

## UC Irvine

### UC Irvine Previously Published Works

**Title**

Increased intelligence is a myth (so far).

**Permalink**

<https://escholarship.org/uc/item/5kh923ms>

**Journal**

Frontiers in systems neuroscience, 8(MAR)

**ISSN**

1662-5137

**Author**

Haier, Richard J

**Publication Date**

2014

**DOI**

10.3389/fnsys.2014.00034

Peer reviewed



# Increased intelligence is a myth (so far)

Richard J. Haier\*

Emeritus, Pediatrics, School of Medicine, University of California, Irvine, CA, USA

\*Correspondence: rjhaier@uci.edu

**Edited by:**

Mikhail Lebedev, Duke University, USA

**Reviewed by:**

James M. Broadway, University of California Santa Barbara, USA

Michael Linderman, Norconnect Inc, USA

**Keywords:** intelligence, *g*-factor, brain imaging, cognitive training, ratio scales, IQ testing

On one hand, intelligence testing is one of the great successes of psychology (Hunt, 2011). Intelligence test scores predict many real world phenomena and have many well-validated practical uses (Gottfredson, 1997; Deary et al., 2010). Intelligence test scores also correlate to structural and functional brain parameters assessed with neuroimaging (Haier et al., 1988; Jung and Haier, 2007; Deary et al., 2010; Penke et al., 2012; Colom et al., 2013a) and to genes (Posthuma et al., 2002; Hulshoff Pol et al., 2006; Chiang et al., 2009, 2012; Stein et al., 2012). On the other hand, intelligence test scores are often misunderstood and can be misused. This paper focuses on a basic misunderstanding that permeates many of the recent reports of increased intelligence following short-term cognitive training. Several of these reports have been published in prominent journals and received wide public attention (Jaeggi et al., 2008, 2011; Mackey et al., 2011).

The basic misunderstanding is assuming that intelligence test scores are units of measurement like inches or liters or grams. They are not. Inches, liters and grams are ratio scales where zero means zero and 100 units are twice 50 units. Intelligence test scores estimate a construct using interval scales and have meaning only relative to other people of the same age and sex. People with high scores generally do better on a broad range of mental ability tests, but someone with an IQ score of 130 is not 30% smarter than someone with an IQ score of 100. A score of 130 puts the person in the highest 2% of the population whereas a score of 100 is at the 50th percentile. A change from an IQ score from 100 to 103 is not the same as a change from 133 to 136. This makes

simple interpretation of intelligence test score changes impossible.

Most recent studies that have claimed increases in intelligence after a cognitive training intervention rely on comparing an intelligence test score before the intervention to a second score after the intervention. If there is an average change score increase for the training group that is statistically significant (using a dependent t-test or similar statistical test), this is treated as evidence that intelligence has increased. This reasoning is correct if one is measuring ratio scales like inches, liters or grams before and after some intervention (assuming suitable and reliable instruments like rulers to avoid erroneous Cold Fusion-like conclusions that apparently were based on faulty heat measurement); it is not correct for intelligence test scores on interval scales that only estimate a relative rank order rather than measure the construct of intelligence. Even though the estimate has considerable predictive value and correlates to brain and genetic measures, it is not a measurement in the same way we measure distance, liquid, or weight even if individual change scores are used in a pre-post design.

SAT scores, for example, are highly correlated to intelligence test scores (Frey and Detterman, 2004). Imagine a student takes the SATs when quite ill. The scores likely are a bad estimate of the student's ability. If the student retakes the test sometime later when well, does an increase in score mean the student's intelligence has increased, or that the newer score is now just a better estimate? The same is true for score changes following SAT preparatory courses. Many colleges and universities allow applicants to submit multiple

SAT scores and the highest score typically carries the most weight; there are many spurious reasons for low scores but far fewer for high scores. Change scores from lowest to highest carry little if any weight. By contrast, change in a person's weight after some intervention is unambiguous.

In studies of the effect of cognitive training on intelligence, it is also important to understand that all intelligence test scores include a certain amount of imprecision or error. This is called the standard error of measurement and can be quantified as an estimate of a "true" score based on observed scores. The standard error of measuring inches or liters is usually zero assuming you have perfectly reliable, standard measurement devices. Intelligence tests generally show high test-retest reliability but they also have a standard error, and the standard error is often larger for higher scores than for lower scores. Any intelligence test score change after an intervention needs to be considered relative to the standard error of the test. Studies that use a single test to estimate intelligence before and after an intervention are using less reliable and more variable scores (bigger standard errors) than studies that combine scores from a battery of tests.

Change scores are never easy to interpret and require sophisticated statistical methods and research designs with appropriate control groups. If you try a training intervention in individuals all of whom have pre-intervention scores below the population mean, for example, re-testing with or without any intervention, may result in higher scores due to the statistical phenomenon of regression to the mean, or due to simple test practice, especially if equivalent alternative forms of

the test are not used. Quasi-experimental designs like post-test only with large samples and random assignment do not have all the same interpretation difficulties as pre-post designs. They have promise but most reviewers are more inclined to value pre-post changes. Latent variable techniques also avoid many of the difficulties of pre-post interval scale changes and they have promise in large samples (Ferrer and McArdle, 2010).

When change scores are used, it is important to identify individual differences even within a group where the average change score statistically increases after an intervention. Imagine a group of 100 students received cognitive training and 100 others received some control intervention. The mean change score in the training group may statistically show a greater increase than the controls. How many of the 100 individuals who received the training actually show an increase? Do they differ in any way from the individuals in the same group who do not show an increase? Does item analysis show whether increased scores are due more to easy test items or hard ones? What about any individuals in the control group that show change score increases as large as shown in the training group? If all 200 participants ultimately get the same training, will the rank order of individuals based on the post-training score be any different than the rank order based on the pre-training scores? If not, what has been accomplished? Most studies do not report such analyses, although newer training studies are addressing issues of multiple measure assessment of intelligence and individual differences (Colom et al., 2013b; Jaeggi et al., 2013). Burgaleta et al provide a good example of showing IQ changes subject-by-subject (Burgaleta et al., 2014).

Nonetheless, the main point is that to make the most compelling argument that intelligence increases after an intervention, a ratio scale of intelligence is required. None yet exists and meaningful progress may require a new way of defining intelligence based on measureable brain or information processing variables. For example, gray and white matter density in specific brain regions assessed by imaging and expressed as a profile of standard scores based on a normative group might substitute for intelligence test scores

(Haier, 2009). Work by Engle and colleagues suggests that working memory capacity and perceptual speed are possible ways to assess fluid intelligence (Broadway and Engle, 2010; Redick et al., 2012) based on a large body of research that shows faster mental processing speed and increased memory capacity are related to higher intelligence.

Jensen has written extensively about an evolution from psychometrics to mental “chronometrics”—the use of response time in milliseconds to measure information processing in a standard way (Jensen, 2006). He argued that the construct of intelligence could be replaced in favor of ratio scale measures of speed of information processing assessed during standardized cognitive tasks like the Hick paradigm. Such measures, for example, would help advance research about the underlying neurophysiology of mental speed and might lead to a more advanced definition of intelligence. Jensen concluded his book on chronometry with this call to action: “. . . chronometry provides the behavioral and brain sciences with a universal absolute scale for obtaining highly sensitive and frequently repeatable measurements of an individual’s performance on specially devised cognitive tasks. Its time has come. Let’s get to work!” (p. 246).

This is a formidable challenge and a major priority for intelligence researchers. Collaboration among psychometricians and cognitive psychologists will be key. There are now a number of studies that fail to replicate the claims of increased intelligence after short-term memory training and various reasons are proposed (Colom et al., 2013b; Harrison et al., 2013). Given our narrow focus here, we note one failure to replicate also assessed working memory capacity and perceptual speed; no transfer effects were found (Redick et al., 2013) and there is reason to suggest that other positive transfer studies may be erroneous (Tidwell et al., 2013). For now, cognitive training results are more inconsistent than not, especially for putative intelligence increases. Nonetheless, it is encouraging that cognitive researchers are working on these issues despite a pervasive indifference or negativity to intelligence research in Psychology in general and for many funding agencies.

In the broader context, intelligence includes more than one component. However, the construct of interest usually is defined by psychometric methods as a general factor common to all mental abilities called the *g*-factor (Jensen, 1998). Fluid intelligence, the focus of several cognitive training studies, is one of several broad intelligence factors and it is highly correlated to *g*. The *g*-factor is estimated by intelligence tests but it is not synonymous with IQ or any other test score; some tests are more *g*-loaded than others. As noted, a score on an intelligence test has little meaning without comparing it to the scores of other people. That is why all intelligence tests require normative groups for comparison and why norm groups need to be updated periodically, as demonstrated by the Flynn Effect of gradual generational increases in intelligence test scores; although whether *g* shows the Flynn effect is still unsettled (te Nijenhuis and van der Flier, 2013). Psychometric estimations of *g* and other intelligence factors have generated strong empirical findings about the nature of intelligence and individual differences, mostly based on correlation studies. These interval assessments, however, are not sufficient to take research to the next step of experimental interventions to increase intelligence.

Speaking about science, Carl Sagan observed that extraordinary claims require extraordinary evidence. So far, we do not have it for claims about increasing intelligence after cognitive training or, for that matter, any other manipulation or treatment, including early childhood education. Small statistically significant changes in test scores may be important observations about attention or memory or some other elemental cognitive variable or a specific mental ability assessed with a ratio scale like milliseconds, but they are not sufficient proof that general intelligence has changed. As in all branches of science, progress depends on ever more sophisticated measurement that drives more precise definitions—think about the evolution of definition for a “gene” or an “atom”. Even with sophisticated interval-based assessment techniques (Ferrer and McArdle, 2010), until we have better measures, especially ratio scales, we need to acknowledge the basic measurement problem and exercise abundant restraint when

reporting putative intelligence increases or decreases.

In the future, there may be strong empirical rationales for spending large sums of money on cognitive training or other interventions aimed at improving specific mental abilities or school achievement (in addition to the compelling moral arguments to do so), but increasing general intelligence is quite difficult to demonstrate with current tests. Increasing intelligence, however, is a worthy goal that might be achieved by interventions based on sophisticated neuroscience advances in DNA analysis, neuroimaging, psychopharmacology, and even direct brain stimulation (Haier, 2009, 2013; Lozano and Lipsman, 2013; Santarnecchi et al., 2013; Legon et al., 2014). Developing equally sophisticated ratio measurement of intelligence must go hand-in-hand with developing promising interventions.

## ACKNOWLEDGMENTS

A version of this paper was presented at the annual meeting of the International Society for Intelligence Research, San Antonio, Texas, December 15, 2012 in a symposium on Improving IQ (chaired by S. Jaeggi and R. Colom).

## REFERENCES

- Broadway, J. M., and Engle, R. W. (2010). Validating running memory span: measurement of working memory capacity and links with fluid intelligence. *Behav. Res. Methods* 42, 563–570. doi: 10.3758/BRM.42.2.563
- Burgaleta, M., Johnson, W., Waber, D. P., Colom, R., and Karama, S. (2014). Cognitive ability changes and dynamics of cortical thickness development in healthy children and adolescents. *Neuroimage* 84, 810–819. doi: 10.1016/j.neuroimage.2013.09.038
- Chiang, M. C., Barysheva, M., McMahon, K. L., de Zubicaray, G. I., Johnson, K., Montgomery, G. W., et al. (2012). Gene network effects on brain microstructure and intellectual performance identified in 472 twins. *J. Neurosci.* 32, 8732–8745. doi: 10.1523/JNEUROSCI.5993-11.2012
- Chiang, M. C., Barysheva, M., Shattuck, D. W., Lee, A. D., Madsen, S. K., Avedissian, C., et al. (2009). Genetics of brain fiber architecture and intellectual performance. *J. Neurosci.* 29, 2212–2224. doi: 10.1523/JNEUROSCI.4184-08.2009
- Colom, R., Burgaleta, M., Roman, F. J., Karama, S., Alvarez-Linera, J., Abad, F. J., et al. (2013a). Neuroanatomic overlap between intelligence and cognitive factors: morphometry methods provide support for the key role of the frontal lobes. *Neuroimage* 72, 143–152. doi: 10.1016/j.neuroimage.2013.01.032
- Colom, R., Roman, F. J., Abad, F. J., Shih, P. C., Privado, J., Froufe, M., et al. (2013b). Adaptive n-back training does not improve fluid intelligence at the construct level: gains on individual tests suggest that training may enhance visuospatial processing. *Intelligence* 41, 712–727. doi: 10.1016/j.intell.2013.09.002
- Deary, I. J., Penke, L., and Johnson, W. (2010). The neuroscience of human intelligence differences. *Nat. Rev. Neurosci.* 11, 201–211. doi: 10.1038/nrn2793
- Ferrer, E., and McArdle, J. J. (2010). Longitudinal modeling of developmental changes in psychological research. *Curr. Dir. Psychol. Sci.* 19, 149–154. doi: 10.1177/0963721410370300
- Frey, M. C., and Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability (vol 15, pg 373, 2004). *Psychol. Sci.* 15, 641. doi: 10.1111/j.0956-7976.2004.00687.x
- Gottfredson, L. S. (1997). Why g matters: the complexity of everyday life. *Intelligence* 24, 79–132. doi: 10.1016/S0160-2896(97)90014-3
- Haier, R. J. (2009). Neuro-intelligence, neuro-metrics and the next phase of brain imaging studies. *Intelligence* 37, 121–123. doi: 10.1016/j.intell.2008.12.006
- Haier, R. J. (2013). *The Intelligent Brain*. Chantilly, VA: The Great Courses Company. Available online at: [http://www.thegreatcourses.com/tgc/courses/course\\_detail.aspx?cid=1642](http://www.thegreatcourses.com/tgc/courses/course_detail.aspx?cid=1642)
- Haier, R. J., Siegel, B. V., Nuechterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., et al. (1988). Cortical Glucose metabolic-rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence* 12, 199–217. doi: 10.1016/0160-2896(88)90016-5
- Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., and Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychol. Sci.* 24, 2409–2419. doi: 10.1177/0956797613492984
- Hulshoff Pol, H. E., Schnack, H. G., Posthuma, D., Mandl, R. C. W., Baare, W. F., van Oel, C., et al. (2006). Genetic contributions to human brain morphology and intelligence. *J. Neurosci.* 26, 10235–10242. doi: 10.1523/JNEUROSCI.1312-06.2006
- Hunt, E. B. (2011). *Human Intelligence*. Cambridge, NY: Cambridge University Press.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., and Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6829–6833. doi: 10.1073/pnas.0801268105
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., and Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10081–10086. doi: 10.1073/pnas.1103228108
- Jaeggi, S. M., Buschkuhl, M., Shah, P., and Jonides, J. (2013). The role of individual differences in cognitive training and transfer. *Mem. Cognit.* doi: 10.3758/s13421-013-0364-z. [Epub ahead of print].
- Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability*. Westport, CT: Praeger.
- Jensen, A. R. (2006). *Clocking the Mind: Mental Chronometry and Individual Differences*. New York, NY: Elsevier.
- Jung, R. E., and Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav. Brain Sci.* 30, 135–154. doi: 10.1017/S0140525X07001185
- Legon, W., Sato, T. F., Opitz, A., Mueller, J., Barbour, A., Williams, A., et al. (2014). Transcranial focused ultrasound modulates the activity of primary somatosensory cortex in humans. *Nat. Neurosci.* 17, 322–329. doi: 10.1038/nn.3620
- Lozano, A. M., and Lipsman, N. (2013). Probing and regulating dysfunctional circuits using deep brain stimulation. *Neuron* 77, 406–424. doi: 10.1016/j.neuron.2013.01.020
- Mackey, A. P., Hill, S. S., Stone, S. I., and Bunge, S. A. (2011). Differential effects of reasoning and speed training in children. *Dev. Sci.* 14, 582–590. doi: 10.1111/j.1467-7687.2010.01005.x
- Penke, L., Maniega, S. M., Bastin, M. E., Hernandez, M. C. V., Murray, C., Royle, N. A., et al. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. *Mol. Psychiatry* 17, 1026–1030. doi: 10.1038/mp.2012.66
- Posthuma, D., De Geus, E. J., Baare, W. F., Hulshoff Pol, H. E., Kahn, R. S., and Boomsma, D. I. (2002). The association between brain volume and intelligence is of genetic origin. *Nat. Neurosci.* 5, 83–84. doi: 10.1038/nn0202-83
- Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., et al. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *J. Exp. Psychol. Gen.* 142, 359–379. doi: 10.1037/a0029082
- Redick, T. S., Unsworth, N., Kelly, A. J., and Engle, R. W. (2012). Faster, smarter? Working memory capacity and perceptual speed in relation to fluid intelligence. *J. Cogn. Psychol.* 24, 844–854. doi: 10.1080/20445911.2012.704359
- Santarnecchi, E., Polizzotto, N. R., Godone, M., Giovannelli, F., Feurra, M., Matzen, L., et al. (2013). Frequency-dependent enhancement of fluid intelligence induced by transcranial oscillatory potentials. *Curr. Biol.* 23, 1449–1453. doi: 10.1016/j.cub.2013.06.022
- Stein, J. L., Medland, S. E., Vasquez, A. A., Hibar, D. P., Senstad, R. E., Winkler, A. M., et al. (2012). Identification of common variants associated with human hippocampal and intracranial volumes. *Nat. Genet.* 44, 552–561. doi: 10.1038/ng.2250
- te Nijenhuis, J., and van der Flier, H. (2013). Is the Flynn effect on g? A meta-analysis. *Intelligence* 41, 802–807. doi: 10.1016/j.intell.2013.03.001
- Tidwell, J. W., Dougherty, M. R., Chrabaszcz, J. R., Thomas, R. P., and Mendoza, J. L. (2013). What counts as evidence for working memory training? Problems with correlated gains and dichotomization. *Psychon. Bull. Rev.* doi: 10.3758/s13423-013-0560-7. [Epub ahead of print].

Received: 29 January 2014; accepted: 23 February 2014; published online: 12 March 2014.

Citation: Haier RJ (2014) Increased intelligence is a myth (so far). *Front. Syst. Neurosci.* 8:34. doi: 10.3389/fnsys.2014.00034

This article was submitted to the journal *Frontiers in Systems Neuroscience*.

Copyright © 2014 Haier. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.