexxica.co.jp

Copyright 2014 Lexxica

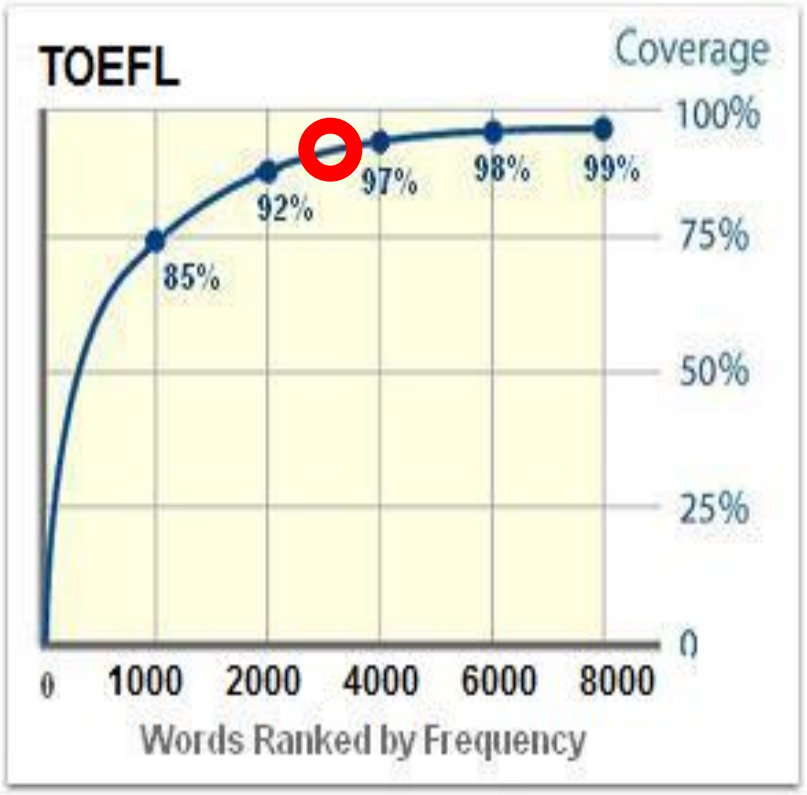
Word Engine

Presentation Outline

- **TOEIC/TOEFL High frequency vocabulary**
- **WordEngine system introduction**
- **TOEIC and TOEFL Score Data Showing WordEngine Increases Test Scores**
- **How and why scores go up**

Copyright 2014 Lexxica

Vocabulary - how much is enough?



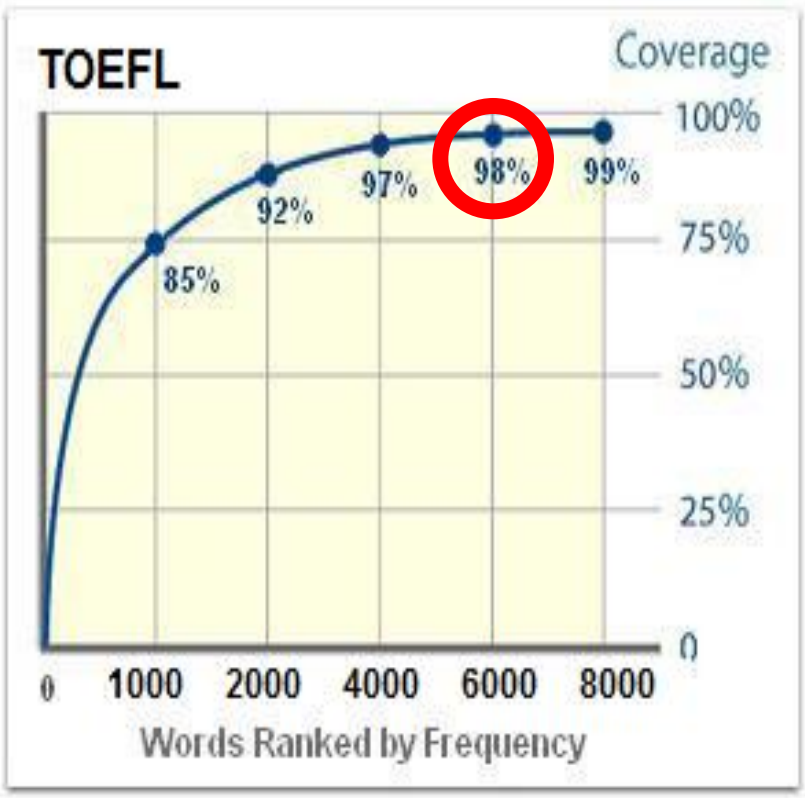
95% coverage of vocabulary is the tipping point for successful comprehension.

Laufer (1989)

"Tipping point = more readers than non-readers at 95%."

Copyright 2014 Lexxica

Vocabulary - how much is enough?



98% coverage is needed for both adequate comprehension and to be able to guess or skip over unknown words based on context.

Hu and Nation (2000)

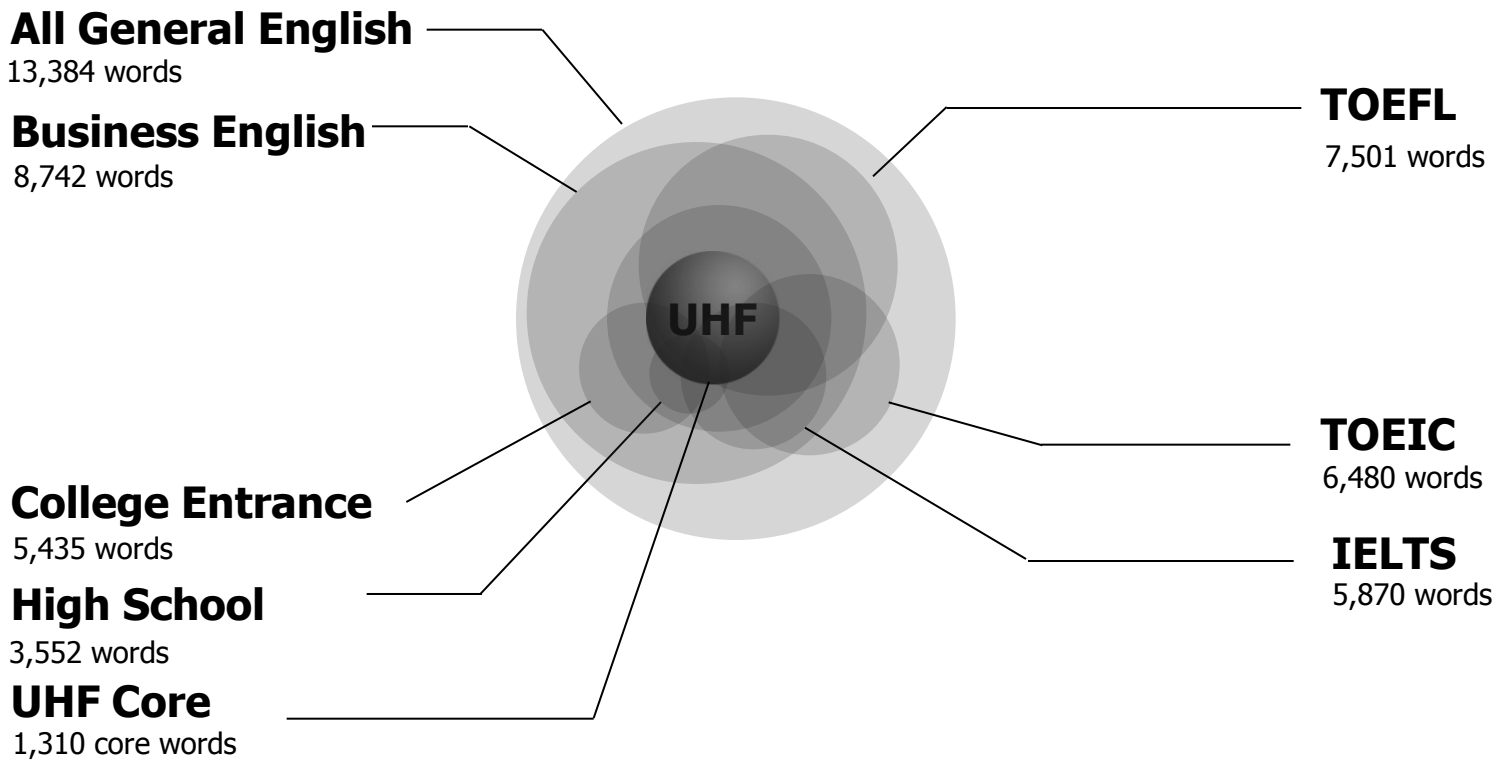
"WordEngine provides 99% coverage."

Copyright 2014 Lexxica

Word Engine

Domain-specific HFV

Lexxica identifies the specific words that cover 99% of each subject.



"The specific words are quite different in each domain."

Copyright 2014 Lexxica

TOEIC Corpus Analysis

1,250,000 total words

14,652 different words

457 UHF words constitute 80% of all occurrences.

1,715 SHF words constitute 95% of all occurrences.

6,480 HF words constitute 99% of all occurrences.

"Based on scans of actual and practice tests 50/50."

Copyright 2014 Lexxica

TOEFL Corpus Analysis

1,250,000 total words

16,736 different words

655 UHF words constitute 80% of all occurrences.

2,875 SHF words constitute 95% of all occurrences.

7,501 HF words constitute 99% of all occurrences.

"Based on scans of actual and practice tests 50/50."

Copyright 2014 Lexxica

Word Engine

"Absolutely essential!"

TOEIC	TOEFL
Words	Words
457	655

0% to 80% = Ultra High Frequency Range

"1,800 to 2,300 total item occurrences in all parts of every test are words from within the Ultra High Frequency range."

Copyright 2014 Lexxica

Word Engine

"The difference between okay and good"

TOEIC Words	TOEFL Words
457	655
1258	2220

91% to 95% = Super High Frequency Range

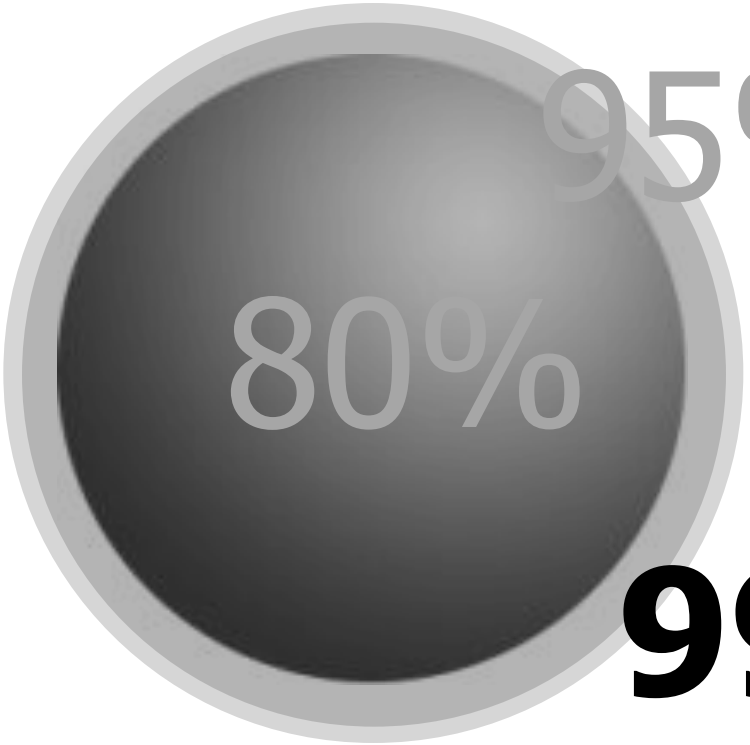
"80 to 200 total item occurrences in all parts of every test are words from within the Super High Frequency range."

Copyright 2014 Lexxica

Word Engine

"The difference between good and great"

TOEIC Words	TOEFL Words
457	655
1258	2220
4765	**4626**

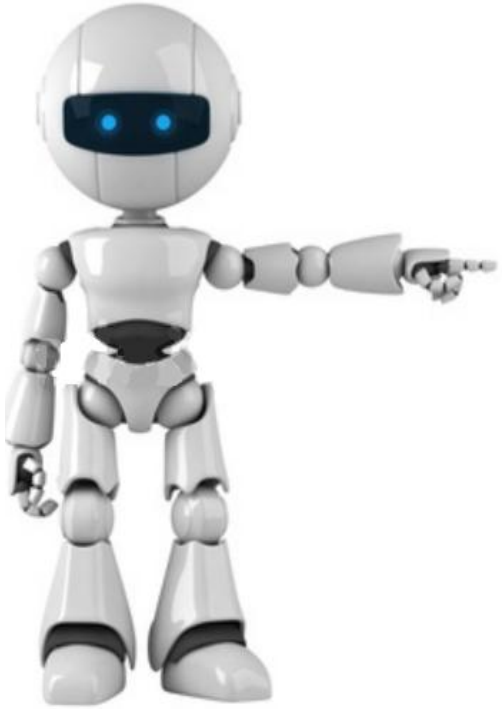


96% to 99% = High Frequency Range

"5 to 75 total item occurrences in all parts of every test are words from among the High Frequency range."

Copyright 2014 Lexxica

ETS Secrets #2 and #3



1/3 of the SHF words in all parts of TOEIC and TOEFL are not SHF in General English.

95% of test questions are recycled.
5% are new questions that are in the process of being measured for difficulty.

Copyright 2014 Lexxica

What does it mean for learners?

EFL students can't learn the words they need because they aren't in their study and reading materials.

"Because their materials are simplified."

Copyright 2014 Lexxica

ETS's advice for scoring higher on TOEIC and TOEFL is to read authentic texts.

"Graded readers can't help because the vocabulary is simplified."

Copyright 2014 Lexxica

How much authentic text is needed?

Based on incidence of occurrence research by Rob Waring, they'll need to read 12,500 hours of authentic text in order to meet the lower frequency test words often enough to know them.

A statistical analysis of the number of English words you need to meet (at given recurrence rates) to 'learn' that number of words

A	B	C (= 100 / B)	D (= x times C)				E (= D / Book length)				
Word rank	Percentage of general English that this word covers	Number of running words needed to be met to meet all these words once	Volume of text you need to read to meet the words at these recurrence rates				Number of books to cover this volume given these recurrence rates				
			5 times	10 times	20 times	50 times	Book length	5 times	10 times	20 times	50 times
1st most frequent (*the*)	5.83898%	17 (1)	86	171	343	856	4,500	0.0	0.0	0.1	0.2
2nd most frequent (*be*)	5.12332%	20	98	195	390	976	4,500	0.0	0.0	0.1	0.2
25th (*as*)	0.44382%	225	1,127	2,253	4,506	11,266	4,500	0.3	0.5	1.0	2.5
50th (*like*)	0.24109%	415	2,074	4,148	8,296	20,739	4,500	0.5	0.9	1.8	4.6
100th (*hear*)	0.10505%	952	4,759	9,519	19,038	47,595	4,500	1.1	2.1	4.2	10.6
500th (*present*)	0.02477%	4,037	20,183	40,366	80,732 (4)	201,829	4,500	4.5	9.0	17.9	44.9
1000th (*blood*)	0.01172%	8,533 (3)	42,665	85,329	170,658	426,645	10,000	4.3	8.5	17.1	42.7
1500th (*intent*)	0.00677%	14,773	73,864	147,727	295,455	738,636	15,000	4.9	9.8	19.7	49.2
2000th (*stumble*)	0.00432% (2)	23,103	115,625	231,250	462,500	1,156,250	20,000	5.8	11.6	23.1	57.8
3000th (*sergeant*)	0.00211%	47,343	236,713	473,425	946,850	2,367,126	30,000	7.9	15.8	31.6	78.9
5000th (*satellite*)	0.00076%	132,143	660,714	1,321,429	2,642,857	6,607,143	80,000	8.3	16.5	33.0	82.6
10,000th (*relativity*)	0.00016%	632,000	3,164,474	6,328,947	12,657,895	31,644,733	80,000	39.6	79.1 (5)	158.2	395.6

Examples: (1) The most frequent word in English (*the*) covers 5.839% of any general English text (i.e. it occurs once in every 17 words).
(2) The 2000th most frequent word in English covers 0.00432% of any general English text (and occurs once every 23,103 words).
(3) To meet all the 1000 most frequent words in English once, you'd need to read 8,533 words.
(4) To meet all the 500 most frequent words in English 20 times, you'd need to read 80,732 words.
(5) To meet all the 10,000 most frequent words in English 10 times, you'd need to read 79.1 books that are 80,000 words long.

© R. Waring 2007

Copyright 2014 Lexxica

Reading at 100 authentic words per minute it will require…

3 hours each day for 8.5 years

"Sorry, no holidays."

Copyright 2014 Lexxica

Reading at 100 graded reader words per minute it will require until…

Never

"Because graded readers are simplified."

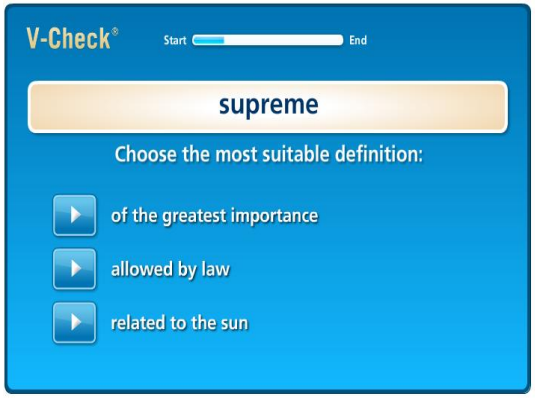
Copyright 2014 Lexxica

Introducing

Word Engine

Main Attributes - 1

WordEngine starts with an ability test called V-Check



Part 1: LDT Questions
Quick estimate

Part 2: Definitions
Refined estimate

Part 3: Actual Study
Confirmed score

Copyright 2014 Lexxica

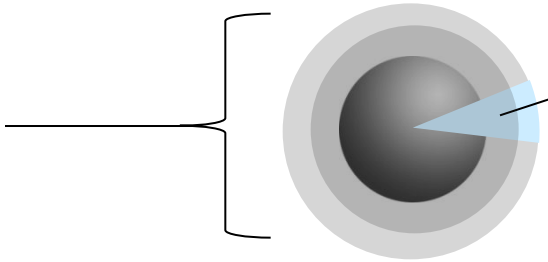
Word Engine

High Speed Vocabulary Learning System

WordEngine identifies each student's missing UHF, SHF and HF words, and then teaches them in the order of importance.

TOEIC

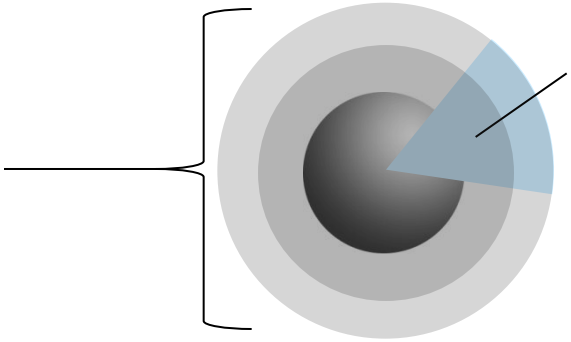
6480 words required



1320 missing

TOEFL

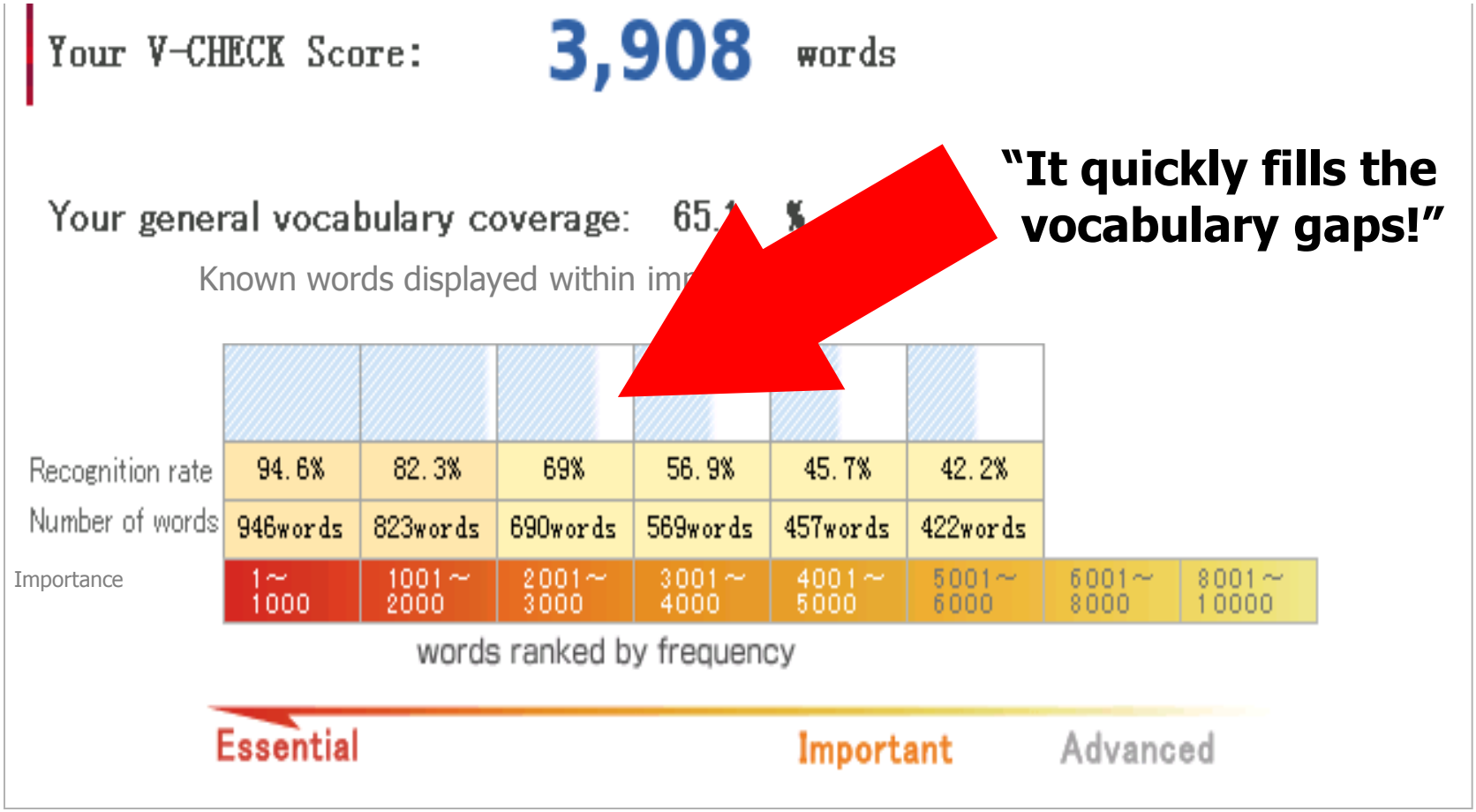
7501 words required



2487 missing

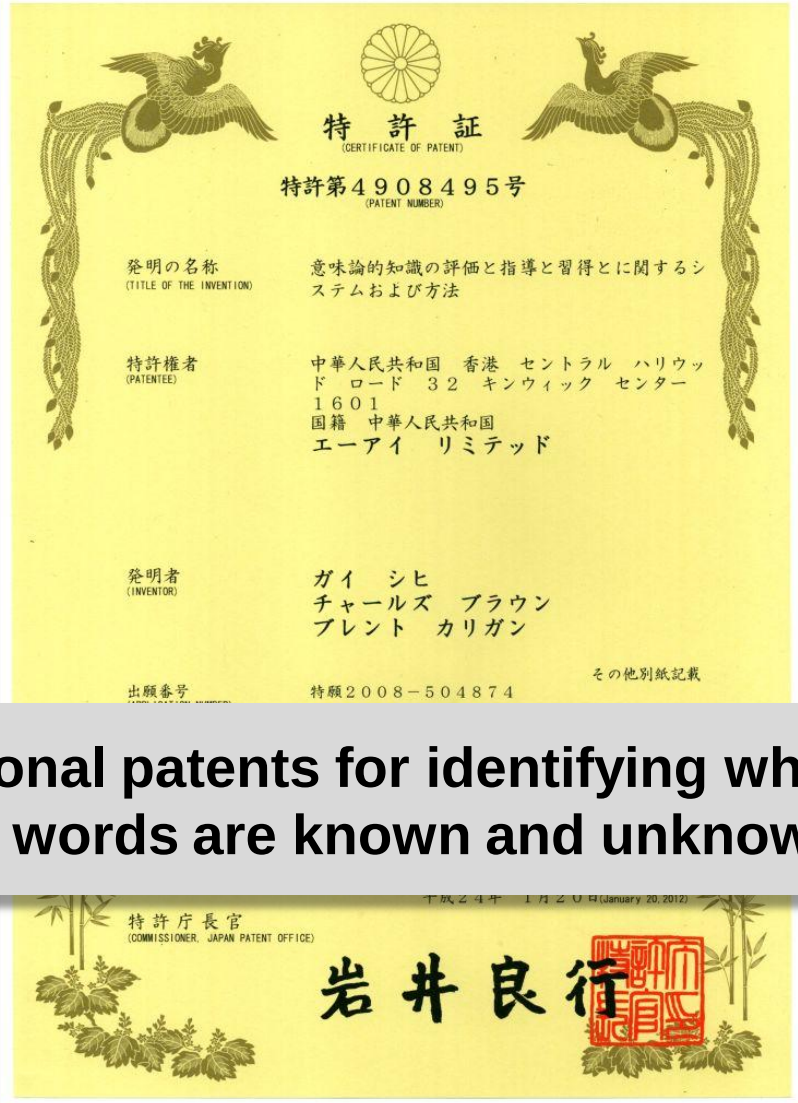
Copyright 2014 Lexxica

Word Engine



Unknown Ultra High and Super High Frequency words are taught first.

Copyright 2014 Lexxica



特　許　証
(CERTIFICATE OF PATENT)

特許第４９０８４９５号
(PATENT NUMBER)

発明の名称
(TITLE OF THE INVENTION)
意味論的知識の評価と指導と習得とに関するシステムおよび方法

特許権者
(PATENTEE)
中華人民共和国　香港　セントラル　ハリウッド　ロード　３２　キンウィック　センター　１６０１
国籍　中華人民共和国
エーアイ　リミテッド

発明者
(INVENTOR)
ガイ　シヒ
チャールズ　ブラウン
ブレント　カリガン

その他別紙記載

出願番号
特願２００８－５０４８７４

特許庁長官
(COMMISSIONER, JAPAN PATENT OFFICE)
岩井良行

One-parameter Logistic Model

Using the Rasch model, we can calculate the probability of an examinee answering an item correctly with the following formula:

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}$$

where $P_i(\theta)$ is the probability of a randomly chosen examinee with ability θ answering item i correctly.

e is the base of natural logarithms (2.718)

θ is the person ability measured in logits

b_i is the difficulty parameter of the item measured in logits

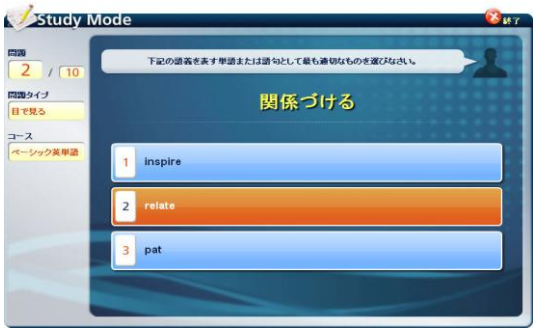
International patents for identifying which specific words are known and unknown.

Copyright 2014 Lexxica

Word Engine

Main Attributes - 2

High speed games for computers and mobile devices



The <u>average</u> study pace is 300 words per hour

Copyright 2014 Lexxica

Current Perspectives on Teaching the Four Skills

ELI HINKEL
Seattle University
Seattle, Washington, United States

"A vast body of research has established that direct and explicit teaching represents the most effective and efficient means of teaching vocabulary, and incidental learning through reading leads to significantly lower rates of vocabulary retention"

Comprehension problems, at all ages, are directly associated with vocabulary naming-speed deficits; the lack of automaticity.

Ackerman, & Dykman (1993); Badian (1995, 1996); Bowers Steffy & Tate (1988); Denckla, & Rudel (1976); Grigorenko (1997); Lovett (1992, 1995); McBride, & Manis (1996)

"Problem solved!"

Copyright 2014 Lexxica

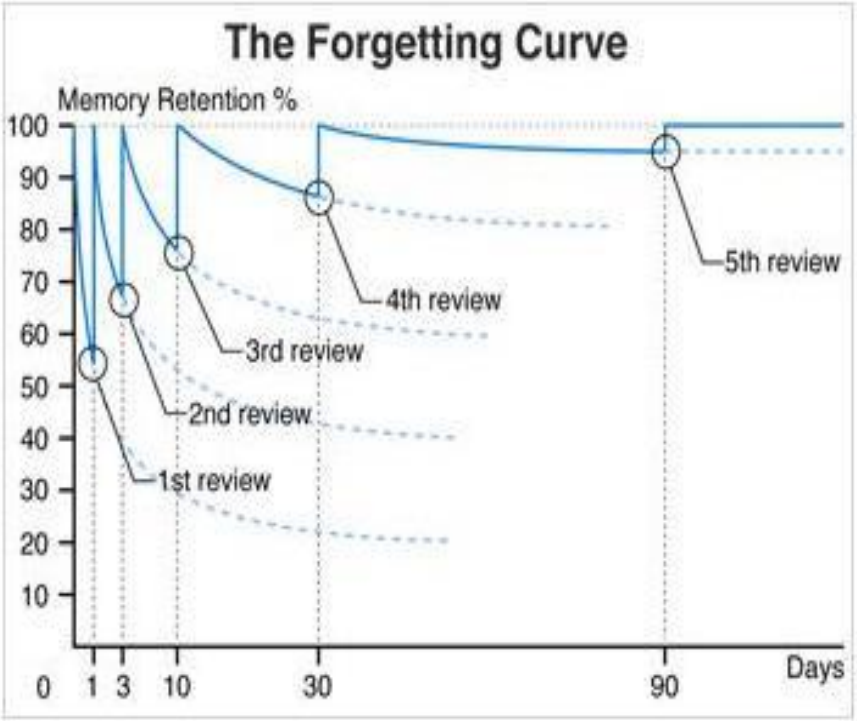
Word Engine

Main Attributes - 3

Inculcates long-term memory retention

Spaced Repetition

A proven way to transfer knowledge from short-term to long-term memory.



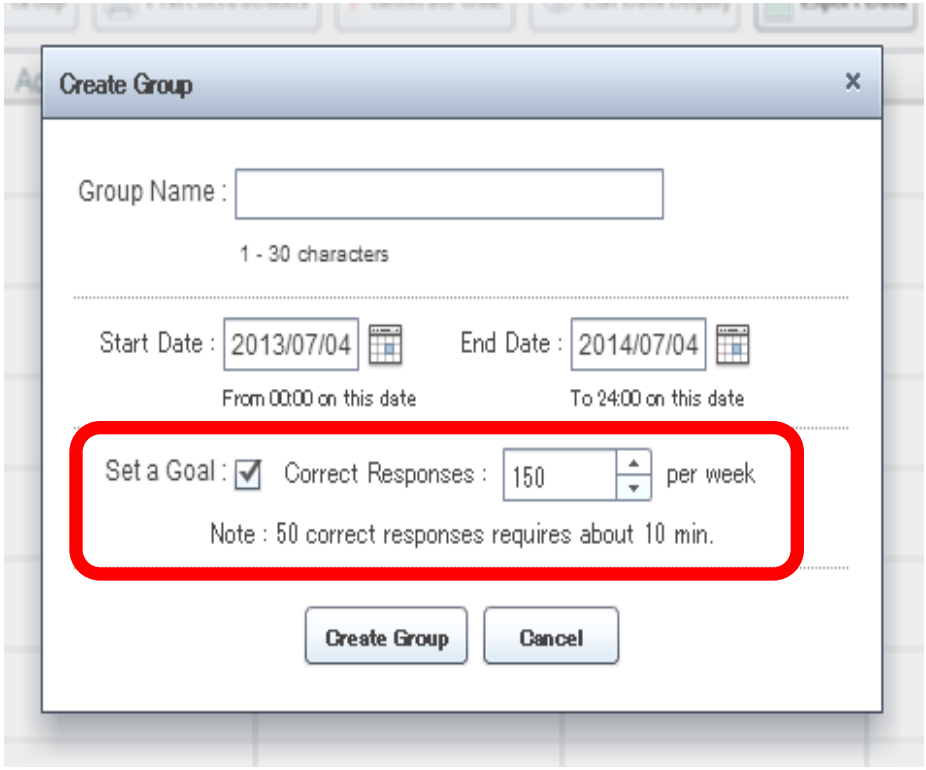
The Forgetting Curve. Dr. Hermann Ebbinghaus

Copyright 2014 Lexxica

Word Engine

Main Attributes - 4

Teacher sets the weekly study goal – set and forget!



Copyright 2014 Lexxica

Word Engine

Main Attributes - 5

Teacher receives automatic weekly email progress reports

Goal hit!

Goal miss

OC-212
Weekly Goal 300 Correct Responses (CR)

Goal Hit: Name: Login ID: Minutes: CR

Yes: Kitoh Ayami: g1211772: 94: 508
Yes: fujita takahiro: g1215219: 45: 324
Yes: Nozue Shuhei: g1214599: 56: 314
Yes: Yamano Moeka: g1216425: 47: 309
Yes: kondo masamitsu: g1212249: 45: 306
No: Oda Kenta: g1211196: 36: 184
No: nakhara asato: g1214166: 29: 151
No: kanazawa kenta: g1211402: 27: 141
No: yamamoto akio: g1216452: 20: 141

Copyright 2014 Lexxica

Word Engine

Main Attributes - 6

An automatic quiz generator produces unique quizzes for each student yet with a common answer key.

Word Engine Quiz
Group : E2 2013
Course : All Courses
Student : 20202 / **** *********

These are words that you recently learned with the WordEngine. Write the letter of the correct word in the space at right.

1. **one who seeks to be elected to a position**
   a. candidate b. bundle c. loaf d. adverb 1.____

2. **the process of making products**
   a. goal b. essay c. industry d. reform 2.____

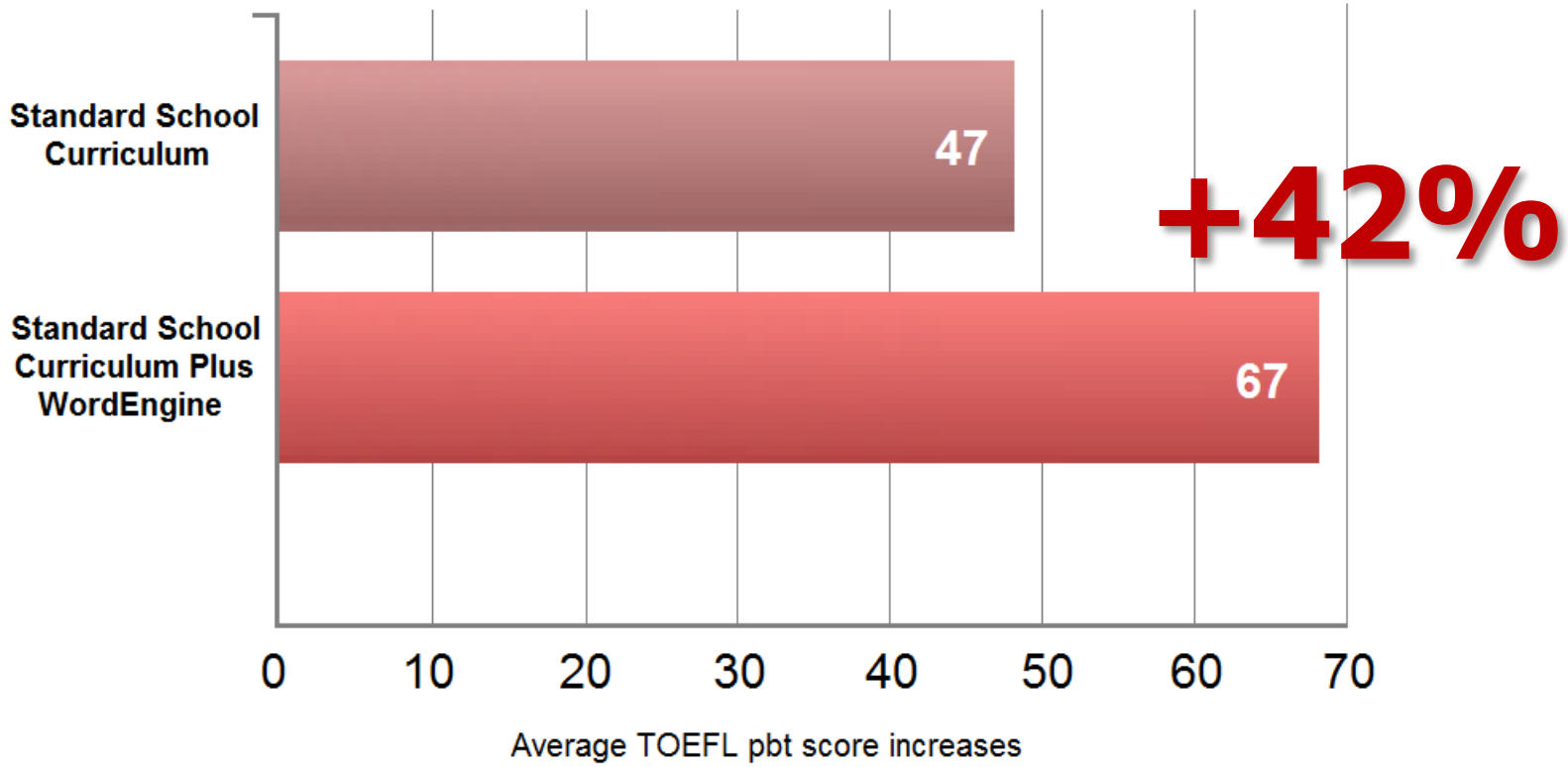
3. **to identify something or someone**
   a. doubt b. assure c. recognize d. avoid 3.____

1. a
2. c
3. c

Copyright 2014 Lexxica

Actual TOEFL pbt score improvements

for students who scored below 400 on their first TOEFL

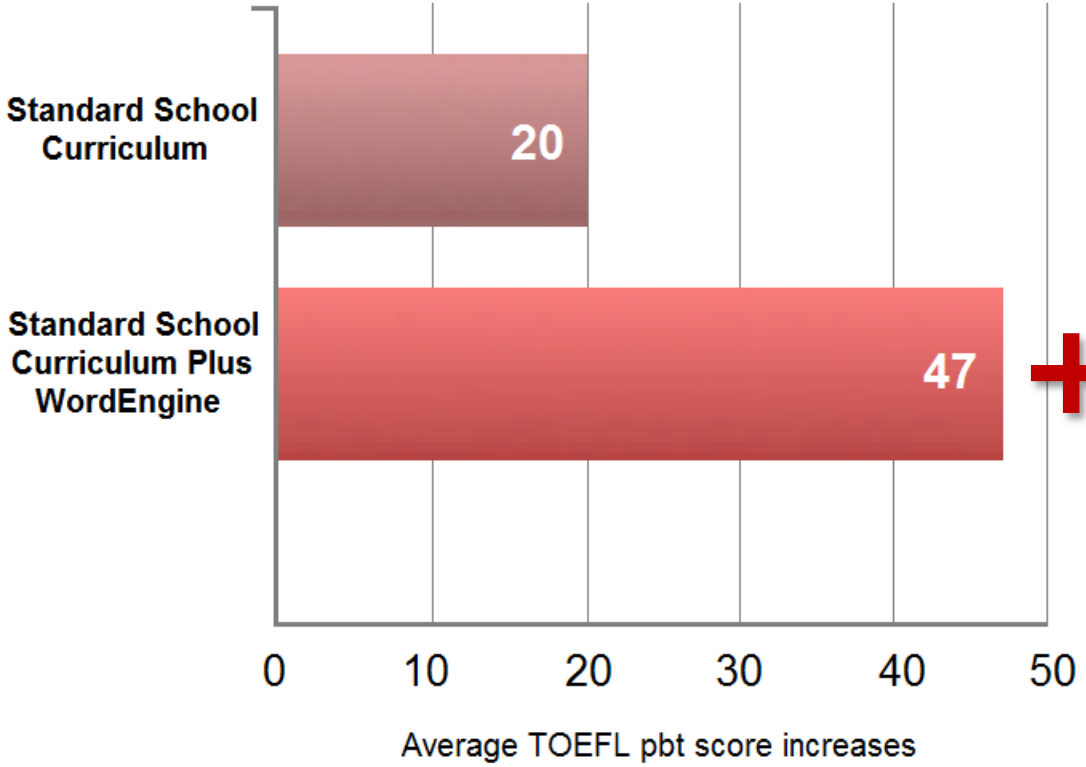


Source: Dr. David Coulson, University of Nigata Prefecture 2011

Copyright 2014 Lexxica

Actual TOEFL pbt score improvements

for students who scored above 400 on their first TOEFL



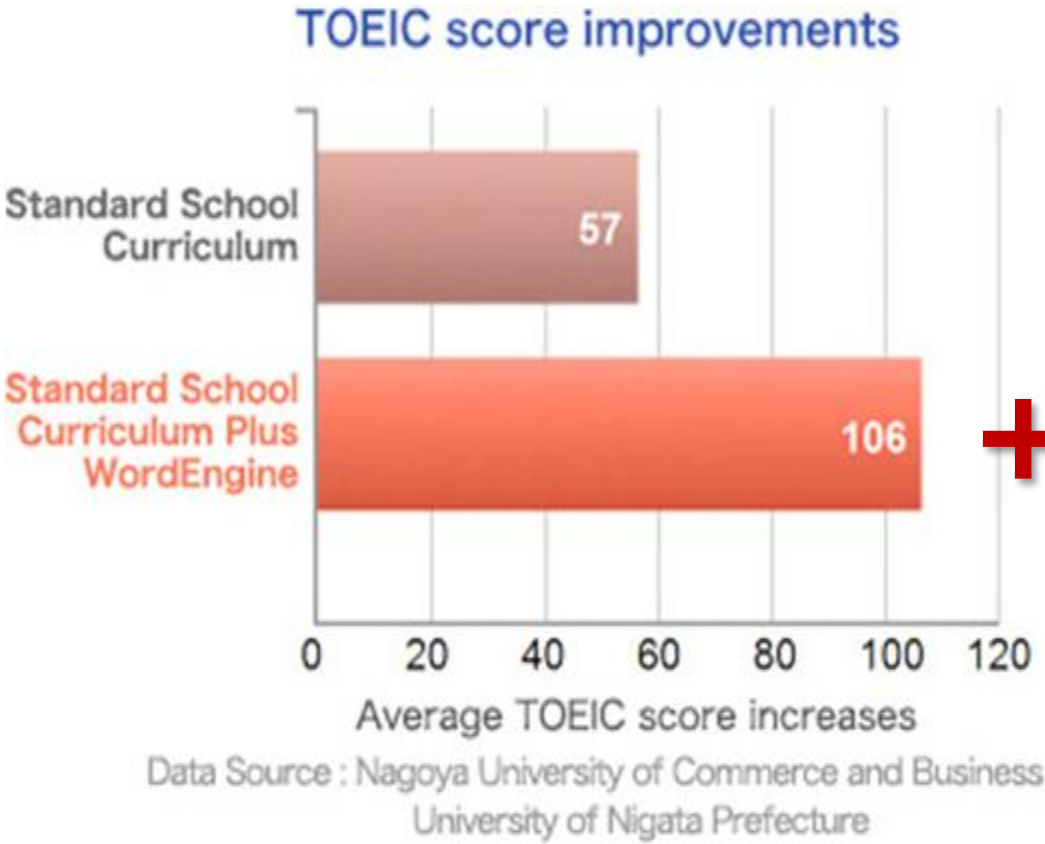
+135%

Source: Dr. David Coulson, University of Nigata Prefecture 2011

Copyright 2014 Lexxica

Word Engine

Actual TOEIC Results

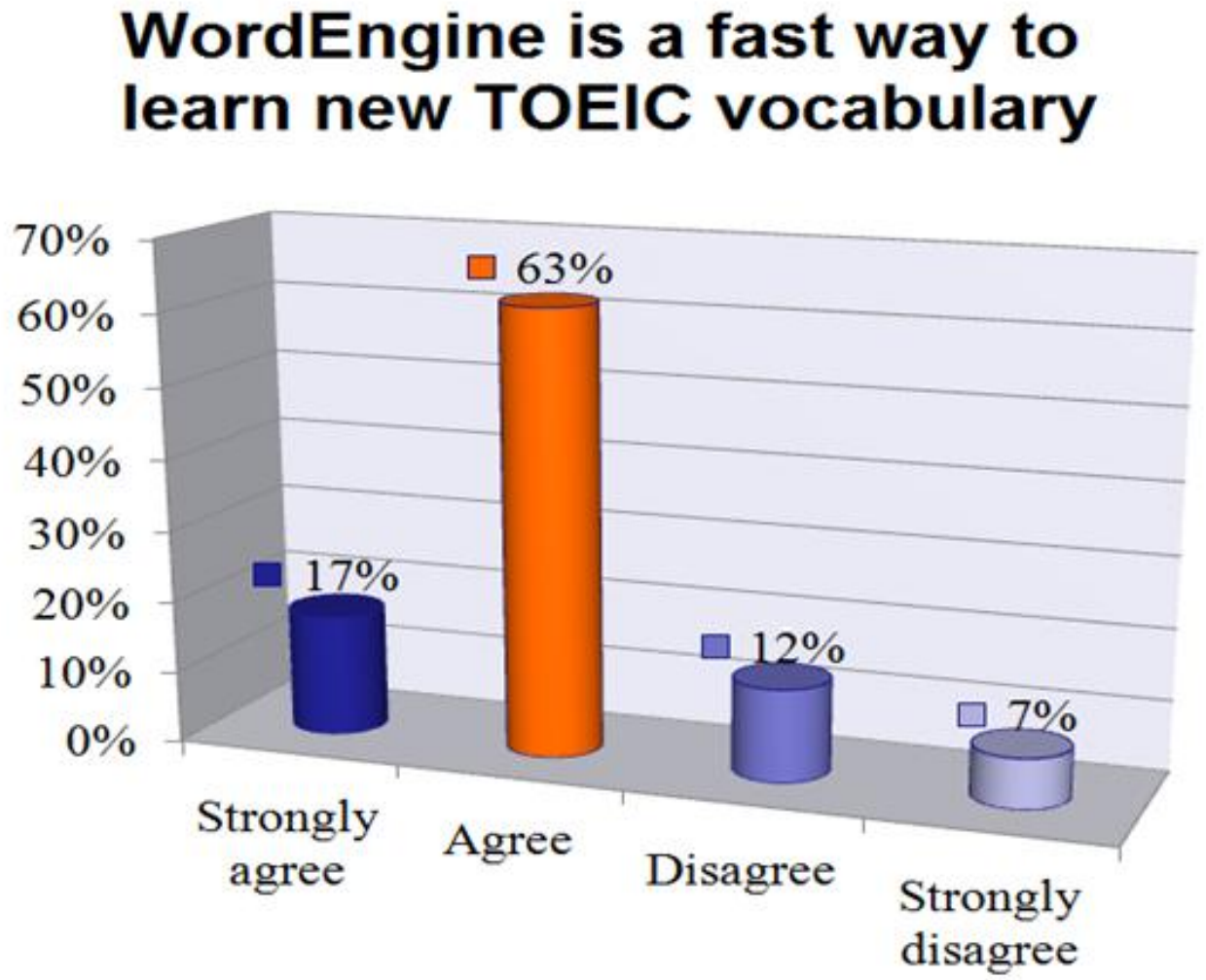


+86%

Copyright 2014 Lexxica

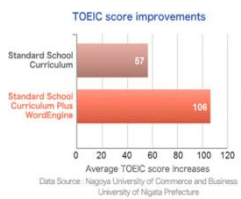
WordEngine™ ワードエンジン

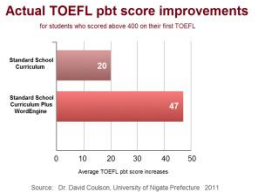
Students like WordEngine because it's fast, and they know it works



Copyright 2014 Lexxica

Word Engine





Scores go up fast because WordEngine teaches the specific words that occur most often on these tests.

Scores go up fast because students process <u>all</u> English faster after regularly exercising their lexical recognition skills with WordEngine.

Copyright 2014 Lexxica

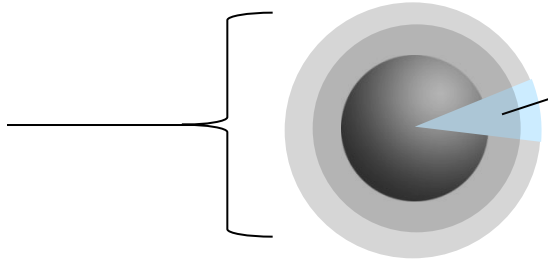
Word Engine

High Speed Vocabulary Learning System

Learn your missing UHF, SHF and HF words

TOEIC

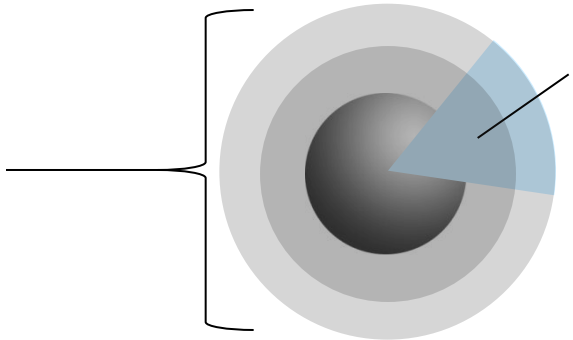
6480 words required



1320 missing

TOEFL

7501 words required



2487 missing

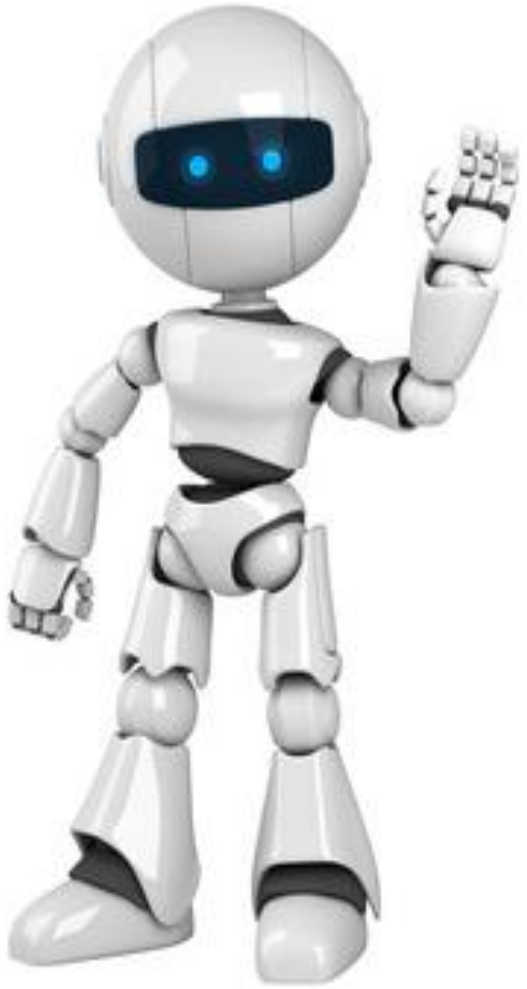
WordEngine teaches the most important words <u>that you don't already know</u>.

Copyright 2014 Lexxica

Contact Lexxica to start a trial program at your school.

"Costs less than 30¢ a week!"

Lexxica R&D
2-7-8 Shibuya 5F
Shibuya-ku, Tokyo 150-0002
info@lexxica.co.jp



Copyright 2014 Lexxica