Changes in Word Usage Frequency May Hamper Intergenerational Comparisons of Vocabulary Skills: An Ngram Analysis of Wordsum, WAIS, and WISC Test Items

Eka Roivainen

Journal of Psychoeducational Assessment 2014 32: 83 originally published online 23 April 2013
DOI: 10.1177/0734282913485542

Journal of Psychoeducational Assessment
2014, Vol 32(1) 83–87

DOI: 10.1177/0734282913485542
jpa.sagepub.com
SAGE

Changes in Word Usage Frequency May Hamper Intergenerational Comparisons of Vocabulary Skills: An Ngram Analysis of Wordsum, WAIS, and WISC Test Items

Eka Roivainen[1]

Abstract

Research on secular trends in mean intelligence test scores shows smaller gains in vocabulary skills than in nonverbal reasoning. One possible explanation is that vocabulary test items become outdated faster compared to nonverbal tasks. The history of the usage frequency of the words on five popular vocabulary tests, the GSS Wordsum, Wechsler Adult Intelligence Scale (WAIS), Wechsler Adult Intelligence Scale–Revised (WAIS-R), Wechsler Intelligence Scale for Children (WISC), and Wechsler Intelligence Scale for Children–Revised (WISC-R) IQ tests, was analyzed by means of the Google ngram viewer. Usage frequency had a 0.38 to 0.73 correlation with item difficulty. In the period between test standardizations, the median change in usage frequency was −17% for WISC words, −8% for Wordsum, −5% for WISC-R, −4% for WAIS, and 0% for WAIS-R words. The correlation between median change in usage frequency and gain in vocabulary score was 0.33. Further studies with a larger set of vocabulary tests are needed to analyze in more detail the magnitude of the effect of changing word usage frequencies.

Keywords

vocabulary test, ngram analysis, Flynn effect

Previous studies indicate that scores on cognitive tests increase at a rate of roughly 0.3 IQ points per year in developed nations (Flynn, 2012). However, it seems that this trend, known as the Flynn effect, is stronger for nonverbal cognitive skills, such as matrix reasoning tasks, than it is for verbal skills, such as vocabulary knowledge. In the United States, the results of the general social survey (GSS) Wordsum vocabulary test, which has been administered to over 25,000 respondents since 1974, shows a stagnant or decreasing word knowledge in younger cohorts (Alwin & Pacheco, 2012). In Britain, Lynn (2009) reported a gain rate of 0.2 to 0.3 points per year on Raven's matrix task between 1979 and 2008 among schoolchildren aged 4 to 15 years

[1]Verve Rehabilitation Center, Oulu, Finland

Corresponding Author:
Eka Roivainen, Verve Rehabilitation Center, Kasarmintie 13, 90101 Oulu, Finland.
Email: eka.roivainen@verve.fi

while indicating no significant change on Raven's vocabulary scales (the Mill Hill and Crichton vocabulary tests). According to Sundet, Barlaug, and Torjussen (2004), test scores of Norwegian conscripts ($n > 960,000$) rose on a matrix reasoning test but not on a vocabulary test from 1970 to the 1990s.

Possible explanations for the Flynn effect include urbanization, better childhood nutrition, more stimulating childhood environment, better health care, and rising educational levels (Flynn, 2012). However, these factors should affect both verbal and nonverbal skills. Furthermore, it seems that vocabulary skills improve particularly slowly in children. Flynn (2010) reports divergent gain rates on the subtests of the Wechsler adult (WAIS; Wechsler, 1955, 1981, 1997, 2008) and Wechsler children's (WISC; Wechsler 1949, 1974, 1991, 2004) intelligence scales. From 1949 (WISC) to 2004 (WISC IV), the gain on the children's vocabulary subtest was only 4.4 IQ points, while the gain on the adult vocabulary test was 17.0 points from 1955 (WAIS) to 2008 (WAIS IV). This is strange because we might expect rising educational levels to have a strong effect especially on children's verbal skills. Flynn explained that the rise of the teenage subculture might have caused the divergence between vocabularies used by parents and their children.

Based on the then available historical word frequency counts and ranks for English words, such as those provided by Kucera and Francis (1967), Hauser and Huang (1997) concluded that prior to 1967, the frequency of the Wordsum words had been decreasing in English texts. Thus, the test items had become more difficult over time. Alwin and Pacheco (2012) refuted this conclusion because changes in the difficulty of Wordsum items in terms of intertest difficulty ranking have been small since 1974.

In the present study, the Google ngram viewer and Google search were used to estimate the frequency of the use of the Wordsum, WAIS, and WISC vocabulary test terms, and the correlation with item difficulty was analyzed.

Method

The WISC was published in 1949. Subsequently, the revised version, WISC-R, was released in 1974. The original WISC vocabulary subtest consists of 40 words and WISC-R vocabulary has 32 words. WISC-R vocabulary words comprised 21 WISC and 11 new words. On the Wechsler tests, the subjects are instructed to explain the meaning of each of the words that are presented in the order of difficulty. The tester reads the words and subjects provide their response orally. The testing discontinues after five consecutive failures. The responses are scored 0, 1, or 2 according to standardized criteria. The WAIS was published in 1955. The vocabulary subtest has 40 words. The WAIS-R was released in 1981. The vocabulary subtest has 35 words, two new and 33 taken from the WAIS. The WAIS III and the WISC III were not included in the present study, because the quality of the ngram database is highest for English texts up to year 2000.

The Wordsum (GSS, 2009) is a 10-word multiple-choice vocabulary test that has been included in the GSS administered to representative national samples of American adults since 1974. From 1972 to 2008, 25,555 persons took the vocabulary test. The GSS interviews are being conducted face to face. The respondents are handed a card with five response choices for each test word. They are instructed to choose the alternative that is closest to the meaning of the test word.

In the present study, half of the test words of each test were categorized as easy and the other half was categorized as hard. As described above, the vocabulary words are presented in the order of difficulty in the Wechsler tests and thus, the WAIS and WISC easy words are those numbered 1 to 20 in the test manual. Words 1 to 18 on WAIS-R and words 1 to 16 on WISC-R were categorized as easy. The Wordsum words commonly labeled as A, B, D, E, and F, known by 85% to 97% of responders (Alwin & Pacheco, 2012) are the easier ones and those labeled C, G, H, I, and J, known by 27% to 77% of responders (Alwin & Pacheco, 2012) were categorized as the hard words.

Table 1. Wordsum, WAIS and WISC Vocabulary Word Popularity Estimates.

	Google Hits 2012			Google Ngram Frequency			
	$<10^7$	10^7-10^8	$>10^8$	$<10^{-5}$	10^{-5}-10^{-4}	10^{-4}-10^{-3}	$>10^{-3}$
WAIS easy words	0	8	12	0	0	12	8
WAIS hard words	7	12	1	2	4	14	0
WISC easy words	1	5	14	0	1	12	7
WISC hard words	10	10	0	4	8	8	0
WISC-R easy words	0	3	13	0	0	11	5
WISC-R hard words	4	10	2	1	3	11	1
Wordsum easy words	1	3	1	0	0	4	1
Wordsum hard words	3	2	0	0	3	2	0
Total easy words	2	19	40	0	1	39	21
Total hard words	24	34	3	7	18	35	1

Note. Easy words = the easiest 50% of test items; hard words = the hardest 50% of test items; ngram frequency for the year of test publication; year 2000 for Wordsum. Google hits September 3, 2012.

The Google ngram viewer is a recently introduced phrase-usage graphing tool, which charts the yearly count of selected combinations, such as words or phrases, found in 5.2 million books digitized by the Google Corporation (Michel et al., 2011). At the moment, the database contains over 500 billion words. Matches found in a minimum of 40 books are indexed in the database. In the present study, American English texts were searched with 1 year smoothing. For example, the ngram data for 1997 are based on the average for the years 1996 to 1998.

Results

Google and Google ngram analysis of the Wordsum, WAIS, and WISC words showed easy words were more popular compared to difficult words (Table 1). The correlation between Google hits (9/2012) and difficulty rank was −0.81 (Spearman's r) for WISC, −0.77 for WISC-R, −0.78 for WAIS, and −0.52 for Wordsum (difficulty rank for 2000-2008). WAIS-R was not analyzed because its vocabulary test is essentially the same as the WAIS original version. The correlations between ngram frequency and item difficulty for WAIS, WISC, WISC-R, and Wordsum were −0.74, −0.69, −0.78, and −0.38, respectively (ngram frequency in the year of test publication and 1974 for Wordsum).

The figures in Table 2 show that test words tended to become less rather than more popular over time. The decrease in use was most pronounced for WISC items while little change was observed for WAIS-R items. The largest decrease in WISC, from 0.000020 to 0.000009, was observed for word #35. WISC-R word #28 fell from 0.00051 to 0.00036, while WAIS word #33 rose from 0.00025 to 0.00038. WAIS-R word #28 fell from 0.000038 to 0.000030, and Wordsum word H fell from 0.00010 to 0.000055.

The correlation between vocabulary score gain and median change in usage was 0.33. The correlation between median change in usage and the difference between observed and expected (estimated by the full-scale IQ gains) vocabulary gains was −0.55. As there is no item-level longitudinal data available for the Wechsler tests, correlation between change in usage frequency and change in difficulty for each test item separately could be calculated for Wordsum items only (−0.21).

Discussion

The analyses support the first impression that the easy words of the vocabulary tests are more popular and more frequently used compared to the difficult words. The ngram analyses further

Table 2. Changes in Word Usage Frequency and Vocabulary Scores.

	Period	Usage Change %	Vocabulary Score Gain %	Full-Scale IQ Gain %	FSIQ-Vocabulary Gain %
WAIS	1955-1981	–4	18	7.5	–10.5
WAIS-R	1981-1997	0	6	4.2	–1.8
WISC	1949-1974	–17	4	7.6	3.6
WISC-R	1974-1991	–5	4	5.4	1.4
GSS	1974-2000	–8	4	7.5[a]	3.5

Note. Median change in usage of vocabulary test words. Method: Google ngram viewer, 1 year smoothing. Vocabulary and IQ gains from Flynn (2010). [a]estimated IQ gain for U.S. adults 1974-2000.

suggest that many of the test words have become less popular over time. We can hypothesize that the WAIS, WISC, WISC-R, and GSS vocabulary tests have become somewhat more difficult. Because of the short time period between test standardizations, the changes are fairly small. However, the size of the gap between observed and expected vocabulary gains seems to be correlated with median change in test word usage. The item-level correlation between change in usage and change in difficulty was weak (–0.21) for the Wordsum items, but it should be noted that the correlation between ngram usage frequency and difficulty is fairly low (–0.38) for the Wordsum words.

Assumably, old-fashioned words function well as vocabulary test items in an IQ test. Such words are probably better known to persons who read books, because changes in written language lag behind changes in spoken language (Curzan, 2009). Book-reading is strongly correlated with general intelligence. The standardization samples of the new test versions are more educated, but old-fashioned words favor older cohorts. This may lead to slow IQ gains in vocabulary tests. However, the observed absence of a strong Flynn effect on other verbal subtests, such as Information, indicates that there must also be other factors at play. Google search and Google ngram are not rigorously validated scientific instruments and it is unclear how well the database and the search algorithm's results reflect general language use (Brysbaert, Keuleers, & New, 2011). Additional analyses with a larger selection of vocabulary tests, such as those used in student achievement studies are needed.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

References

Alwin, D. F., & Pacheco, J. (2012). Population trends in verbal intelligence in the United States. In P. W. Marsden (Ed.), *Social trends in American life. Findings from the General Social Survey since 1972* (pp. 340-363). New Jersey, NJ: Princeton University press.

Brysbaert, M., Keuleers, E., & New, B. (2011). Assessing the usefulness of Google Books' word frequencies for psycholinguistic research on word processing. *Frontiers in Psychology*, *2*, 27.

Curzan, A. (2009). Historical corpus linguistics and evidence of language change. In A. Luedeling & M. Kytö (Eds.), *Corpus linguistics* (pp 1091-1108). Berlin, Germany: Gruyter.

Flynn, J. R. (2010). Problems with IQ gains: The huge vocabulary gap. *Journal of Psychoeducational Assessment*, *28*, 412-433.

Flynn, J. R. (2012). *Are we getting smarter. Rising IQ in the 21st century*. Cambridge, UK: Cambridge University Press.

GSS (General Social Survey). (2009). *Cumulative file for wordsum 1972-2006*. Chicago, IL: National Opinion Research Center.

Hauser, R., & Huang, M. H. (1997). Verbal ability and socioeconomic success. *Social Science Research*, *26*, 331-376.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University press.

Lynn, R. (2009). Fluid intelligence, but not vocabulary has increased in Britain, 1979–2008. *Intelligence*, *37*, 249-255.

Michel, J.B., Shen, Y. K., Aiden, A. P., Veres, A., & Gray, M. K., The Google Books Team, . . . Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*, 176-182.

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, *32*, 349-362.

Wechsler, D. (1949). *Wechsler intelligence scale for children*. New York, NY: Psychological Corporation.

Wechsler, D. (1955). *Wechsler adult intelligence scale: Manual*. New York, NY: Psychological Corporation.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children—Revised*. New York: Psychological Corporation.

Wechsler, D. (1981). *Wechsler adult intelligence scale–Revised: Manual*. New York, NY: Psychological Corporation.

Wechsler, D. (1991). *The Wechsler intelligence scale for children—Third edition*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997). *Wechsler adult intelligence scale–Third edition: Manual*. San Antonio, TX: Psychological Corporation.