Intelligence 56 (2016) 65–71

Contents lists available at ScienceDirect

Intelligence

Does the raising IQ-raising *g* distinction explain the fadeout effect?

John Protzko

University of California, Santa Barbara, United States

ARTICLE INFO

Article history:
Received 4 November 2015
Received in revised form 29 January 2016
Accepted 18 February 2016
Available online xxxx

Keywords:
Intelligence
Cognitive development
Factor analysis
g

ABSTRACT

Previous investigations into raising IQ show that after an intervention ends, the effects fade away. This paper is an attempt to understand one possible reason for this fadeout; the idea that the effects fade because they were not to the underlying construct *g*. A large (N = 985) randomized controlled trial is re-analyzed to investigate whether the intervention, which began at birth and lasted for the first three years of the children's life, raised the underlying cognitive factor of IQ tests. This was done under strict measurement invariance. The intervention indeed raised the *g* factor at age three. No effects were seen at follow-up assessments at ages five and eight after the intervention ended. Therefore, the raising IQ/raising *g* distinction is insufficient as an explanation for the fadeout effect, as changes to the environment can improve *g* and still fade.

1. Introduction

No psychological intervention is permanent; after a time, the experiment ends. While the results of randomized controlled trials (RCTs) allow us to understand causality (e.g. Shadish, Cook, & Campbell, 2002)—following children after an early intervention ends allows us to test theories of causality in development. One phenomenon that has come out of the research on raising IQ is the fadeout effect (Protzko, 2015; Bailey, Duncan, Odgers and Yu, 2015; Bailey, Watts, Littlefield and Geary, 2015). After the intervention ends, the beneficial effects on IQ fade away (see also Herrnstein & Murray, 1994; Howe, 1997). The question is: why does the fadeout effect for IQ occur?

One possible reason is that when an intervention raises IQ scores it does not raise the underlying trait of intelligence (Jensen, 1997). The general cognitive ability factor, general intelligence factor, or *g*, is the highest factor that occurs from a battery of cognitive tests. Since almost all cognitive ability tests correlate with one another (Jensen, 1998), *g* is meant to represent the latent construct that each test is measuring. This theory differentiates effects on IQ scores (test scores) from *g* (trait of intelligence) and from here on is referred to as the raising IQ/raising *g* theory. A corollary would be teaching a sprinter to lean forward at the end of a race to have their torso cross the finish line earlier. One did not make them any faster (no change to the trait of speed), while still lowering their time (manifest test score). The purpose of this paper is to examine whether the raising IQ/raising *g* theory explains the fadeout effect.

We first examine some of the assumptions of the theory. The first assumption is that the *g* substantially differs from full-scale IQ scores. People who do well on one cognitive test also do well on other tests; this high correlation indicates the tests may all be measuring the same underlying construct (Jensen, 1998). This construct, the name for the single factor extracted via factor analysis, is *g* (or general intelligence; Cattell, 1971). While there are other ways to analyze the intercorrelations of multiple cognitive tests (Gf-Gc, Horn & Cattell, 1967; VPR, Johnson & Bouchard, 2005; CHC, McGrew, 2009), each method still entails a *g* factor because of residual and group factor correlations (see McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002; Blair, 2007, 2010; for well written critiques of *g*). A complete review of the merits of whether *g* differs enough from IQ is far beyond the scope of this paper. Many intelligence researchers believe they differ enough to be considered related but distinct (e.g. Gottfredson, 1997; Jensen, 1998), a view we share here. If we assume the two are not distinct then there is no purpose to examine the raising IQ/raising *g* theory.

A second assumption of the theory is that raising IQ does not ensure raising *g*. An IQ test is often composed of different subtests and these subtests correlate with the *g* factor to different extents. The assumption is mainly from IQ to *g*. One can raise IQ scores without raising *g* (as opposed to raising *g* without affecting IQ scores). When this occurs, the gains are "hollow" (like the runner who leans forward to increase his time). One way to see if *g* is raised is to examine the subtest scores following an intervention and see if it is related to the subtest's g-loading. If one raised *g*, subtests that correlate heavily with *g* will increase more than subtests that are less g-loaded; more *g*, more movement. If one did not raise *g* then the more g-loaded a subtest, the less it is affected (Jensen, 1997).

There are two known instances where IQ goes up but *g* does not: practice/retest effects and teaching to the test. When an individual takes an IQ test—they receives a score that places their level of cognitive ability in relation to others. When they take the test again, their score increases due to practice/retest effects. For subjects new to IQ testing, these gains are around 3–6 IQ points and come from myriad factors—including decreased anxiety, better understanding of instructions, better knowledge

http://dx.doi.org/10.1016/j.intell.2016.02.008
0160-2896/

of how to allocate time to certain questions (Vernon, 1954, 1960). The gains represent an increase in specific elements to the test-taking situation. Appropriately, across 64 studies on retest effects, *g* is unmoved by practice and retest effects despite increases in IQ scores (te Nijenhuis, van Vianen, & van der Flier, 2007; see also Estrada, Ferrer, Abad, Román, & Colom, 2015).

For teaching to the test, Mediated Learning (Feuerstein, 1979) is a program that exposes underprivileged individuals to the knowledge and skills tested on IQ tests. Training procedures include direct training on matrix-reasoning questions that differ only trivially from IQ tests (Feuerstein, Rand, Hoffman, & Miller, 1980, p. 115–116). This experimental paradigm is purposefully and admittedly teaching to the tests. Appropriately, Mediated Learning increases IQ scores but has no effect on *g* (te Nijenhuis et al., 2007). Retest effects and teaching to the test represent ways to increase IQ scores without affecting *g*. We argue these two paradigms are categorically different from interventions that try to raise intelligence through intense environmental change or pharmacological intervention.

The raising IQ/raising *g* theory is still strong. In 2014, a meta-analysis of non-experimental studies of Head Start on IQ indicated that the gains were largely unrelated to *g*. The authors concluded "The finding that the IQ gains from Head Start were mostly on the non-*g* variance might explain why IQ gains from such programs fade with time" (te Nijenhuis, Jongeneel-Grimen, & Kirkegaard, 2014). Therefore, this argument must be addressed with experimental methods and tested to better understand the nature of the fadeout effect. In this paper, we examine whether a three-year intervention that started at birth raised *g*.

Using a paradigm that attempts to raise intelligence and not just IQ scores, we examine the fadeout effect. A possible reason for the fadeout effect is the gains in IQ were never to the underlying construct *g*; the gains were merely to extraneous variance in the IQ test (Jensen, 1997). The gains were not real or appreciable in the first place, they were not gains to *g*. This implies that the fadeout occurred because without effects on *g*—there can be no lasting effect.

To fully disprove this theory we would have to show that interventions that do not raise *g* fade, and those that do not fade never raised *g*. This, however, seems unlikely to occur because there is as yet no intervention studied that does not exhibit the fadeout effect. We propose this theory is still testable because it invokes a difference between raising IQ and raising *g*. It suggests that something different would happen if an intervention were able to raise *g*. Namely: it would not experience the fadeout effect.

This provides us with two hypotheses that we explore in this paper:

H1) a genuine increase in IQ scores not from teaching to the test or practice/retest effects will also raise *g*

H2) there should be no fadeout from an intervention that raised *g*, if the raising IQ/raising *g* theory is correct.

We explore the relationship between the effects of an intervention from birth on *g*. We use an intervention that did not try to manipulate retest effects, which did not teach directly to the tests, but instead tried to make a genuine and permanent gain in the lives of children.

We organize the analysis and results around our two hypotheses. To test the first hypothesis, we re-analyze an intervention known to exhibit the fadeout effect. We test whether the intervention raised *g* by the time the intervention ended. If the intervention did affect *g* it would decrease the likelihood that the fadeout effect is due to the raising IQ/raising *g* distinction. If the intervention did not affect *g*, this increases the likelihood that the fadeout effect is due to the raising IQ/raising *g* distinction. To test the second hypothesis we explore whether there is still fadeout after an intervention raises *g*. Given that the intervention affected *g*, we explore whether there are continuing effects on *g* despite the fadeout of IQ scores. If the intervention continues to have an effect on *g* and does not fade, this increases the likelihood of the raising IQ/raising *g* explanation of the fadeout effect. If, instead, there is still a fadeout effect for results on *g*, the raising IQ/raising *g* distinction is an unlikely explanation of the fadeout effect. It follows that the second hypothesis is only testable in light of an initial increase to *g*.

2. Method

In a large national intervention, 985 low birth weight (≤2500 g; LBW) children were provided an intense and cognitively demanding environment during the first three years of their lives (see *Intervention* section, also see Infant Health and Development Project, IHDP; Ramey et al., 1992). The study aimed to ameliorate the negative effects of being born at LBW. One of the many outcomes studied was the IQ of the children.

Infants were randomly assigned at birth based on strata of birth weight (high- and low-LBW—above and below 2000 g, respectively) to the experimental or control group. By the end of the three year intervention, children who received the intervention had significantly higher IQs (9.4 points, Ramey et al., 1992). At ages five and eight, two and five years after the intervention ended, there was no more effect of the intervention on the whole sample's IQ (Brooks-Gunn et al., 1994; McCarton et al., 1997).

Unlike instances of teaching to the test (e.g. Skuy et al., 2002), this intervention showed parallel gains in mathematics ability as well as IQ five years after the intervention ended (McCarton et al., 1997), suggesting, though not proving, the IQ gains were not empty.[1] To test whether the fadeout effect could be because of the raising IQ/raising *g* distinction, we look at the effect of the intervention directly on *g*. The method used here is a direct test of whether the intervention raised *g*, superior to an indirect method of argument based on transfer.

2.1. Intervention

The intervention had three main components: home visits, attendance at special full day child development centers, and parent support groups. The home visits occurred weekly for the first year of the child's life and then bi-weekly afterwards. Home visits provided emotional, social, and practical support, as well as information to parents regarding their child's development. Children attended the development centers 50 weeks/year, 5–6 days per week for the second and third years of their lives. Qualified and trained personnel staffed the development centers and constructed educational activities. Parents met in support groups seasonally for the last two years of the intervention (see Ramey et al., 1992 for further details).

2.2. Measures

Children took the Stanford–Binet Intelligence Scale, form L–M, 3rd edition (SB; Terman & Merrill, 1972) at three when the intervention ended; the Peabody Picture Vocabulary Test-R (PPVT; Dunn & Dunn, 1981) and the Wechsler Preschool and Primary Scale of Intelligence-R (WPPSI; Wechsler, 1989) at five; and the Wechsler Intelligence Scale for Children-III (WISC; Wechsler, 1991), the PPVT, and the Matrices subtest from the Differential Ability Scales (DAS; Elliot, 1990) at eight years old. The SB consists of groups of six questions administered in ascending order starting at the age of the child. Each question is marked as either correct or incorrect. The test was constructed without subtests. The PPVT consists of series of words and four accompanying pictures. Individuals reason out which picture exemplifies the definition of the word. The WPPSI consists of 12 subtests that load onto performance (Object Assembly, Animal Pegs, Geometric Design, Mazes, Block Design,

[1] Children did not receive mathematics instruction during their time in the intervention (first three years of life). It is possible that the intervention caused math gains by age 8 through non-*g* pathways, such as causing long-term increased motivation. We thank Dr. Drew Bailey suggesting this.

and Picture Completion) and verbal (Arithmetic, Sentences, Information, Vocabulary, Comprehension, and Similarities) factors. The DAS provides children with a pattern with a missing element and asks children to complete the pattern. The WISC consists of ten standard subtests divided into Verbal (Information, Similarities, Arithmetic, Vocabulary, Comprehension) and Performance (Picture Completion, Coding, Picture Arrangement, Block Design, Object Assembly) factors.

2.3. Statistical analyses

To investigate whether the intervention directly affected *g*, we use multi-group confirmatory factor analysis (MGCFA). MGCFA tests whether the factor structure of the tests is the same between the groups, and whether there are any differences on the factors. This allows us to create a *g* factor and test whether there were any differences between the experimental and control groups. This method is superior to correlating gains and *g*-loadings as MGCFA has been shown to control for type I error rates in this type of investigation (Lubke, Dolan, & Kelderman, 2001). This procedure is chosen because previous methods of investigating the raising IQ/raising *g* hypothesis have relied on correlations between subtest gains and the *g*-loading of those gains (e.g. te Nijenhuis et al., 2007, 2014). Those previous analyses have been shown to be extremely biased in the presence of even minor violations of the factor analytic model (Ashton & Lee, 2005; Dolan, 2000; Lubke et al., 2001).

MGCFA tests whether the scores on higher order factors (such as *g*) are the same between groups. We also test measurement invariance between the groups, which tests if the factor structure, factor loadings, means, and variances are the same between the two groups (Dolan, 2000). It is possible that the intervention could actually change the nature of the *g* factor, if so strict invariance (same factor structure, factor loadings, and mean variance are the same) would not hold between the groups.

Hypothesis 1. We use the individual items for this analysis because the SB was created with no subtests and as a 1 factor test (McNemar, 1942). We use MGCFA with strict factorial invariance (Dolan, 2000) on the age three data to confirm: a) the same construct is being measured in both the experimental and control groups, and b) the differences between the groups exists on the *g* factor.

Hypothesis 2. To investigate whether there is still fadeout on the *g* factor after the intervention ends we perform a MGCFA with the age five variables in a single *g* factor model, the age eight data in a hierarchical *g* factor model, and the *g* factor at age eight regressed on the age five *g* factor. We again use strict factorial invariance. If the fadeout effect still occurs there will be no effect on *g* at either age five, eight, or the path between them.

3. Results

Hypothesis 1. Does an intervention that alters a child's life and raises his or her IQ also raise *g*?

We used a tetrachoric single-factor analysis using theta parameterization (Millsap & Yun-Tein, 2004) with strict factorial invariance (Dolan, 2000) to investigate model fit.[2] In the strict invariance model there were no improper estimates (all factor loadings and residual variances positive, for example). The latent variable *g* was properly identified for both experimental and control groups. All of the factor loadings were positive, were coherent, and were significantly larger than zero for both experimental and control groups. The single factor model conformed to strict invariance (groups held invariant across factor loadings, thresholds, and error variances) across the control and experimental groups (See Table 1). The excellent fit indicates that the strict invariance model adequately described the data; the same *g* is being measured in both experimental and control groups.

Loosening the invariance constraints by letting groups vary across error variances, factor loadings, or thresholds showed no change in model fit (all $\Delta CFI \leq .01$; see Cheung & Rensvold, 2002; Chen, 2007 for standards of model fit and comparison of statistics in invariance testing); this means we have strong reason to believe the same construct (*g*) was being measured in both experimental and control groups. Strict measurement invariance was established (Vandenberg & Lance, 2000).

For dichotomous item-level data, measurement invariance as shown here means not only that the same measure (*g*) is being measured in both groups, all error variances are the same, and all items load onto the same factor equally, it also means any increase in *g* is equal across item difficulties. If an intervention only increases ability on easier or more difficult items (see Esposito et al., 2013 for an example), measurement invariance would be violated. Strict invariance means a stable increase across all item difficulties (See Table 2 in Appendix A for item-level descriptives).

Having ascertained the same *g* is being measured in both groups, we can investigate whether one group scored differently on that construct. Using MGCFA, we see that the experimental group scored significantly higher on the *g* factor than did the control group ($\beta = .496$, 95%CI = .356 to .636). Thus, we can see the IHDP caused an increase in the *g* factor, confirming our first hypothesis.

Hypothesis 2. There should be no fadeout from an intervention that raised *g*, if the raising IQ/raising *g* theory is correct.

Age five: Two years after the intervention ended, the basic increase in IQ seen at age three had faded out for the whole sample (Brooks-Gunn et al., 1994). While the experimental group still had higher IQs, the result was not statistically significant. Unfortunately, the individual subtest data for the WPPSI is not available for analysis, so we derive the *g* factor from the PPVT and the verbal and performance factors of the WPPSI. We again use strict factorial invariance (Dolan, 2000) to investigate model fit. The single factor model conformed to strict invariance (groups held invariant across factor loadings, and thresholds) across the control and experimental groups (CFI = .989, RMSEA = .079). As before, we have strong reason to believe the same construct is being measured in both experimental and control groups.

Age eight: Five years after the intervention ended, the basic increase in IQ seen at age three was still absent in the full sample (McCarton et al., 1997). We derive a *g* factor from the PPVT, the DAS, and all ten subtests of the WISC. For the age 8 data we repeat the same procedures above. The analysis first revealed single factor model (CFI = .919) did not fit as well as a two-factor (CFI = .966) hierarchical model (consistent with the factor structure of the WISC-III; $\Delta CFI = .047$).

We investigated whether the intervention had enduring effects on children at age five and eight using MGCFA (see Fig. 1). The fadeout effect was still present on *g*, children in the experimental group were no different in *g* than children in the control group at either age five ($\beta = .048$, 95%CI = −.089 to .185) or age eight ($\beta = .035$, 95%CI = −.098 to .169).

Therefore, the IHDP intervention raised *g* but still faded out.

Table 1
Fit indices for Strict invariance model and free error variance models.

Fit index	Strict invariance	Separate error variance	Separate factor loadings	Separate thresholds
CFI	.955	.954	.955	.956
RMSEA	.038	.039	.038	.038

[2] We follow Byrne, Shavelson, and Muthén (1989) in model order. We begin with an invariant model first followed by testing source of model misfit. This was done for the a priori reason that the data here come from a randomized controlled trial. Thus, unlike other instances of measurement invariance of testing separate groups, we are dealing with counterfactual and not comparison groups. This follows from the logic of random subpopulations and invariance (Meredith, 1964). See Horn & McArdle, 1992 for a similar approach.

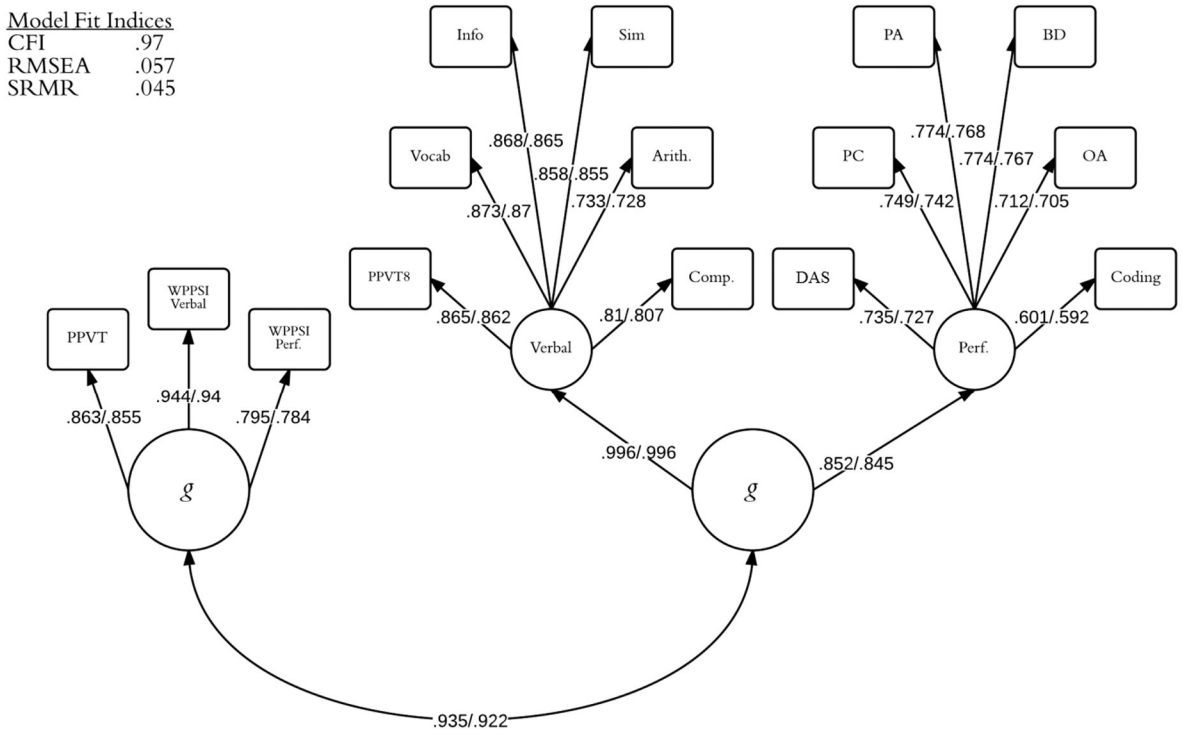

Fig. 1. Model used for the ages 5–8 MGCFA. PPVT = Peabody Picture Vocabulary Test, DAS = Matrices subtest of the Differential Ability Scales. All other abbreviations belong to WISC subtests: Vocab = Vocabulary, Info = Information, Sim = Similarities, Arith = Arithmetic, Comp = Comprehension, PC = Picture Completion, PA = Picture Arrangement, BD = Block Design, OA = Object Assembly. Note in the final model PPVT at ages 5 and 8 were allowed to covary, along with the WPPSI performance test at age 5 and the performance latent variable at age 8. This was done to improve model fit and it did not alter the results. Loadings are for the control group/experimental group, respectively. Note that a latent growth-curve model (or longitudinal modeling in general) cannot be modeled on this data, as there is no homogeneity of index (indicators for age 3 is item-level data as there are no subtests, for age 5 is a test, and for age 8 is a latent variable).

4. Discussion

The purpose of this article was to better understand the fadeout effect, wanting to see if the raising IQ/raising *g* distinction was an explanation. The IHDP was a nationally based intervention using LBW children. The intervention began at birth and continued through the first three years of the children's lives. Previous investigations show the intervention raised the children's IQ. Now we see that not only did it raise their IQ, but also raised *g*. This allowed us to explore whether the effects on *g* faded out. If the effects did not fade, this would increase the likelihood of the raising IQ/raising *g* theory as an explanation to the fadeout effect: fade out only occurs when an intervention raises IQ but fails to raise *g*. If the effects of the intervention faded on *g*, this would undermine the purpose of the raising IQ/raising *g* distinction and decrease its likelihood as an explanation of the fadeout effect.

When the intervention ended, the children had higher IQs and higher *g*. Two years later, the beneficial effects on IQ and on *g* faded out. Three years later still, when the children were eight, the effects on IQ and on *g* remained absent.

The IHDP was a study of massive scope and complexity that saw positive IQ gains for children—gains which then faded. A possibility for explaining any fadeout effect is that the original benefits to IQ from an intervention were hollow; they were not gains to *g*. Here, for the IHDP intervention, this was not the case. The intervention did raise *g* and still faded. Therefore, the raising IQ/raising *g* distinction is insufficient as an explanation for the fadeout effect.

This does not mean that an intervention can only affect the underlying construct or else be an example of teaching to the test. An intervention can affect test-specific variance that is independent of an underlying construct. However, if a scientific question is aimed at the construct level, such as does an intervention raise intelligence (as a construct), analysis should be aimed at the latent variable level and not test-specific variance (Borsboom & Mellenbergh, 2002).

4.1. Alternate theories and stability and the fadeout effect

It could be argued that there was no fadeout, but instead *g* measured at age three is substantively different from *g* measured at age five. In this study, as *g* is indexed by different tests, there could be a possibility that *g* may not be the same construct between the ages. We find this unlikely. A supplementary analysis reveals that the correlation between *g* at age three and *g* at age five is so strong ($r = .883$, $r = .86$ for experimental and control groups, respectively) that it is difficult to say *g* at the different age groups are substantively different, when the correlations between them indicate they are similar. This high correlation is a necessary but not sufficient condition for the claim that *g* is not substantively different at different ages (see Borsboom, Mellenbergh, & van Heerden, 2004). It is difficult to explain how the correlation between the two processes can be so high while the trait under question not only 'substantively changes', but also shows the fadeout effect. Both the experimental and control groups demonstrate the same stability of latent *g* between ages 3–5. In addition, developmental work on the nature of intelligence indicates that a unitary *g* fits developmental models better than a new *g* at each age (Žebec, Demetriou, & Kotrla-Topić, 2015).

In addition, systematic investigations into whether *g* is different when extracted from different batteries has shown the *g* factor to be invariant to the type of tests that make up the battery (Jensen & Weng, 1994; Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004; Thorndike, 1987). Also, in investigations when the tests used remain the same over ages, there is no change in factor structure over early development (e.g. Bickley, Keith, & Wolfle, 1995). Such an argument does not explain why the "changing *g*" would differ between groups. If the intervention was targeted to age 3-specific cognitive processes, the initial

increase in *g* would not be present as such teaching to the tests do not transfer (te Nijenhuis et al., 2007).

This supplementary analysis also highlights the early stability of *g* across ages. The correlation of ages 3–5 on *g* was over .86, and the correlation of ages 5–8 was over .92. This suggests that, while the nature of the test batteries may change and grow with development, the underlying factor common among them is exceptionally robust. Therefore, we can reject the idea that the nature of *g* changes substantially and is responsible for the fadeout effect. To focus on the tests that make up the battery and ignore the structure of those tests and correlations between factors undermines the idea of latent variables and modern measurement theory. It places the individual test above the construct, which leads, in our opinion, inevitably back to operationalism. While the nature of the test batteries may change and grow with development, the underlying factor common among them remains robust. A full discourse on the meaning of stability of *g* in development is beyond the scope of this paper, though we will wade into it somewhat here. This treatment is not exhaustive.

There are four types of explanations for developmental stability as high as we observed in this study: i) the correlation is spurious; ii) there is a causal connection between *g* at ages three to five (either direct or indirect); iii) the same construct is being measured twice; or iv) there is a common cause to both *g* at age three and five. As before, the high correlation is a necessary but not sufficient for all of these explanations.

We find the first explanation, that the relationship is spurious, to be unlikely. The reason is not the magnitude of the correlation; it is based on theoretical and causal knowledge of the development of intelligence. After all, the divorce rate in Maine correlated with the per capita consumption of margarine at $r = .993$ between 2000 and 2009.[3] Spurious correlations can take any magnitude.

It is also plausible that the high correlation is indicative of a causal relationship. The correlation between lightning and thunder is nearly perfect. This is not because thunder and lightning are the same thing, but instead because lightning causes thunder (the rapid expansion of air from being heated by lightning). Similarly, there may be a causal connection between *g* at each age, explaining the high correlation. These causal connections could be either direct (*g* at age three causes *g* at age five; see Žebec et al., 2015) or indirect (*g* at age three causes individuals to self-select and be selected into environment which cause *g* at age five; e.g. Bronfenbrenner & Ceci, 1994). This account has the most immediate intuitive appeal, but runs into problems with the data presented here. If *g* at age three causes *g* at age five, then the intervention effects should still be present. Instead, we observed not only a complete fadeout on *g* at age five, but no overlap whatsoever in the effect sizes (lower bound at age three = .356, upper bound at age five = .185). In fact, *g* at age three explains 78% of the variance of *g* at age five. If we observed a stability of intervention effects and attribute any loss to the less than perfect stability, we would not observe the drop in effect seen here. The intervention effects at age five given 78% of intervention effect size is .278, outside of the CI of age five effects. This suggests that the intervention did not have a stable effect and merely faded because of the less than perfect stability between ages 3–5, something else is at work (see Fig. 2). It also suggests there may not be a causal connection between *g* at different ages (see also Protzko, 2015).

It is possible that the stability shown here is because the same construct *g* is being measured twice. For a long time the concept of *g* has been considered a general mental capacity and capability (e.g. 52 signatories in Gottfredson, 1997; Gottfredson, 2009) and specifically not a skill. This stands in contrast to other latent constructs, such as mathematical ability (see Bailey, Watts, Littlefield & Geary, 2014) which contain both trait-like ability and state-like skill components. Since *g* is considered by many to be a capacity, we might say that *g* at even the earliest ages is the same as *g* in adulthood. Though manifested differently, the *capability* remains the same. Though intuitively unappealing, the notion of capacity has the ability to remain the same over development. More theoretical work must be done to better enlighten the debate on this point.

Finally, the high correlation could be due to a common cause. The stability of height is exceptionally robust as well, the correlation of height between age nine and adulthood is .83 (Tanner, Whitehouse, Marshall, & Carter, 1975). This could mean that the 'height' at age nine and adulthood is the same thing, with height conceptualized as: the *inter-individual difference* in centimeters from the bottom of one's feet to the top of one's head. Though mean values of that distance will grow with development, the *measurement of the inter-individual differences* remains the same. This reflects a kind of differential continuity (opposed to absolute, structural, ipsative, or heterotypic continuity; which concern, respectively, overall level, factor structure, within-person processes, and coherence across multiple behaviors; see Caspi & Roberts, 2001). This could also be because height at age nine has a common cause as height at adulthood, the unfolding of a genetic predisposition. A similar process could be at play in the development of intelligence, *g* at age three correlates with *g* at age five due to a common cause, the unfolding of common genes (e.g. Haworth et al., 2010). We find this common-cause the most likely explanation for the stability of *g*, though the same construct measured twice cannot be discounted.

In the end, however, any explanation of the fadeout effect which posits changing *g* as a cause of the fade (for whatever reason) carries with it the implication that interventions which raise *g* in adulthood would be permanent. If development is the cause of the fading, when that cause is removed, the effects would go as well. While there has not been a systematic investigation of the fadeout effect in adulthood, we find this implication unlikely. In addition, the rate of developmental change is highly nonlinear, yet interventions which start later in a child's life fadeout at the same rate as those which start earlier (Protzko, 2015).

5. Conclusion

An early solution proposed for the fadeout effect was that the raising IQ/raising *g* distinction was at play, but instead while the gains to IQ scores faded, the gains to *g* remained. Specifically: "For instance, those who introduce such an argument never apply it in order to support the view that 'fading' or 'wash-out' effects do not indicate 'real' decrements in intelligence, although that would be quite a reasonable view to hold, especially since in a number of the studies in which fading has taken place children have continued to show other kinds of gains that point to increases in intelligence, such as improved school performance." (Howe, 1997; p.60). In this study we see no such continuation of *g* gains.

Even though we can raise *g*—that does not mean those gains will be permanent. The remaining gains to both IQ and *g* fade. The fadeout effect is unlikely due to a given intervention raising IQ test-specific elements and not general intelligence. Whatever reason for the fadeout effect, the raising IQ/raising *g* distinction is not it.

The other possibility thus far proposed for the fadeout effect concerns where the children end up after an intervention. Most interventions, this one included, use children from underprivileged backgrounds. After the intervention ends, these children, who have just had their IQs raised, return to their impoverished homes. They attend subpar schools and are exposed to more violence (Lee & Loeb, 1995). This is not an effect of the intervention; this is merely a result of who is chosen to intervene on. In this intervention, children were low-birthweight infants from predominantly impoverished homes. That children ended back in their impoverished homes explain the fadeout effect far better than the raising IQ/raising *g* hypothesis. This explanation, however, has been criticized as unable to fully account for the fadeout effect (see Protzko, 2015).

Because of the existence of the fadeout effect at the latent level presented here, one may (mistakenly) believe that interventions

[3] http://www.tylervigen.com/view_correlation?id=1703

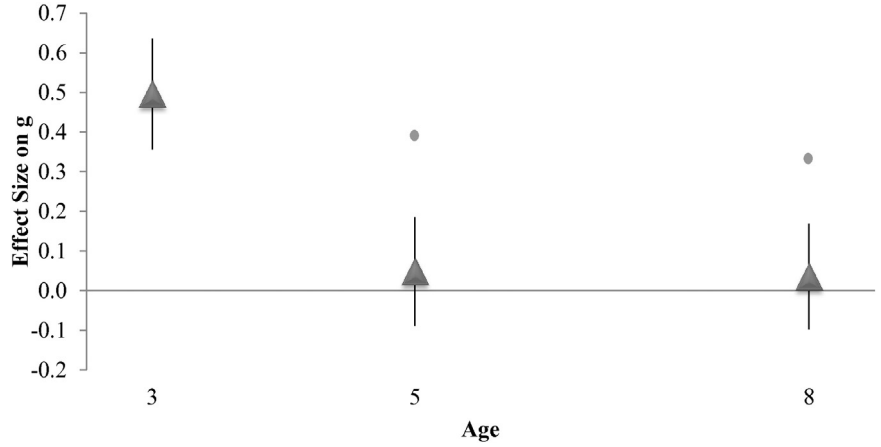

Fig. 2. Effect sizes on *g* at the end of the intervention (age three), two and five years later with 95%CI. Dots above ages five and eight represent effect size estimates based on instability estimates of *g*.

aimed at enhancing intellectual development are useless or doomed to fail. This is not merited by the data here. What has yet to be forwarded, and indeed is a position we do not endorse, is why improvements on *g* should be permanent. We know of no reason to presuppose such permanency other than assumption.

We hope to show here that intelligence at the latent level can be raised through targeted environmental interventions. Raising IQ is not an instance of raising test scores with no concomitant effects on the latent underlying trait. We show this using strict measurement invariance under MGCFA, a technique built for testing such theories which should make its way into more intervention work. Despite these effects on *g*, the salutary effects of an improvement to early environments will still fade. Even when a temporary intervention raises *g*, sustained effects require a sustained intervention (see Protzko, 2015, also Papageorgiou, Christou, Spanoudis & Demetriou, 2016).

Here we see that an intervention can increase even the underlying latent construct *g*, and yet still fade away. Therefore, this paper contributes to the refutation of the argument that the fadeout effect is due to intervention effects not occurring on *g*.

Acknowledgment

The Authors would like to thank Jim Jaccard for his technical and theoretical advice on this project. We would also like to thank Andrew Maul for his theoretical discussions as well. We would like to thank Clancy Blair for access to the data. We would like to thank Drs. Richard Haier, Drew Bailey, and Ricardo Colom for their helpful reviews and criticisms in making this a stronger paper. We would also like to thank Andreas Demetriou for continued discussion on the fadeout effect.

Appendix A

Table 2
Factor loadings of items and thresholds from the strict invariance model.

	Control		Experimental	
Item code	Factor loading	Threshold	Factor loading	Threshold
1	0.438	−0.960	0.403	−0.977
2	.979	−1.298	.975	−1.428
3	.883	−1.457	.863	−1.573
4	.824	−1.479	.796	−1.58
5	.649	−.467	.611	−.486
6	.922	−1.108	.907	−1.205
7	.76	−.112	.727	−.118
8	.884	−.79	.864	−.853
9	.534	−.82	.496	−.842

Table 2 (*continued*)

	Control		Experimental	
Item code	Factor loading	Threshold	Factor loading	Threshold
10	.761	−.303	.727	−.32
11	.394	−.552	.361	−.56
12	.784	−.316	.752	−.335
13	.464	.086	.428	.087
14	.563	−.044	.524	−.045
15	.516	−.009	.478	−.009
16	.197	−.8	.178	−.803
17	.641	.164	.602	.17
18	.659	.249	.621	.259
19	.849	.007	.824	.008
20	.476	.393	.44	.401
21	.728	.182	.693	.191
22	.754	.501	.72	.529
23	.628	.952	.59	.988
24	.384	.42	.352	.426
25	.81	.82	.78	.873
26	.647	.438	.609	.456
27	.633	.021	.594	.022
28	.711	1.357	.675	1.424
29	.735	1.03	.7	1.085
30	.894	1.401	.875	1.516
31	.758	.83	.724	.877
32	.745	1.149	.71	1.212
33	.849	1.072	.824	1.15
34	.851	1.344	.826	1.443

References

Ashton, M. C., & Lee, K. (2005). Problems with the method of correlated vectors. *Intelligence*, *33*(4), 431–444.

Bailey, D. H., Duncan, G., Odgers, C., & Yu, W. (2015). *Persistence and fadeout in the impacts of child and adolescent interventions. Working paper No. 2015–27* (Retrieved from the Life Course Centre working paper series website:http://www.lifecoursecentre.org.au/working-papers/persistence-and-fadeout-in-the-impacts-of-child-and-adolescent-interventions).

Bailey, D. H., Watts, T. W., Littlefield, A. K., & Geary, D. C. (2014). State and trait effects on individual differences in children's mathematical development. *Psychological Science*, *25*(11), 2017–2026.

Bailey, D. H., Watts, T. W., Littlefield, S. K., & Geary, D. C. (2015). State and trait effects on individual differences in children's mathematical development. *Psychological Science*.

Bickley, P. G., Keith, T. Z., & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the life span. *Intelligence*, *20*(3), 309–328.

Blair, C. (2007). Inherent limits on the identification of a neural basis for general intelligence. *Behavioral and Brain Sciences*, *30*(2), 154–155.

Blair, C. (2010). Fluid cognitive abilities and general intelligence. *The handbook of life-span development*.

Borsboom, D., & Mellenbergh, G. J. (2002). True scores, latent variables, and constructs: A comment on Schmidt and Hunter. *Intelligence*, *30*(6), 505–514.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071.

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature–nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review*, *101*(4), 568.

Brooks-Gunn, J., McCarton, C., Casey, P., McCormick, M., Bauer, C., Bernbaum, J., ... Meinert, C. (1994). Early intervention in LBW, premature infants: Results through age 5 years from the infant health and development program. *Journal of the American Medical Association*, *272*(16), 1257–1262.
Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466.
Caspi, A., & Roberts, B. W. (2001). Personality development across the life course: The argument for change and continuity. *Psychological Inquiry*, *12*(2), 49–66.
Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin, 343.
Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504.
Cheung, G., & Rensvold, R. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255.
Dolan, C. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research*, *35*(1), 21–50.
Dunn, L. M., & Dunn, L. M. (1981). *Manual for the peabody picture vocabulary test-revised*. Circle Pines, MN: American Guidance Service.
Elliot, C. D. (1990). *Manual for the differential ability scales. san* Antonio, TX: Psychological Corp.
Esposito, R., Cilli, F., Pieramico, V., Ferretti, A., Macchia, A., Tommasi, M., ... Sensi, S. L. (2013). Acute effects of modafinil on brain resting state networks in young healthy subjects. *PloS One*, *8*, 7.
Estrada, E., Ferrer, E., Abad, F. J., Román, F. J., & Colom, R. (2015). A general factor of intelligence fails to account for changes in tests' scores after cognitive practice: A longitudinal multi-group latent-variable study. *Intelligence*, *50*, 93–99.
Feuerstein, R. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques*. Baltimore: University Park Press.
Feuerstein, R., Rand, Y., Hoffman, M. B., & Miller, R. (1980). *Instrumental enrichment: An intervention program for cognitive modifiability*. Baltimore: University Park Press.
Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79–132.
Gottfredson, L. S. (2009). Chapter 1: Logical Fallacies Used to Dismiss the Evidence on Intelligence Testing. In R. F. Phelps (Ed.), *Correcting Fallacies about Educational and Psychological Testing*. Washington (DC): American Psychological Association, ISBN, 978-1.
Haworth, C. M. A., Wright, M. J., Luciano, M., Martin, N. G., De Geus, E. J. C., Van Beijsterveldt, C. E. M., ... Plomin, R. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry*, *15*(11), 1112–1120.
Herrnstein, R., & Murray, C. (1994). *The bell curve*. New York, NY: Free Press.
Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, *26*, 107–129.
Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*, 117–144.
Howe, M. J. A. (1997). *IQ in question: The truth about IQ*. London: SAGE (1997).
Jensen, A. R. (1997). Adoption data and two g-related hypotheses. *Intelligence*, *25*(1), 1–6.
Jensen, A. R. (1998). *The g factor: The science of mental ability*. London: Praeger.
Jensen, A. R., & Weng, L. J. (1994). What is a good g? *Intelligence*, 231–258.
Johnson, W., & Bouchard, T. J., Jr. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*(4), 393–416.
Johnson, W., Bouchard, T. J., Jr., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one *g*: Consistent results from three test batteries. *Intelligence*, *32*(1), 95–107.
Lee, V. E., & Loeb, S. (1995). Where do head start attendees end up? One reason why preschool effects fade out. *Educational Evaluation and Policy Analysis*, *17*, 62–82.
Lubke, G. H., Dolan, C. V., & Kelderman, H. (2001). Investigating group differences on cognitive tests using Spearman's hypothesis: An evaluation of Jensen's method. *Multivariate Behavioral Research*, *36*(3), 299–324.
McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*, 115–142.
McCarton, C. M., Brooks-Gunn, J., Wallace, I. F., Bauer, C. R., Bennett, F. C., Bernbaum, J. C., & Meinert, C. L. (1997). Results at age 8 of early intervention for low-birth-weight premature infants: The infant health and development program. *JAMA*, *277*(2), 126–132.
McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10.
McNemar, Q. (1942). *The revision of the Stanford–Binet scale: An analysis of the standardization data*. Boston: Houghton Mifflin.
Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, *29*(2), 177–185.
Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, *39*, 479–515.
te Nijenhuis, J., Jongeneel-Grimen, B., & Kirkegaard, E. O. (2014). Are Headstart gains on the *g* factor? A meta-analysis. *Intelligence*, *46*, 209–215.
te Nijenhuis, J., van Vianen, A. E. M., & van der Flier, H. (2007). Score gains on g-loaded tests: No g. *Intelligence*, *35*, 283–300.
Papageorgiou, E., Christou, C., Spanoudis, G., & Demetriou, A. (2016). Augmenting intelligence: Developmental limits to learning-based cognitive change. *Intelligence*, *56*, 16–27.
Protzko, J. (2015). The environment in early intelligence: A Meta-Analysis of the fadeout effect. *Intelligence*.
Ramey, C. T., Bryant, D. M., Wasik, B. H., Sparling, J. J., Fendt, K. H., & LaVange, L. M. (1992). Infant health and development program for LBW, premature infants; program elements, family participation, and child intelligence. *Pediatrics*, *89*, 454–465.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth Cengage learning.
Skuy, M., Gewer, A., Osrin, Y., Khounou, D., Fridjon, P., & Rushton, J. P. (2002). Effects of mediated learning experience on Raven's matrices scores of African and non-African university students in South Africa. *Intelligence*, *30*, 221–232.
Tanner, J. M., Whitehouse, R. H., Marshall, W. A., & Carter, B. S. (1975). Prediction of adult height from height, bone age, and occurrence of menarche, at ages 4 to 16 with allowance for midparent height. *Archives of Disease in Childhood*, *50*(1), 14–26.
Terman, L. M., & Merrill, M. A. (1972). *Stanford-binet intelligence scale: Manual for the third revision-form L-M*. Boston, MA: Houghton Mifflin.
Thorndike, R. L. (1987). Stability of factor loadings. *Personality and Individual Differences*, *8*(4), 585–586.
Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70.
Vernon, P. E. (1954). Symposium on the effects of coaching and practice in intelligence tests. V: Conclusions. *British Journal of Educational Psychology*, *24*, 57–63.
Vernon, P. E. (1960). *Intelligence and attainment tests*. London: University of London Press.
Wechsler, D. (1989). *Wechsler preschool and primary scale of intelligence-revised*. Psychological Corporation.
Wechsler, D. (1991). *Manual for the Wechsler intelligence scale for children—Third edition*. San Antonio, TX: Psychological Corp.
Žebec, M. S., Demetriou, A., & Kotrla-Topić, M. (2015). Changing expressions of general intelligence in development: A 2-wave longitudinal study from 7 to 18 years of age. *Intelligence*, *49*, 94–109.