Intelligence 39 (2011) 443–455

Contents lists available at ScienceDirect

Intelligence

The Flynn effect puzzle: A 30-year examination from the right tail of the ability distribution provides some missing pieces

Jonathan Wai *, Martha Putallaz

Duke University, Talent Identification Program, 1121 West Main Street, Durham, NC 27701, United States

ARTICLE INFO

Article history:
Received 8 October 2010
Received in revised form 22 February 2011
Accepted 25 July 2011
Available online 26 August 2011

Keywords:
Flynn effect
Environmental factors
Sex differences
Historical examination
Gifted

ABSTRACT

The Flynn effect is the rise in IQ scores across the last eighty or more years documented in the general distribution of both industrialized and developing nations primarily on tests that require problem solving and non-verbal reasoning. However, whether the effect extends to the right tail (i.e., the top 5% of ability) remains unknown. The present study uses roughly 1.7 million scores of 7th-grade students on the SAT and ACT as well as scores of 5th- and 6th-grade students on the EXPLORE from 1981 to 2010 to investigate whether the effect operates in the right tail. The effect was found in the top 5% at a rate similar to the general distribution, providing evidence for the first time that the *entire curve* is likely increasing at a constant rate. The effect was also found for females as well as males, appears to still be continuing, is primarily concentrated on the mathematics subtests of the SAT, ACT, and EXPLORE, and operates similarly for both 5th and 6th as well as 7th graders in the right tail. These findings help clarify the *nature* of the effect and may suggest ways that potential *causes* can now be more meaningfully offered and evaluated.

1. Introduction

When data provide researchers with a scientific puzzle – a paradoxical finding that cannot easily be placed in the current network – this provides an opportunity for the field to advance in an attempt to resolve the contradictions, apparent or real. For the field of psychology and specifically intelligence, one such conundrum has been named the *Flynn effect* after Flynn (1984, 1987), a political scientist who convinced psychologists that it needed explanation. This phenomenon is the rise in scores on intelligence tests over the last eight or more decades at the rate of about 10 IQ points per 30 years (Rodgers, 1998). The effect is noteworthy partly because large *differences* in scores are demonstrated in just those situations where *similarity* would be expected (Deary, 2001). The rises occur primarily on those tests with content that does not appear easily learned, such as the Raven's Progressive Matrices (Raven, 1941; Raven, 1981), which includes no words or numbers but requires the completion of abstract patterns. The Raven's primarily measures *g* or general intelligence (Jensen, 1998), hence the paradox: Why should the rise occur on precisely those measures where we would not expect it? There has been disagreement over whether the gains are genuine intelligence increases, due to artifact, or potentially both, as well as why they are occurring. Flynn (1999, p. 6) marveled at the almost magical monotonic rate of gain, saying "It is as if some unseen hand is propelling scores upward." Perhaps before trying to understand why scores are rising, we should learn more about all the places this invisible hand is at work.

1.1. A brief history of research on the Flynn effect

Since Flynn's (1984, 1987) initial synthetic papers, others have confirmed the effect operates in the general distribution in industrialized nations including the United States, Britain, and Denmark (e.g., Flynn, 1984, 1987, 2007; Lynn, 2009;

* Corresponding author at: Duke University, Talent Identification Program, 1121 West Main Street, Durham, NC 27701, United States.
E-mail address: jon.wai@duke.edu (J. Wai).

0160-2896/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2011.07.006

Teasdale & Owen, 1987, 1989) as well as developing countries including Brazil, Kenya, and Sudan (e.g., Colom, Flores-Mendoza, & Abad, 2007; Daley, Whaley, Sigman, Espinosa, & Neumann, 2003; Khaleefa & Lynn, 2009). Many have thus concluded that the Flynn effect is well established (e.g., Deary, 2001; Jensen, 1998).

Deary (2001, p. 112) said that "If there was a prize to be offered in the field of human intelligence research, it would be for the person who can explain" the effect. Researchers have not been shy in proposing causal explanations. The American Psychological Association published *The Rising Curve* (Neisser, 1998) wherein multiple scholars proposed causes. None of them were definitive and at best remain unproven hypotheses (Jensen, 1996, 1998). More recently, a special issue in the *Journal of Psychoeducational Assessment* (Kaufman & Weiss, 2010) was devoted to the topic. Aligned with Jensen's (1996, 1998) view, in this special issue Ceci and Kanaya (2010, p. 446) note that "It is clear that we still have a long way to go before we understand the exact nature and magnitude of the [Flynn Effect] for all types of individuals, across time, and across tests. Until we do, discussions regarding its cause(s) and interpretation may be futile." We think it has been helpful to the field for scholars to propose multiple explanations so that relevant data can be used to test or help inform these perspectives. Some of these explanations include an improvement in education (Blair, Gamson, Thorne, & Baker, 2005; Ceci, 1991; Flynn, 1984, 2007), increased test sophistication and more confident test taking attitudes (Brand, 1987), more cognitive stimulation arising from the greater complexity of more recent environments, for example, the broad exposure to television and video games (Schooler, 1998; Sundet, Barlaug, & Torjussen, 2004; Williams, 1998), the idea that environmental and social factors may serve as multipliers of cognitive abilities in a model of reciprocal causality — the *individual* or *social multiplier* (Dickens & Flynn, 2001; Flynn, 2007), improvements in nutrition (Lynn, 1990, 2009), and heterosis or hybrid vigor (Mingroni, 2004, 2007). Jensen (1998) stated that although the proponents of the various theories often treat the individual explanations as if they were mutually exclusive, they are not incompatible. Therefore, he suggested a *multiplicity hypothesis* — the idea that each of the proposed explanations "is involved to some (as yet undetermined) degree in producing the secular rise in scores" (p. 323).

1.1.1. What research has been missing

In contrast to those who have taken the Flynn effect as well established, there have been others who have expressed their concern over the quick acceptance of the effect. For example, Rodgers (1998), (Rodgers & Wänström, 2007) stressed that there are still areas in which the nature and legitimacy of the effect remains unclear, for example, within the right tail of the ability distribution. Some of these areas, in particular for the right tail, include whether the effect holds for females as well as males, the current time period, some subtests over others, more than one measure, and for different age groups.

First, the effect consistently appears in the general distribution, but to our knowledge no studies have been able to adequately test whether the Flynn effect operates throughout the right tail (i.e., the top 5%). Flynn (1996, p. 25) has suggested that "IQ gains extend to every IQ level," but it is surprising this assertion has been generally assumed without empirical verification. Prior studies have likely had ceiling effects on measures that prevented individual differences in the right tail from being adequately captured and likely have not explicitly sampled from a right tail population. We are now able to bring some evidence to help evaluate whether gains extend to every level using samples from the top 5% on measures that have enough headroom. Whether the effect operates in the right tail may provide clues as to which causal hypotheses appear most promising (and for what parts of the ability distribution). Also, as some have suggested (Micceri, 1989; Rodgers, 1998, p. 351), "we need to understand in much more detail the nature of the whole distribution of IQ scores to unravel the puzzles underlying the Flynn effect." Second, research has primarily examined males and whether the effect operates similarly for females remains unknown for the right tail (Rodgers & Wänström, 2007), although it has been demonstrated for the general distribution (Ang, Rodgers, & Wänström, 2010). If the effect operated only for males, this would suggest different causes than if the effect also operated for females. Third, some studies suggest the effect may have ended in Denmark, Norway, and Britain (Shayer, 2007; Sundet et al., 2004; Teasdale & Owen, 2008). Therefore, it is important to assess whether the effect continues, at least in the United States. Fourth, the effect primarily appears on measures that require problem solving and other non-verbal components of IQ. Thus, does the effect operate on similar non-verbal measures in the right tail? Fifth, if the effect appears on multiple tests measuring similar constructs in the right tail, this would provide validation that the effect in the right tail is not test specific. We use the Scholastic Assessment Test (SAT) and the American College Test (ACT) to examine whether the effect is test specific for 7th graders and the EXPLORE test to address this question for 5th and 6th graders, and we also use these tests to examine whether the effect operates in the right tail for different age groups.

Rodgers (1998, p. 338) has said that "research addressing the *legitimacy* and *meaning* of the effect should precede research *testing* for and evaluating *causes* of the effect." Therefore, certain aspects of the nature of the effect – some of the boundary conditions – need to be clarified before the many explanations that researchers have proposed can be more fully considered.

2. Method

2.1. The present study

With a focus on the right tail of the ability distribution, the aims of the present study include: 1. determining whether the Flynn effect operates in the top 5%, and if so, whether it does to the same degree as in other parts of the distribution, 2. determining whether the effect operates similarly for males and females, 3. determining whether the effect continues, 4. determining whether the effect appears at the composite level or on particular subtests, 5. determining whether the effect appears on both the SAT and ACT as well as the EXPLORE, and 6. determining whether the effect operates similarly for different age groups (i.e., 5th and 6th versus 7th graders).

We provide a broad historical analysis of average SAT, ACT and EXPLORE scores across 30 years – one generation – using