

Secular Declines in Spearman's *g*: Some Evidence From the United States

HARRISON KANE

*College of Education
University of Nevada, Las Vegas*

THOMAS D. OAKLAND

*Educational Psychology Department
University of Florida*

ABSTRACT. The Flynn Effect (J. Flynn, 1987) refers to the apparent increases in intelligence quotient (IQ) observed over the past few decades. A related phenomenon is that the variance in test scores accounted for by Spearman's *g* (C. Spearman, 1904) varies according to IQ level. That is, *g* accounts for less variance in high IQ groups than in low IQ groups. Spearman termed this variant aspect of *g* the "law of diminishing returns." This study extends prior research on the Flynn Effect and the law of diminishing returns by examining changes in the statistical importance of Spearman's *g* that may accompany secular increases in IQ. Based on the standardization data from the United States versions of the Wechsler scales (i.e., Wechsler Preschool and Primary Scales of Intelligence, D. Wechsler, 1967; Wechsler Intelligence Scale for Children, D. Wechsler, 1949; Wechsler Adult Intelligence Scale, D. Wechsler, 1955), this study indicated that, in most cases, the statistical significance of Spearman's *g* has indeed declined over the past several years. A. R. Jensen (1998) suggested that the components of the Flynn Effect warrant special investigation. The present study addresses this issue, in part by analyzing the Verbal and Performance subscales of the Wechsler scales. Additionally, this study further confirms the law of diminishing returns as applicable to different ages and time periods.

CHARLES E. SPEARMAN (1904) introduced the construct of general intelligence (*g*) to describe the common variance shared by a battery of cognitive tests. Usually represented as the first unrotated principal component in a factor analysis of a test battery (see Jensen & Weng, 1994), Spearman's *g* routinely accounts for around 50% of the total variance in scores from cognitive tests (Jensen, 1998). Although the topic of considerable scientific and societal debate (e.g., Gould, 1981; Herrnstein & Murray, 1994), *g* has endured as a viable scientific phenom-

Address correspondence to Harrison Kane, College of Education, Department of Special Education, University of Nevada, 4505 Maryland Parkway, Las Vegas, Las Vegas, NV 89154.

enon and is compatible with most psychometric theories of intelligence (Carroll, 1993; Jensen, 1998).

On a related topic, the mean level of intelligence in a number of countries seemingly has increased significantly over the past few decades, with total accumulated gains usually on the order of 15 intelligence quotient (IQ) points. These increases in measured IQ have been termed "massive" (e.g., Flynn, 1987). The term usually associated with these secular gains is the Flynn Effect, referencing the author of the most cited review addressing this phenomenon. In most countries where the Flynn Effect has been observed, economic growth and development have accompanied measured increases in IQ.

The underlying causes of the Flynn Effect are open to debate. Some researchers (e.g., Teasdale & Owen, 1987) have attributed the increases to greater cognitive stimulation afforded through environmental improvements (e.g., television, toys, and schooling). Other researchers (e.g., Lynn, 1992) have underscored improved nutrition and health care. Still other researchers (e.g., Neisser, 1998) believe these impressive gains simply represent the increasing overlap between material on intelligence tests and information common to a specific culture and society.

One means of gaining insight into the underlying nature of the Flynn Effect is to investigate the factor structures of intelligence tests from different periods of time. A theoretical justification for this procedure may be found in Spearman's law of diminishing returns. In *The Abilities of Man*, Spearman (1927) noted that correlations between test scores vary according to intellectual ability, with correlations being higher among individuals with low IQ than among those with high IQ. Spearman remarked:

Now, all the changes we have been considering follow a general rule. The correlations always become smaller—showing the influence of *g* on any ability to grow less—in just the classes of person which, on the whole, possess this *g* more abundantly. The rule is, then, that the more "energy" a person has available already, the less advantage accrues to his ability from further increments of it. (p. 219)

Spearman did not write extensively on this finding, mentioning it only in light of the "diminishing returns of *g*" (p. 127). This peculiar aspect of *g* remained largely uninvestigated until Detterman and Daniel (1989) divided the standardization samples of the U.S. versions of the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) and the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) into five ability groups and calculated the intercorrelations between subtests. They found that the average intercorrelation score between subtests declined significantly as ability increased. For those with IQs below 78, intercorrelations averaged .56, uncorrected for range restriction; for those with IQs above 122, intercorrelations averaged .22. These results confirmed Spearman's law of diminishing returns and suggested that the amount of variance attributable to *g* varies inversely with intelligence. Other studies

(Brand, 1996) have found similar results, confirming that g exerts a greater influence at the lower end of the intellectual spectrum. In particular reference to the Flynn Effect, Spearman's law of diminishing returns offers the intriguing possibility that, as scores on standardized intelligence tests increase over time, there may be a concomitant decrease in the variance accounted for by g . In other words, Spearman's law of diminishing returns also should occur on a secular level. The magnitude of test intercorrelations serves as a barometer of the influence of g ; therefore, information on test intercorrelations across time should provide reasonably straightforward information as to the nature of possible secular changes in cognitive abilities.

The purpose of the present study was to investigate the hypothesis of a decline in g using standardization data obtained from the U.S. versions and revisions of the Wechsler Preschool and Primary Scales of Intelligence (WPPSI; Wechsler, 1967), the WISC (Wechsler, 1949), and the WAIS (Wechsler, 1955).

Method

The WAIS was standardized on 1,700 American adults selected as representative of the population on the basis of the 1950 U.S. census. There were seven different age groups, ranging from 16 to 64 years, with 200 participants at the three age levels between 16 and 24 years (i.e., 16-17, 18-19, and 20-24), 300 participants at the age levels between 25 and 54 years (i.e., 25-34, 35-44, and 45-54), and 200 participants in the 55- to 64-year-old group. An equal number of males and females participated in the standardization.

The WAIS-III (Wechsler, 1997) is the latest revision of the original WAIS and represents an attempt to sample a greater array of cognitive abilities. The WAIS-III standardization sample included 1,250 individuals ranging from 16 to 89 years old. Thirteen age levels were represented, with between 75 and 100 individuals at each level. The WAIS-III standardization sample was stratified to ensure that it was representative of the 1990 U.S. census population.

The WISC was designed as a downward extension of Form II of the Wechsler-Bellevue Intelligence Scale (1939) and intended for use with children between 5 and 15 years old. The WISC standardization sample was representative of the 1940 U.S. census population and consisted of 2,200 American children, with 100 boys and 100 girls at each of 11 age levels. Unlike the WAIS, which included non-Whites (e.g., Blacks, American Indians, Asians, and Puerto Ricans) in its standardization sample, the WISC relied exclusively on White children. The exclusion of minorities serves to reduce the variance of WISC test scores and therefore attenuates subtest intercorrelations. However, the present study was conducted with the assumption that the Flynn Effect, and the accompanying decline in the statistical significance of Spearman's g , would be sufficiently pronounced as to overcome this inexplicable shortcoming in the WISC standardization.

The standardization procedures for the WISC-III (Wechsler, 1991) were similar to those of the WAIS and WAIS-III. A sample of 2,200 children was distributed evenly among 11 age levels from 6 through 16 years old, with an equal number of boys and girls at each level. Within each age level, the sample matched the 1988 U.S. census data according to race, socioeconomic status (SES), gender, geographic region, and parents' education.

The WPPSI represents the final downward extension of the Wechsler scales and is intended for use with children between 4 and 6½ years old. The WPPSI was standardized on 1,200 American children, with boys and girls distributed equally in each of six age levels, separated by half years. The 1960 U.S. census data were used to select a representative standardization sample based on race, SES, gender, geographic region, and father's education.

The WPPSI-R (Wechsler, 1989) has a more extensive age range than its predecessor, from ages 3 to 7 years. The WPPSI-R was standardized on 1,700 American children, stratified by age, gender, SES, and geographic region. There were 100 boys and 100 girls at each of eight age groups, from 3 through 7 years old, and one group of 50 boys and 50 girls at 7 years old. The 1986 U.S. census data were used to select representative children for the normative sample.

The subtests of the various Wechsler scales have shown satisfactory to outstanding reliability. Median split-half reliability coefficients ranged from .67 to .87 for the verbal tests and from .63 to .87 for the performance tests. The reliability estimates were relatively the same across all age levels. Standard errors of the intercorrelations were small, averaging about .05. The sample sizes were more than adequate to ensure adequate estimates of reliability. Thus, the correlations between subtests are reasonably accurate and close to the true correlations between the subtests. No corrections were made for unreliability.

The WPPSI and WPPSI-R have 11 subtests in common, as do the WAIS and WAIS-III. The WISC and WISC-III have 12 subtests in common. For each pair of scales, average intercorrelations among the common subtests were normalized using Fisher transformations. The means of these transformed correlations then were compared across (a) correlations between all subtests, (b) the verbal subscales, (c) the performance subscales, and (d) the verbal and performance scales.

Comparisons then were made using independent sample *t* tests at the .05 alpha level. Because a decline in test correlations was predicted, all tests for significance were one-tailed. These analyses then were followed by maximum likelihood analysis, to compare the correlation matrices for equality.

Results

Mean normalized intercorrelations between the common subtests are contained in Table 1. The table is divided into three sections, with each section devoted to one of the Wechsler scales. The predicted decline in subtest intercorrelation scores occurs for the WPPSI, WISC, and WAIS and their respective revisions. On

TABLE 1
Mean Normalized Intercorrelations Between Subtests of the Wechsler Scales

Test	<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>
		<i>WPPSI</i>		<i>WPPSI-R</i>		
All	55	.47	.10	.40	.11	3.0**
Verbal	15	.58	.08	.42	.08	6.0**
Performance	10	.46	.10	.58	.07	5.5**
Verbal-Performance	30	.40	.09	.34	.07	3.3**
		<i>WISC</i>		<i>WISC-III</i>		
All	66	.44	.13	.39	.17	2.0*
Verbal	15	.56	.16	.58	.18	.3
Performance	15	.44	.08	.34	.15	2.5*
Verbal-Performance	36	.39	.11	.33	.12	2.0*
		<i>WAIS</i>		<i>WAIS-III</i>		
All	55	.65	.14	.54	.16	3.9**
Verbal	15	.77	.19	.70	.21	1.0
Performance	10	.64	.07	.50	.11	3.5**
Verbal-Performance	30	.59	.08	.47	.08	6.0**

Note. WPPSI = Wechsler Preschool and Primary Scales of Intelligence. WISC = Wechsler Intelligence Scale for Children. WAIS = Wechsler Adult Intelligence Scale.

*Statistical significance using one-tailed tests at the .05 alpha level. **Statistical significance using one-tailed tests at the .01 alpha level.

the WPPSI and WPPSI-R, mean intercorrelation scores decreased significantly from .47 to .40, $t(54) = 3.0$, $p < .01$. For the WISC and WISC-III, mean intercorrelation scores significantly decreased from .44 to .39, $t(65) = 2.0$, $p < .05$. The WAIS and WAIS-III displayed the numerically largest decline in mean intercorrelation scores, from .65 to .54, $t(54) = 3.9$, $p < .01$.

Mean intercorrelation scores on Verbal subtests for the WISC and WISC-III, as well as the WAIS and WAIS-III, showed no significant declines. However, mean intercorrelation scores on the WPPSI and WPPSI-R declined significantly, $t(14) = 6.0$, $p < .01$, from .58 to .42.

The Performance subscale mean intercorrelation scores decreased significantly on the WISC and WISC-III, from .44 to .34, $t(14) = 2.5$, $p < .01$. They also decreased significantly for the WAIS and WAIS-III, from .64 to .50, $t(9) = 3.5$, $p < .01$. In contrast, mean intercorrelation scores on the WPPSI and WPPSI-R increased significantly, from .46 to .58, $t(29) = 5.5$, $p < .01$.

Also reported are the means of cross-correlations between the Verbal and Performance subtests from the Wechsler scales. Each revision of the three Wechs-

ler scales evidenced a significant decrease: a .06 decrease on the WPPSI-R, $t(29) = 3.3$, $p < .01$; a .06 decrease on the WISC-III, $t(35) = 2.0$, $p < .05$; and a .12 decrease on the WAIS-III, $t(29) = 6.0$, $p < .01$.

Maximum likelihood analysis supported the findings. The WISC, WAIS, WPPSI, and their subsequent revisions were tested for equality between the correlation matrices. Comparison of the WISC and WISC-III yielded significant differences in structure, $\chi^2(72) = 240$, $p < .01$. Analysis of the WAIS and WAIS-III also produced a significant result, $\chi^2(66) = 916$, $p < .01$. Finally, comparison of the WPPSI and WPPSI-R was significant, $\chi^2(66) = 125$, $p < .01$. Thus, for each of the Wechsler scales, significant differences exist between the sets of correlations from the original and its most recent revision.

Discussion

On measures of intelligence administered over one or more decades, the magnitude of g appears to decrease while the average IQs within a nation concomitantly increase. Using the original and latest revisions of the various Wechsler scales as a point of reference, we noted a statistical decline in the magnitude of g on 9 of the 12 pairs of comparisons. Between 1967 and 1989, children at the age levels in the WPPSI and WPPSI-R standardization sample experienced an average increase in IQ of 8 points. This populational increase in IQ was accompanied by a 15% decline in the statistical influence of g . Between 1949 and 1994, increases in adult intelligence evidenced by the Flynn Effect were accompanied by an 11% decline in the statistical magnitude of g . Similarly, increases in adult intelligence as measured by the WAIS and WAIS-III saw a concomitant decrease in the variance attributable to g , a drop on the order of 17% of its 1955 magnitude.

On measures of intelligence administered over one or more decades, IQ increases are larger on measures of nonverbal ability and tests that do not emphasize scholastic content. For example, Wechsler scales show IQ changes of 7.8 on Performance and 4.2 on Verbal (Jensen, 1998). Thus, in contrast to the Verbal subscales, the Performance subscales were expected to show a stronger decline in g . This expectation was confirmed in comparisons between the original and revised editions of the WISC and WAIS. On both measures, the statistical decline in g was significant on the Performance subscales and was not significant on the Verbal subscales.

The singular unexpected finding was that the statistical magnitude of g increased across the WPPSI and WPPSI-R Performance subscales. This change was in the direction opposite that expected, prompting further examination. The magnitude of g was examined for each of the six age levels common to the WPPSI and WPPSI-R. The results followed no strong developmental trends, either when we compared age levels within each test or when we compared age levels across tests. Thus, rather than purely reflecting changes in the population of preschool children between 1967 and 1989, these unexpected findings may be attributable

to changes in WPPSI items during renorming. The WPPSI-R was published 22 years after the original version. In addition, the age range is 2 years wider, with a 1-year extension both upward and downward. Although the WPPSI-R is highly reliable, about 40% of its items are different from those of the WPPSI. Therefore, although a statistical decline in the influence of g may have occurred across the populations tested by the versions of the WPPSI, this decline is likely obscured by the dramatic changes in the test format.

In addition, few studies have systematically investigated the structure of mental abilities of young children. The findings of studies using very young children as participants have suggested models of adult intelligence cannot adequately account for the intellectual ability of children (Miller & Vernon, 1996). The WPPSI simply may not provide an adequate representation of children's abilities. Replication of this study, using other intelligence tests designed to measure the cognitive abilities of young children, is strongly recommended.

In a comprehensive review of the malleability of IQ, Bouchard and Segal (1985) concluded that "no single environmental factor appears to have a large influence on IQ" (p. 452). Therefore, secular declines in the statistical magnitude of g , as evidenced in this study, are probably related to various conditions, each possibly contributing to the observed cumulative effect. Assuming that fluid abilities are strongly influenced by biological factors, the significant decline among the Performance subtest intercorrelations suggests that these changes may be related to variations in the biological substrate of intelligence. A number of other biological changes found in the population parallel the patterns observed in this study. For example, gains in height mirror increases in IQ, when compared in standard deviation units (e.g., Martorell, 1998).

Of the many explanations offered for the secular increases in IQ and the concomitant statistical decrease in g , a few deserve to be underscored. First, virtually every industrialized country, including the United States, has experienced improvements in nutrition (Jensen, 1998; Lynn, 1992). People in the United States consume vast amounts of processed foods, which frequently are augmented and fortified with nutrients. Improved nutrition may enhance the development of specialized cognitive abilities, a process that would lead to differentiation of cognitive abilities and a statistical decline in g . Second, the mandatory inoculation of children has eliminated or greatly reduced the frequency and intensity of many childhood illnesses (e.g., rubella, scarlet fever) that may inflict a toll on cognitive growth. Third, better prenatal care has decreased the incidence of fetal loss and infant mortality, and presumably, has contributed to healthier and more intelligent babies. Fourth, social mobility and geographic migration occur commonly within the United States. Fewer adults spend their lives in the same location in which they were born. Thus, regional and class barriers have relaxed and there is an increase in outbreeding. Outbreeding (i.e., hybrid vigor) has a positive effect on IQ. Most notably, the effect of outbreeding on IQ has been readily observed in the southern regions of the United States. Wheeler (1942) noted that

between 1930 and 1940, there was a 10-point increase in IQ among individuals residing in the southern United States.

Certainly, any inferences regarding the results of the present study are limited to the conditions that may have affected the standardization samples analyzed. The Flynn Effect, as observed over the past few decades, cannot be extrapolated forward or backward more than a few years. If the Flynn Effect occurred in this magnitude over time, with a 15-point increase in IQ per generation, then the average IQ of an individual born around 1900 would be 65. Taken to absurd extremes, persons living at the time of Leonardo da Vinci would have an IQ on the order 5, if tested with the WAIS-III. Similarly, if the Flynn Effect were extrapolated forward, virtually every first world country would be in the midst of a cultural renaissance (Flynn, 1987). The results of this study should not be generalized to other populations or time periods.

REFERENCES

- Bouchard, T. J., & Segal, N. L. (1985). Environment and IQ. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurement, and applications* (pp. 391-464). New York: Wiley.
- Brand, C. R. (1996). *The g factor: General intelligence and its implications*. New York: Wiley.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.
- Detterman, D., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence, 13*, 349-359.
- Flynn, J. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*(2), 171-191.
- Gould, S. J. (1981). *The mismeasure of man*. New York: Norton.
- Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Jensen, A. R., & Weng, L. (1994). What is a good g? *Intelligence, 18*(3), 231-258.
- Lynn, R. (1992). Does Spearman's g decline at high IQ levels? Some evidence from Scotland. *The Journal of Genetic Psychology, 153*, 229-230.
- Martorell, R. (1998). Nutrition and worldwide rise in IQ scores. In U. Neisser (Ed.), *The rising curve* (pp. 183-206). Washington, DC: American Psychological Association.
- Miller, L. T., & Vernon, P. A. (1996). Intelligence, reaction time, and working memory in 4- to 6-year-old children. *Intelligence, 22*, 155-190.
- Neisser, U. (1998). *The rising curve: Long-term gains in IQ and related measures*. Washington, DC: American Psychological Association.
- The Psychological Corporation. (1997). *Wechsler Adult Intelligence Scale-III*. San Antonio, TX: Author.
- Spearman, C. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology, 15*, 201-293.
- Spearman, C. E. (1927). *The abilities of man: Their nature and measurement*. London: Macmillan.
- Teasdale, T. W., & Owen, D. R. (1987). National secular trends in intelligence and education: A 20-year cross-sectional study. *Nature, 325*, 119-121.

- Wechsler, D. (1939). *Wechsler-Bellevue Intelligence Scale*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1949). *Manual for the Wechsler Intelligence Scale for Children*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1955). *Wechsler Adult Intelligence Scale*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1967). *Manual for the Wechsler Preschool and Primary Scales of Intelligence*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1989). *Manual for the Wechsler Preschool and Primary Scales of Intelligence-Revised*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: The Psychological Corporation.
- Wheeler, L. R. (1942). A comparative study of the intelligence of East Tennessee mountain children. *Journal of Educational Psychology*, 33, 321-334.

Received January 19, 2000