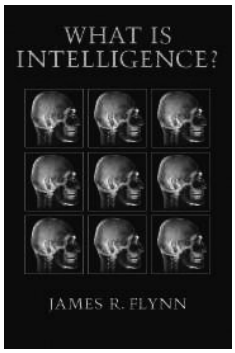
BOOK REVIEW

What is intelligence? Beyond the Flynn effect

James R. Flynn (2007)
Cambridge, UK: Cambridge University Press. 216 pages
ISBN-13: 978-0-521-88007-7
ISBN-10: 0-521-88007-6

Jelte M. Wicherts



Around their 18th birthday, basically all Dutch males born between 1934 and 1964 unknowingly took part in a study of the malleability of intelligence. When these young men appeared before the Dutch military draft board, they took a nonverbal IQ test based on Raven's (1960) Progressive Matrices. With a little help from Piet Vroon, James Flynn (1987) discovered that those born in 1934 (cohort of 1952) scored on average 20 IQ points lower on the test than those born in 1964 (cohort of 1982). This suggested that in only 30 years, the Dutch male population had shown an increase of more than one standard deviation in average IQ. Flynn (1987) also documented this gain in average IQ in 13 other countries over the course of the 20th century and the effect is now commonly known as the Flynn effect. The Flynn effect raises many questions: How can IQ be substantially heritable, yet show such strong gains that appear to be due to environmental factors? Were Dutch males in 1982 so much smarter than Dutch males in 1952?

University of Amsterdam
Department of Psychology, Psychological Methods, University of Amsterdam, Roetersstraat 15, NL 1018 WB Amsterdam, e-mail: j.m.wicherts@uva.nl

In his latest book, James Flynn (2007) discusses the gains that bear his name. He poses four paradoxes and sets out to solve them. His *Factor analysis paradox* is clearly related to our study (Wicherts et al., 2004) in which we found that the gains in five datasets were not factorially invariant with respect to cohorts. For instance, we studied gains of Dutch adults on the Wechsler Adult Intelligence Scale (WAIS; Stinissen et al., 1970) from 1967 to 1999. We found that all subtests showed an increase, but that the Similarities and Comprehension subtests showed gains that were too large to be explained in terms of the latent variables that the WAIS is supposed to measure (e.g., verbal intelligence). Apparently, the Dutch have become better at taking the Similarities and Comprehension subtests, but not necessarily more proficient in 'true' verbal intelligence. In factor analytic terms, this change in subtest-specific ability represents measurement bias, and hence we denoted these gains as artifactual in our paper. However, a subtest-specific effect might still be meaningful. An intelligence subtest measures something general (general intelligence, verbal intelligence, spatial intelligence, etc.), but also something that is specific to this subtest (cf. Spearman, 1904). So, the scores on the Similarities subtest not only reflect verbal intelligence (which in turn is influenced by general intelligence), but also the ability to solve similarity problems such as 'What do automobiles and airplanes have in common?' As we found in our study, these subtest-specific abilities appear to be partly responsible for the Flynn effect. Although we did not study the gains in Raven's tests, it is quite likely that these gains are due in part to subtest-specific effects as well. Ap-

parently, IQ test scores have changed in meaning.

Flynn's second and third paradox are related to the question of whether one can interpret the gains as indicating that more recent generations are relatively 'smart' (Paradox 2) or that the older generations were relatively 'stupid' (Paradox 3). He solves these paradoxes by suggesting that over the course of the 20th century, society has evolved from demanding practical (concrete) world views from its people to demanding more scientific (abstract) world views. According to Flynn, people were not necessarily less intelligent in those days, but they were just not used to viewing the world through scientific spectacles. One benefits greatly from a scientific viewpoint when taking the Similarities and Comprehension subtests. Although he does not consider them as such, Flynn's explanation of the score gains on the Similarities and Comprehension subtests and Raven's tests is that all are gains of subtest-specific abilities. Flynn states that such effects cannot be studied by performing factor analysis. I beg to differ, because that is how we found these effects to begin with! In addition, multi-group factor analysis can be employed to study whether the gains on certain subtests are due to the same common cause. I went back to our WAIS data and indeed found that the Dutch gain on the Similarities and Comprehension subtests can be explained in terms of one additional latent factor. This is consistent with Flynn's hypothesis, but also with alternative hypotheses. For example, in the scoring scheme of the Similarities and Comprehension subtests, points can be gathered by simply talking a lot (which does not apply to other WAIS subtests). Maybe society has changed such that test-takers have become more talkative. While many 'old-fashioned' test-takers did not dare to speak if they were uncertain, contemporary test-takers talk even when they are not sure that they know the right answer. This would result in higher scores on these subtests. A subtest-specific effect found by factorial invariance analyses could also be due to Differential Item Functioning (DIF). For instance, because of their increased use, people may have become more knowledgeable on what airplanes and automobiles have in common, making the item easier. Flynn provides some interesting insights, but more research is needed to understand the psychometric and substantive meaning of the Flynn effect. Besides, test-specific abilities are only part of the story, because there appear to be gains in broad intelligence factors as well (Wicherts et al., 2004). It is likely that the introduction of television and the broad dissemination of computer games has had a positive effect on spatial visualisation skills (Greenfield, 1998). For example, playing the popular computer game of Tetris could enhance one's nonverbal IQ (Okagaki & Frensch, 1994). The use of factor analyses and DIF analyses can shed light on the specific nature of such increases.

Flynn also discusses the 'identical twin paradox', or the question of how high heritability of IQ can be reconciled with the Flynn effect. His solution to this paradox is based on the model developed by Dickens and Flynn (2001) and Flynn suggests some interesting ways to test the model. Flynn goes on to propose that intelligence should be defined or at least studied by focusing on what subtests measure. Although this is an important bottom-up approach, it might not explain why IQ subtests show invariantly positive correlations. Recently, Dickens (2007) and Van der Maas et al. (2006) have proposed developmental models that could explain the existence of the positive manifold without posing an underlying unitary physiological or psychological cause.

In later chapters, Flynn discusses the implications of the Flynn effect. First, the use of obsolete IQ norms normally inflates people's IQs. Flynn nicely illustrates that this may result in an overestimation of groups' average IQ on the basis of old norms. Second, Flynn has been an expert witness in several capital crime cases in the United States, in which the actual execution of defendants depended on IQ scores. Namely, under US law, people suffering from mental retardation (i.e., IQ below 70) are not held responsible for their deeds and hence not executed. However, if the defendant's IQ is based on obsolete norms, his or her IQ may incorrectly fall above this threshold for mental retardation, which may have dire consequences. Flynn rightly argues that a correction needs be made for the Flynn effect when (outdated) IQ scores are used for high-stake decisions. Flynn's discussion of many diverse topics related to the Flynn effect clearly illustrates his broad intellectual focus on society. Educated as a moral philosopher, Flynn has no trouble connecting the IQ gains to issues like democracy, equality, and human rights, as well as to the writings of Plato, Aristotle, and Nietzsche.

Flynn makes some peculiar statements, such as 'g measures general intelligence' or that 'there is no evidence that the gains are a matter of [..] cultural bias'. To me, g equals general intelligence and Flynn's explanation of Similarities' gains is an excellent example of cultural bias. However, Flynn is excused, for he understands his effect exceptionally well. Anyone interested in the Flynn effect (including those Dutch males born between 1934 and 1964) will certainly benefit from reading this book.

Acknowledgement

The preparation of this article was supported by a grant from the Netherlands Organisation for Scientific Research (NWO).

References

Dickens, W.T. (2007). *What is g?* Paper presented at the Schelling Symposium, University of Maryland, College Park, MD.

Dickens, W.T. & Flynn, J.R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review, 108*, 346-369.

Flynn, J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101,* 171-191.

Flynn, J.R. (2007). *What is intelligence? Beyond the Flynn Effect.* Cambridge, UK: Cambridge University Press.

Greenfield, P.M. (1998). The cultural evolution of IQ. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Okagaki, L. & Frensch, P.A. (1994). Effects of video game playing on measures of spatial performance: Gender effects in late adolescence. *Journal of Applied Developmental Psychology, 15,* 33-58.

Raven, J.C. (1960). *Guide to the Standard Progressive Matrices.* London: H.K. Lewis & Co. Ltd.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15,* 201-292.

Stinissen, J., Willems, P.J., Coetsier, P., & Hulsman, W.L.L. (1970). *Handleiding bij de Nederlandstalige bewerking van de Wechsler Adult Intelligence Scale (WAIS). [Dutch WAIS Manual].* Lisse, the Netherlands: Swets and Zeitlinger.

van der Maas, H.L.J., Dolan, C.V., Grasman, R.P.P.P., Wicherts, J.M., Huizenga, H.M. & Raijmakers, M.E.J. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review, 113,* 842-861.

Wicherts, J.M., Dolan, C.V., Hessen, D.J., Oosterveld, P., van Baal, G.C.M., Boomsma, D.I., et al. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence, 32,* 509-537.