



Twentieth Century Intercohort Trends in Verbal Ability in the United States

Shawn F. Dorius,^a Duane F. Alwin,^b Julianna Pacheco^c

a) Iowa State University; b) Pennsylvania State University; c) University of Iowa

Abstract: Vocabulary test score trends from the General Social Survey contradict the widespread conclusion that scores on standardized intelligence tests have systematically increased over the past century. We use a vocabulary test included in 20 nationally representative surveys administered since 1974 to test three hypotheses proposed to account for these trends, including changes in the formal measurement properties of the test, over-time changes in the meaning of education, and intercohort differences in exposure to words on the test. We find no support for the idea that test scores have declined because of changes in the structure of the test. Instead, our results show that education selectivity accounts for some cohort differences among prewar cohorts and that cohort-specific differences in exposure to words on the test account for nearly all variation in vocabulary scores of respondents born after 1945, suggesting different causal processes have influenced cohort verbal ability during distinct historical eras.

Keywords: verbal ability; crystallized intelligence; Flynn effect; WORDSUM; cohorts; test scores

RESULTS from standardized intelligence tests indicate that scores have systematically increased over the past century throughout the world, with substantial evidence pointing to generally higher test scores among more recent cohorts across a range of cognitive measures. This historical rise in scores on standardized tests from one generation to the next, often called the “Flynn Effect” (Neisser 1997, 1998), was named for political scientist James Flynn, who discovered and quantified its pervasiveness (Flynn 1984, 1987a, 1987b, 1998, 1999). Results supporting the Flynn Effect square with expectations that increasing levels of education bring about better intellectual performance in a population (Baker et al. 2015)—what has elsewhere been referred to as the *progress model* of societal cognitive development (Alwin 2009; Alwin and Pacheco 2012). Flynn has argued that the phenomenon of rising IQ scores is not genetic in origin but, rather, has more to do with an environmental change toward greater emphasis on scientific, rational thinking (see Flynn 2007, 2012). Others have focused on improvements in health, nutrition, and education (e.g., Tuddenham 1948; Neisser 1997, 1998; Lynn 2009). The general consensus in recent years is that the Flynn effect does not reflect genuine changes in intelligence (e.g., *g* factor; Jensen 1998) but instead reflects the kind of analytic ability at a premium in the developed world (Flynn 2012; te Nijenhuis and van der Flier 2013).

Researchers report generally higher test scores among more recent birth cohorts across a range of cognitive measures (Schaie 2005, 2008; Schaie, Willis, and Pennack 2005), providing further support for expectations that increasing levels of education bring about better population-level performance on cognitive tests. However, it is well known that test scores in the United States have not always risen over time, nor have they risen in all domains. Indeed, during the 1960s and 1970s, there

Citation: Dorius, Shawn F., Duane F. Alwin and Julianna Pacheco. 2016. “Twentieth Century Intercohort Trends in Verbal Ability in the United States.” *Sociological Science* 3: 383-412.

Received: January 13, 2016

Accepted: January 28, 2016

Published: June 13, 2016

Editor(s): Jesper Sørensen, Stephen Morgan

DOI: 10.15195/v3.a18

Copyright: © 2016 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited.

were serious concerns in the United States about the declining performance by the young on measures of verbal ability. Average verbal scores on standardized tests like the Scholastic Aptitude Test (SAT) and the American College Testing Service (ACT) tests declined systematically from the mid-1960s through the mid-1980s (see Harnischfeger and Wiley 1976; Wirtz et al. 1977; Ferguson 1976, Cleary and McCandless 1976; Congressional Budget Office 1986, 1987; Gardner et al. 1983).

Although there are serious problems with interpreting changes in college admissions test scores as if they reflect true aggregate levels of verbal and quantitative skills of high school students, given the changing composition of the test-taking population (Wirtz et al. 1977), most observers of test score declines concede that net of these compositional shifts, the test score decline was and is real (Blake 1989; Alwin 1991; for a counterargument, see Wilson and Gove 1999). SAT test scores on the verbal dimensions have recently continued to decline, dropping to their lowest point in four decades (Chandler 2011; Layton and Brown 2012). Birth cohort trends in verbal ability derived from the General Social Survey (GSS) vocabulary test score data reproduce the SAT score declines almost exactly with respect to their direction and timing (see Alwin 1991; Alwin and McCammon 1999, 2001; Glenn 1994), adding further support to the test score decline thesis but also contradicting the cohort patterns implied by the literature surrounding the Flynn Effect.

A number of plausible arguments have been advanced to account for the declining levels of vocabulary knowledge among post-World War II birth cohorts in the United States, which we only briefly mention here. Explanations have ranged from changes in societal values (Jencks 1979), declines in motivation and effort in school (Winter 1977; Menard 1988), increased time watching television and declines in reading (Wirtz et al. 1977; Schramm 1977; Morgan 1986; Glenn 1994, 1999), changes in family structure (Zajonc 1976), increased substance usage among students (Wirtz 1977), simplification of curriculum (Chall 1983, 1996; Chall and Conard 1991; Hayes, Wolfer, and Wolfe 1996; Stedman 1996), and aging (Wilson and Gove 1999). We extend this body of work by examining three additional hypotheses: changes in the measurement properties of the test (Malhotra, Krosnick, and Haertel 2007), macrosocial changes in word usage (Alwin & McCammon 1999), and education selection (Huang 2003).

The first hypothesis that has been advanced to account for the intercohort decline in the GSS data is that the vocabulary test cohort trends (hereafter referred to by the scale name, WORDSUM) are due to changes in the measurement properties of the test (Malhotra, Krosnick, and Haertel 2007). This explanation leverages advances in psychometric methods to consider intercohort differences in item functioning and factorial invariance in WORDSUM. Evidence of either differential item functioning or measurement invariance between birth cohorts might explain the observed trends.

A second hypothesis advanced to account for the intercohort WORDSUM decline in the post-1940s birth cohorts posits that some or all of the words used in the GSS vocabulary test (and, presumably many words used in other standardized tests) have simply gone out of fashion (Alwin 1991:628). The declining prevalence of WORDSUM test items, which we here refer to as the *word exposure hypothesis*,

might explain the intercohort decline in the GSS data as well as more general test score results.

A third argument that is relevant to intercohort variation in vocabulary knowledge—but which applies mainly to differences among the pre-1940 birth cohorts—argues that the expansion of formal schooling over the course of industrialization has been attended by changes in the degree to which students are “selected” into mass schooling. Variation in certain respondent characteristics known to be associated with verbal ability, including individual genetic factors, mortality risk, and socioeconomic status, might select high verbal ability students into formal schooling (Huang 2003). These same selection forces also influence educational attainments. In practical terms, this implies that the meaning of educational attainment has changed as formal schooling became compulsory and universal, such that schooling may not be as selective of high cognitive ability students as it once was. We refer to this phenomenon as the *education selection hypothesis*.

We systematically investigate these explanations of the intercohort test score decline using three interconnected analytic strategies. In the first step, we conduct a series of statistical tests to address the question of whether between-cohort differences in verbal ability are due to changes in the formal measurement properties of the test.¹ We also establish the intercohort comparability of WORDSUM. We employ the Item Response Theory (IRT) model framework (Embretson and Reise 2000), which is a logit-based form of the common factor model (Glöckner-Rist and Hoijtink 2003), to test changes in the difficulty parameters of the test questions over time and across cohorts. We then use the multi-group confirmatory factor analysis (MGCFA) framework to examine the psychometric properties of the GSS test score. This framework provides a formal test of the invariance of the measurement model and item loadings across cohorts. Not only is factorial invariance an alternate test of change in the measurement properties of the verbal ability index, but it is also a necessary step for establishing the comparability of the test across birth cohorts.

In the second step, we construct a measure of cohort-specific education selectivity and use it in a series of multivariate regression models in which WORDSUM scores are regressed on a measure of education selection and control variables in an effort to account for intercohort differences in verbal ability.

And third, we develop an innovative measure of word exposure drawn from the Google Books corpus (Michel et al. 2011) to assess the effect of intercohort differences in exposure to words in published textual material. We use this measure of cohort exposure in multivariate regression models to understand the extent to which the effect of education selection is diminished or amplified. Before presenting results, we first discuss the theoretical and substantive background to the problem and present a description of our data and measures.

Theoretical Background

For many decades, social and cognitive scientists have argued for potentially enduring differences in the childhood and adolescent socialization experiences of birth cohorts, owing to the impact of historical events and exposures during “critical periods” or “impressionable phases of development” (Mannheim 1927; Ryder

1965; Schaie 1984). Each new cohort, the argument goes, is exposed to distinctive experiences during childhood and adolescence that are thought to affect a wide range of behavioral and ideational outcomes and continue to have an influence throughout the life course because of processes that promote stability of individual differences (Alwin and McCammon 2003, 2007). Because of the distinctive and indelible nature of cohort experiences, each cohort has the potential to differ from those that precede and follow it, and such differences have a tendency to persist. As a consequence, cohort turnover is partly responsible for change in the composition of society (Mannheim 1952; Ryder 1965; Firebaugh 1992). From these perspectives, social change is by no means assumed to be entirely because of cohort replacement. Changes that occur within cohorts, including aging and period influences, operate in tandem with changes that occur between cohorts to effect the macrosocial trend in a response variable (e.g., IQ test scores). The twin forces of social change that proximately reside within and between cohorts contribute to an understanding of the dynamic forces that shape individual lives and the composition of society.

One area in which cohort experiences appear to leave enduring effects on young people is in their *verbal ability*, a phenomenon that has received substantial attention, due in part to the inclusion of a vocabulary test in the GSS (Alwin 1991, 2009; Glenn 1994, 1999; Hauser and Huang 1997; Alwin and McCammon 1999, 2001; Bowles, Grimm, and McArdle 2005; Alwin and Pacheco 2012). Amidst confirming evidence in the test score decline, a longstanding puzzle in the literature on this subject is the existence of a counterintuitive trend toward declining scores on the GSS vocabulary test (WORDSUM), beginning with cohorts born after WWII. By contrast, vocabulary scores steadily increased for cohorts born between 1885 and 1945, thus following the pattern documented by Flynn (2007, 2012), namely monotonically rising test scores for each new birth cohort. The post-1940s trend in the GSS data is not unique, and as noted, it parallels test score declines during the 1960s and 1970s.

There is ample evidence suggesting general cognitive decline in older age (Hofer and Alwin 2008), but it is important to realize that although most cognitive abilities peak in early adulthood and then decline with age, vocabulary knowledge is unique in that it is one of the few skill sets that remain relatively intact over the adult lifespan (Verhaeghen 2003). Research suggests that vocabulary knowledge peaks in the 50s, with only a very slow decline into older age (Denny 1982; Baltes 1987; Salthouse 1991; Park et al. 1996; Alwin and McCammon 2001; Huang 2003; Verhaeghen 2003; Bowles et al. 2005; Schaie 2005). The general conclusion is that while there are detectable aging effects in the GSS data (Alwin and McCammon 2001; Alwin and Pacheco 2012), their magnitude is insufficient to account for the effects of cohort experiences, especially those witnessed among cohorts born after WWII. Leveraging recent methodological developments involving the use of hierarchical age-period-cohort and ridge-regression models, Yang and Land (2006, 2008; see also Yang et al. 2008) found that cohort effects in the GSS data were robust to the presence of controls for both aging and period factors. They concluded that the intercohort decline in vocabulary test scores in the GSS were real, rather than a methodological artifact (Yang and Land, 2006, 2008; see also Yang et al. 2008).

Research has shown that intercohort results reported for the GSS data reproduce the SAT score declines almost exactly with respect to their direction and timing and

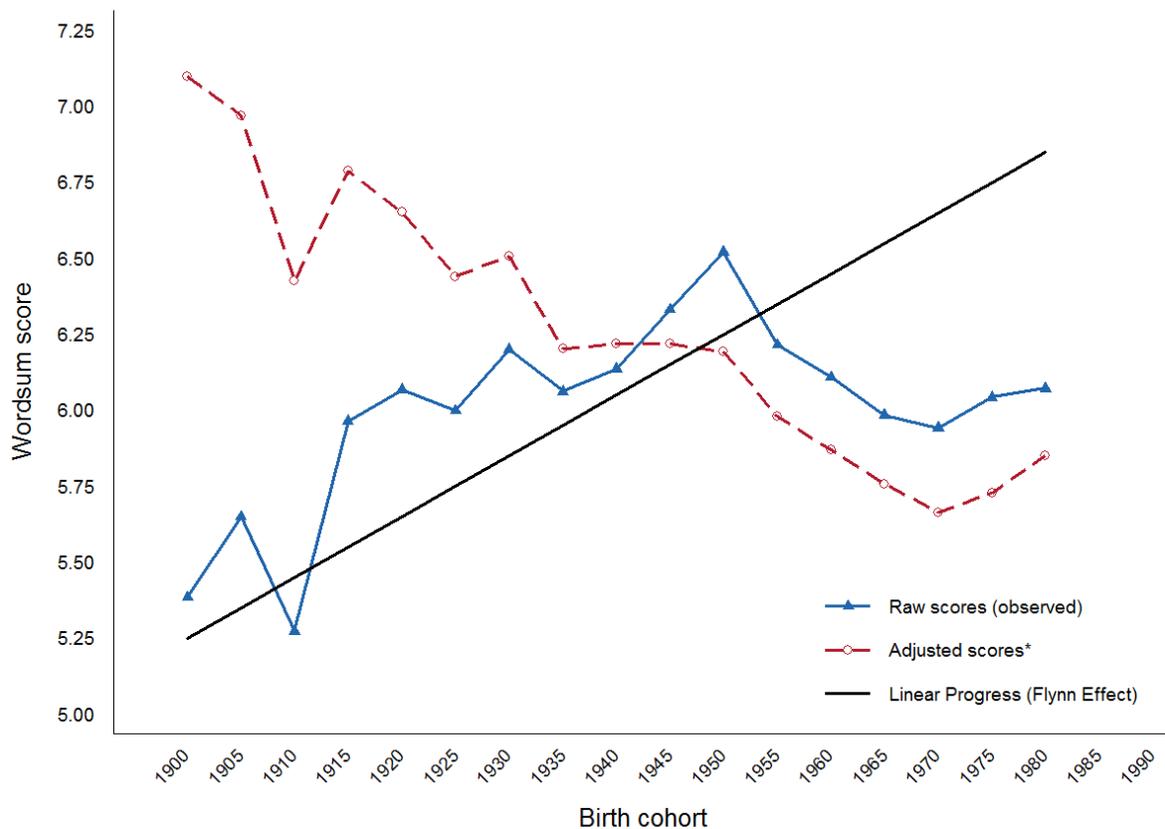


Figure 1: Raw and adjusted WORDSUM test scores, by birth cohort: General Social Survey, 1974–2012.

Note: Estimates sample weighted and adjusted for respondent age and education. First and last birth cohort categories were collapsed as follows: 1900 = 1885–1900; 1980 = 1980–1988.

also in contradiction to the cohort patterns implied by the literature surrounding the Flynn Effect. In Figure 1, we report the raw, cohort-specific vocabulary test scores derived from 20 social surveys collected from 1974 to 2012 for cohorts born between 1885 and 1988. We report trends using five-year cohort categories, which we superimposed on two additional GSS test score patterns. The first superimposed pattern shows the cohort trends after adjusting for respondent age and education. These data indicate that once verbal ability scores are adjusted for differences in age and level of schooling at time of testing, the overall cohort trend is toward lower vocabulary for each new birth cohort. Another set of patterns in Figure 1 depicts the theoretical expectations derived from the Flynn Effect (linear progress model), a pattern that shows a systematic over-time increase in performance on cognitive tests. Taken together, these trends produce some rather surprising conclusions.

First, if one assumes the linear progress model is the correct model for predicting average vocabulary scores over cohorts, the raw scores are consistent with the model up to about 1945 and dramatically incompatible thereafter. Second, once

the cohort-specific test scores are adjusted to constant age and schooling levels of the population as a whole, the earlier-born cohorts have systematically higher scores, and vocabulary test scores decline across cohorts. This set of adjusted scores provides evidence that is completely reversed from expectations based on the Flynn Effect. Although there is likely more than a single explanation for the adjusted test score trends, and quite possibly a different explanation for cohort disparities at different points across the cohort axis (Alwin 2009), the results contradict test score trends based on other measures of verbal abilities (e.g., Wechsler Adult Intelligence Test [WAIS]; Flynn 2010). In the next section, we discuss several hypotheses that have been developed to explain the adjusted WORDSUM test score decline.

Prior Explanations

Here, we briefly review a number of hypotheses that have been proposed to account for the intercohort decline in test scores. Jencks (1979:13), for example, suggested that schools and parents failed to instill proper academic values in children during this period, concluding that students “lost respect for the values of reason.” Others have proposed different mechanisms generally related to the same theme—that children of the 1960s and 1970s were exposed to a different learning environment, with less motivation and fewer hours spent in homework (Winter 1977; Menard 1988), increased time spent watching television, and less reading (Wirtz et al. 1977; Schramm 1977; Morgan 1986; Glenn 1994, 1999). Zajonc (1976) argued that the changes were due to family configuration (family size, birth order, and birth spacing) brought about by the post-WWII demographic baby boom (see also Alwin 1991). Some also suggested that drug and alcohol use by high school students may have contributed to test score declines during the 1960s and 1970s, but few suggest this as a principal explanation (Wirtz et al. 1977). Education researchers have offered other explanations involving school composition and curriculum. One of these arguments has to do with the systematic “dumbing down” of learning materials in schools, the other an overall tendency toward declining complexity in schooling content (Chall 1983, 1996; Chall and Conard 1991; Hayes, Wolfer, and Wolfe 1996; Stedman 1996). Despite these changes, public school apologists have maintained that these test score declines were not a reflection on the job schools were doing (Alexander 1997).

Word Exposure

The exposure hypothesis posits that the growing obsolescence of words used in standardized tests accounts for intercohort vocabulary test score declines (i.e., rising item-level difficulty parameters and pass rates), rather than objective declines in general verbal ability. Because the English language is dynamic and subject to changing usage and centrality in various forms of communication (Brysbaert, Keuleers, and New 2011), word usage varies by place, time, and across generations (Flynn 2012). Variation in the frequency of word usage is positively associated with word recall and test performance (Segui et al. 1982), suggesting that the difficulty of vocabulary test items is associated with word usage (Laufer 1997) and word age (Woodley et al. 2015). If later-born cohorts have not been exposed to the words

used in the GSS vocabulary test to the same extent as earlier-born cohorts, it may appear that they have less vocabulary knowledge when, in fact, they simply have different vocabulary knowledge than earlier-born cohorts (Flynn 2012).

The exposure hypothesis has primarily been applied to post-WWII cohorts, in which the effectiveness of the American education system in developing students' basic verbal skills has been questioned (Alwin 1991; Alwin and McCammon 1999, 2001). For example, an analysis of 800 textbooks used in elementary, middle, and high schools between 1919 and 1991 (Hayes, Wolfer, and Wolfe 1996) reported that words in textbooks became progressively easier after WWII. What they referred to as the "systematic dumbing down" of U.S. reading textbooks was embodied in the daily use of simplified instructional materials that resulted in "a cumulating deficit in students' knowledge base and advanced verbal skill" (Hayes et al. 1996:483). In another test of the exposure hypothesis, Hauser and Huang (1997) conducted detailed analyses of English-language dictionaries over the early part of the twentieth century to measure word use frequencies. Based on their study of word counts and rankings of English-language words in dictionaries for 1921, 1931, 1944, and 1967, Hauser and Huang (1997:345) concluded that decreased exposure "occurs to some degree across the lists of 1921, 1931, and 1944, and it appears strongly when we include the 1967 list in the comparison" (see also Thorndike 1921, 1931; Thorndike and Lorge 1944; Kučera and Francis 1967). While these results are suggestive of an effect of differential exposure to word content on vocabulary knowledge, limitations in both studies leave it an open question as to whether declining word exposure is responsible for declining WORDSUM test scores.

One of the major mechanisms for the transmission of knowledge is formal schooling, and there is a strong relationship between time spent in school and vocabulary knowledge, as shown in all of the research on this topic. However, the education payoff for vocabulary knowledge in later-born cohorts may not be reflected in the words included in common tests of verbal ability, due in part, as we argue here, to declining exposure to the words in the test. The experiences of the post-WWII cohorts, both in school and in the extracurricular world, may have systematically lowered vocabulary knowledge, but this does not account for the relatively higher scores of persons in the earliest-born cohorts.

We leverage recent developments in "culturomics" (Michel et al. 2011) to develop an innovative measure of word exposure that allows us to provide the most rigorous test to date of the word exposure hypothesis. This measure is derived from a corpus of more than 500,000,000,000 words used in English-language books published between 1500 and 2000. This corpus makes it possible to directly measure the central mechanism theorized to cause intercohort differences in vocabulary knowledge, rather than relying on birth year as a proxy for these causal mechanisms. In doing so, we are able to establish a much stronger test of the word exposure hypothesis than has heretofore been possible (see Chall et al. 1977 for an early analysis of test scores using the content of books).

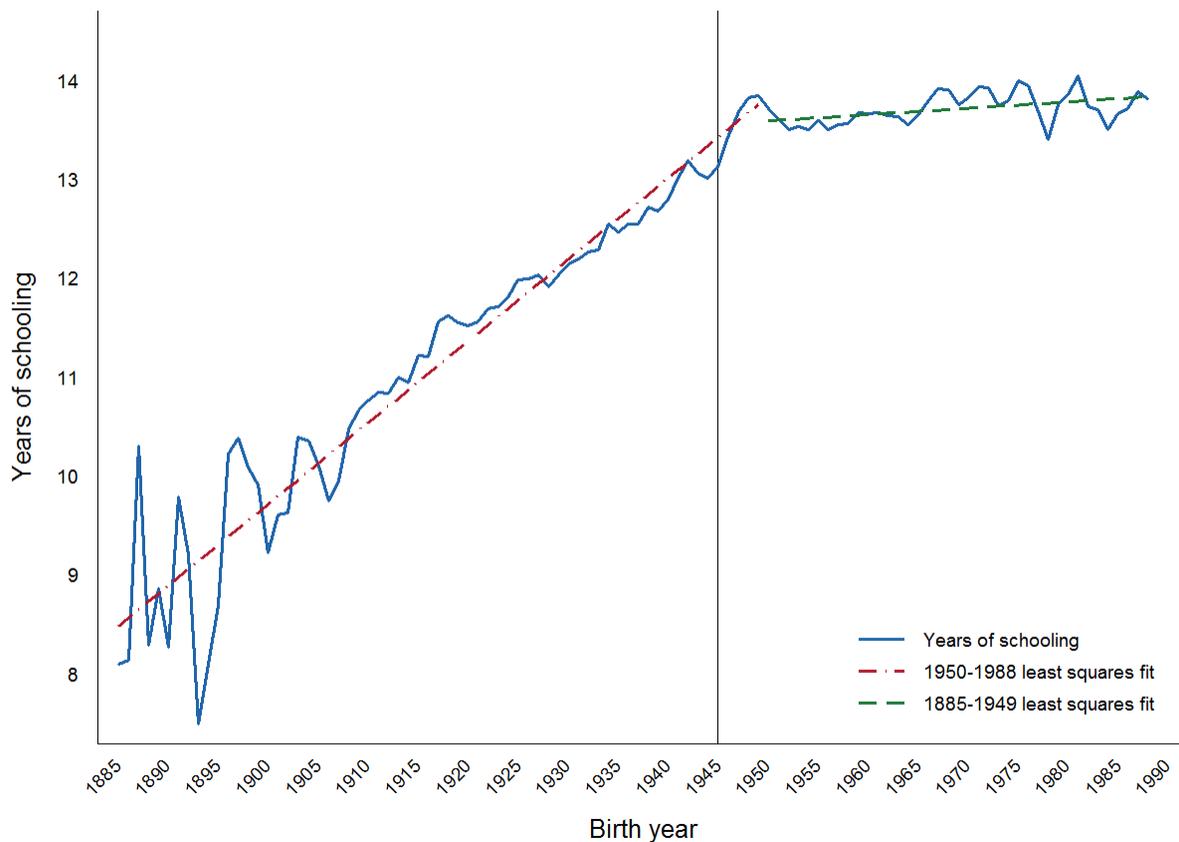


Figure 2: Schooling trends by birth cohort: General Social Survey, 1974–2012.

Note: Estimates are sample weighted. A three-year moving average was applied to the schooling trend line. Birth cohort reference category (1946–1950) identified by vertical line.

Selectivity of Education

The GSS data provide evidence of a strong relationship between cohort membership and levels of schooling, based on calculations from 38 years of GSS data spanning the years 1974–2012. As can be seen in Figure 2, systematic cohort differences in educational attainment reveal a strong association between year of birth and completed years of schooling among cohorts born in the first half of the twentieth century. This trend, however, leveled off considerably beginning with cohorts born after WWII, after which cohorts do not appreciably differ in schooling attainments. These trends imply that education was far more selective on factors associated with verbal ability prior to 1950 than afterward.

A high school degree was substantially less common in 1900 than in 2000, suggesting that the meaning of educational attainment has changed as formal schooling became compulsory and universal. Early in the century, only the top students went on to college, whereas now the public is exposed to at least some

postsecondary schooling. In essence, the scale that we use to measure education—years of schooling—may have changed, such that a “year” of schooling measured in, say, 1920 may mean something very different from a “year” of schooling measured in 2000.

This also suggests that comparing individual-level educational attainment across cohorts is insufficient to account for the relationship between formal schooling and vocabulary test scores. Respondent-level variation in characteristics associated with verbal ability, such as latent cognitive abilities, health status, socioeconomic status, and other family background variables, might over-select children with high verbal ability into formal schooling and under-select children with low verbal ability into school (Huang 2003), though some question this assertion (Baker et al. 2015).

Still, despite the over-time rise in educational attainment, the evidence from the GSS cited earlier suggests that education has diminishing returns among more recent cohorts, at least with respect to vocabulary knowledge. If so, then it is necessary to adjust vocabulary test scores not only for respondent education but also by cohort-specific educational attainments. We expect that the potential for education selectivity to account for the test score decline is greatest in the cohorts born before WWII, as this was the period when intercohort differences in educational attainment were largest.

Changes in Test Functioning

Of central concern in this research is the apparent increase in the difficulty of the GSS vocabulary test among more recently born cohorts. Researchers have devised several strategies to measure over-time changes in the difficulty of WORDSUM, but the results are mixed. In their analysis of English-language dictionaries over the early part of the twentieth century, Hauser and Huang (1997) concluded that “WORDSUM has become somewhat more difficult across time, independent of any other change in verbal ability in the general population” (Hauser and Huang 1997:344-345). While the Hauser and Huang (1997) results are suggestive, they do not go beyond the measurement of lexical word frequencies, nor do they provide a direct test of the formal measurement properties of WORDSUM.

Another approach is to investigate the psychometric properties of the test items, including the systematic analysis of item difficulties (Malhotra, Krosnick, and Haertel 2007; Beaujean and Sheng 2010; Alwin and Pacheco 2012). This work has shown that WORDSUM item difficulties, measured by the proportion of the population that correctly answers the question, have not appreciably changed. They concluded that “some words in the WORDSUM score are getting easier, some are getting harder, but none have changed substantially; the predominant pattern is toward decreasing rather than rising item difficulty” (p. 353). These findings led Alwin and Pacheco (2012) to draw the opposite conclusion from that of Hauser and Huang (1997), indicating that changes in the difficulty of the GSS items was not in the direction of increasing difficulty.

Some recent literature suggests the content of the GSS WORDSUM score may be more complex than heretofore acknowledged and that the specific content of the test score needs to be taken into account in assessing the nature of these trends. Bowles

et al. (2005) found that rather than a univocal measure of verbal ability, the GSS test items were composed of two underlying factors: one reflecting general knowledge (basic words) and one reflecting specialized knowledge (advanced words). These findings are compatible with difficulty-related vocabulary factors in other sets of intelligence tests (Bailey and Federman 1979; Beck et al. 1989; Gustafsson and Holmberg 1992). Alwin and Pacheco (2012) compared the difficulties of WORD-SUM items with those of the same items used in a 1941 Gallup survey. They found that words that were comparatively easy in 1941 became easier (as assessed in the 2004–2008 GSS samples) and that those that were more difficult in 1941 became more challenging in the more recent 2004–2008 GSS (p. 352–53). Using this broader compass, Alwin and Pacheco (2012) conclude that the items measuring basic knowledge have become somewhat more difficult while those items measuring advanced knowledge have become somewhat easier. But because they did not directly assess intercohort factorial invariance, it is unclear whether the patterns they reported reflect genuine changes in the population.

We replicated the Alwin and Pacheco (2012) analyses using a more powerful set of research strategies to compare word difficulty over time and across birth cohorts, providing a direct quantitative assessment of whether the GSS vocabulary test is becoming more difficult. In short, rather than assessing item difficulty using pass rates (proportion correct), we gauge item difficulty using IRT difficulty parameter estimates, as this approach yields a more statistically justified measure. For those unfamiliar with IRT models, we note that the difficulty parameters in IRT models estimate the point on the underlying continuum of word knowledge where half the population is expected to correctly answer the question. To save space, we will only briefly summarize results of formal tests of the measurement properties of the GSS vocabulary test, but we present a detailed discussion, including point estimates and other model results, in the online supplement.

Methods

Samples

Our analysis relies on 20 nationally representative, cross-sectional samples of U.S. adults, collected at roughly biannual intervals from 1974 to 2012 as part of the GSS (Smith et al. 2012). In each of the early GSS surveys, approximately 1,500 respondents were interviewed. For several years, the vocabulary test was only administered to two-thirds of the GSS sample. In more recent years, the GSS employed a larger overall sample size (~3,000), so the number of respondents who completed the test since the 1990 has averaged roughly 1,500 per wave.

The inclusion of a vocabulary test in the GSS data has a number of advantages. Respondents in each sample completed an identical 10-item vocabulary test, allowing researchers to conduct more rigorous tests regarding cohort differences in verbal ability. Another strength of nationally representative data such as the GSS is that they contain a large amount of variation in the age, birth year, and educational attainment of research subjects, each of which is strongly associated with cognitive test scores. Because much of what we know about over-time, aggregate changes in

performance on cognitive tests is based on studies of student populations, measurements of cognitive abilities in nationally representative samples offer researchers unique opportunities to test social and demographic propositions regarding human abilities.

In keeping with the precedent set in previous research, we restrict the composition of the GSS sample to the native-born, English-speaking population (see Alwin 1991; Alwin and McCammon 1999, 2001). The GSS sample is typically biased in the lower adult age ranges because the GSS does not interview eligible respondents on college campuses or military reservations (see Alwin 1991:628). For this reason, we restrict our sample to those aged 24 or older. We employ weights provided by the GSS study staff (variable name: WTSALL) that adjust for household size, oversampled subgroups, and experimental design effects of the GSS samples (see Appendix A 2932 of GSS documentation for details of sample and design weights).

Verbal Ability

Measures of vocabulary knowledge correlate highly with tests of general intelligence—usually in the range of 0.7 to 0.8 or higher (see Miner 1957:28–31)—and are considered good indicators of the verbal component of general intelligence tests (see also Wolfe 1980; Hauser and Huang 1997). The vocabulary test in the GSS data is one such test that has been widely used as a measure of verbal ability, with more than 100 published articles and chapters using these data in recent decades (Cor et al. 2012). WORDSUM is composed of 60 words or short phrases contained in ten questions consisting of the target word and five answer choices from which respondents must identify the correct synonym (one correct answer and four distractors).² We identify the individual items from the GSS WORDSUM score using the mnemonics WORD A, WORD B, WORD C, et cetera, in keeping with the GSS study's desire not to publicize these words. Representative content from a related vocabulary test (Thorndike 1942) is given by Miner (1957:53). We recoded "Don't Know" responses to incorrect answers so that a 1 indicates a correct answer and 0 identifies answers that are not correct. A small amount of additional incomplete data was removed from the analysis using listwise deletion (0.004 percent, 104 cases) for a final sample of 22,684.

An important distinction between the GSS vocabulary test and the WAIS verbal ability test is that WAIS uses a "free recall paradigm," whereas the GSS test uses a "recognition paradigm" (Flynn 2012). The difference between the two approaches is that WAIS is designed to measure active vocabulary, which people typically acquire through conversation, while the GSS test measures passive vocabulary of the kind typically acquired by reading books.

Education Selectivity

We measure education selection using a cohort-level estimate of average years of educational attainment in our samples. Self-report education was used to generate estimates of mean cohort education levels (see Table 1 and Figure 2 for quantitative and qualitative presentations of the measure, respectively). This measure reflects cohort differences in the selectivity of the education system over historical time. Net

of this measure, the person-level years of completed schooling captures respondents' within-cohort relative ranking in the education distribution.

Measures of Word Exposure

The data source from which we developed a cohort-specific measure of word exposure was the Google Books corpus (Michel et al. 2010). The now-public data, most commonly known through the Google Books Ngram Viewer (<https://books.google.com/ngrams>), currently contains roughly 500,000,000,000 English words dating back to the 1500s. In total, each of the more than 15,000,000 scanned and digitized books represents approximately 12 percent of books published between 1500 and 2009 (Michel et al. 2010).³ Validity tests of these data demonstrate that compared to American English-language subtitle lexicon data, the Google Books corpus have similar predictive power in explaining word recall (Brysbaert et al. 2011). Because of post-2000 changes in the composition and method of digitizing books in the corpus⁴, researchers should use caution in interpreting changes in word frequencies during this period, as they may reflect compositional changes in the database rather than functional changes in the word usage.

A case-insensitive data query was used to extract raw frequencies, counts of each word comprising the WORDSUM test, year in which the word appeared in a published book, the number of times it was used, and the total number of words in the corpus for every year from 1885 to 2003. In keeping with precedent, we converted the annual frequencies into percentages to account for the substantial increase in the overall size of the corpus from one year to the next.⁵

Cohort theories posit that socialization occurring during a person's formative years, typically understood to span the period from birth through adolescence, tend to leave an indelible mark that persists over the life course (Mannheim 1952; Ryder 1965). Likewise, crystallized intelligence, including verbal ability, undergoes substantial development during a person's early years, due in part to the learning environments of the home, community, and school (Cattell 1963, 1971a, 1971b; Horn 1982a, 1982b; Horn and Cattell 1967; Horn and Donaldson 1976, 1980). For the purposes of this research, we define "early life" as the first 15 years of life. Using this definition, we averaged the word exposure percentages from year of birth to year in which respondent reached age 15 using a 15-year, forward-moving average. We then standardized and averaged these 60 variables into three summary measures of cohort word exposure, including a 46-item measure of *basic* vocabulary knowledge (6 questions * 6 words), a 24-item measure of *advanced* vocabulary knowledge (4 questions * 6 words), and a 60-item measure of *general* vocabulary knowledge (all 10 questions * 6 words).⁶ The summary measures of cohort word exposure contain 104 unique values—one for each birth cohort in the study—with ranges and means approximating a z-score distribution. Descriptive information on education selection, word exposure, WORDSUM, and other key variables are reported in Table 1.

Table 1: Descriptive Statistics (N = 22,684)

Measure	Mean	Standard Deviation	Minimum	Maximum
General Vocabulary Knowledge	6.13	2.08	0	10
Basic Vocabulary Knowledge	5.02	1.30	0	6
Advanced Vocabulary Knowledge	1.10	1.23	0	4
Respondent Schooling	12.96	3.01	0	20
Respondent Age	46.35	15.30	24	89
Cohort Schooling	0.06	1.00	7.65	1.31
General Word Exposure	0.03	0.20	0.41	0.39
Basic Word Exposure	0.03	0.16	0.36	0.35
Advanced Word Exposure	0.03	0.26	0.51	0.48

Note: All statistics are sample weighted. Measures of word exposure include all words and phrases comprising the WORDSUM scale (target words and all choices).

Results

Factorial Structure and Item Functioning

We begin by briefly summarizing our analysis of the measurement structure of WORDSUM, including exploratory factor analysis, IRT analysis, and MGCFA analysis. Full results, including point estimates and visual analysis, can be found in the online supplement. These analyses allowed us to assess the viability of making intercohort comparisons in mean scores on the GSS vocabulary test.

Model results (see online supplement Table A2) provided unequivocal evidence that WORDSUM is composed of two factors, one based on easy words and one based on advanced words (Bowles et al. 2008). A two-factor solution vastly improved the fit of the model to the data compared to a single factor solution. Variation in item difficulties—an individual ability level ranging from -3 to $+3$ that identifies the point at which the respondent has a 50 percent chance of answering the test question correctly⁷—provide further evidence of two factors in this measure of verbal ability. With item difficulties below -0.5 , words A, B, D, E, F, and I measure basic verbal ability. Words C, G, H, and J, with item difficulties above 0.5 , measure advanced verbal ability.

Period and cohort trend analysis of item difficulty parameters identified a slight decrease in item difficulty on the basic words among pre-1945 cohorts and a slight increase in item difficulty on the advanced words beginning with the post-1945 cohorts, but these trends are not linear. After 1945, IRT difficulty trends on the basic words was relatively flat, though several items did increase slightly in difficulty among the most recent birth cohorts.

MGCFA model results, which were estimated in Mplus (Muthén and Muthén 1998–2011; Muthén 2002), indicate that the formal measurement properties of the two-factor measure of vocabulary knowledge do not appreciably vary by birth cohort. Applying equality constraints to the factor loadings and item thresholds (scalar equivalence) resulted in a 0.009 (0.986–0.977) change in the Conditional Fit

Index (CFI) from the unconstrained model (configural equivalence). Other model fit statistics, including the Root Mean Squared Error of Approximation (RMSEA) and Tucker-Lewis index (TLI), show very modest differences between configural and scalar invariant models (0.005, and 0.003, respectively). The period trend in the difficulty parameters is consistent with prior IRT analyses of the GSS verbal ability scores (Beaujean and Sheng 2010; Alwin and Pacheco 2012). These results indicate that changes in the formal measurement properties of the GSS measure of verbal ability are not the cause of the intercohort decline in verbal ability.

By documenting the measurement invariance of the two-factor model for vocabulary measures, the results reported here provide the first evidence for *intercohort* comparability of WORDSUM. This is important because previous research has implicitly assumed factorial invariance without formally testing this assumption. These results also confirm the existence of two factors, basic and advanced vocabulary knowledge, in WORDSUM and establish the intercohort comparability of the GSS measure that is assumed in the regression analyses presented below (Horn, McArdle, and Mason 1983).

We next report results from trend analysis of word exposure and regression analysis of the effect of education selection and word exposure on cohort differences in verbal ability.

Word Exposure

Figure 3 reports exposure to the words comprising the GSS vocabulary test separately for each item and cumulatively in the form of two indexes: one each for the basic and advanced vocabulary factors identified in previous research (Malhotra, Krosnick, and Haertel 2007). As can be seen from the word usage trends, some items have become more common while others have become less common. For example, the 12 words and phrases comprising items A and B have increased in frequency over the twentieth century, while the words comprising items E and I have decreased in frequency. Others, such as Word C and Word F, exhibit nonlinear patterns of exposure.

Exposure to the words comprising the basic vocabulary test items on WORDSUM has become less common over the course of the twentieth century. Exposure to advanced words, on the other hand, declined only slightly prior to 1950, after which these words also became less common (bottom right two panels of Figure 3).

Multivariate Models

A number of modelling strategies have been utilized in prior research studying age, period, and cohort effects (APC) on WORDSUM scores in the GSS data, including linear, hierarchical, and ridge regression models (Yang and Land 2006, 2008; Yang et al. 2008). This body of work failed to detect period influences in the GSS data, instead finding that cohort effects in the GSS data are robust to the presence of measures of age and period factors. The growing consensus from this research is that changes in WORDSUM test scores are not related to factors associated with historical time. These findings, coupled with the absence of a clear theoretical justification for assuming the presence of period influences that affect the general cognitive

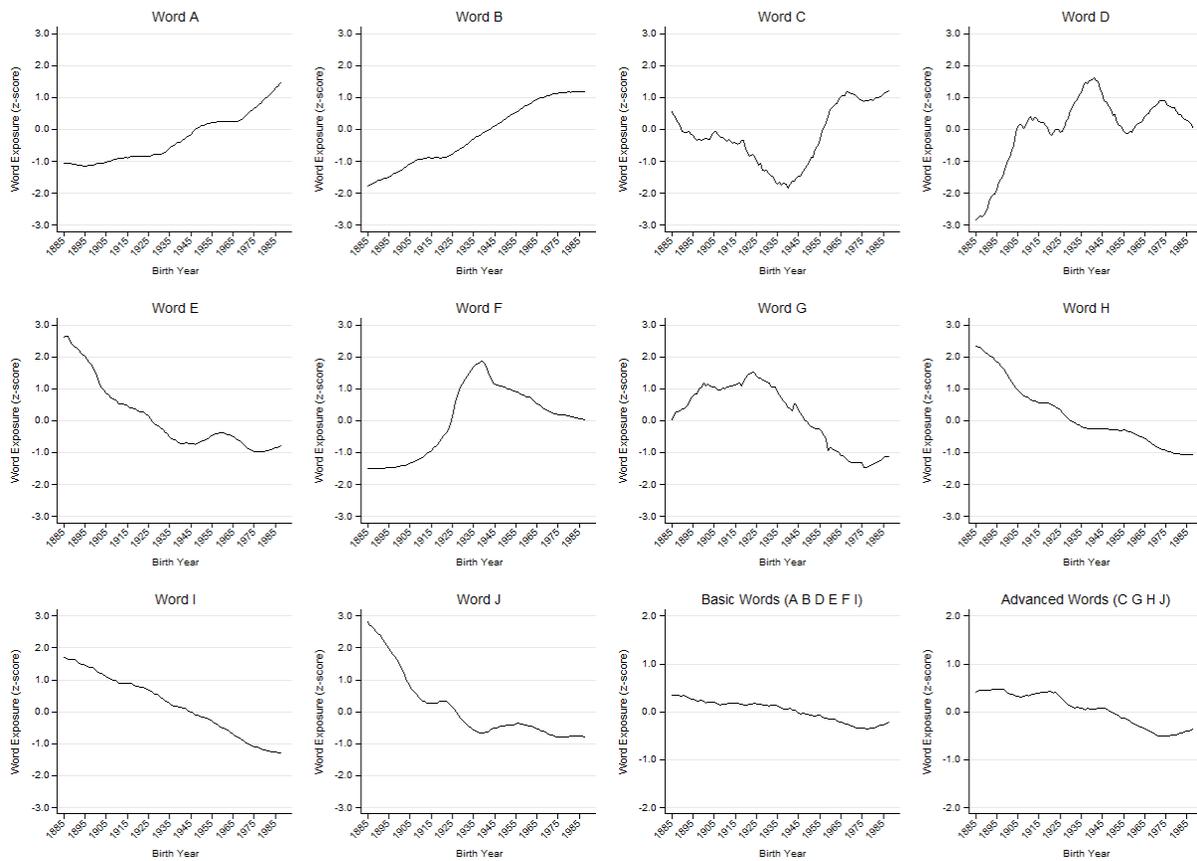


Figure 3: Trends in cohort exposure to WORDSUM test items in books: 1885–1988.

Note: Word exposure measured using the standardized share of WORDSUM test words relative to all words in the Google corpus. We use a 15-year, forward-moving average to smooth the exposure data. See methods section for detailed explanation.

abilities of the entire test-taking population from one occasion of measurement to the next, suggest that linear regression models are appropriate for testing our hypotheses. In the analysis that follows, we deploy a mechanisms-based approach (Winship and Harding 2008) to measure the effect of word exposure and education selection on intercohort differences in verbal ability. Rather than use birth year as a proxy for cohort effects, we deploy measures of the mechanisms hypothesized to differentially transmit verbal ability across birth cohorts, including exposure to formal education and early-life exposure to words on the GSS vocabulary test.

Table 2 contains regression coefficients and model statistics in which we measured the effect of word exposure and education selection on intercohort differences in verbal ability. The relevant comparisons in Table 2 are changes in the magnitude, direction, and statistical significance of the cohort coefficients from the baseline models that control only for respondent schooling and age (models 1, 4, and 7) to the models that include additional controls for education selection (models 2, 5, and 8) and word exposure (models 3, 6, and 9).

Table 2: Regression coefficients expressing the effect of word exposure on basic, advanced, and general vocabulary knowledge

Model	Basic Vocabulary Knowledge			Advanced Vocabulary Knowledge			General Vocabulary Knowledge		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Centered Schooling	0.183* (0.005)	0.183* (0.005)	0.183* (0.005)	0.194* (0.005)	0.195* (0.005)	0.195* (0.005)	0.369* (0.007)	0.370* (0.007)	0.370* (0.007)
Age	0.008 (0.013)	0.008 (0.013)	0.008 (0.013)	0.012 (0.010)	0.011 (0.010)	0.009 (0.010)	0.020 (0.019)	0.018 (0.019)	0.016 (0.020)
Age2	0.029* (0.005)	0.029* (0.005)	0.029* (0.005)	0.019* (0.004)	0.019* (0.004)	0.019* (0.004)	0.048* (0.008)	0.048* (0.008)	0.048* (0.008)
Cohort Schooling		0.028 (0.065)	0.029 (0.064)		0.109* (0.052)	0.105* (0.050)		0.184* (0.088)	0.183* (0.083)
Exposure to Basic Words			0.062 (0.556)					0.596 (1.131)	
Exposure to Advanced Words						0.999* (0.305)			1.574* (0.644)
Birth cohort									
<1900	0.168 (0.134)	0.060 (0.273)	0.072 (0.298)	0.745* (0.104)	0.331 (0.223)	0.027 (0.234)	0.882* (0.192)	0.187 (0.401)	0.267 (0.443)
1901–1905	0.172 (0.185)	0.094 (0.273)	0.102 (0.286)	0.602* (0.112)	0.302 (0.162)	0.045 (0.168)	0.753* (0.251)	0.249 (0.363)	0.075 (0.388)
1906–1910	0.174 (0.124)	0.255 (0.206)	0.249 (0.218)	0.400* (0.071)	0.087 (0.160)	0.152 (0.165)	0.209 (0.166)	0.316 (0.279)	0.633* (0.303)
1911–1915	0.118 (0.079)	0.055 (0.163)	0.063 (0.180)	0.400* (0.049)	0.157 (0.126)	0.135 (0.145)	0.569* (0.109)	0.162 (0.224)	0.223 (0.274)
1916–1920	0.066 (0.060)	0.016 (0.127)	0.023 (0.140)	0.351* (0.063)	0.159 (0.114)	0.173 (0.146)	0.434* (0.108)	0.111 (0.189)	0.349 (0.262)
1921–1925	0.101 (0.080)	0.059 (0.121)	0.066 (0.137)	0.119* (0.059)	0.042 (0.097)	0.336* (0.116)	0.222 (0.127)	0.050 (0.177)	0.442 (0.235)
1926–1930	0.100 (0.066)	0.069 (0.101)	0.076 (0.117)	0.168* (0.060)	0.049 (0.082)	0.058 (0.076)	0.290* (0.125)	0.090 (0.158)	0.009 (0.184)
1931–1935	0.061 (0.051)	0.081 (0.069)	0.075 (0.083)	0.051 (0.068)	0.028 (0.080)	0.035 (0.080)	0.015 (0.098)	0.149 (0.119)	0.104 (0.168)

Table 2 continued

Birth Cohort	Basic Vocabulary Knowledge			Advanced Vocabulary Knowledge			General Vocabulary Knowledge		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1936–1940	0.046 (0.066)	0.059 (0.069)	0.056 (0.070)	0.014 (0.049)	0.033 (0.063)	0.019 (0.066)	0.001 (0.078)	0.079 (0.083)	0.029 (0.112)
1941–1945	ref	ref	ref	ref	ref	ref	ref	ref	ref
1946–1950	0.001 (0.049)	0.016 (0.062)	0.011 (0.080)	0.034 (0.053)	0.023 (0.060)	0.066 (0.052)	0.025 (0.093)	0.071 (0.106)	0.098 (0.103)
1951–1955	0.017 (0.039)	0.007 (0.044)	0.013 (0.077)	0.228* (0.039)	0.189* (0.043)	0.032 (0.062)	0.239* (0.073)	0.174* (0.077)	0.015 (0.124)
1956–1960	0.065 (0.038)	0.055 (0.045)	0.064 (0.104)	0.292* (0.049)	0.251* (0.052)	0.014 (0.096)	0.349* (0.075)	0.279* (0.083)	0.048 (0.192)
1961–1965	0.074 (0.043)	0.062 (0.051)	0.074 (0.129)	0.389* (0.051)	0.341* (0.056)	0.033 (0.139)	0.460* (0.086)	0.380* (0.095)	0.092 (0.264)
1966–1970	0.115 (0.067)	0.098 (0.076)	0.115 (0.191)	0.445* (0.039)	0.377* (0.051)	0.106 (0.153)	0.555* (0.087)	0.442* (0.103)	0.156 (0.341)
1971–1975	0.016 (0.065)	0.005 (0.079)	0.016 (0.217)	0.479* (0.044)	0.398* (0.058)	0.166 (0.177)	0.489* (0.083)	0.352* (0.105)	0.330 (0.388)
1976–1980	0.068 (0.048)	0.087 (0.064)	0.065 (0.221)	0.439* (0.086)	0.364* (0.092)	0.187 (0.188)	0.366* (0.108)	0.240* (0.116)	0.412 (0.395)
1981–1988	0.174 (0.108)	0.162 (0.110)	0.181 (0.209)	0.485* (0.075)	0.439* (0.076)	0.052 (0.166)	0.654* (0.152)	0.577* (0.151)	0.013 (0.358)
Intercept	5.086	5.094	5.096	1.208	1.239	1.174	6.295	6.347	6.259
R2	0.181	0.181	0.181	0.216	0.216	0.217	0.270	0.270	0.271
n	22,684	22,684	22,684	22,684	22,684	22,684	22,684	22,684	22,684

* p < 0.05. Basic vocabulary knowledge measured using six WORDSUM items. Advanced vocabulary knowledge measured using four WORDSUM items. General vocabulary knowledge measured using all 10 WORDSUM items. Results are sample weighted. Models estimated in Stata 13.

The key finding, which can be seen by comparing point estimates in models 1 and 2 and models 5 and 6, is that cohort differences in educational attainment accounted for virtually all of the prewar cohort decline in basic and advanced vocabulary knowledge. The inclusion of a measure of education selection effectively eliminated intercohort test score differences, as measured by cohort specific slopes. In line with our assertion, education selection had a negligible effect on postwar cohort mean test scores. The twin effects of respondent education and cohort education levels on the intercohort trend in verbal ability offers support for arguments regarding the influence of mass education on population trends on standardized cognitive tests (Baker et al. 2015).

We also find that controlling for word exposure substantially influenced the adjusted cohort test score trends in these data compared to the baseline cohort test scores. Exposure to advanced vocabulary words effectively accounted for all of the postwar cohort test score deficit (compare models 4 and 5 to 6). Controlling for exposure to advanced vocabulary words reduced the number of statistically significant cohort coefficients from nine to one (compare models 5 and 6). In nearly all instances, changes in the statistical significance of the cohort coefficients was due to substantial changes in the estimated slopes in the predicted direction, attended by relatively smaller increases in the standard errors.⁸

Exposure to the WORDSUM test content found in English-language printed books exerted substantively meaningful effects on cohort test scores, net of respondent age, education, and average cohort schooling. Controlling for word usage in books contributed to a 50 percent or greater decline in the size of postwar cohort slopes on the full-scale measure of verbal ability (compare models 8 and 9), net of the effects of other variables in the model.

Thus, the substantial cohort increase in average completed schooling during the *first* half of the twentieth century accounts for cohort differences in advanced vocabulary knowledge prior to 1950, and declining exposure to advanced words during the *second* half of the century effectively accounts for cohort variation in advanced vocabulary knowledge after 1950. These findings agree with the Hauser and Huang (1997) assertion for time-varying effects of word exposure on vocabulary knowledge, whereby education selection operates among earlier-born cohorts and word exposure operates among both earlier- and later-born cohorts, though with greater effect on recent birth cohorts. These results provide direct empirical support for both the exposure and selection hypotheses and offer additional evidence in support of environmental accounts of intercohort test score patterns in the GSS data (Flynn 2012).

These results also demonstrate that virtually all of the explanatory power for the cohort slopes on the general vocabulary factor resides in exposure to advanced vocabulary words, leading us to conclude that declining exposure to advanced vocabulary words is an important determinant of macrosocial test score trends in the GSS data. One reason that prewar cohorts perform better on the GSS vocabulary test than postwar birth cohorts is that many of the difficult words on standardized tests such as WORDSUM are, by definition, “old” words (Roivainen 2013). As words age and fall out of fashion, as a great many words invariably do, they become more “difficult” in an empirical sense by virtue of their decreasing usage in various

forms of communication.⁹ The declining prevalence of words gives earlier-born cohorts an advantage on vocabulary tests containing words that have fallen out of favor, as is the case in the GSS test (see Figure 3). Our models show that once we account for the prewar advantage in exposure to advanced words, differences in age, education, and cohort schooling, recent cohorts perform *better* on the test than earlier-born cohorts.

We performed various robustness checks by including additional control variables—such as respondent sex, race, number of siblings, and occupational prestige—and controls for rural upbringing, early life spent in the South, a measure of family intactness, and the educational attainment of the respondent's father and mother. The inclusion of these control variables resulted in a modest improvement in model fit, and all effects were statistically significant, with the exception of family intactness. The presence of additional covariates did not, however, have a meaningful effect on our inferences. If anything, their inclusion in the models *increased* the effect of exposure to advanced words on both general and advanced vocabulary knowledge, providing further support for the conclusions we draw from these analyses.

To summarize the key findings from Table 2 and further establish the main findings of this article, we report the cohort coefficients visually in Figure 4. The substantial over-time decrease in exposure to advanced words on the WORDSUM test accounts for the intercohort decline in verbal ability among post-1950 cohorts. These conclusions are most clearly seen in panels B and C of Figure 4, where we graph cohort test score trends for advanced and general vocabulary. Once we accounted for the substantial increase in schooling during the first half of the century, the cohort trend in advanced verbal ability was effectively flat for the entire period from 1900 to 1988. The effect of differential exposure to vocabulary words was even more pronounced on general verbal ability, which we report in panel C to facilitate comparisons with prior research based on a single factor. In contradiction to prior research (see Figure 1), a model that accounts for education selection and cohort differences in exposure to words provide support for the progress mode (Flynn Effect).

By way of illustration, consider the cohort-specific test scores reported in Figure 4, panel C. The pre-1900 cohorts scored an average of 71 percent on the WORDSUM test ($\mu = 7.1$), compared to an average score of 58 percent for the 1976–1980 cohorts ($\mu = 5.8$) in the baseline models that control for age and time in school. After adjusting for education selection and word exposure, the test score gap between the earliest and latest born cohorts not only decreased by 13 percentage points (7.1–5.8) but reversed direction to one favoring the later-born cohorts by three percentage points (5.9–6.2) for a gross change of 16 percentage points. These data suggest that if recently-born cohorts had been exposed to the test content at equivalent levels as were prewar cohorts, their test scores would be higher than early-born cohorts. In other words, the counterfactual results we present here suggest that the generational rise in crystallized intelligence has not yet abated in America. When properly adjusted for cohort differences in word exposure, the GSS data indicate that the verbal ability of today's young people is not in decline. Rather, they have simply had less exposure to "old" words. Flynn (2012) suggested that one possibility for generational differences on verbal ability tests is a growing gap in the active

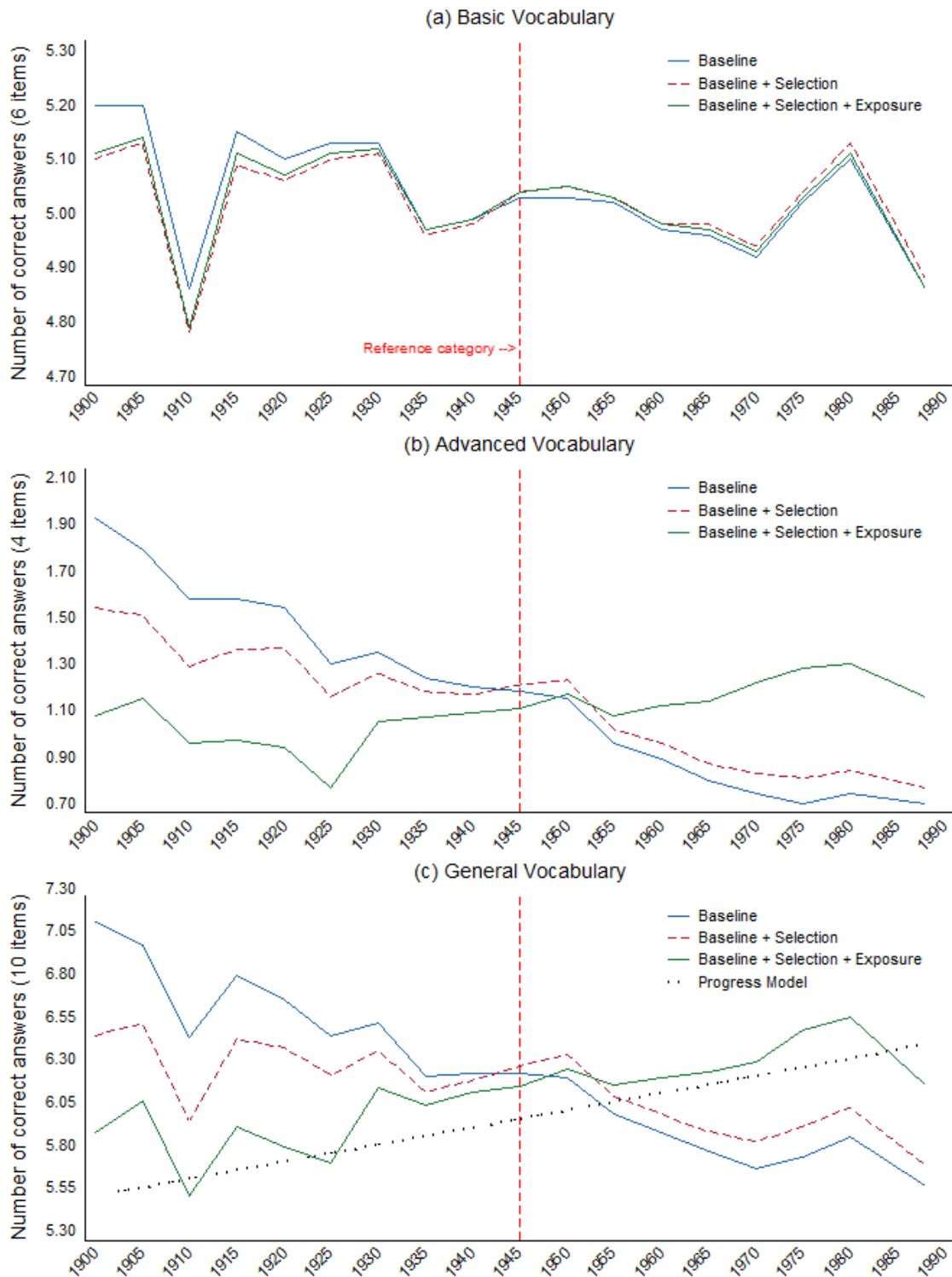


Figure 4: Adjusted cohort test scores in basic, advanced, and general vocabulary knowledge.

Note: Baseline model includes controls for respondent age and education. Selection includes an additional control for mean cohort education levels. Exposure also controls for cohort-specific exposure to test items. Trend line point estimates are derived from Table 2.

vocabulary of old and young people. Our results provide tacit empirical support for these claims.

Discussion and Conclusion

We began this article by drawing attention to a longstanding puzzle in the literature regarding national trends in verbal ability, which indicated a counterintuitive pattern of declining scores on the GSS vocabulary test beginning with cohorts born after WWII. The trends in the GSS vocabulary test score data are not unique, in that they parallel test score declines observed on other standardized tests during the 1960s and 1970s. However, they do not follow the pattern documented in other tests of cognitive ability (Flynn 2007, 2012), namely monotonically rising test scores for each new birth cohort. The Flynn Effect is observed in the raw WORDSUM scores for cohorts born between 1885 and 1950 but not for postwar birth cohorts. The GSS data provide evidence that is completely reversed from the expectations based on a general consensus in the literature that not only is the Flynn Effect real but that it can be interpreted as reflecting an increasing aptitude for cognitive skills being taught in schools and utilized in a growing number of cognitively demanding occupations (Flynn 2007, 2012).

A thorough review of all of the interpretations that have been advanced for these cohort differences was beyond the scope of this article. Instead, we have focused on the possibility that the intercohort WORDSUM trends derive from changing exposure to words and to cohort-specific education selection. Our analysis shows that one of the major mechanisms for the transmission of knowledge is formal schooling, and there is a strong relationship between time spent in school and vocabulary knowledge (Baker et al. 2015). However, the education payoff for vocabulary knowledge in later-born cohorts is not reflected in the words included in common tests of verbal ability, due in part to their declining usage in contemporary books.

Our analysis considered the role of education selection. That education experiences are so intimately tied to what we mean by intellectual ability—certainly in the sense of *crystallized abilities*—makes it virtually impossible to talk about cognitive abilities without taking education into account. Observed test score differences among age groups must therefore also account for age differences in education. Decades ago, Lorge (1956:133) suggested that because “schooling makes for a difference in intelligence test scores...it must follow that the older [earlier born] members of the population are at some disadvantage not only because of their remoteness from formal schooling but also because they had less of it.” More recently, Bandura (1989:734) argued that a major part of age (cohort) differences in test scores “seems to be due to differences in educational experiences across generations rather than to biological aging. It is not so much that the old have declined in intelligence but that the young have enjoyed the benefit of richer intellectual experiences enabling them to function at a higher level.”

Our examination of the selection phenomenon argued that the schooling experiences of members of different cohorts measured in the GSS samples do not reflect intertemporal homogeneity. What we found was somewhat counterintuitive

because cohorts experiencing the highest levels of schooling had the lowest vocabulary test scores (i.e., mean cohort education has a negative effect in our models). So, while the meaning of “years of schooling” may have changed over the twentieth century, the evidence from the GSS data cited earlier suggests education has diminishing returns among more recent cohorts, at least with respect to vocabulary knowledge. The education selection process can be used to account for intercohort differences among prewar birth cohorts but not those cohorts born later.

The research presented here points to a cohort advantage on tests using words that were once relatively common but then fell out of fashion (Alwin 1991). We found strong support for the exposure hypothesis, whereby declining frequency of WORDSUM items in published texts accounts for virtually all of the postwar cohort decline in vocabulary ability. The measure of word exposure introduced in this research is needed to account for the declines in GSS test scores among those post-WWII birth cohorts. For this reason, we recommend that researchers using WORDSUM adjust for differential cohort exposure to the test items.

Our model results provide support for environmental causes of the intercohort differences in verbal ability, including the word exposure hypothesis originally formulated by Alwin (1991) and the education selection argument made by Huang (2003). These results provide a more compelling and theoretically grounded explanation for intercohort variation in vocabulary knowledge than previous explanations that have looked to age and period factors (e.g., Wilson and Gove 1999; Yang and Land 2006, 2008).

Birth cohorts of the twentieth century differ from one another in several key ways, but just two factors, exposure to words and exposure to formal schooling, account for virtually all of the observed between-cohort differences in verbal ability on the GSS test. Our results also suggest that there may be different causal processes at different periods in the macrosocial trend (i.e., different explanations at different historical periods). A promising line of future research would be to replicate our modelling strategy among the populations of advanced economies to see if recent declines in test scores are due to changes in exposure and selection that we observed in American samples (for a summary of recent findings, see Dutton and Lynn 2013).

Notes

- 1 We remind reviewers that the relationship between vocabulary test scores and intelligence only holds within-cohorts.
- 2 The GSS protocol for administering the vocabulary test is as follows: we would like to know something about how people go about guessing words they do not know. On this card are listed some words—you may know some of them, and you may not know quite a few of them. The card handed to the respondent included instructions as follows: Please look first at the word in capital letters on each line. Then look at the other words in smaller type on the same line and tell me which one of these words comes closest in meaning to the one in capital letters. EXAMPLE BEAST 1. Afraid 2. Words 3. Large 4. Animal 5. Bird The correct answer in this example is No. 4 because the word “animal” comes closer to “beast” than any of the other words. On each line, the first word is in capital letters—like BEAST. Then, there are five other words. Tell me the number of the word that comes closest to the meaning of the word in capital letters. For example, if

the word in capital letters is BEAST, you would say “4” since “animal” comes closer to BEAST than any of the other words. If you wish, I will read the words to you. These words are difficult for almost everyone—just give me your best guess if you are not sure of the answer.

- 3 The Ngram database contains all words appearing at least 40 times in the corpus. To date, Google has supplied six searchable American English databases (one-gram, two-gram, three-gram, four-gram, and five-gram databases). The one-gram database contains single words (e.g., fish), the two-gram database contains combinations of up to two words that were adjacent to each other in the scanned books (e.g., catching fish), the three-gram database contains combinations of three adjacent words (likes catching fish), and so on.
- 4 Many books were provided to the Google Books project directly in digital format, thus eliminating the scanning and optical character recognition stage of data collection. Post-2000 books skew heavily toward fictional texts, which have substantially different readership among the general public (Michel et al. 2011).
- 5 An appropriate test of the word exposure hypothesis calls for a measure of exposure that accounts for change in WORD usage relative to all other words in the database rather than a simple change in the absolute occurrence of each WORD. Since WORDSUM weights each word equally, our summary measure of exposure needed to do the same. We accomplished this through a standardization of each of the 60 WORD scores. This is important because overall usage of each WORD varied considerably, with some words being very common and others obscure (in some cases by an order of magnitude of two or more). In a composite scale, using unstandardized frequencies would allow a small number of high frequency words to overwhelm the effects of less frequently used words.
- 6 Multiple choice tests such as the one employed in the GSS contain an inherent dependency among the target word and answer choices. That is, a respondent that knows the meaning of the incorrect choices might ascertain the correct meaning of the target word by process of elimination, even if they have not been exposed to the target word itself.
- 7 Item difficulty parameters are not to be confused with item pass rates, or proportion correct. The former is a model-based parametric estimate, while the latter is a nonparametric statistic.
- 8 As is the case in models that contain a large number of dummy variables, multicollinearity is present in these data. While this does not affect the slopes, collinearity does inflate the standard errors, making it difficult to interpret tests of statistical significance. Collinearity was substantially lower in the models involving advanced and total vocabulary knowledge.
- 9 Related analysis of the relationship between word frequencies in the Google corpus and GSS test scores by Woodley et al. (2015) lead to the same conclusion, which is that old words have lower pass rates than new words.

References

- Alexander, Karl L. 1997. “Public Schools and the Public Good.” *Social Forces* 76:1-30. <http://dx.doi.org/10.1093/sf/76.1.1>
- Alwin, Duane F. 1991. “Family of Origin and Cohort Differences in Verbal Ability.” *American Sociological Review* 56:625-38. <http://dx.doi.org/10.2307/2096084>
- Alwin, Duane F. 2009. “History, Cohorts, and Patterns of Cognitive Aging.” Pp. 9-38 in *Aging and cognition: Research methodologies and empirical advances*, edited by H.B. Bosworth and

- C. Hertzog. Washington DC: American Psychological Association. <http://dx.doi.org/10.1037/11882-001>
- Alwin, Duane F., and Ryan J. McCammon. 1999. "Aging vs. Cohort Interpretations of Inter-Cohort Differences in GSS Verbal Scores." *American Sociological Review* 64:272-286. <http://dx.doi.org/10.2307/2657532>
- Alwin, Duane F., and Ryan J. McCammon. 2001. "Aging, Cohorts, and Verbal Ability." *Journal of Gerontology: Social Sciences* 56B:S1-S11. <http://dx.doi.org/10.1093/geronb/56.3.s151>
- Alwin, Duane F., and Ryan J. McCammon. 2003. "Generations, Cohorts and Social Change." Pp. 23-49 in *Handbook of the Life Course*, edited by Jeylan T. Mortimer and Michael J. Shanahan. New York, NY: Kluwer Academic/Plenum Publishers. http://dx.doi.org/10.1007/978-0-306-48247-2_2
- Alwin, Duane F., and Ryan J. McCammon. 2007. "Rethinking Generations." *Research in Human Development* 4:219-237. <http://dx.doi.org/10.1080/154276007016663072>
- Alwin, Duane F., and Julianna Pacheco. 2012. "Population Trends in Verbal Intelligence in the United States." Pp. 338-368 in *Social Trends in American Life: Findings from the General Social Survey since 1972*, edited by Peter V. Marsden. Princeton, NJ: Princeton University Press. <http://dx.doi.org/10.1515/9781400845569-015>
- Alwin, Duane F., Ryan J. McCammon, Linda A. Wray, and Willard L. Rodgers. 2008. "Population Processes and Cognitive Aging." Pp. 69-89 in *Handbook of Cognitive Aging: Interdisciplinary Perspectives*, edited by Scott M. Hofer and Duane F. Alwin. Thousand Oaks, CA: Sage Publications. <http://dx.doi.org/10.4135/9781412976589.n4>
- Baker, David B, Paul J. Eslinger, Martin Benavides, Ellen Peters, Nathan F. Diechmann, and Juan Leon. 2015. "The Cognitive Impact of the Education Revolutions: A Possible Cause of the Flynn Effect on Population IQ." *Intelligence* 49:144-158. <http://dx.doi.org/10.1016/j.intell.2015.01.003>
- Bailey, Kent G., and Edward J. Federman. 1979. "Factor Analysis of Breadth and Depth Dimensions on Wechsler's Similarities and Vocabulary Subscales." *Journal of Clinical Psychology* 35:341-345. [http://dx.doi.org/10.1002/1097-4679\(197904\)35:2<341::AID-JCLP2270350224>3.0.CO;2-C](http://dx.doi.org/10.1002/1097-4679(197904)35:2<341::AID-JCLP2270350224>3.0.CO;2-C)
- Baltes, Paul B. 1987. "Theoretical Propositions of Life-Span Developmental Psychology: On the Dynamics Between Growth and Decline." *Developmental Psychology* 23:611-626. <http://dx.doi.org/10.1037/0012-1649.23.5.611>
- Bandura, Albert. 1989. "Regulation of Cognitive Processes Through Perceived Self-Efficacy." *Developmental Psychology* 25:729-735. <http://dx.doi.org/10.1037/0012-1649.25.5.729>
- Beaujean, A.A., and Y Sheng. 2010. "Examining the Flynn Effect in the General Social Survey Vocabulary Test using Item Response Theory". *Personality and Individual Differences*, 48:294-298. <http://dx.doi.org/10.1016/j.paid.2009.10.019>
- Beck, Niels C., David Tucker, Robert Frank, Jerry Parker, Rebecca Lake, Susan Thomas, Wemara Lichty, Ellen Horwitz, Bruce Horwitz, and Frank Merritt. 1989. "The Latent Factor Structure of the WAIS-R: A Factor Analysis of Individual Item Responses." *Journal of Clinical Psychology* 45:281-293.
- Blake, Judith. 1989. *Family size and achievement*. Berkeley, CA: University of California Press.
- Bowles, Ryan P., Kevin J. Grimm, and John J. McArdle. 2008. "A Structural Factor Analysis of Vocabulary Knowledge and Relations to Age." *Journal of Gerontology: Psychological Sciences* 60B:234-P241.

- Brysbart, Marc, Emmanuel Keuleers, and Boris New. 2011. "Assessing the Usefulness of Google Books Word Frequencies for Psycholinguistic Research on Word Processing." *Frontiers in Psychology* 2:1-8. <http://dx.doi.org/10.3389/fpsyg.2011.00027>
- Cattell, Raymond B. 1963. "Theory of Fluid and Crystallized Intelligence: A Critical Experiment." *Journal of Educational Psychology* 54:1-22. <http://dx.doi.org/10.1037/h0046743>
- Cattell, Raymond B. 1971a. "The Structure of Intelligence in Relation to the Nature-Nurture Controversy." Pp. 3-30 In *Intelligence: Genetic and Environmental Influences*, edited by Robert Cancro. New York: Grune and Stratton.
- Cattell, Raymond B. 1971b. *Abilities: Their Structure, Growth and Action*. Boston: Houghton-Mifflin.
- Chall, Jeanne S. 1983. "Literacy: Trends and Explanations." *Educational Researcher* 12:3-8. <http://dx.doi.org/10.3102/0013189X012009003>
- Chall, Jeanne S. 1996. "American Reading Achievement: Should We Worry?" *Research in the Teaching of English* 30:303-312.
- Chall, Jeanne S., and Sue S. Conard. 1991. *Should Textbooks Challenge Students? The Case for Easier or Harder Books*. New York, NY: Teachers College Press.
- Chall, Jeanne, S., Sue S. Conard, and Susan Harris. 1977. *An Analysis of Textbooks in Relation to Declining SAT Scores*. Princeton, NJ College Entrance Examination Board and Educational Testing Service.
- Chandler, Michael Alison. 2011. "SAT Reading Scores Drop to Lowest Point in Decades." *The Washington Post*, September 14, 2011.
- Cleary, T. Anne, and Sam A. McCandless. 1976. "Summary of Score Changes (in Other Tests)." Princeton, NJ: College Entrance Examination Board.
- Congressional Budget Office. 1986. *Trends in Educational Achievement*. Washington, D.C.: The Congress of the United States.
- Congressional Budget Office. 1987. *Educational Achievement: Explanations and Implications of Recent Trends*. Washington, D.C.: The Congress of the United States.
- Davis, James A., Tom W. Smith and Peter V. Marsden. 2012. *General Social Surveys, 1972-2012: Cumulative Codebook*. [machine-readable data file]. Chicago: National Opinion Research Center, producer, 2008; Storrs, CT: The Roper Center for Public Opinion Research, University of Connecticut, distributor.
- Denny, Nancy Wadsworth. 1982. "Aging and Cognitive Changes." Pp. 807-27 in *Handbook of Developmental Psychology: Research and Theory*, edited by B.B. Wolman. Englewood Cliffs, NJ: Prentice-Hall.
- Dutton, Edward, and Richard Lynn. 2013. "A Negative Flynn effect in Finland, 1997-2009". *Intelligence*. 41: 817-820. <http://dx.doi.org/10.1016/j.intell.2013.05.008>
- Embretson, Susan E., and Steven P. Reise. 2000. *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ferguson, Richard L. 1976. "The Decline in ACT Test Scores: What Does it Mean?" *Educational Technology* 16:21-27.
- Firebaugh, Glenn. 1992. "Where Does Social Change Come From? Estimating the Relative Contributions of Individual Change and Population Turnover." *Population Research and Policy Review* 11:1-20. <http://dx.doi.org/10.1007/BF00136392>
- Flynn, James R. 1984. "The Mean IQ of Americans: Massive Gains." *Psychological Bulletin* 95:29-51. <http://dx.doi.org/10.1037/0033-2909.95.1.29>
- Flynn, James R. 1987a. "Causal Factors in Generational IQ Gains 1932 to 1978." *Nature* 328 (6133): 765-765. <http://dx.doi.org/10.1038/328765b0>

- Flynn, James R. 1987b. "Massive IQ Gains in 14 Nations: What IQ Tests Really Measure." *Psychological Bulletin* 101:171-191. <http://dx.doi.org/10.1037/0033-2909.101.2.171>
- Flynn, James R. 1998. "IQ Gains Over Time: Toward Finding the Causes". Pp. 25-66 in *The Rising Curve: Long-term Gains in IQ and Related Measures*, edited by Ulrich Neisser. Washington D.C.: American Psychological Association. <http://dx.doi.org/10.1037/10270-001>
- Flynn, James R. 1999. "Searching for Justice: The Discovery of IQ Gains Over Time." *American Psychologist* 54:5-20. <http://dx.doi.org/10.1037/0003-066X.54.1.5>
- Flynn, James R. 2007. *What is Intelligence? Beyond the Flynn Effect*. Cambridge, UK: Cambridge University Press. <http://dx.doi.org/10.1017/CB09780511605253>
- Flynn, James R. 2012. *Are We Getting Smarter? Rising IQ in the Twenty-First Century*. Cambridge, UK: Cambridge University Press. <http://dx.doi.org/10.1017/CB09781139235679>
- Gardner, David P., Yvonne W. Larsen and others. 1983. *A Nation at Risk: The Imperative for Educational Reform*. Washington, D.C.: U.S. Government Printing Office.
- Glenn, Norval D. 1994. "Television Watching, Newspaper Reading, and Cohort Differences in Verbal Ability." *Sociology of Education* 67:216-230. <http://dx.doi.org/10.2307/2112792>
- Glenn, Norval D. 1999. "Further Discussion of the Evidence for an Inter-Cohort Decline in Education-Adjusted Vocabulary." *American Sociological Review* 64:267-271. <http://dx.doi.org/10.2307/2657531>
- Glöckner-Rist, Angelika, and Herbert Hoijtink. 2003. "The Best of Both Worlds: Factor Analysis of Dichotomous Data Using Item Response Theory and Structural Equation Modeling." *Structural Equation Modeling* 10:544-565. http://dx.doi.org/10.1207/S15328007SEM1004_4
- Gustafsson, Jan-Eric, and Lena M. Holmberg. 1992. Psychometric Properties of Vocabulary Test Items as a Function of Word Characteristics. *Scandinavian Journal of Educational Research* 36:191-210. <http://dx.doi.org/10.1080/0031383920360303>
- Hauser, Robert M., and Min-Hsiung Huang. 1997. Verbal Ability and Socioeconomic Success: A Trend Analysis. *Social Science Research* 26:331-76. <http://dx.doi.org/10.1006/ssre.1997.0604>
- Hayes, Donald, Loreen Wolfer, and Michael Wolfe. 1996. Textbook Simplification and its Relation to the Decline in SAT-Verbal Scores. *American Educational Research Journal* 33:489-508. <http://dx.doi.org/10.3102/00028312033002489>
- Harnischfeger, Annegret, and David E. Wiley. 1976. "The Marrow of Achievement Test Scores Declines." *Educational Technology* 16:5-14.
- Herrnstein, R.J., and Murray, C. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.
- Hofer, Scott M., and Duane F. Alwin. 2008. *Handbook of Cognitive Aging: Interdisciplinary Perspectives*. Thousand Oaks, CA: Sage Publications.
- Horn, J.L., McArdle, J.J., and R. Mason. 1983. "When is Invariance not Invariant: A Practical Scientist's View of the Ethereal Concept of Factorial Invariance." *The Southern Psychologist*, 1, 179-188.
- Horn, John L. 1982a. "The Theory of Fluid and Crystallized Intelligence in Relation to Concepts of Cognitive Psychology and Aging in Adulthood." Pp. 237-278 in *Aging and Cognitive Processes*, edited by F.I.M. Craik and S. Trehub. New York: Plenum. http://dx.doi.org/10.1007/978-1-4684-4178-9_14

- Horn, John L. 1982b. "The Aging of Human Abilities." Pp. 847-870 in *Handbook of Developmental Psychology: Research and Theory*, edited by B.B. Wolman. Englewood Cliffs, NJ: Prentice-Hall.
- Horn, John L., and Raymond B. Cattell. 1967. "Age Differences in Fluid and Crystallized Intelligence." *Acta Psychologica* 26:107-129. [http://dx.doi.org/10.1016/0001-6918\(67\)90011-X](http://dx.doi.org/10.1016/0001-6918(67)90011-X)
- Horn, John L., and Gary Donaldson. 1976. "On the Myth of Intellectual Decline in Adulthood." *American Psychologist* 31:701-719. <http://dx.doi.org/10.1037/0003-066X.31.10.701>
- Horn, John L., and Gary Donaldson. 1980. "Cognitive Development in Adulthood." Pp. 445-529 in *Constancy and Change in Human Development*, edited by Orville G. Brim, Jr. and Jerome Kagan. Cambridge, MA: Harvard University Press.
- Huang, Min-Hsiung. 2003. "Has Verbal Ability Declined in America?" *EurAmerica* 33:757-800.
- Jencks, Christopher. 1979. "Why Students Aren't Learning." *The Center Magazine*, July/Aug., pp. 12-14.
- Jensen, Arthur R. 1998. *The G Factor: The Science of Mental Ability*. Greenwood Publishing: Westport, CT.
- Kučera, Henry, and W. Nelson Francis. 1967. *Computational Analysis of Present-Day American English*. Providence, R.I.: Brown University Press.
- Laufer, B. 1997. "What's in a Word That Makes it Hard or Easy? Some Intralexical Factors that Affect the Learning of Words". *Vocabulary: Description, Acquisition and Pedagogy*, Cambridge: Cambridge University Press
- Layton, Lyndsey, and Emma Brown. "SAT Reading Scores Hit a Four-Decade Low." *Washington Post*, September 24, 2012.
- Loge, Irving. 1956. "Aging and intelligence." *Journal of Chronic Diseases* 412:131-139. [http://dx.doi.org/10.1016/0021-9681\(56\)90013-3](http://dx.doi.org/10.1016/0021-9681(56)90013-3)
- Lynn, Richard. 2009. "What Has Caused the Flynn Effect? Secular Increases in the Development Quotients of Infants". *Intelligence*, 37:16-24. <http://dx.doi.org/10.1016/j.intell.2008.07.008>
- Malhotra, Neil, Jon A. Krosnick, and Edward Haertel. 2007. "The Psychometric Properties of the GSS Wordsum Vocabulary Test." Unpublished GSS Trend Report: General Social Survey, NORC, Chicago, IL.
- Mannheim, Karl. 1952. "The Problem of Generations." Pp. 276-322 in *Essays in the Sociology of Knowledge*, edited by P. Kecskemeti. Boston: Routledge and Kegan Paul. (Original work published in 1927).
- Menard, Scott. 1988. "Going Down, Going Up: Explaining the Turnaround in SAT Scores." *Youth and Society* 20:3-28. <http://dx.doi.org/10.1177/0044118X88020001001>
- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books". *Science*. 331:176-182. <http://dx.doi.org/10.1126/science.1199644>
- Miner, John B. 1957. *Intelligence in the United States*. New York: Springer Publishing Company.
- Morgan, Michael. 1986. "Television and Adults' Verbal Intelligence." *Journalism Quarterly* 63:537-541. <http://dx.doi.org/10.1177/107769908606300313>
- Muthén, Bengt O. 2002. "Beyond SEM: General Latent Variable Modeling." *Behaviormetrika* 29:81-117. <http://dx.doi.org/10.2333/bhmk.29.81>

- Muthén, Linda K., and Bengt O. Muthén. 1998-2011. *Mplus—Statistical Analysis with Latent Variables: User's Guide*. Version 6.12. Los Angeles, CA: Muthén and Muthén.
- Neisser, Ulrich. 1997. "Rising Scores on Intelligence Tests." *American Scientist* 85:440-47.
- Neisser, Ulrich. 1998. *The Rising Curve: Long Term Gains in IQ and Related Measures*. Washington, D.C.: American Psychological Association. <http://dx.doi.org/10.1037/10270-000>
- Park, Denise C., et al. 1996. "Mediators of Long Term Memory Performance across the Life Span." *Psychology of Aging* 11:621-637. <http://dx.doi.org/10.1037/0882-7974.11.4.621>
- Roivainen, Eka. 2013. "Changes in Word Usage Frequency May Hamper Intergenerational Comparisons of Vocabulary Skills: An Ngram Analysis of Wordsum, WAIS, and WISC Test Items." *Journal of Psychoeducational Assessment*. 32: 83-87. <http://dx.doi.org/10.1177/0734282913485542>
- Ryder, Norman B. 1965. "The Cohort as a Concept in the Study of Social Change." *American Sociological Review* 30:843-861. <http://dx.doi.org/10.2307/2090964>
- Salthouse, Timothy A. 1991. *Theoretical Perspectives on Cognitive Aging*. Hillsdale, NJ: Lawrence Erlbaum.
- Schaie, K. Warner. 1984. "Historical Time and Cohort Effects." Pp. 1-15 in *Lifespan Developmental Psychology: Historical and Generational Effects*, edited by K.A. McCloskey and H.W. Reese. New York: Academic Press. <http://dx.doi.org/10.1016/B978-0-12-482420-1.50006-4>
- Schaie, K. Warner. 2005. *Developmental Influences on Adult Intelligence: The Seattle Longitudinal Study*. Oxford UK: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780195156737.001.0001>
- Schaie, K. Warner. 2008. "Historical Patterns and Processes of Cognitive Aging." Pp. 368-383 in *Handbook of Cognitive Aging: Interdisciplinary Perspectives*, edited by Scott M. Hofer and Duane F. Alwin. Thousand Oaks, CA: Sage Publications. <http://dx.doi.org/10.4135/9781412976589.n23>
- Schaie, K. Warner, Sherry L. Willis, and Sara Pennak. 2005. "A Historical Framework for Cohort Differences in Intelligence." *Research in Human Development* 2:43-67. <http://dx.doi.org/10.1080/15427609.2005.9683344>
- Schramm, Walter. 1977. *Television and the Test Scores*. Princeton, NJ: College Entrance Examination Board.
- Segui, J., Jacques Mehler, Uli Frauenfelder, and John Morton. 1982. "The Word Frequency Effect and Lexical Access", *Neuropsychologia* 20:615-627. [http://dx.doi.org/10.1016/0028-3932\(82\)90061-6](http://dx.doi.org/10.1016/0028-3932(82)90061-6)
- Stedman, Lawrence C. 1996. "An Assessment of Literacy Trends, Past and Present." *Research in the Teaching of English* 30:283-302.
- Stern, Yaakov. 2007. *Cognitive Reserve: Theory and Applications*. New York, NY: Taylor and Francis Group.
- te Nijenhuis, J., and van der Flier, H. 2013. "Is the Flynn effect on G?: A Meta-Analysis" *Intelligence*, 41:802-807. <http://dx.doi.org/10.1016/j.intell.2013.03.001>
- Thorndike, Robert L. 1921. *The Teacher's Word Book*. New York, NY: Teachers College, Columbia University.
- Thorndike, Robert L. 1931. *The Teacher's Word Book of the Twenty Thousand Words Found Most Frequently and Widely in General Reading for Children and Young People*. New York, NY: Teachers College, Columbia University.

- Thorndike, Robert L. 1942. "Two Screening Tests of Verbal Intelligence." *Journal of Applied Psychology* 26:128-135. <http://dx.doi.org/10.1037/h0060053>
- Thorndike, Robert L., and George H. Gallup. 1944. "Verbal Intelligence of the American Adult." *Journal of General Psychology* 30:75-85. <http://dx.doi.org/10.1080/00221309.1943.10544458>
- Tuddenham, R.D. 1948. "Soldier Intelligence in World Wars I and II." *American Psychologist*, 3: 54-56. <http://dx.doi.org/10.1037/h0054962>
- Verhaeghen, Paul. 2003. "Aging and Vocabulary Scores: A Meta-Analysis." *Psychology and Aging* 18:332-339. <http://dx.doi.org/10.1037/0882-7974.18.2.332>
- Wilson, James, and Walter Gove. 1999. "The Inter-Cohort Decline in Verbal Ability: Does it Exist?" *American Sociological Review* 64:253-66. <http://dx.doi.org/10.2307/2657530>
- Winter, David G. 1977. *Motivational Factors in the SAT Score Decline*. Princeton, NJ: College Entrance Examination Board.
- Winship, Christopher, and David Harding. 2008. "A Mechanism-Based Approach to the Identification of Age-Period-Cohort Models." *Sociological Methods and Research* 36:362-401. <http://dx.doi.org/10.1177/0049124107310635>
- Wirtz, Willard, Harold Howe II, and others. 1977. *On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline*. New York: College Entrance Examination Board.
- Wolfe, Lee M. 1980. "The Enduring Effects of Education on Verbal Skills." *Sociology of Education* 53:104-114. <http://dx.doi.org/10.2307/2112492>
- Woodley of Menie, M.A., H.B.F. Fernandes, A.J. Figueredo, and Meisenberg, G. 2015. "By Their Words Ye Shall Know Them: Evidence of Genetic Selection Against General Intelligence and Concurrent Environmental Enrichment in Vocabulary Usage Since the Mid-19th Century." *Frontiers in Psychology* 6:361.
- Yang, Yang, and Kenneth C. Land. 2006. "A Mixed Models Approach to the Age-Period-Cohort Analysis of Repeated Cross-Section Surveys, With an Application to Data on Trends in Verbal Test Scores." *Sociological Methodology* 36:75-97. <http://dx.doi.org/10.1111/j.1467-9531.2006.00175.x>
- Yang, Yang, and Kenneth C. Land. 2008. "Age-Period-Cohort Analysis of Repeated Cross-Section Surveys: Fixed or Random Effects?" *Sociological Methods and Research*, 36, 297-326. <http://dx.doi.org/10.1177/0049124106292360>
- Yang, Yang, Sam Schulhofer-Wohl, Wenjiang J. Fu, and Kenneth C. Land. 2008. "The Intrinsic Estimator for Age-Period-Cohort Analysis: What it is and How to Use It." *American Journal of Sociology* 113:1697-736. <http://dx.doi.org/10.1086/587154>
- Zajonc, Robert. 1976. "Family Configuration and Intelligence." *Science* 192:227-236. <http://dx.doi.org/10.1126/science.192.4236.227>

Acknowledgements: This research was supported by National Institute of Child Health and Human Development Training Grant from the Population Studies Center at the University of Michigan (T32 HD007339). Duane Alwin was supported in part by a grant from the National Institute of Aging (R01AG021203), a grant from the National Science Foundation (SES-1331454), and the McCourtney endowment, College of the Liberal Arts, Pennsylvania State University, during the writing of this article. Please direct correspondence to Shawn F. Dorius (sdorius@iastate.edu).

Shawn F. Dorius: Department of Sociology, Iowa State University.
E-mail: sdorius@iastate.edu.

Duane F. Alwin: Department of Sociology and Criminology, Pennsylvania State University.
E-mail: dfa2@psu.edu.

Julianna Pacheco: Department of Political Science, University of Iowa.
Email: julianna-pacheco@uiowa.edu.