

Does Reading Cause Later Intelligence? Accounting for Stability in Models of Change

Drew H. Bailey
University of California, Irvine

Andrew K. Littlefield
Texas Tech University

This study reanalyzes data presented by Ritchie, Bates, and Plomin (2015) who used a cross-lagged monozygotic twin differences design to test whether reading ability caused changes in intelligence. The authors used data from a sample of 1,890 monozygotic twin pairs tested on reading ability and intelligence at five occasions between the ages of 7 and 16, regressing twin differences in intelligence on twin differences in prior intelligence and twin differences in prior reading ability. Results from a state–trait model suggest that reported effects of reading ability on later intelligence may be artifacts of previously uncontrolled factors, both environmental in origin and stable during this developmental period, influencing both constructs throughout development. Implications for cognitive developmental theory and methods are discussed.

Understanding the genetic and environmental pathways in the development of children’s cognitive skills has important implications for theories of cognitive development and for educational practice. Studies using longitudinal, genetically informed data sets have led to important progress in our understanding of theories of cognitive development. For example, the heritability of children’s cognitive skills increases with age. This counterintuitive yet robust finding suggests that children’s active selection of environments is an important process during their cognitive development (Scarr & McCartney, 1983; Tucker-Drob & Briley, 2014).

In a recent investigation, Ritchie, Bates, and Plomin (2015; henceforth, RBP) used a cross-lagged monozygotic twin differences design to test whether reading ability influenced changes in intelligence. The authors utilized data from a sample of monozygotic twin pairs tested on reading ability and intelligence at five occasions between the ages of 7 and 16, regressing twin differences in intelligence on twin differences in prior intelligence and twin differences in prior reading ability (see Figure 1). The model also allowed for effects of

twin differences in intelligence on subsequent twin differences in reading ability, statistically adjusting for prior twin differences in reading ability. This method’s unique feature is that, because monozygotic twins are genetically identical, phenotypic differences in reading ability between cotwins must be attributed to environmental experiences unique to each cotwin. This is a major advantage over studies using nongenetically informed data to study development, as correlated changes during development in these samples may be attributable to genetic differences among children causing changes in both intelligence and other factors.

RBP found consistent effects of twin differences in reading ability on later twin differences in intelligence, concluding that learning to read may improve children’s intelligence (Figure 1). If true, these findings have important implications for children’s cognitive development, as improving children’s early reading ability may have lasting effects on their intelligence. More generally, rich, genetically informed data sets with data from several waves per participant may provide answers to many enduring questions from educational and differential psychology and further our theoretical understanding of cognitive development.

Although combining longitudinal and behavior genetic methods may help illuminate the pathways underlying cognitive development, we stress the

The authors thank Stuart Ritchie for helpful comments about this project. Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under Award P01HD065704. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Correspondence concerning this article should be addressed to Drew H. Bailey, School of Education, 3200 Education, University of California, Irvine, Irvine, CA 92697-5500. Electronic mail may be sent to dhbailey@uci.edu.

© 2016 The Authors

Child Development © 2016 Society for Research in Child Development, Inc. All rights reserved. 0009-3920/2016/xxxx-xxxx

DOI: 10.1111/cdev.12669

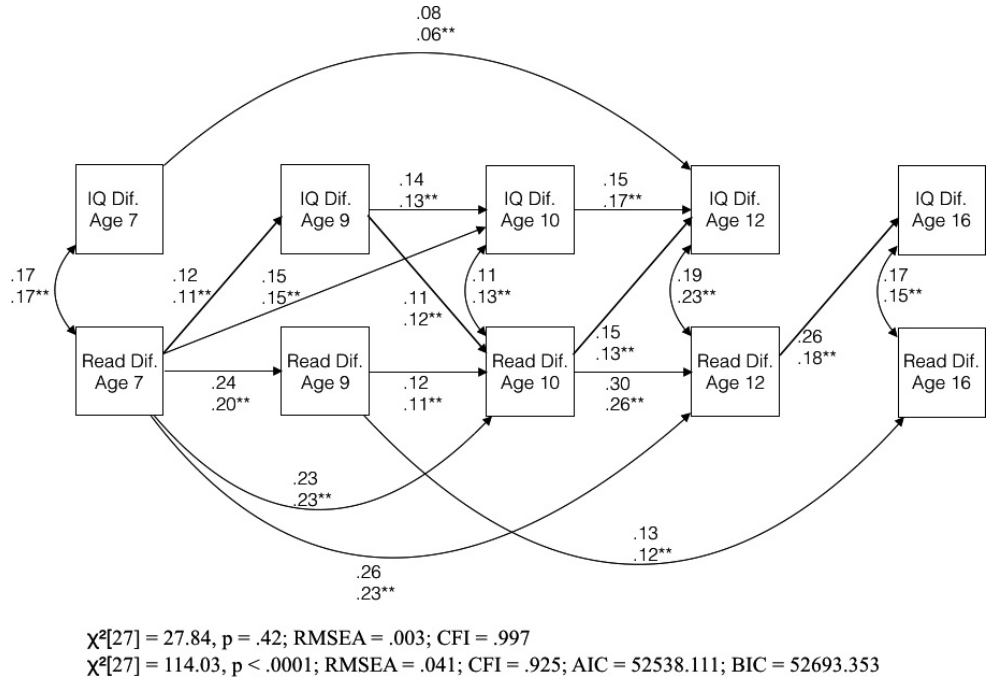


Figure 1. Model of the codevelopment of twin differences in intelligence and reading ability from Ritchie et al. (2015). Note. All top paths are from Ritchie et al. (2015), and are statistically significant ($p < .05$); figure reproduced with permission from Stuart Ritchie. Bottom estimates and fit statistics are from the same specification, using the correlation matrix published in the same article, instead of the raw data. For these paths, * $p < .05$. ** $p < .01$.

importance of testing for the robustness of these findings under different model specifications (Duncan, Engel, Claessens, & Dowsett, 2014; Tomarken & Waller, 2003). Importantly, we also highlight that patterns of differences in results across specifications may inform theories of cognitive development. Below, we discuss the theoretical and statistical assumptions of RBP’s model, and which alternative models might yield equivalent, or even improved, fit to the data.

Is This How Cognitive Development Works?

RBP’s model, displayed in Figure 1, points to a specific process through which IQ and reading ability develop. First, three autoregressive paths are constrained to be 0: the path from IQ difference at age 7 to IQ difference at age 9, the path from IQ difference at age 12 to IQ difference at age 16, and the path from reading ability difference at age 12 to reading ability difference at age 16. If accurate, this model has curious implications for children’s cognitive development. It is difficult to understand why, for example, IQ at age 7 would directly influence children’s IQ at age 12, but IQ at age 7 would not influence IQ at ages 9 or 10. Similarly, this model posits that reading ability at age 16 is caused by

reading ability at age 9 but not by reading ability at ages 10 or 12. In our appraisal, the verisimilitude of such hypothesized processes is suspect. Is it plausible that changing 7 year olds’ intelligence significantly impacts their intelligence at age 12 but that the effect on their intelligence at ages 9 or 10 is nil? Although developmentalists have hypothesized the existence of critical periods in children’s cognitive development (e.g., Almond & Mazumder, 2011; Gale, O’Callaghan, Godfrey, Law, & Martyn, 2004), the theories of critical periods of which we are aware point to earlier time frames than the ages covered in RBP’s study. Furthermore, these theories do not account for latent periods during which individual differences in intelligence can be both well measured but unaffected by events occurring during these periods.

Perhaps this model reveals something about important critical periods during children’s cognitive development. Indeed, the model RBP present fit the data very well (root mean square error of approximation [RMSEA] = .003, comparative fit index [CFI] = .997, Tucker Lewis Index [TLI] = .996). However, we are skeptical of this hypothesized model for two primary reasons. First, this model predicts that early childhood interventions should have nonmonotonically emerging

effects on children's intelligence (e.g., an intervention influencing age 7 intelligence should impact intelligence at age 12 but not ages 9 or 10). Conflicting with this prediction, the treatment effects of early childhood interventions on children's intelligence test scores, such as those documented in the Perry Preschool and Abecedarian Projects, appear to peak immediately following treatment and subsequently decay monotonically (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Schweinhart et al., 2005; though the effect appears to reach an asymptote at a value above zero in the case of the Abecedarian Project). Similarly, the model predicts similar or growing effects of children's age 7 reading achievement on their achievement at ages 9, 10, and 12. However, treatment effects of early interventions aimed at children's academic skills have been shown to decay quickly and monotonically over time (see Bailey, Duncan, Odgers, & Yu, 2016 for review, or Bus & van IJzendoorn, 1999 for review of early reading intervention effects more specifically).

The second reason that we are skeptical of the assumptions about cognitive development made by RBP's model has to do with the processes through which this model was derived. RBP fit a complex cross-lagged panel regression model, in which all possible paths from earlier skills to later skills were freely estimated, and then dropped all of the paths that were not statistically different from 0. Altogether, more than half of the original paths between observed variables were dropped during this process. This approach may result in overfitting the model to the data, and it is unlikely that the same model would fit as well in an independent sample (MacCallum, 1986). Model fit must be interpreted along with information about the process via which the model is generated (Tomarken & Waller, 2003). It is useful that RBP provided such a detailed explanation of their model selection procedure; however, based on the contingency of this procedure on the data, we think that a better index of the cross-lagged panel model's fit to the data would result from a plausible, prespecified model, which would reduce the likelihood of overfitting.

Cross-Lagged Panel Regression and Cognitive Development

Importantly, the findings that RBP chose to highlight were related to the cross-lagged effects of reading ability difference on later IQ difference scores not the developmental process underlying either IQ or reading ability. Furthermore, the cross-

lagged effects of reading ability in RBP's model consistently differed significantly from zero.

However, the validity of causal inferences drawn from the cross-lagged panel regression model depends on important assumptions. Primarily, the autoregressive model must adequately model the stability of the developing phenotypes in the model. In other words, twin differences in intelligence and reading ability must develop primarily in a path-dependent fashion (Hamaker, Kuiper, & Grasman, 2015).

This assumption appears to be significantly violated in the development of children's mathematics achievement (Bailey, Watts, Littlefield, & Geary, 2014). One way of testing the plausibility of a purely state-dependent model of development is to compare the test-retest correlation of a given characteristics across various time lags between measurements. An autoregressive model predicts that the correlations between measurements should decay exponentially as the time between assessments increases. Bailey et al. (2014) found that the correlations in a large, geographically diverse U.S. sample between measures of mathematics achievement decayed from .72 between Grades 1 and 3 to .66 between Grade 1 and age 15. This suggests that individual differences in children's mathematics achievement are largely (though not completely) stable over time and that cross-lagged panel regression models, which adjust for the same construct at the previous time point, do not fully account for all of the factors influencing stability of individual differences over time and likely produce overestimates of the causal pathways of early mathematics achievement on later outcomes of interest.

Are factors producing correlations among twin differences in cognitive skills stable across time? We predict that they are for the most part. Unique environmental effects may occur very early during development and influence cognition across the life span. On the other hand, it is also possible that these show a different pattern than the mathematics achievement data described above. The state-trait study of mathematics achievement did not separate out genetic and environmental factors, and there is evidence for much stronger stability of genetic influences on cognition across development than unique environmental stability (Tucker-Drob & Briley, 2014). Still, unique environmental effects appear to have stability greater than zero across development, and their test-retest stability does not appear to decay as the distance between measurement occasions increases (Tucker-Drob & Briley, 2014, p. 968), consistent with the hypothesis that most of

the (small) longitudinal relations between identical twin differences in intelligence are attributable to stable factors, compared with state-dependent factors, during childhood.

The correlations among the twin difference scores (RBP, table S4) are informative. The correlations between reading ability difference at age 7 and reading ability difference at ages 9, 10, 12, and 16 were .20, .26, .31, and .08, respectively. The correlations between intelligence difference at age 7 and intelligence difference at ages 9, 10, 12, and 16 were .04, .03, .08, and .08, respectively. These correlations show a great deal of stability over time; that is, to the extent that twin differences persist over time, they appear to reflect trait-like differences, rather than a state-dependent, autoregressive process. If this is true, then the cross-lagged paths reported by RBP are likely overestimates of the effects of occasion-specific reading ability on later intelligence, because they fail to account for the stability of reading ability and intelligence across time (Hamaker et al., 2015).

Current Study

The goals of the current study were (a) to present a plausible model of the development of individual differences in intelligence and reading ability during middle childhood and adolescence, and (b) to re-estimate the effects of reading ability on later intelligence and intelligence on later reading ability using this model.

Following Bailey et al. (2014) and recommendations by Hamaker et al. (2015), we used a state–trait modeling approach (Steyer, 1987) to partition the variance in monozygotic twin differences in reading ability and intelligence throughout development into trait effects (i.e., effects of factors that influence individual differences in cognition similarly across development) and state effects (i.e., effects of individual differences in cognition on subsequent individual differences in cognition). Thus, levels of reading ability and intelligence at a given time point were modeled as a function of (a) a trait-like factor, (b) the *same* cognitive ability at the immediately preceding measurement occasion (the state effect), (c) the *other* cognitive ability at the immediately preceding measurement occasion (the cross-lagged effect), and (d) unique sources of variation (e.g., measurement error; Jackson, Sher, & Wood, 2000). This model is similar to the random intercept cross-lagged panel model (RI-CLPM) described by Hamaker et al. (2015), though the factors are interpreted as latent variables or traits within a state–

trait model rather than as random intercepts (as in multilevel modeling; see Hamaker et al., 2015, p. 105). We compared the fit of this model to the fit of a standard cross-lagged panel regression model and investigated the robustness of conclusions based on the latter model when between-subject stable factors were adjusted.

Method

We used the correlation matrix of IQ and reading difference scores reported in RBP's online supporting information (RBP, table S4; for more detailed information on the sample and measures, see RBP) for the following analyses. The correlation matrix was calculated from data on 1890 monozygotic twin pairs from the Twins Early Development Study (Haworth, Davis, & Plomin, 2013). Twins in the sample were born between January 1994 and December 1996. Participants were tested on measures of reading ability and intelligence at five waves (ages 7, 9, 10, 12, and 16). Reading ability and intelligence were measured with between 1 and 6 tests per construct per wave. For waves at which more than a single test was used to measure a construct (all construct–wave combinations except for age 9 reading ability), constructs were estimated as the summed *z* scores on each test corrected for age and sex, and standardized. The correlation matrix consists of correlations among twin differences on these constructs across the five waves.

We estimated three models: the specification used by RBP, a simpler cross-lagged panel model, in which reading and IQ difference scores are influenced only by reading and IQ difference scores at the previous measurement occasion and unique sources of variation and a state–trait model, in which reading and IQ difference scores are influenced by latent reading and intelligence traits, respectively, reading and intelligence difference scores at the immediately preceding measurement occasion, and unique sources of variation. Models were estimated in Mplus version 7 (Muthén & Muthén, 1998–2012).

Other longitudinal models, such as the RI-CLPM (Hamaker et al., 2015) and the STARTS model (Kenny & Zautra, 1995), were not considered, because accurate parameter estimates and model fit statistics depend on the variance–covariance matrix (rather than the correlation matrix alone). Due to large changes in measurement methods across waves, we were not confident in our abilities to produce comparable variance estimates across

waves. However, both the cross-lagged panel model and state-trait model produce equivalent results whether a variance-covariance matrix or correlation matrix is used as input (Mplus demonstration of this is available from authors by request).

Results

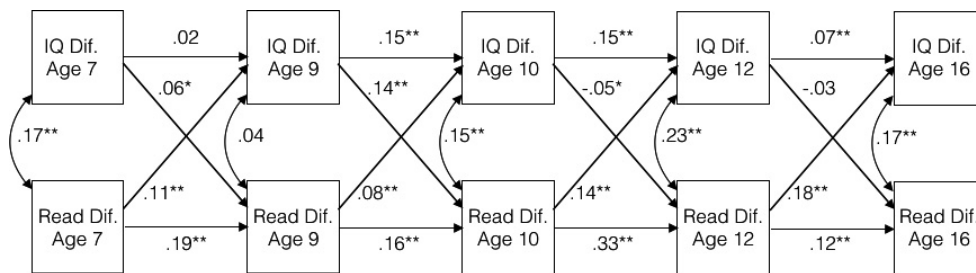
RBP's model appears in Figure 1. Estimates were very similar to those RBP obtained using the raw data. All of the paths retained their statistical significance and did not differ much from those reported by RBP. The model fit remained acceptable but was notably worse than the model fit obtained from the raw data (RMSEA = .041 vs. .003).

The cross-lagged panel model appears in Figure 2. The pattern of results is similar to those in RBP's final model (Figure 1): Most, but not all, autoregressive paths are statistically significant, all four of the cross-lagged paths from reading ability to intelligence are statistically significant, whereas only two paths from intelligence to reading ability are statistically significant and positive (one is statistically significant and negative). The key difference between this model and the model presented by RBP is that its fit is much worse (RMSEA = .081; CFI = .743). This difference in model fit is attributable to some combination of including plausible but nonsignificant paths and from not including implausible paths that would result in the prediction of discontinuous effects during the development of intelligence and reading achievement, explained above. This finding casts doubt on the hypothesis that the cross-lagged panel model accurately reflects the developmental process of

environmental contributions to intelligence and reading ability.

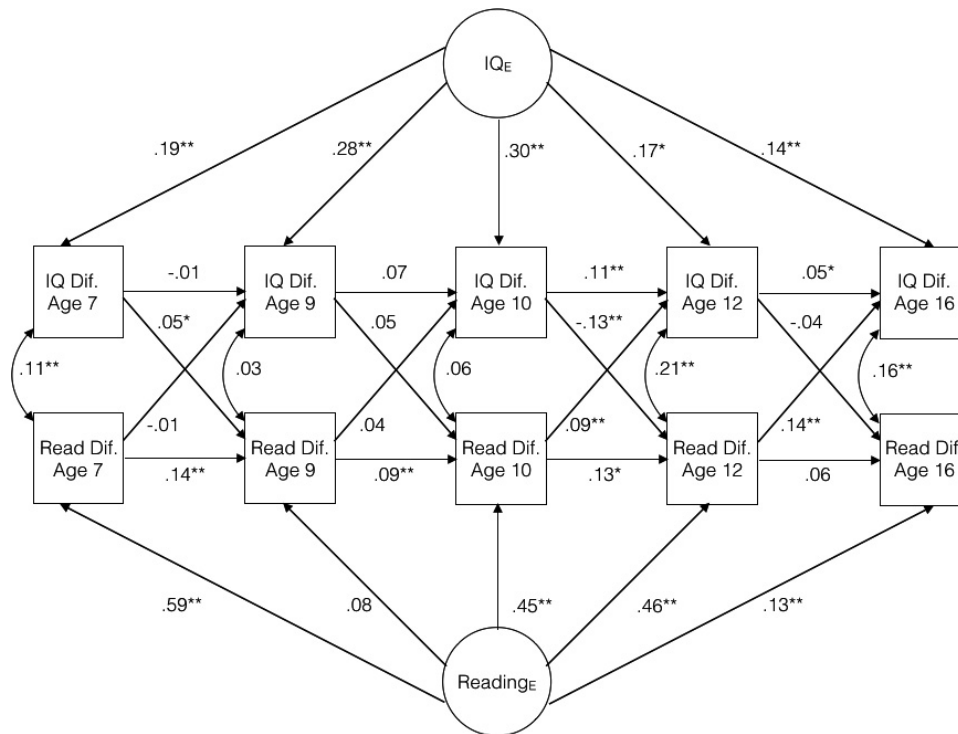
The state-trait model appears in Figure 3. We imposed a single constraint on this model, namely that the loadings of the reading factor on twin reading differences at ages 10 and 12 were constrained to be equal. We did this because, when the model was estimated without this constraint, the loading of reading difference at age 12 was greater than 1. We suspect this very large trait loading resulted from the particularly good measurement of reading ability at age 12 (this was the only construct-wave combination during which six total assessments were used). The model fit the data very well (RMSEA = .043; CFI = .960). In contrast to the cross-lagged panel model, only five of the eight autoregressive paths were positive and statistically significant, and these were generally smaller than the paths from the trait factors (9 of the 10 of which were statistically significant).

Of the eight cross-lagged paths in the state-trait model, three were positive and statistically significant. Both of these paths involved the age 12 reading ability measure, the trait loading for which was constrained in this model because it was greater than 1 in the unconstrained model. The positive cross-lagged path from age 12 reading ability to age 16 IQ may partially reflect the incomplete statistical control of reading ability score at age 12 for trait effects. Furthermore, the negative path from IQ at age 10 to reading ability at age 12 is difficult to interpret but may partially compensate for a larger trait effect on reading ability at age 10 and subsequently larger autoregressive path from age 10 to age 12 reading ability than would otherwise be observed, due to the constrained equality of the trait effect on reading ability at ages 10 and 12. Regardless, the evidence for a causal effect of



$$\chi^2[24] = 321.83, p < .0001; \text{RMSEA} = .081; \text{CFI} = .743; \text{AIC} = 52751.902; \text{BIC} = 52923.776$$

Figure 2. Cross-lagged panel regression model of the codevelopment of twin differences in intelligence and reading ability. Note. Coefficients are standardized regression weights; * $p < .05$. ** $p < .01$.



$\Phi = .76^{**}$; $\chi^2[14] = 62.09$, $p < .0001$; RMSEA = .043; CFI = .960; AIC = 52512.170; BIC = 52739.488

Figure 3. State-trait model of the codevelopment of twin differences in intelligence and reading ability.

Note. Coefficients are standardized regression weights; * $p < .05$. ** $p < .01$. Φ is the correlation between the two latent variables.

reading ability on intelligence, especially early in development, is reduced in the better fitting state-trait model, relative to the cross-lagged panel model. Notably, the two remaining statistically significant paths from reading ability to IQ occur between the final two waves, perhaps indicating increasing effects of specific abilities on intelligence scores in adolescence than in earlier childhood.

The correlation between the trait factors, .76, was notably high, suggesting that the relatively stable, environmentally caused factors differing between individuals contributing to the development of reading ability and intelligence during this developmental period are substantially overlapping.

The residual covariances between intelligence and reading ability were positive and statistically significant at three of the five waves, indicating that these factors changed together. These are consistent with the existence of an environmentally caused relation between within-individual changes in reading ability and intelligence. This could be a causal pathway running from reading ability to intelligence, as posited by RBP, a pathway running in the opposite direction, or an influence of within-individual changes on some other variable (e.g.,

personality or mood) that influences test performance and possibly construct level variation in reading ability and intelligence.

Discussion

The current study compared the fit and inferences resulting from models of the codevelopment of monozygotic twin differences in reading ability and intelligence using a cross-lagged panel regression model and a state-trait model. Replicating the findings of RBP's previous analysis using the cross-lagged panel regression model on the same data set, we found that previous reading ability predicted later IQ consistently across development. However, we found that the model exhibited poor fit to the data.

The state-trait model fit the data very well but showed less evidence for effects of individual differences in children's environmentally caused reading ability on their later intelligence. Longitudinal relations between monozygotic twin difference scores in reading ability and intelligence, both within and between constructs, were largely

explained by highly correlated factors influencing these difference scores throughout the observed time period.

Implications

Results are consistent with very high, previously reported relations between constructs underlying performance on cognitive and academic assessments (Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012) but extend these findings to measures that are independent of variance caused by genetic differences. Why might some unique environmental factors contribute similarly to children's cognitive development throughout this developmental period, despite the fact that children's environments change very much throughout this period? Developmental theories often posit lasting effects brought on by early developmental experiences (for review, see Fraley, Roisman, & Haltigan, 2013). Empirical examples are numerous, including the masculinized psychosocial characteristics of girls and adult women with congenital adrenal hyperplasia, a condition characterized by high levels of exposure to prenatal and early postnatal androgens (Berenbaum & Hines, 1992; Zucker et al., 1996). More directly relevant to the current study, critical periods during which children's brain development can be environmentally influenced have been supported with strong quasi-experimental designs (e.g., Almond & Mazumder, 2011).

The large correlation of trait-like, environmentally caused reading ability and intelligence factors may appear to some to contradict the *generalist genes hypothesis* (Kovas & Plomin, 2007; Plomin & Kovas, 2005), which posits that variance common to a wide range of cognitive skills is largely attributable to variation in the same genes, whereas variance unique to specific cognitive skills is caused by specialist environments, unique to particular skills. However, we see our results as consistent with the generalist genes hypothesis. That is, because of the low correlations between twin difference scores the majority of test score variance across time and constructs caused by unique environmental factors are indeed due to separate environmental effects on each test. Our results modify the generalist genes hypothesis, however, by providing the qualification that the small correlations observed across occasions attributable to nonshared environmental factors are largely due to stable nonshared environmental influences, perhaps during children's very early development. In other words, longitudinal associations between nonshared environmental

influences on cognition appear to be *small* but *stable* as the time between measurement occasions increases and across at least two constructs (intelligence and reading).

We encourage future work on the extent to which such factors are identifiable or might be targeted by intervention. Although improving reading ability during the specific time we observed is not a promising candidate for boosting intelligence, perhaps early language and literacy skills compose some of the stable variance in the unique environmental trait variance observed in the current study (but see Paris, 2005, for an alternative view). We note that results only apply to the range of variation in reading ability and intelligence observed in the current study, and we firmly recommend against educators lessening their frequency or intensity of reading instruction based on these findings.

Limitations

We think it would be premature to rule out the possibility of transactional effects in the codevelopment of reading ability and intelligence during the observed developmental range. One reason for this is that monozygotic twin difference scores may be imprecisely estimated to a degree that we could not reliably detect clinically relevant cross-lagged panel effects in these data. Difference scores sometimes contain much more measurement error than raw scores (Rogosa & Willett, 1983), a fact consistent with the sometimes low loadings of the twin difference scores on their respective trait factors (Figure 3). Notably, the construct-wave combination from which we found the largest positive cross-lagged effect (age 12 reading ability; Figure 3) was the construct-wave combination with the best measurement (though see Results for a possible alternative explanation). We encourage the development of methods to account for the measurement error in these scores, but did not use them here, to make our analysis comparable to that presented by RBP and because of the unclear impact of commonly used adjustments on bias of estimated unique environmental effects (cf. Tucker-Drob & Briley, 2014).

Another reason we should not rule out transactional effects in the codevelopment of reading ability and intelligence during the observed developmental range is that they might occur along genetic or shared environmental pathways; these pathways account for a substantial amount of the variance in children's cognitive ability (Tucker-Drob & Briley, 2014) but are ruled out as sources of variation when cognitive skills are operationalized as

monozygotic twin differences. We encourage work that pursues these possibilities further, though we provide two methodological recommendations below that we hope researchers will consider.

Conclusion

In conclusion, we agree with Tucker-Drob and Briley's (2014) assertion that:

future research efforts to collect multivariate longitudinal genetically informative data on cognition, personality, and environments, combined with sophisticated genetic and longitudinal statistical approaches will be valuable for testing and delineating specific transactional processes of gene-environment correlation and their implications for the stability of individual differences in cognition over time and for cognitive development more generally. (p. 972)

However, we stress two methodological points that will improve progress toward this goal. First, we recommend that researchers use limited, predefined specification searches (e.g., those outlined by Wood, Steinley, & Jackson, 2015) to avoid overfitting. An alternative approach would involve cross-validating a model attained via an exploratory specification search in another sample or a subset of the sample one used to estimate the exploratory model (Byrne, 1994). Second, we recommend that authors consider models that account for stable between-subject differences across development. Although explaining the origins of these preexisting differences is also an important goal, failing to account for them in developmental models when they exist will hinder progress in this area of study.

References

- Almond, D., & Mazumder, B. (2011). Health capital and the prenatal environment: The effect of Ramadan observance during pregnancy. *American Economic Journal: Applied Economics*, 3, 56–85. doi:10.1257/app.3.4.56
- Bailey, D. H., Duncan, G., Odgers, C., & Yu, W. (2016). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*. doi:10.1080/19345747.2016.1232459
- 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, 2017–2026. doi:10.1177/0956797614547539
- Berenbaum, S. A., & Hines, M. (1992). Early androgens are related to childhood sex-typed toy preferences. *Psychological Science*, 3, 203–206. doi:10.1111/j.1467-9280.1992.tb00028.x
- Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, 91, 403–414. http://dx.doi.org/10.1037/0022-0663.91.3.403
- Byrne, B. M. (1994). Testing for the factorial validity, replication, and invariance of a measuring instrument: A paradigmatic application based on the Maslach Burnout Inventory. *Multivariate Behavioral Research*, 29, 289–311. http://dx.doi.org/10.1207/s15327906mbr2903_5
- Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology*, 37, 231–242. http://dx.doi.org/10.1037/0012-1649.37.2.231
- Duncan, G. J., Engel, M., Claessens, A., & Dowsett, C. J. (2014). Replication and robustness in developmental research. *Developmental Psychology*, 50, 2417–2425. doi:10.1037/a0037996
- Fraley, R. C., Roisman, G. I., & Haltigan, J. D. (2013). The legacy of early experiences in development: Formalizing alternative models of how early experiences are carried forward over time. *Developmental Psychology*, 49, 109–126. doi:10.1037/a0027852
- Gale, C. R., O'Callaghan, F. J., Godfrey, K. M., Law, C. M., & Martyn, C. N. (2004). Critical periods of brain growth and cognitive function in children. *Brain*, 127(2), 321–329.
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20, 102–116. doi:10.1037/a0038889
- Haworth, C. M. A., Davis, O. S. P., & Plomin, R. (2013). Twins Early Development Study (TEDS): A genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Research and Human Genetics*, 16, 117–125. doi:10.1017/thg.2012.91
- Jackson, K. M., Sher, K. J., & Wood, P. K. (2000). Prospective analysis of comorbidity: Tobacco and alcohol use disorders. *Journal of Abnormal Psychology*, 109, 679–694. doi:10.1037//0021-843X.109.4.679
- Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive g and academic achievement g one and the same g? An exploration on the Woodcock-Johnson and Kaufman tests. *Intelligence*, 40, 123–138. http://dx.doi.org/10.1016/j.intell.2012.01.009
- Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology*, 63, 52–59.
- Kovas, Y., & Plomin, R. (2007). Learning abilities and disabilities generalist genes, specialist environments. *Current Directions in Psychological Science*, 16, 284–288. doi:10.1111/j.1467-8721.2007.00521.x

- MacCallum, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*, 107–120. <http://dx.doi.org/10.1037/0033-2909.100.1.107>
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.
- Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, *40*, 184–202. doi:10.1598/RRQ.40.2.3
- Plomin, R., & Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological Bulletin*, *131*, 592–617. doi:10.1037/0033-2909.131.4.592
- Ritchie, S. J., Bates, T. C., & Plomin, R. (2015). Does learning to read improve intelligence? A longitudinal multivariate analysis in identical twins from age 7 to 16. *Child Development*, *86*, 23–36. doi:10.1111/cdev.12272
- Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement*, *20*, 335–343. doi:10.1111/j.1745-3984.1983.tb00211.x
- Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype → environment effects. *Child Development*, *54*, 424–435. doi:10.2307/1129703
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry Preschool study through age 40*. Ypsilanti, MI: High/Scope Press.
- Steyer, R. (1987). Konsistenz und Spezifität: Definition zweier zentraler Begriffe der Differentiellen Psychologie und ein einfaches Modell zu ihrer Identifikation [Consistency and specificity: Definition of two central concepts of differential psychology and a simple model for their identification]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *8*, 245–258.
- Tomarken, A. J., & Waller, N. G. (2003). Potential problems with “well fitting” models. *Journal of Abnormal Psychology*, *112*, 578–598. doi:10.1037/0021-843X.112.4.578
- Tucker-Drob, E. M., & Briley, D. A. (2014). Continuity of genetic and environmental influences on cognition across the life span: A meta-analysis of longitudinal twin and adoption studies. *Psychological Bulletin*, *140*, 949–979. doi:10.1037/a0035893
- Wood, P. K., Steinley, D., & Jackson, K. M. (2015). Right-sizing statistical models for longitudinal data. *Psychological Methods*, *20*, 470–488. doi:10.1037/met0000037
- Zucker, K. J., Bradley, S. J., Oliver, G., Blake, J., Fleming, S., & Hood, J. (1996). Psychosexual development of women with congenital adrenal hyperplasia. *Hormones and Behavior*, *30*, 300–318. doi:10.1006/hbeh.1996.0038