

Vol. 25, Issue 1, 2017

# High Stakes IQ Testing: The Flynn Effect and its Clinical Implications

Stephanie C. Black\*

Mar 31, 2017

Tags: april 2017

# Abstract

Intelligence testing has become ubiquitous in many sectors: In the education sector it is frequently used for determinations of learning disabilities and giftedness. Intelligence testing is affected by the Flynn Effect (FE), the highly debated long-term trend of rising average intelligence test scores within the population. The FE has been observed in many nations and varies with regard to ability, intelligence types, age, location and prosperity. Recent observations of a potential reversal have added to the complexity. FE research centres on four hypotheses: testing artefacts, biological factors, environmental changes and multi-driver models. The FE affects the periodical re-norming of intelligence tests and the comparability of IQ scores over time. This has clinical implications for the assessment of intellectual ability for various client populations, including forensic, disabled, juvenile and ethnic minority clients, as well as social implications concerning ethical intelligence, developing nations and the concept of age-related cognitive decline. Examples of high-stakes decisions based on IQ tests include fitness to stand trial, access to social disability services and, of particular interest for the education sector, access to special education services and accelerated programs. The present article underlines the importance of clinical practitioners and staff in the education sector being aware of how the FE affects the determination of intellectual ability for diverse client populations and ensuring that the most current editions of IQ tests (i.e., correctly normed versions) are used. Given the high stakes involved in IQ testing, there needs to be greater awareness and more research on the FE. Areas for possible future research include optimised methodologies, novel factors, diverse cultures and older participants.

## High stakes IQ testing: The Flynn Effect and its clinical implications

The Flynn Effect (FE) is the highly debated, long-term trend of rising average intelligence test scores within the population. Standardised psychometric intelligence measures were developed in the early 1900s, and the intelligence quotient (IQ) metric quickly became widely used in many settings, including determinations of special education needs and academic entrance exams. Scholars soon noticed increasing average IQ scores in specific populations (Lynn, 1982; Merrill, 1938; Schaie & Strother, 1968; Smith, 1942; Tuddenham, 1948; Wechsler, 1981). However, Flynn (1984, 1987) first observed a long-term trend of rising average intelligence test scores within the general population. The Flynn Effect (FE; named such by Herrnstein and Murray in 1994) has since been observed in many populations worldwide, but scholars disagree on its variability with various factors, and some populations even exhibit a reverse FE - that is, decreasing IQ scores over time (Pietschnig & Voracek, 2015; Williams, 2013).

The scores of a given population on a given IQ test are, by definition, always re-normed to a mean of 100 (Ceci & Williams, 2016). The FE, therefore, affects the periodical re-norming of intelligence tests and the comparability of test scores over time. As will be shown in this article, this has clinical implications for the assessment of intellectual ability for various client populations, including forensic, disabled, juvenile and ethnic minority clients, as well as social implications concerning ethical intelligence, developing nations and the concept of age-related cognitive decline. Examples of particular high-stakes decisions involving IQ tests include exemptions from the death penalty, fitness to stand trial, access to social disability services and, of particular interest for the education sector, access to special education services and accelerated programs. It becomes clear then that IQ test results can have a potentially life-changing impact on clients, and the FE, in turn, can substantially affect these high-impact test results. In fact, clients' life trajectories can be forever altered if the FE is not taken into account correctly during IQ tests and the resulting determinations of intellectual ability. Consequently, it is vitally important that clinical practitioners and staff in the education

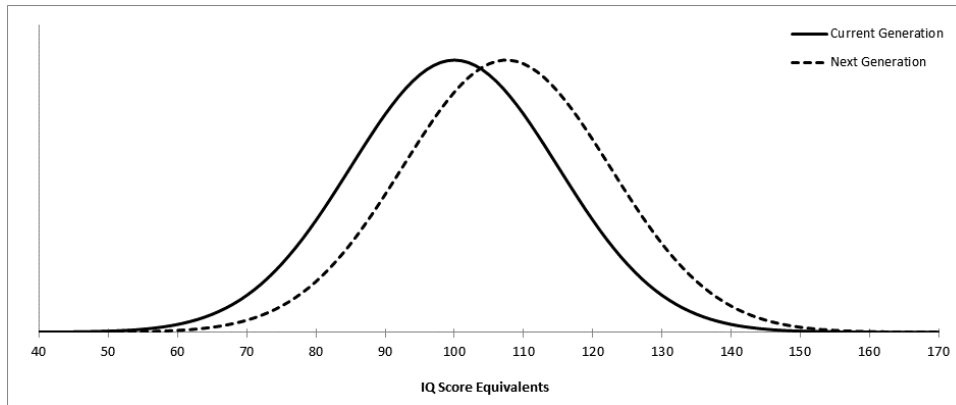
sector are aware of how the FE affects IQ tests and determinations of intellectual ability for diverse client populations. Most importantly, practitioners need to ensure that the most current editions of IQ tests (i.e., correctly normed versions) are used. Moreover, given that research on some aspects of the FE is still inconclusive, more studies are required, and practitioners need to keep themselves informed of new developments regarding the FE.

The aim of this article is not to attempt comprehensive coverage of the research on the FE (as provided by Pietschnig & Voracek, 2015; Trahan, Stuebing, Fletcher & Hiscock, 2014), but rather to emphasise the far-reaching consequences that a lack of awareness of the FE by practitioners can have on a diverse range of clients. With this in mind, the present article first provides a brief overview of observations and hypotheses regarding the FE, then discusses the implications of the FE for various client populations in clinical practice and society and, finally, concludes with suggestions for future research.

## Description of the FE

### Magnitude of the FE

Research converges regarding the mean magnitude of the FE. The magnitude of population IQ gains over time (i.e., the FE) is typically reported in points per decade ( $\Delta IQ$ ); the current consensus is an average of  $\Delta IQ = 3$  (Williams, 2013). The first FE was reported at  $\Delta IQ = 2.91$  (Flynn, 1984). While reported gains for various populations range widely (Must & Must, 2013; Nijenhuis, Cho, Murphy & Lee, 2012), a recent meta-analysis spanning over a century and including nearly four million participants from 31 countries arrived at an average  $\Delta IQ = 2.8$  (Pietschnig & Voracek, 2015). Another meta-analysis yielded a similar  $\Delta IQ = 2.93$  (Trahan et al., 2014). While these meta-analyses are limited by imperfect equivalence of the included primary studies and the compared test editions, evidence quality is excellent and results support the consensus of  $\Delta IQ = 3$  (Pietschnig & Voracek, 2015). To put these values into context, a mean magnitude of the FE of  $\Delta IQ = 3$  would indicate that the next generation would have a mean IQ of 107.5, that is half a standard deviation above the current mean IQ (see Figure 1).



**Figure 1. Estimated IQ Distributions of the Current Generation ( $M = 100$ ,  $SD = 15$ ) and the Next Generation ( $M = 107.5$ ,  $SD = 15$ ).**

For illustration purposes only; assuming (1) that a generation is 25 years, (2) that the FE will continue throughout the next generation at the current magnitude of  $\Delta IQ = 3$  and (3) disregarding variability of the FE.

## Variability of the FE

*Variability of the FE with intellectual ability.* To date, research on variability of the FE with intellectual ability level is inconclusive. Habets, Jeandarme, Uzieblo, Oei and Bogaerts (2015) and Zhou, Zhu and Weiss (2010) observed that more intelligent individuals experience stronger IQ gains. In contrast, Habets et al. (2015) and Williams (2013) observed the largest IQ gains at the low end of the IQ distribution, while Trahan et al. (2014) observed no significant variation with ability level at all. Psychometric tests perform most reliably within a given range of the measured construct (Habets et al., 2015). Therefore, the reason for the variability of the FE could be explained by the psychometric nature of the IQ tests, with the FE potentially following a U-curve, showing high variations at extremes but little variation closer to the average range of the distribution.

*Variability of the FE with intelligence type.* The FE varies depending on the type of intelligence concerned; fluid intelligence is more strongly affected by the FE than crystallised intelligence. Fluid intelligence is the capacity for logical problem-solving, including inductive and deductive reasoning, and crystallised intelligence is the ability to use skills and experience, including vocabulary and knowledge (Cattell, 1971). Fluid intelligence is usually measured with Raven's Progressive Matrices (Raven & Court, 1998; Raven, 2000). Flynn (1987) originally found some

of the largest IQ gains on tests of fluid intelligence. Conversely, an Australian study found little change in fluid intelligence for Victorian school children (Cotton et al., 2005); however, the FE is typically smaller in children (Pietschnig & Voracek, 2015). The most recent international meta-analyses support larger gains for fluid intelligence than crystallised intelligence (Brouwers, Vijver & Van Hemert, 2009; Pietschnig & Voracek, 2015; Trahan et al., 2014). This is in line with the consensus that the FE is stronger for fluid intelligence than crystallised intelligence (Williams, 2013).

*Variability of the FE with age.* The FE varies with age. An FE is found in all age ranges from childhood to adulthood (Ceci & Williams, 2016) and can appear before school age (Williams, 2013). However, the FE is more pronounced for adults than for children (Pietschnig & Voracek, 2015).

*Variability of the FE with socioeconomic conditions.* The FE correlates with socioeconomic status and prosperity. The FE has been observed in as many as 31 countries, both industrialised and pre-industrialised nations (Ceci & Williams, 2016; Pietschnig & Voracek, 2015). A recent meta-analysis evidenced a positive association with gross domestic product and socioeconomic conditions (Pietschnig & Voracek, 2015). Accordingly, recent high IQ gains were observed in developing countries where prosperity is improving, including Kenya, Brazil, Turkey and Argentina (Flynn & Rossi-Casé, 2012; Flynn, 2012). In the developed world IQ gains were highest after the Second World War, which was a period of the steepest prosperity increase in history (Pietschnig & Voracek, 2015).

## **Recent observations of diminishing or reversing FE**

The latest data suggest that the FE may be diminishing or even reversing in some nations (Pietschnig & Voracek, 2015). This is supported by evidence from Norway (Sundet, Barlaug & Torjussen, 2004; Sundet, Eriksen, Borren & Tambs, 2010; Sundet, 2014), Denmark (Teasdale & Owen, 2005, 2008), Finland (Dutton & Lynn, 2013) and France (Dutton & Lynn, 2015). While reversals are also reported for toddlers in the United States (Kaufman, 2010), a large recent meta-analysis did not yet support a decline in

the United States (Trahan et al., 2014). In Britain, Germany and South Korea, the FE is still strongly ongoing (Flynn, 2012).

The FE has been studied in many nations. However, variability with regard to ability, intelligence types, age, location and prosperity is still being debated, and recent observations of a potential reversal have added to the complexity.

## **Past and current research: Explanations for the FE**

The FE literature is extensive, and numerous hypotheses have been proffered to explain the FE. Research centres on four clusters of hypotheses for the rising secular IQ scores: testing artefacts, biological factors, environmental changes (Neisser, 1997) and multi-driver hypotheses (Williams, 2013).

### **Explaining the FE with artefacts of IQ testing**

The first cluster of hypotheses, in line with Flynn's (1984, 1987) original conclusions, attributes the FE to psychometric testing artefacts instead of real intelligence gains. While IQ typically correlates with cognitive resources, it is simply an operationalisation of intelligence, which has associated uncertainties (Pietschnig & Voracek, 2015). IQ measures purport to measure pure intelligence, but influences of skill and culture cannot be fully excluded (Habets et al., 2015; Schooler, 1998). The FE might result from increased test sophistication due to increased exposure to testing and increased guessing (Brand, 1996; Neisser, 1997; Woodley, Nijenhuis, Must & Must, 2014). The high magnitude of the FE on newer culture-free IQ tests (e.g., Raven's matrices) might be explained by lower test sophistication for these tests due to less exposure (Brouwers et al., 2009). Additionally, subsequent test editions differ not just in norms, but also in composition and instructions; therefore, the FE could result from suboptimal concurrent validities (Kaufman, 2010). Few FE studies employ item response analysis (which better exposes testing artefacts; Beaujean & Osterlind, 2008), but those that do, have reported minimal FE (Williams, 2013), supporting testing artefact explanations. The evidence suggests, then, that the FE might at least partially be attributed to testing artefacts, including test

sophistication, test changes and overreliance on classical test theory.

## **Explaining the FE with biological factors**

The second cluster of hypotheses interprets the FE as real intelligence gains through biological factors. Physiological factors include improved nutrition (Lynn, 1989), decreased lead exposure (Nevin, 2000), decreased pathogen load (Eppig, Fincher & Thornhill, 2010) and artificial lighting (Williams, 2013) – all resulting from improved socioeconomic conditions. Genetic factors potentially influencing the FE include heterosis (i.e., increased hybrid vigour through globalisation; Mingroni, 2007) and epigenetics (i.e., improved gene expression through environmental adaptation; Storfer, 1999), although they are thought to be limited by dysgenics (i.e., low infertility rates of intelligent people; Wang, Fuerst & Ren, 2016; Preston, 1998). Overall, biological explanations of the FE are well supported.

## **Explaining the FE with environmental factors**

The third cluster of hypotheses explains the FE with environmental changes, including parenting, schooling and technology. Families have become smaller and professional childminders more prevalent, thereby giving children more contact with adults (Sundet et al., 2010), parenting increasingly emphasises cognitive development (Neisser, 1997), and the early childhood environment has become more complex (Schooler, 1998), which facilitates early cognitive development (Flynn, 2012; Rodgers, 2015). Children are more extensively schooled (Cahan & Cohen, 1989), own a larger repertoire of problem solving techniques (Schooler, 1998) and are better trained in hypothetical reasoning and abstraction than any other generation (Flynn, 2012), which further improves cognitive development (Ceci & Williams, 2016; Kaufman, 2010). Urbanisation and modern technology increase socio-environmental complexity (Flynn, 2012; Schooler, 1998), employment has become more cognitively demanding (Schooler, 1998), and increasingly visual-technological environments improve processing speed and visual cognition (Greenfield, 1998), which develops cognitive resources and flexibility via increased demand (Flynn, 2012; Schooler, 1998). Consequently, there is convincing evidence for



environmental explanations of the FE.

All three hypothesis clusters discussed so far – testing artefacts, biological and environmental explanations – are potentially subject to ceiling effects (Sundet et al., 2004; Teasdale & Owen, 2008). This could explain observations of both increasing and diminishing IQ gains.

## Explaining the FE with multi-factor models

The fourth cluster of hypotheses concerns multiple drivers (Williams, 2013). Many of the aforementioned factors are backed by solid evidence, and, similar to the nature-nurture debate, it appears unreasonable to expect any single factor to explain fully the FE. The FE is more likely to be the result of multiple factors. The Social Multipliers Model explains that multiple small environmental changes can amplify each other (i.e., small environmental advantages improve performance, which, in turn, generates further environmental advantages, thus creating an ever accelerating loop) until a critical mass in environmental change is reached which produces the FE (Dickens, 2001). The Life History Model amalgamates biological and environmental factors: the improving biological and environmental conditions increase contraception, education and longevity, which slow life history speed; this results in fewer offspring experiencing more parental effort and better biological and environmental conditions, which facilitates cognitive development and increases intellectual ability (Woodley, 2012). There is good evidence for multi-driver models, but the Life History Model is especially well supported by a large recent meta-analysis (Pietschnig & Voracek, 2015).

Existing research, therefore, provides convincing evidence to support all four clusters of hypotheses of the FE. However, the variability explained by each factor remains unclear.

## Clinical implications of the FE

The FE has important implications for clinical practice. The FE becomes salient whenever intelligence assessments are performed and IQ cut-off points determine high-stakes decisions (Trahan et al., 2014). This can affect forensic, disabled, juvenile and minority clients.

## Implications of the FE for forensic clients

*Death penalty.* The most dramatic implication of the FE concerns the United States, where, in 2002, Daryl Atkins' death sentence was converted to life in prison upon diagnosis of intellectual disability (Kaufman, 2010). Over 80 intellectually disabled offenders have since been spared execution (Trahan et al., 2014). Therefore, assessing intellectual disability can literally mean life or death (Habets et al., 2015). Borderline intellectually disabled offenders could potentially slide below the IQ cut-off if retested with new norms, thereby escaping execution (Flynn, 1999). Some scholars argue that correcting IQ scores for the FE is unscientific due to the FE's variability, and that it violates standardisation and test guidelines (Hagan, Drogin & Guilmette, 2010, 2008). However, these arguments are based on small studies and are not reflective of recent large meta-studies, which support FE corrections being applied to all high-stakes intellectual assessments (Fletcher, Stuebing & Hughes, 2010). Sternberg (2010) offers a different perspective, pronouncing the FE to be irrelevant to criminal proceedings, as ethical intelligence rather than cognitive intelligence matters more in these situations. Nevertheless, the expert consensus is that, given the high stakes involved, the FE should be considered when assessing defendants' IQs (Kaufman, 2010).

*Other court proceedings.* IQ cut-off points, and therefore the FE, are also important for other court proceedings. For instance, there are jurisdictions that declare minors with learning disabilities incompetent to stand trial (Kanaya & Ceci, 2012) or intellectually disabled offenders not guilty by reason of insanity (Habets et al., 2015). Moreover, intelligence assessments also affect interrogations, court rulings, parole assessments, mandatory offender treatment programs (Habets et al., 2015), civil commitment evaluations (Melton, Petrila, Poythress & Slobogin, 2007) and child custody determinations (Crossman, Powell, Principe & Ceci, 2002; Erickson, Lilienfeld & Vitacco, 2007). All these intelligence assessments for forensic purposes are influenced by the FE.

## Implications of the FE for intellectually disabled clients

*Intellectual disability assessments.* Another dramatic example of the impact of the FE is the assessment of intellectual disability. Intellectual disability assessments determine eligibility for social services, disability insurance payments and in the United States even access to organ transplants (Hagan et al., 2008). While clinical judgement is considered, an IQ cut-off point two standard deviations below the mean usually demarks intellectual disability (Trahan et al., 2014). Assessments using outdated test norms could push vulnerable clients above the cut-off. Several studies found large discrepancies in intellectual disability assessments between differently normed tests. In one example only 58% of clients assessed as intellectually disabled, were still assessed as intellectually disabled by re-normed tests, which could significantly disadvantage vulnerable clients (Habets et al., 2015).

*Stronger FE at lower ability.* The FE may be stronger at the extreme ends of the IQ distribution; that is, intellectually disabled and borderline clients may be especially affected (Habets et al., 2015). This would discredit the emerging clinical practice of uniformly adding 0.3 IQ points per year to correct for the FE regardless of ability level (Zhou et al., 2010). Additionally, it has not been excluded that different domains of intelligence may vary differently with ability level, which casts doubt upon the clinical practice of using abbreviated IQs or subscales as substitute for an unobtainable full scale IQ (Zhou et al., 2010).

## Implications of the FE for clients with learning disabilities and the gifted

*Learning disabilities.* Intellectual assessments have high impacts on school children and students, since IQ scores inform the diagnosis of learning disability (i.e., if achievement scores are substantially lower than the corresponding IQ score; Kanaya & Ceci, 2011). Children's IQ scores decline with re-normed tests (e.g., retesting on WISC-III after testing on WISC-R), while achievement scores remain stable; this decreases the probability of a learning disability diagnosis, as the gap between IQ and achievement scores decreases, potentially denying or removing

access to special education (Kanaya & Ceci, 2012). Additionally, the FE varies across children's age, ability levels and subtests used, casting further doubt on the reliability of IQ testing for access to special education (Kanaya & Ceci, 2011). Moreover, historically Australian children were tested using United States norms until the Australian Standardisation Project for WISC-IV, WIAT-II and CELF-4 established Australian normative data (Hannan, 2005). United States WISC norms, however, were consistently lower than Australian norms, resulting in inflated IQ scores, underdiagnoses of learning disabilities and under-allocation of special education resources (Kamieniecki & Lynd-Stevenson, 2002). In other words, before the Australian Standardisation Project established Australian normative data, IQ test results of Australian children were consistently inflated and learning disabilities underdiagnosed. This example further highlights the impact the FE can have on clients with learning disabilities.

*Gifted students.* Gifted pupils and students are also affected by the FE. Outdated norms increase the probability of labelling children and young adults 'gifted' (Kamieniecki & Lynd-Stevenson, 2002). This creates inflated expectations from caregivers and educators and excessive performance pressure for the child or young adult (Neihart, 1999). Additionally, as IQ scores have been observed to decline with retesting (Kanaya & Ceci, 2012), clients previously classified as 'gifted' may subsequently lose the 'gifted' label (Ceci & Williams, 2016). While there is little research on effects of reclassification from 'gifted' to 'normal', reclassification might have negative impacts on perceived self-efficacy (Bandura, 1977), making it harder to succeed in many life activities.

## **Implications of the FE for ethnic minority clients**

The FE also impacts on IQ testing of ethnic minority clients. One example of ethnic minority clients affected by the FE is migrants. Analogous to the inappropriateness of testing Australian children using United States norms (Kamieniecki & Lynd-Stevenson, 2002), it might be inappropriate to test migrants with norms originating outside their native culture. Another group of ethnic minority clients affected by the FE are Indigenous Australians. Studies comparing Indigenous and non-Indigenous Australians' IQs show a difference of 0.3–0.4 standard deviations, attributed

largely to socioeconomic differences (Leigh & Gong, 2009), which likely increases the impact of the FE. More research is needed regarding the FE in migrants and Indigenous Australians (McDonald, Comino, Knight & Webster, 2012; Pearson, 2012).

## Implications of the FE for the general population

*Ethical intelligence.* In addition to clients in clinical practice, the FE has implications for society in general. The FE does not generalise to ethical intelligence, and increasing cognitive intelligence in the absence of increasing ethical intelligence might even have negative moral implications for society (Sternberg, 2010). There is little research into whether ethical intelligence is subject to the FE, but it appears that ethical behaviour has not increased with the FE. Sternberg (2010) argues that less effort should be directed to researching magnitude and patterns, and more effort should be directed to ethical implications of the FE. Moreover, as ethical intelligence is a skill rather than an ability, increased effort should be made to teach ethical thinking. Ultimately, more research into and better education about ethical intelligence is needed.

*The developing world.* The FE is likely to be accelerating in the developing nations. There is support for the hypothesis that the FE is partially caused by industrialisation (Pietschnig & Voracek, 2015). As developing nations are embarking upon industrialisation they will probably experience explosive IQ gains similar to those previously experienced by the developed countries, which is supported by recent studies (Flynn, 2012). This factor has extensive implications for the developing world.

*Age-related cognitive decline.* The FE may change the discourse regarding age-related cognitive decline. Dickinson and Hiscock (2010) found that 85% of cognitive decline with age is explained by the FE. A recent meta-study supported this: age-related IQ decline is markedly reduced after adjustment for the FE (Trahan et al., 2014). Therefore, age-related cognitive decline might need to be reinterpreted as a mere result of the FE. This strategy might have implications for how society treats aging citizens.

## Conclusion

The present article described observations regarding the magnitude and variability of the FE; research was summarised around four clusters of hypotheses: testing artefacts, biological factors, environmental changes and multi-driver hypotheses. Implications for clinical populations including forensic, disabled, juvenile and minority clients were discussed and societal implications were examined. Several examples of high-stakes decisions based on IQ tests underlined the importance of clinical practitioners being aware of how the FE affects the periodic re-norming of IQ tests and the determination of intellectual ability for diverse client populations. Most importantly, practitioners need to ensure that the most current editions of IQ tests (i.e., correctly normed versions) are used. Moreover, practitioners need to keep themselves informed of new developments regarding the FE.

As the Flynn Effect can have significant implications for the lives of potentially millions of clients undergoing intelligence testing across the world each year (Kanaya & Ceci, 2011), and given that research on some aspects of the FE is still inconclusive, further research is important, and more studies are required. Most FE studies utilise limiting testing methodologies (Zhou et al., 2010), so new surveys could be designed utilising the more precise item response theory (Beaujean & Osterlind, 2008). Additionally, older data from earlier test editions could be reanalysed (Habets et al., 2015), and data from health and social service agencies could be explored (Williams, 2013). Further research might explore if the FE follows a U-curve similar to the reliability curves of psychometric tests (Habets et al., 2015). Diagnosis protocols for learning disabilities may need investigation to determine whether to rely more on achievement and skill and less on IQ testing (Kanaya & Ceci, 2012).

The majority of FE research is conducted on younger populations (Pietschnig & Voracek, 2015), inconsistent with demographic trends towards an older population (Pink, 2009); therefore, more research should be conducted within older populations. Additionally, more research is needed to establish if age-related cognitive decline needs to be reinterpreted as a mere artefact of

the FE. Furthermore, there are insufficient longitudinal studies for clinical and forensic populations (Habets et al., 2015). More data from a diverse range of societies are required (Zhou et al., 2010), and further research is needed into how the FE affects minorities, including Indigenous Australians and migrants.

More recently identified factors, including the impact of computing, mobile communications and visual media should be explored (Schooler, 1998). As effects similar to the FE exist for memory and attention span, research should determine if the FE generalises across other cognitive domains and affects neurological parameters (Rönnlund & Nilsson, 2009). Lastly, more research on ethical intelligence and the impact of the FE on it is needed (Sternberg, 2010).

In summary, IQ testing involves very high stakes and clients' life trajectories can be forever altered if the FE is not taken into account correctly. Accordingly, there needs to be greater awareness amongst practitioners and more research on the FE is required.

The author may be contacted via:

[https://www.researchgate.net/profile/Stef\\_Black](https://www.researchgate.net/profile/Stef_Black)

## References

- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191. [doi.org/10.1037/0033-295X.84.2.191](https://doi.org/10.1037/0033-295X.84.2.191)
- Beaujean, A. A. & Osterlind, S. J. (2008). Using item response theory to assess the Flynn effect in the National Longitudinal Study of Youth 79 Children and Young Adults Data. *Intelligence*, 36(5), 455-463. [doi.org/10.1016/j.intell.2007.10.004](https://doi.org/10.1016/j.intell.2007.10.004)
- Brand, C. (1996). *The g factor: General intelligence and its implications*. Chichester, England: Wiley.
- Brouwers, S. A., Van de Vijver, F. J. & Van Hemert, D. A. (2009). Variation in Raven's Progressive Matrices scores across time and place. *Learning and Individual Differences*, 19(3), 330-338. [doi.org/10.1016/j.lindif.2008.10.006](https://doi.org/10.1016/j.lindif.2008.10.006)

Cahan, S. & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development*, 60, 1239-1249.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Oxford, England: Houghton Mifflin.

Ceci, S. J. & Williams, W. M. (2016). A qualitative synthesis of the Flynn Effect. *Measurement: Interdisciplinary Research and Perspectives*, 14(2), 56-63. <http://dx.doi.org/10.1080/15366367.2016.1173949>

Cotton, S. M., Kiely, P. M., Crewther, D. P., Thomson, B., Laycock, R. & Crewther, S. G. (2005). A normative and reliability study for the Raven's Coloured Progressive Matrices for primary school aged children from Victoria, Australia. *Personality and Individual Differences*, 39(3), 647-659. <http://dx.doi.org/10.1016/j.paid.2005.02.015>

Crossman, A. M., Powell, M. B., Principe, G. F. & Ceci, S. J. (2002). Child testimony in custody cases: A review. *Journal of Forensic Psychology Practice*, 2(1), 1-31. [http://dx.doi.org/10.1300/J158v02n01\\_01](http://dx.doi.org/10.1300/J158v02n01_01)

Dickens, J. R. W. T. & Flynn. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review*, 108, 346-369. <http://dx.doi.org/10.1037/0033-295X.108.2.346>

Dickinson, M. D. & Hiscock, M. (2010). Age-related IQ decline is reduced markedly after adjustment for the Flynn effect. *Journal of Clinical and Experimental Neuropsychology*, 32(8), 865-870. <http://dx.doi.org/10.1080/13803391003596413>

Dutton, E. & Lynn, R. (2013). A negative Flynn effect in Finland, 1997-2009. *Intelligence*, 41(6), 817-820. <http://dx.doi.org/10.1016/j.intell.2013.05.008>

Dutton, E. & Lynn, R. (2015). A negative Flynn Effect in France, 1999 to 2008-9. *Intelligence*, 51, 67-70. <http://dx.doi.org/10.1016/j.intell.2015.05.005>

Eppig, C., Fincher, C. L. & Thornhill, R. (2010). Parasite



prevalence and the worldwide distribution of cognitive ability. *Proceedings of the Royal Society B: Biological Sciences*, 277, 3801-3808. <http://dx.doi.org/10.1098/rspb.2010.0973>

Erickson, S. K., Lilienfeld, S. O. & Vitacco, M. J. (2007). A critical examination of the suitability and limitations of psychological tests in family court. *Family Court Review*, 45(2), 157-174.

Fletcher, J. M., Stuebing, K. K. & Hughes, L. C. (2010). IQ scores should be corrected for the Flynn effect in high-stakes decisions. *Journal of Psychoeducational Assessment*, 28(5), 469-473. <http://dx.doi.org/10.1177/0734282910373341>

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95(1), 29. <http://dx.doi.org/10.1037//0033-2909.95.1.29>

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101(2), 171. <http://dx.doi.org/10.1037//0033-2909.101.2.171>

Flynn, J. R. (1999). Searching for justice: the discovery of IQ gains over time. *American Psychologist*, 54(1), 5.

Flynn, J. R. (2012). *Are we getting smarter?: Rising IQ in the twenty-first century*. Cambridge: Cambridge University Press.

Flynn, J. R. & Rossi-Casé, L. (2012). IQ gains in Argentina between 1964 and 1998. *Intelligence*, 40(2), 145-150. <http://dx.doi.org/10.1016/j.intell.2012.01.006>

Greenfield, P. M. (1998). The cultural evolution of IQ. In U. Neisser (Ed) *The rising curve: Long-term gains in IQ and related measures* (pp. 81-123). Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/10270-003>

Habets, P., Jeandarme, I., Uzieblo, K., Oei, K. & Bogaerts, S. (2015). Intelligence is in the eye of the beholder: Investigating repeated IQ measurements in forensic psychiatry. *Journal of Applied Research in Intellectual Disabilities*, 28(3), 182-192. <http://dx.doi.org/10.1111/jar.12120>

Hagan, L. D., Drogin, E. & Guilmette, T. (2010). IQ scores should not be adjusted for the Flynn effect in capital punishment cases. *Journal of Psychoeducational Assessment*, 28(5), 474-476. <http://dx.doi.org/10.1037/a0012693>

Hagan, L. D., Drogin, E. Y. & Guilmette, T. J. (2008). Adjusting IQ scores for the Flynn effect: Consistent with the standard of practice? *Professional Psychology: Research and Practice*, 39(6), 619. <http://dx.doi.org/10.1177/073428291373343>

Hannan, T. (2005). Assessing children: Hits and myths. *InPsysc*, 27(3), 14-17.

Herrnstein, R. J. & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Simon and Schuster.

Kamieniecki, G. W. & Lynd-Stevenson, R. M. (2002). Is it appropriate to use United States norms to assess the "intelligence" of Australian children? *Australian Journal of Psychology*, 54(2), 67-78.

Kanaya, T. & Ceci, S. (2012). The impact of the Flynn effect on LD diagnoses in special education. *Journal of Learning Disabilities*, 45(4), 319-326. <http://dx.doi.org/10.1177/0022219410392044>

Kanaya, T. & Ceci, S. J. (2011). The Flynn effect in the WISC subtests among school children tested for special education services. *Journal of Psychoeducational Assessment*, 29(2), 125-136. [doi.org/10.1177/0734282910370139](http://dx.doi.org/10.1177/0734282910370139)

Kaufman, A. S. (2010). In what way are apples and oranges alike? A critique of Flynn's interpretation of the Flynn Effect. *Journal of Psychoeducational Assessment*, 28(5), 382-398. [doi.org/10.1177/0734282910373346](http://dx.doi.org/10.1177/0734282910373346)

Leigh, A. & Gong, X. (2009). Estimating cognitive gaps between Indigenous and non-Indigenous Australians. *Education Economics*, 17(2), 239-261. <http://dx.doi.org/10.1080/09645290802069418>

Lynn, R. (1982). IQ in Japan and the United States shows a

growing disparity. *Nature*, 297, 222-223. <http://dx.doi.org/10.1038/297222a0>

Lynn, R. (1989). Positive correlation between height, head size and IQ: A nutrition theory of the secular increases in intelligence. *British Journal of Educational Psychology*, 59, 372-377. <http://dx.doi.org/10.1111/j.2044-8279.1989.tb03112.x>

McDonald, J. L., Comino, E., Knight, J. & Webster, V. (2012). Developmental progress in urban Aboriginal infants: A cohort study. *Journal of Paediatrics and Child Health*, 48(2), 114-121. <http://dx.doi.org/10.1111/j.1440-1754.2011.02067.x>

Melton, G. B., Petrila, J., Poythress, N. G. & Slobogin, C. (2007). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers*. Guilford Press.

Merrill, M. A. (1938). The significance of IQ's on the revised Stanford-Binet scales. *Journal of Educational Psychology*, 29(9), 641-651. <http://dx.doi.org/10.1037/h0057523>

Mingroni, M. A. (2007). Resolving the IQ paradox: heterosis as a cause of the Flynn effect and other trends. *Psychological Review*, 114(3), 806. <http://dx.doi.org/10.1037/0033-295X.114.3.806>

Must, O. & Must, A. (2013). Changes in test-taking patterns over time. *Intelligence*, 41(6), 780-790. <http://dx.doi.org/10.1016/j.intell.2013.04.005>

Neihart, M. (1999). The impact of giftedness on psychological well-being: What does the empirical literature say? *Roeper Review*, 22(1), 10-17.

Neisser, U. (1997). Rising scores on intelligence tests. *American Scientist*, 85(5), 440-447.

Nevin, R. (2000). How lead exposure relates to temporal changes in IQ, violent crime, and unwed pregnancy. *Environmental Research Section A*, 83, 1-22. <http://dx.doi.org/10.1006/enrs.1999.4045>

Pearson, C. (2012). Recruitment of indigenous australians with

linguistic and numeric disadvantages. *Research & Practice in Human Resource Management*, 20(1), 1.

Pietschnig, J. & Voracek, M. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn Effect (1909-2013). *Perspectives on Psychological Science*, 10(3), 282-306. [doi.org/10.1177/1745691615577701](https://doi.org/10.1177/1745691615577701)

Pink, B. (2009). Australian social trends: Using statistics to paint a picture of Australian society. *Canberra: Australian Bureau of Statistics*.

Preston, S. H. (1998). Differential fertility by IQ and the IQ distribution of a population. In U. Neisser (Ed.) *The rising curve: Long-term gains in IQ and related measures* (pp. 377-387). Washington, DC, US: American Psychological Association. <http://dx.doi.org/10.1037/10270-014>

Raven, J. C. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1), 1-48. <http://dx.doi.org/10.1006/cogp.1999.0735>

Raven, J. C. & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales*. Oxford Psychologists Press Oxford, UK.

Rodgers, J. L. (2015). Methodological issues associated with studying the flynn effect: Exploratory and confirmatory efforts in the past, present, and future. *Journal of Intelligence*, 3(4), 111-120. <http://dx.doi.org/10.3390/jintelligence3040111>

Rönnlund, M. & Nilsson, L.-G. (2009). Flynn effects on sub-factors of episodic and semantic memory: Parallel gains over time and the same set of determining factors. *Neuropsychologia*, 47(11), 2174-2180. <http://dx.doi.org/10.1016/j.neuropsychologia.2008.11.007>

Schaie, K. W. & Strother, C. R. (1968). A cross-sequential study of age changes in cognitive behavior. *Psychological Bulletin*, 70(6), 671-680. <http://dx.doi.org/10.1037/h0026811>

Schooler, C. (1998). Environmental complexity and the Flynn effect. In U. Neisser (Ed.) *The rising curve: Long-term gains in IQ*

and related measures (pp. 67-79). Washington, DC, US: American Psychological Association. <http://dx.doi.org/10.1037/10270-002>

Smith, S. (1942). Language and non-verbal test performance of racial groups in Honolulu before and after a fourteen-year interval. *The Journal of General Psychology*, 26(1), 51-92.

Sternberg, R. J. (2010). The Flynn Effect: So what? *Journal of Psychoeducational Assessment*, 28(5), 434-440. <http://dx.doi.org/10.1177/073428291373349>

Storfer, M. (1999). Myopia, intelligence, and the expanding human neocortex: Behavioral influences and evolutionary implications. *International Journal of Neuroscience*, 98, 153-276. [doi.org/10.3109/00207459908997465](http://dx.doi.org/10.3109/00207459908997465)

Sundet, J. M. (2014). The Flynn Effect in families: Studies of register data on Norwegian military conscripts and their families. *Journal of Intelligence*, 2(3), 106-118. <http://dx.doi.org/10.3390/jintelligence2030106>

Sundet, J. M., Barlaug, D. G. & Torjussen, T. M. (2004). The end of the Flynn effect?: A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32(4), 349-362. <http://dx.doi.org/10.1016/j.intell.2004.06.004>

Sundet, J. M., Eriksen, W., Borren, I. & Tambs, K. (2010). The Flynn effect in sibships: Investigating the role of age differences between siblings. *Intelligence*, 38(1), 38-44.

Teasdale, T. W. & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn Effect in reverse. *Personality and Individual Differences*, 39(4), 837-843. <http://dx.doi.org/10.1016/j.paid.2005.01.029>

Teasdale, T. W. & Owen, D. R. (2008). Secular declines in cognitive test scores: A reversal of the Flynn Effect. *Intelligence*, 36(2), 121-126. <http://dx.doi.org/10.1016/j.intell.2007.01.007>

te Nijenhuis, J., Cho, S. H., Murphy, R. & Lee, K. H. (2012). The Flynn effect in Korea: Large gains. *Personality and Individual Differences*

*Differences*, 53(2), 147-151. <http://dx.doi.org/10.1016/j.paid.2011.03.022>

Trahan, L. H., Stuebing, K. K., Fletcher, J. M. & Hiscock, M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin*, 140(5), 1332. <http://dx.doi.org/10.1037/a0037173>

Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist*, 3(2), 54-56.

Wang, M., Fuerst, J. & Ren, J. (2016). Evidence of dysgenic fertility in China. *Intelligence*, 57, 15-24. <http://dx.doi.org/10.1016/j.intell.2016.04.001>

Wechsler, D. (1981). *WAIS-R manual: Wechsler adult intelligence scale-revised*. New York: Psychological Corporation.

Williams, R. L. (2013). Overview of the Flynn effect. *Intelligence*, 41(6), 753-764. <http://dx.doi.org/10.1016/j.intell.2013.04.010>

Woodley, M. A. (2012). A life history model of the Lynn-Flynn effect. *Personality and Individual Differences*, 53, 152-156. <http://dx.doi.org/10.1016/j.paid.2011.03.028>

Woodley, M. A., te Nijenhuis, J., Must, O. & Must, A. (2014). Controlling for increased guessing enhances the independence of the Flynn effect from g: The return of the Brand effect. *Intelligence*, 43, 27-34. <http://dx.doi.org/10.1016/j.intell.2013.12.004>

Zhou, X., Zhu, J. & Weiss, L. G. (2010). Peeking inside the "black box" of the Flynn effect: Evidence from three Wechsler instruments. *Journal of Psychoeducational Assessment*, 28(5), 399-411. [http://dx.doi.org/http://dx.doi.org/10.1016/j.intell.2010.12.004](http://dx.doi.org/10.1016/j.intell.2010.12.004)