

A Word Frequency Index of Word Familiarity: A Tool for Item Writers

Joel P. Wiesen, Ph.D.

jwiesen@appliedpersonnelresearch.com

IPAC 2011 Conference

Washington, D.C.

July 19, 2011

Wiesen (2011), International Personnel Assessment Council Conference

1

Conclusion

- Topic: New, free web tool to identify words uncommonly found in reading material from various cultural groups
- <http://wordsly.com>
- Call for research partners

Wiesen (2011), International Personnel Assessment Council Conference

2

Overview of This Presentation

- Challenge
- Approach
- Status
- Preliminary Findings
- Plans for the Future

Wiesen (2011), International Personnel Assessment Council Conference

3

Challenge

- Writing questions clearly for all audiences
- Challenges
 - Differential word knowledge can be a source of construct irrelevant variance
(1999 Standards for Educational and Psychological Tests)
 - *g* is measured indirectly by learned verbal and quantitative skills
(SIOP, 2011)

Wiesen (2011), International Personnel Assessment Council Conference

4

Examples of Readability Issues

- Unnecessarily difficult words used in test questions
 - “Aligned” in a spatial ability question for Firefighter
- Non-testing usage
 - Airline safety instructions words:
 - “Placard” “Lavatories”
 - “Obstruction” “Criteria”
 - “Corresponding” “Impart”

Wiesen (2011), International Personnel Assessment Council Conference

5

Traditional Readability Formulas

- Focus on
 - Number of words per sentence
 - Number of syllables per word
- Crude measures
 - Assume longer is harder
 - Yet have utility
 - Formulas can be “defeated”

Wiesen (2011), International Personnel Assessment Council Conference

6

Develop a New Approach

- Readability formulas **pre-date computers**
- Use current technology

Wiesen (2011), International Personnel Assessment Council Conference

7

Content Validity Approach

- Problem
 - What words are likely to be familiar to various cultural groups?
 - What words are likely to be equally familiar to various cultural groups?
- Approach
 - Find and analyze examples of reading material from various cultural groups.

Wiesen (2011), International Personnel Assessment Council Conference

8

Consider Various Word Sources

- Newspapers
- Magazines
- High School Reading Lists

Wiesen (2011), International Personnel Assessment Council Conference

9

Magazines Used

- Ebony
- Jet
- Essence
- Sister
- Time
- Parent
- This list will be expanded

Wiesen (2011), International Personnel Assessment Council Conference

10

Newspapers Used

- New York Times
- This list will be expanded

Wiesen (2011), International Personnel Assessment Council Conference

11

HS Reading Lists Used

- Atlanta
- Baltimore
- Jackson MS
- Miami
- New York City
- Washington D.C.
- This list will be expanded

Wiesen (2011), International Personnel Assessment Council Conference

12

HS Books in Database So Far

- Achebe, *Things Fall Apart*
- Albom, *Tuesdays With Morrie*
- Alexie, *The Absolutely True Diary Of A Part Time Indian*
- Alvarez, *The Garcia Girls Lost Their Accent*
- Bloor, *Tangerine*
- Cormier, *T Chocolate War*
- DiCamillo, *Because of Winn Dixie*
- Douglas, *Narrative of the Life of Frederick Douglas*
- Dowell, *Chicken Boy*
- Ehrenreich, *Nickel and Dime*
- Green, *Looking For Alaska*
- Griffin, *Black Like Me*

Wiesen (2011), International Personnel Assessment Council Conference

13

Status

- Have assembled word frequencies from above sources
- Have web site that helps identify infrequently used words: <http://wordsly.com>
- Plan additional functions for web site

Wiesen (2011), International Personnel Assessment Council Conference

14

Preliminary Findings

Wiesen (2011), International Personnel Assessment Council Conference

15

HS Books

- 12 HS Books in database now
- 680,000 total words
- 33,500 unique words (sum across books)
- Only 607 unique words in common to all 12 books

Wiesen (2011), International Personnel Assessment Council Conference

16

Sample of Unique Words in Common Across Books

- a
- able
- about
- across
- act
- afraid
- after
- afternoon
- again
- against
- ago
- ahead
- air
- alive
- all
- almost
- alone
- along

Wiesen (2011), International Personnel Assessment Council Conference

17

Unique “A” Words in Common Across Books

- already
- also
- always
- am
- an
- and
- another
- answer
- any
- are
- arms
- around
- as
- ask
- asked
- asking
- at
- attention
- away

Wiesen (2011), International Personnel Assessment Council Conference

18

Few “Common” Words

- Only 235 words appear at least 5 times in all 12 HS Books

Unique “A” Words Appearing 5+ Times in Each Book

- a
- about
- after
- again
- all
- almost
- already
- always
- an
- and
- another
- any
- are
- around
- as
- ask
- asked
- at
- away

NY Times

- 6 months of NYT summaries in database now
- Read as 20 files (due to size)
- 2,000,000 total words
- 63,300 unique words (sum across 20 files)
- Only 2,600 unique words in common to all 20 files

Wiesen (2011), International Personnel Assessment Council Conference

21

Implications of Little Overlap

- Writing questions using only frequent words will be challenging

Wiesen (2011), International Personnel Assessment Council Conference

22

Website

- Identifies unique words in user supplied text
- Returns frequency of each word for each word source (NYT, HS books, etc.)
- <http://wordsly.com>

Wordsly.com Entry Page

[Send as Fax](#) [Send as PDF](#) [Txt Message](#) [Plagiarism Check](#) [Text to Voice](#)

WordSly.com

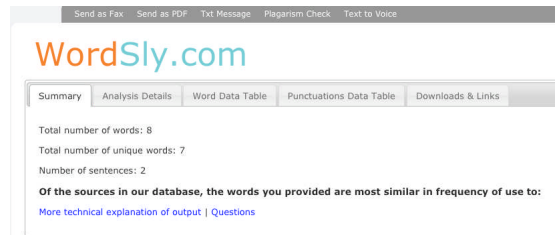
Enter text below

OR

upload file

[Examples](#) [Technical Information](#)

Initial Summary



Send as Fax Send as PDF Txt Message Plagiarism Check Text to Voice

WordSly.com

Summary Analysis Details Word Data Table Punctuations Data Table Downloads & Links

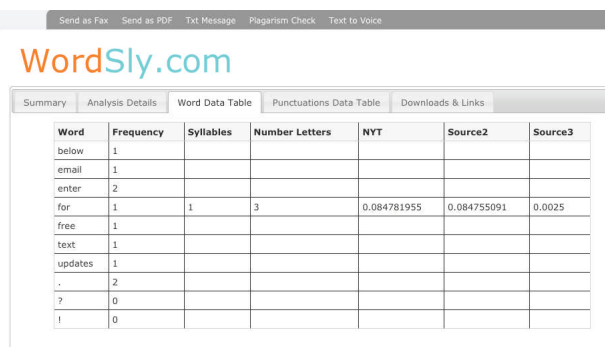
Total number of words: 8
Total number of unique words: 7
Number of sentences: 2

Of the sources in our database, the words you provided are most similar in frequency of use to:
[More technical explanation of output](#) | [Questions](#)

Wiesen (2011), International Personnel Assessment Council Conference

25

Word Frequency Table



Send as Fax Send as PDF Txt Message Plagiarism Check Text to Voice

WordSly.com

Summary Analysis Details Word Data Table Punctuations Data Table Downloads & Links

| Word | Frequency | Syllables | Number Letters | NVT | Source2 | Source3 |
|---------|-----------|-----------|----------------|-------------|-------------|---------|
| below | 1 | | | | | |
| email | 1 | | | | | |
| enter | 2 | | | | | |
| for | 1 | 1 | 3 | 0.084781955 | 0.084755091 | 0.0025 |
| free | 1 | | | | | |
| text | 1 | | | | | |
| updates | 1 | | | | | |
| . | 2 | | | | | |
| ? | 0 | | | | | |
| ! | 0 | | | | | |

Wiesen (2011), International Personnel Assessment Council Conference

26

Downloads Available



Wiesen (2011), International Personnel Assessment Council Conference

27

Cautions

- The approach ignores
 - different meanings of words
 - lead: a gray element
 - lead: what leaders do
 - different parts of speech, spelled the same
- Still fine tuning the definition of a word
 - plurals, possessives, gerunds counted separately

Wiesen (2011), International Personnel Assessment Council Conference

28

Plans for the Future

- Add traditional readability measures
- Develop new readability measures
 - Call for research partners
 - Seeking **cloze score** reading passages
 - Seeking other reading material with measures of readability
- Tweak word count function

Wiesen (2011), International Personnel Assessment Council Conference

29

Your Suggestions/Questions?

- Other Features/Functions
- Research Topics
- Research Partners

Wiesen (2011), International Personnel Assessment Council Conference

30

Call me any time to talk about this
subject:
(617) 244-8859

Copies of this presentation are (or will be) available at:
<http://ipac.org>

Wiesen (2011), International Personnel Assessment Council Conference

31

References

- (SIOP, 2011) An Interview With Frank L. Schmidt. *The Industrial-Organizational Psychologist*, April.
(Downloaded 5/22/2011 from
<http://www.siop.org/tip/april11/04schmidt.aspx>)

Wiesen (2011), International Personnel Assessment Council Conference

32