Published in final edited form as:

Intelligence. 2010 ; 38(4): 385–392. doi:10.1016/j.intell.2010.05.002.

More than just IQ: A longitudinal examination of self-perceived abilities as predictors of academic performance in a large sample of UK twins

Tomas Chamorro-Premuzic^{a,*}, Nicole Harlaar^b, Corina U. Greven^c, and Robert Plomin^c

^aGoldsmiths, University of London, Department of Psychology, New Cross, SE14 6NW, United Kingdom

^bMind Research Network, 1101 Yale Blvd NE, Albuquerque, NM 87106, USA

^cKing's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, De Crespigny Park, London, SE5 8AF, United Kingdom

Abstract

This paper examines the longitudinal causal relationship between self-perceived abilities (SPA) and academic achievement (Ach) while controlling for cognitive ability (CA). In all, 5957 UK school children were assessed on SPA, Ach and CA at ages 9 and 12. Results indicated that SPA and Ach at age 9 independently affected both SPA and Ach at age 12, even when CA was considered. Moreover the effects of previous Ach on subsequent SPA were of similar magnitude to the effects of prior SPA on subsequent Ach, suggesting that the link between SPA and Ach independent of CA is reflective of both “insight” (children’s accounts of their previous performance) and self-efficacy (the self-fulfilling or motivational effects of self-beliefs). Practical and theoretical implications for the study of SPA are discussed.

Keywords

Self-perceived abilities; Academic achievement; Intelligence; Self-assessed intelligence; Motivation; Intellectual competence

In recent years there has been a marked interest in self-perceived abilities (SPA)¹. Unlike traditional ability tests, which measure individual differences in the ability to perform a variety of standardized reasoning problems, individual differences in SPA are assessed via self-report inventories. Typically, these ask participants to rate their different abilities (e.g., mathematical, verbal, and spatial) – how good or able they *think* they are in these domains – on a given scale and in comparison to a reference group (e.g., “most people”, “your

*Corresponding author: Department of Psychology, Goldsmiths, University of London, New Cross, London SE14 6NW, United Kingdom. Tel.: +44 207 919 7885; fax: +44 207 919 7873. pss02tc@gold.ac.uk (T. Chamorro-Premuzic).

¹Given the number of interchangeable labels for self-perceived abilities (e.g., academic self-confidence/competence/competencies/esteem/concept/efficacy/belief/beliefs, self-estimated/perceived/assessed/perceptions of/views of/evaluations of/estimates of intelligence/ability/abilities, etc) we will use “SPA” throughout this paper. Empirical and conceptual differences between these constructs are discussed in detail by Marsh (2007).

classmates”, “people your age”, etc.). Thus SPA provide a subjective account of one’s ability to learn, reason, and solve problems within different domains of intelligence.

Differential psychologists have attempted to validate SPA in relation to cognitive ability (CA) tests (Alicke, 1985; Chamorro-Premuzic & Furnham, 2005; Cogan et al., 1915; Furnham et al, 1999; Paulhus et al, 1998), but they have rarely looked at the effects of SPA on academic achievement (Ach) (Chamorro-Premuzic, 2007). Correlations between SPA and CA tests typically fall in the range of .2 to .5 and have been used to both support and reject the hypothesis that people are able to assess their own intelligence quite accurately; however, the validity of SPA – like that of CA – might be better judged against Ach (Chamorro-Premuzic & Furnham, 2006). Accordingly, educational, social and developmental psychologists have validated SPA using Ach as criterion (Bandura, 1982, 1997; Denissen, Zarrett, & Eccles 2007; Eccles et al, 1998; Marsh Trautwein, Ludtke, Koller, & Baumert, 2005), yet they seldom included measures of CA, which is the most important determinant of Ach (Chamorro-Premuzic & Furnham, 2005; Deary, Whiteman, Starr, Whalley, & Fox, 2004; Gottfredson, 2002; Gustafsson & Undheim, 1996).

Given the overlap between SPA and CA (Ackerman & Wolman, 2007) on one hand, and CA and Ach (Chamorro-Premuzic, 2007; Chamorro-Premuzic & Furnham, 2005; Deary, Whiteman, Starr, Whalley, & Fox, 2004; Gottfredson, 2002; Gustafsson & Undheim, 1996) on the other, it is important to show that any effects between SPA and Ach are independent of CA. In one of the few large-scale studies (N = 1678) to address this question, Spinath, Spinath, Harlaar, and Plomin (2006) found that although *g* was the strongest predictor of Ach ($R^2 = .25$), a substantial percentage of variance in Ach could be explained by the common variance between SPA and *g*, and that SPA had incremental validity (over *g*) in the prediction of Ach ($R^2 = .05$). However, the single wave nature of their study made it difficult to infer any causal links underlying the data (Baumeister, Campbell, Krueger, & Vohs, 2003). Spinath et al therefore concluded that “the role of [SPA] for school Ach cannot be fully understood without developmental studies focussing on the overlap of [SPA] and cognitive abilities” (p. 371).

Previous research suggests that SPA and Ach are not only correlated but also reciprocally and causally inter-related (Bandura, 1997; Eccles et al, 1998; Marsh et al, 2005), overcoming the dichotomous interpretations of SPA as either a *consequence* (skill-development model; see Calsyn & Kenny, 1977) or *cause* (self-enhancement model; see Byrne, 1984) of Ach. Although these bidirectional effects have been documented empirically via longitudinal designs (Marsh et al, 1999; Wigfield & Karpachian, 1991), few studies (e.g., Gose et al, 1980; Marsh, 1990a; Schicke & Fagan, 1994) controlled for CA data. One of the few exceptions is a study of 1456 pupils from Canada, Germany and Hong Kong (Marsh, 1990a), which found that SPA were influenced by prior CA (as measured through 4 standardized tests) but not prior Ach. Thus studies that control for previous Ach but not CA may explain cognitive effects (insight) in terms of conative effects (motivation), assuming that self-enhancement leads to higher Ach when it is in fact caused by CA.

To overcome the above methodological limitations, the present study set out to explore the longitudinal link between SPA and Ach while controlling for previous CA. Data on SPA,

CA and Ach were available at two points in time (ages 9 and 12), which enabled us to examine whether any potential effects of SPA on Ach and Ach on SPA were independent not only of previous SPA and Ach, but also of CA.

Methods

Sample

The sample consisted of 3220 girls and 2737 boys who were ascertained from UK population records of twin births as part of the Twins Early Development Study (see Oliver & Plomin, 2007). Participants were included in the current study if one or both twins in each pair had data on at least one measure (Ach, SPA, CA) available at both ages 9 and 12. Within this sample, 64.9% had data on at least four (out of six) measures, and 88% had data on at least three measures. Children with severe neurological or genetic conditions were excluded. Informed parental consent was obtained at both waves of assessment.

Measures and procedure

Ach—At ages 9 and 12, teachers rated children's Ach based on key stage 2 and key stage 3 National Curriculum criteria, respectively (The National Curriculum, 2004; also see Kovas, Haworth, Dale, & Plomin, 2007). The National Curriculum stipulates learning goals for children from ages 5 to 16 years at state-funded schools in England and Wales. The National Curriculum teacher ratings assess children according to these learning goals, and are based on the teacher's knowledge of the child's achievement over the academic year. Here we focused on the following skill areas that were rated by teachers at key stages 2 and 3: Using and Applying Mathematics; Numbers; Shapes, Space and Measures (Ach in Mathematics); Speaking and Listening; Reading; Writing (Ach in English); Scientific Enquiry; Life Processes and Living Things; Physical Processes (Ach in Science). Teachers rated each area separately. At age 9, assessments were made on five-point scale. The National Curriculum scale at age 12 is an upward extension of the 9-year criteria. Ratings were made on a nine-point scale. Each point on the scales refers to a National Curriculum level. The Qualifications and Curriculum Authority provides teachers with level descriptions, which specify the types and range of performance that children at each level should demonstrate. At both ages, scores on the three skill areas within each subject were highly inter-correlated (average correlation = .81). We therefore created Mathematics, English, and Science composite scores by summing standardized scores for the three skill areas within their respective subjects. The mean age at the time of National Curriculum assessment was 9.04 years (SD: .29) at the 9-year assessment and 12.08 years (SD: .29) at the 12-year assessment.

There is evidence for substantial agreement between teacher National Curriculum ratings and scores on group-administered National Curriculum reading tests. For example, cross-tabulation analyses on a nationwide sample of 600,000 children have shown that agreement between National Curriculum teacher assessments of reading and scores on group-administered National Curriculum reading tests at Key Stage 1 is good (Cohen's kappa = .80; Dale, Harlaar, & Plomin, 2005). Similarly high levels of agreement have been reported for Key Stage 2 teacher and test assessments (Reeves, Boyle, & Christie, 2001).

SPA—Children's SPA were assessed at ages 9 and 12 using the Perceived Ability in School Scale (Spinath, Spinath, Harlaar & Plomin, 2006). The scale contains nine items that ask children how good they think they are at successfully completing activities listed in the UK National Curriculum. There were three items for English, three for Mathematics, and three for Science (see the Appendix for the items). Responses were made on a 5-point Likert Scale, with '1' indicating that the respondent believed that they were "not at all good" at the activity in question, and '5' indicating that they believed they were "very good" at that activity. Internal consistency of each scale is acceptable or good (at age 9: Cronbach's $\alpha = .62$ for English; .82 for Mathematics; .65 for Science; at age 12: Cronbach's $\alpha = .71$ for English; .87 for Mathematics; .75 for Science). At both ages, confirmatory factor analysis of the nine items yielded a clear three-factor solution. Consequently, we created separate English, Mathematics, and Science Self-perceived Ability ratings at each age by summing the three items pertaining to each scale. Mean age at time of assessment of SPA was 9.01 years (SD: .28) at the 9-year assessment and 11.77 years (SD: .32) at the 12-year assessment.

CA—A practical problem in conducting research on CA is that it is expensive to test large samples in-person, especially when the participants are widely distributed geographically, as in the Twins Early Development Study sample. In an attempt to address this problem we adapted standard cognitive tests for completion by participants in their own home. Validity and reliability information is described in full in Haworth et al. (2007).

At age 9, participants received a test booklet containing four CA tests that they completed under the supervision of the parent (guided by an instruction booklet). We used the Vocabulary and General Knowledge Multiple-choice tests from the Wechsler Intelligence Scale for Children, third edition (WISC-III-PI; Kaplan, Fein, Kramer, Delis, & Morris, 1999), and the Figure Classification and Figure Analogies subtests of the Cognitive Abilities Test 3 (CAT3; Smith, Fernandes, & Strand, 2001). A principal component factor analysis of the four measures indicated a clear one-factor solution, with the principal component explaining 53.9% of the variance. We used factor scores derived from this analysis as our measure of 9-year general CA.

At age 12, children completed an internet battery of cognitive tests comprising the Vocabulary and General Knowledge tests from the WISC-III-PI (Kaplan et al., 1999), the WISC-III-UK Picture Completion (Wechsler, 1992) and Raven's Standard Progressive Matrices (Raven, Court, & Raven, 1996). The instructions and stimuli for our web-based adaptations of these tests were identical to the original tests. However, we incorporated adaptive branching rules in each of the tests to enable us to test the full range of ability, while requiring individual children to complete only a relatively small number of items to ascertain their level of performance efficiently. The branching algorithm for each test was developed and revised on 1000 Twins Early Development Study families prior to testing. A principal component factor analysis of the four measures indicated a clear one-factor solution, with the principal component explaining 44.7% of the variance. We used factor scores derived from this analysis as our measure of 12-year general CA.

Analyses

In the first stage of analysis, we sought to identify a measurement model that fitted the data satisfactorily. A four-factor (SPA and Ach at age 9, SPA and Ach at age 12) confirmatory factor analysis model was fitted to the data. We allowed all factors to be inter-correlated. We also allowed errors of the same variable across time to correlate (e.g., SPA in Math at age 9 with SPA in Math at age 12). One loading for each latent factor was fixed to 1 to set its scale.

In the second stage of analysis, we tested the structural relations among the factors, based on the best-fitting confirmatory factor analysis model. Our primary model of the structural relations between SPA and Ach is illustrated in Figure 1. Latent variables of SPA and Ach are symbolized by circles and are labelled SPA₁ and SPA₂ (for SPA at ages 9 and 12, respectively), and Ach₁ and Ach₂ (Ach at ages 9 and 12, respectively). In this model, the developmental stability of Ach and SPA and Ach is represented by b_{11} and b_{22} . The relationship between SPA at time 1 and Ach at time 2 (b_{12}) versus the relationship of Ach of time 1 and SPA at time 2 (b_{21}) are the critical cross-lagged coefficients. If SPA is seminal, then b_{12} should exceed b_{21} . In contrast, b_{21} should be greater than b_{12} if Ach is a precursor to SPA. No difference between cross-lagged coefficients is indicative of no preponderance of causal effect of one variable over the other (there could still be causal relations occurring, but neither variable is “more responsible” for the causal relations among the variables)². It is noteworthy that the cross-lagged paths (b_{12} and b_{21}) were independent of effects of SPA and Ach at age 9. In this way, we accounted for the possibility that the correlation between SPA and Ach at age 9 could have led to indirect effects through this correlation, i.e. we ensured that cross-lagged paths were not driven by pre-existing associations between SPA and Ach at age 9. The square of the standardized path coefficients from SPA₁ and Ach₁ provides an estimate of the percent variance accounted for by the *direct* effect of the latent variables². Thus, our path estimates were somewhat conservative, and the total effects of the SPA and Ach factors are likely to be greater than the square of the standardized path coefficients.

Scores were adjusted for CA at ages 9 and 12 by regressing CA on the SPA and Ach factors at both ages, thus controlling for any variance explained by CA and creating residual latent factor scores that were independent of CA. The analyses accounted for the fact that CA at age 9 and CA at age 12 were correlated, although we omitted the correlations from the Figures for the purpose of clarity.

In a third stage of analysis, with the aim to obtain additional insights into the role of CA in the relations between SPA and Ach, we then repeated the second stage of analysis, but without controlling for CA. As CA has been shown to be correlated with both Ach and SPA, we expected the cross-lagged path coefficients (b_{12} , b_{21}) to be larger when CA was not accounted for. In addition, as we expected CA to be more strongly associated with Ach than with SPA, we predicted that, after controlling for CA, the cross-lagged path from Ach to SPA (b_{21}) would be reduced to a greater extent than the path from SPA to Ach (b_{12}). As this

²We are grateful for the input of an anonymous reviewer on these points.

was not a focus of this paper, we only report relevant coefficients from the analyses on unadjusted scores; however full results are available from the corresponding author.

Causal hypotheses among the manifest variables were tested in the following sequence:

1. SPA₁ and Ach₁ have direct paths to both SPA₂ and Ach₂ (referred to here as the full cross-lagged effects model).
2. SPA₁ has direct paths to SPA₂ and Ach₂; Ach₁ has a direct path to Ach₂ only (SPA₁ → Ach₂ cross-lag effects model).
3. Ach₁ has direct paths to SPA₂ and Ach₂; SPA₁ has a direct path to SPA₂ only (Ach₁ → SPA₂ cross-lag effects model).
4. Ach₁ has a direct path to Ach₂ only; SPA₁ has a direct path to SPA₂ only (stability only effects).

Model 1 was the most general of the four models. If SPA and Ach mutually influence each other, then all structural path coefficients from time 1 to time 2 of Model 1 would be similar in magnitude. However, if SPA is causally related to Ach, then Model 2 would not be significantly worse than Model 1 in terms of overall model fit, and Model 2 would provide a better model fit than Models 3 or 4. Conversely, if Ach was causally related to SPA, then Model 3 would not be significantly worse than Model 1 in terms of overall model fit, and Model 3 would provide a better model-to-data fit than Model 2 or Model 4. The fit of all four models was evaluated against the null model, which only includes correlations among the factors, but no regression coefficients. Thus, the null model makes no assumption of temporal direction.

Prior to analyses, variables were standardized across the whole sample to a mean of zero and unit variance. Analyses were conducted in the program MPlus (Muthén & Muthén, 2007) using a full-information maximum-likelihood fit function (Enders & Bandalos, 2001). This provides parameter estimates while dealing appropriately with missing data by maximizing the casewise likelihood of the observed data (Arbuckle, 1996). Because participants in our study consisted of twins, they are likely to be more similar to one another than a pair of unrelated individuals in a population and cannot, therefore, be considered independent observations. We used the cluster option available in Mplus to take into account the non-independence of twin data. The accuracy of parameter estimates was assessed using likelihood-based 95% confidence intervals. Goodness of fit was also assessed using three statistics: the Bayesian Information Criterion (BIC), the sample-size-adjusted BIC, and the Standardized Root Mean Square Residual (SRMR). Lower BIC and sample-size adjusted BIC values indicate a better (more parsimonious) fit to the data. For the SRMR, values less than .08 are considered to indicate a good fit.

Results

Descriptive statistics

Table 1 shows the means and standard deviations of scores on each measure (standardized on the whole sample to a mean of zero and standard deviation of one) for boys and girls. The

effect size of sex differences was examined using Cohen's *d*. Boys scored somewhat higher than girls on all measures except for Ach and SPA ratings on English. However, the majority of effect sizes could be classed as small ($d < .20$). The largest effect size emerged for SPA ratings on Math ($-.41$ and $-.33$ at ages 9 and 12, respectively), indicating that there was approximately 22–27% non-overlap between boys and girls in their ratings of perceived ability in Math (Cohen, 1988).

Model-fitting analyses

The best-fitting confirmatory factor analysis model excluded the cross-time correlated error for SPA in Science; this parameter was statistically non-significant (full model: BIC: 152366.01, sample-size adjusted BIC: 152200.77, SRMR: .09; reduced model: BIC: 152361.26, sample-adjusted BIC: 152199.20, SRMR: .09). Standardized parameter estimates are shown in Figure 2. Cross-time correlated errors for the measured variables, omitted from Figure 2 for the purpose of clarity, were moderate in magnitude (Ach Maths: .19, CI = .12, .26; Ach English: .22, CI = .16, .29; SPA Maths: .52, CI = .49, .55; SPA Science: .29, CI = .25, .32; SPA English: .48, CI = .44, .52). Factor loadings in Figure 2 show that 69–77% of the variance in Ach ratings, and 20–38% of the variance in SPA ratings were accounted for by the Ach and SPA latent factors, respectively. There was moderate stability between age 9 and age 12 Ach scores (.48), and between age 9 and age 12 SPA scores (.44). At both ages, SPA and Ach scores were moderately correlated (.39 at both ages). Cross-age correlations between SPA and Ach scores were also significant. Specifically, the correlation between Ach at time 1 and SPA scores at time 2 was .40 whereas the correlation between SPA scores at time 1 and Ach at time 2 was .29.

Results from the model-fitting comparisons for the structural models are shown in Table 2. Neither the path connecting SPA₁ to Ach₂ nor the path connecting Ach₁ to SPA₂ could be dropped from the model without resulting in a significant deterioration in model-fit. The model that provided the best fit to the data was the full cross-lagged model.

Path coefficients from the best-fitting full cross-lagged model are shown in Figure 3. 12-year SPA was significantly predicted by 9-year SPA. In addition, 12-year SPA was significantly predicted by 9-year Ach; specifically, Ach at age 9 accounted for approximately 1.7% ($.13^2 = .017$) of the variance in SPA at age 12. In turn, 12-year Ach was significantly predicted by 9-year Ach and by 9-year SPA. Notably, SPA at age 9 accounted for 1.2% ($.11^2 = .012$) of the variance in Ach at age 12. Thus, Ach and SPA had reciprocal influences on each other over time, and this effect was of similar size for Ach and SPA.

When we repeated the model-fitting comparisons in Table 2, but without accounting for CA, the null model and the full cross-lagged model provided the best fit to the data (null model: BIC = 127110.984; sample-size adjusted BIC = 126961.631; RSMR = 0.053; full cross-lagged model: BIC = 127110.984; sample-size adjusted BIC = 126961.631; RSMR = 0.053); however there were conceptual reasons for choosing the cross-lagged model over the null model. The full cross-lagged model revealed that 9-year Ach accounted for 5.3% of the variance in 12-year SPA (standardized path coefficient = .23; 95% confidence interval: .17–.29), whereas 9-year SPA accounted for a smaller, albeit not significantly smaller, proportion of the variance in 12-year Ach, namely 1.4% (standardized path coefficient = .12; 95%

confidence interval: .07-.18). Comparing these results to those in Figure 3 suggests that effects of 9-year Ach on 12-year SPA are more strongly driven by CA than effects of 9-year SPA on 12-year Ach.

Discussion

The present study set out to assess the longitudinal causal relationship between SPA and Ach while considering the effects of CA. Accordingly, it attempted to integrate both SPA-CA and SPA-Ach approaches in order to assess the degree to which SPA are related to Ach via “insight” (in that they are indicative of children’s actual CA or previous Ach) and “self-efficacy” (if they have motivational or self-fulfilling effects on subsequent Ach independently of CA or previous Ach) (Chamorro-Premuzic & Furnham, 2004, 2006). The current findings provide longitudinal support for Spinath et al’s (2006) single-wave study on the validity of SPA in predicting Ach even after controlling for CA. Moreover our results suggest that the effects of previous Ach on subsequent SPA were of similar magnitude to the effects of prior SPA on subsequent Ach, suggesting that the link between SPA and Ach independent of CA is reflective of both “insight” (children’s accounts of their previous performance) and self-efficacy. Thus “feeling competent” on different ability domains appears to be partially due to children’s previous Ach levels at school, whereby high-performing children adjust their SPA upwards and low-performing children adjust them downwards. Likewise, some children may improve their academic performance because they rate their abilities highly; and others may worsen their academic performance because they had no confidence in their abilities. These findings are consistent with reciprocal influences between SPA and Ach. It is important to note, however, that cross-lagged effects were small once CA and the contemporaneous association between Ach and SPA at age 9 were taken into account: age 9 Ach only accounted for only 1.7% of the variance in age 12 SPA, and age 9 SPA for only 1.2% in age 12 Ach. Results also suggested that effects of previous Ach on subsequent SPA are more strongly confounded by CA, than effects of prior SPA on subsequent Ach, which was due to the fact that CA was more strongly associated with Ach than with SPA.

These results have a number of important applied and theoretical implications. From an applied point of view our findings provide support for the longitudinal predictive power of SPA in the prediction of Ach. Spinath et al (2006) had shown that SPA predict Ach even when CA is controlled for, but they assessed all three constructs at one point in time only. Our longitudinal design enabled us to demonstrate that SPA are relatively good predictors of Ach three years later even after controlling for CA. The “complete” predictive relation of SPA predicting Ach independent of CA was the correlation of SPA at age 9 with Ach at age 12 - and this was shown in Figure 2 as being .29. Although this predictive relationship seemed to be largely driven by preexisting associations between SPA and Ach, we argue that given the economical nature of SPA – a measure that can be administered in any form and in less than five minutes – the validity of SPA in predicting future Ach should nonetheless be a point of interest for educational researchers and practitioners interested in the prediction of academic performance. Whether SPA should be the focus of interventions aimed at increasing children’s performance at school (Spinath et al, 2006) is a matter of theoretical discussion, not only in the light of our finding of small effects of SPA on Ach

once previous Ach was taken into account, but also as estimates of relations between SPA and Ach were at the latent variable level, which disattenuated relations for unreliability and unique variance. Observed scores on SPA and Ach would correlate at much lower levels - and it is only observed scores that would be used to characterize individual students for use in interventions.

From a theoretical standpoint our results suggest that the self-enhancing or self-fulfilling effects of SPA in school settings (whether independently of CA or not) are much more modest than previously implied. Only 1.2% of the variance in age 12 Ach was explained by age 9 SPA when CA and age 9 Ach were considered. Accordingly, conceptualizations of SPA as motivational factors (Kanfer, & Ackerman, 2000; Kanfer, & Heggestad, 1997; Middleton & Midgley, 1997; Roeser, Midgley & Urda, 1996; Spinath et al, 2006) are only somewhat supported by the current results.

On the other hand, the results provide some support for the conceptualization of SPA as a valid measure of individual differences in intellectual competence (Chamorro-Premuzic & Furnham, 2006; Chamorro-Premuzic & Arceche, 2008), capturing variance not only from CA (Alicke, 1985; Chamorro-Premuzic & Furnham, 2005; Cogan et al., 1915; Furnham et al, 1999; Paulhus et al, 1998), but also Ach. Given the conceptual overlap between SPA and Ach (SPA are based on Ach-related areas), it may seem unsurprising that SPA are significantly correlated with Ach beyond CA. Yet prior to the current study there was conclusive evidence for the fact that CA is a strong predictor of Ach (Chamorro-Premuzic & Furnham, 2005; Deary, Whiteman, Starr, Whalley, & Fox, 2004; Gottfredson, 2002; Gustafsson & Undheim, 1996), and only limited evidence for the predictive power of SPA (especially when CA is accounted for) (see Spinath et al, 2006).

There are of course limitations to our study that limit the generalizability of the current findings and demand some caution in regards to the above discussed implications. First, our longitudinal design was relatively short-span as data were collected over a 3-year period. Moreover, there were only two stages of data collection, at ages 9 and 12; with more points of assessment during this period or a more extended longitudinal design, we would have no doubt been able to identify more specific patterns of stability, change, and reciprocal influence among the variables measures. Indeed, the implication that interventions aimed at improving Ach through increasing or changing SPA may be of limited value is somewhat harsh given that SPA were assessed only once three years prior to the children's Ach. Motivational theories emphasizing the self-fulfilling nature of competence-related beliefs (Bandura, 1997; Dweck, 1986; Dweck, & Leggett, 1988; Middleton & Midgley, 1997; Roeser, Midgley & Urda, 1996) are at least as situational as dispositional, hence assessing SPA just before children's actual performance (on the school exams) may have yielded different results. This points to an important, second limitation, namely the absence of actual performance data; instead, teachers' ratings were used to operationalize Ach. Although there is evidence for the accuracy of such ratings as indicators of "real" academic performance (Hoge & Coladarci, 1989; Kovas et al., 2007; Oliver, Harlaar, Thomas, Hayiou, Kovas, Walker, & Petrill, 2004), teachers' perceptions of children's performance may be biased by their attitudes towards them, and indeed by children's non-Ach-related behaviors and personality traits. A third important limitation is that the effect sizes reported, especially the

crucial cross-lagged effects, were quite small. Owing to the large sample size, even paths below .1 are significant, but the importance of such effects should not be over-estimated.

Despite these limitations, the current results provide some important empirical support to the conceptualization of SPA as valid measures of intellectual competence (tapping into both CA and Ach). Although the motivational or self-fulfilling effects of SPA beyond previous Ach or measured CA seem marginal, SPA are indicative of children's past Ach and therefore predictive of future Ach. Future studies in this area should therefore take into account that the relation of SPA and Ach over time may only be accurately interpreted if prior levels of Ach are accounted for.

An interesting question to be addressed by future research regards the etiology of SPA, specifically the extent to which SPA can be attributed to genetic and non-genetic factors and to what degree links between CA, SPA, and Ach are mediated genetically and environmentally. The only two studies to address these questions provided evidence for substantial genetic influences of around 40–50% on SPA (Greven, Harlaar, Kovas, Chamorro-Premuzic, & Plomin, 2009; Spinath, Spinath, & Plomin, 2008). Moreover, SPA have been shown to predict a small, albeit significant, proportion of the variance in subsequent Ach independently of CA for genetic rather than environmental reasons (Greven et al., 2009). However, no study to date has examined the complete cross-lagged relationships between SPA and Ach in a genetically sensitive design. Findings from this study are currently extended to explore the genetic and environmental etiologies of these cross-lagged relationships through comparisons of identical and non-identical twins using data from the Twins Early Development Study.

Acknowledgments

The Twins Early Development Study has been funded since 1995 by a program grant from the U.K. Medical Research Council (G9424799, now G050079). Funding has also been received to develop additional areas of research from the U.S. National Institute of Child Health and Human Development (HD) for a quantitative genetic study of school environments (HD44454) and of mathematics (HD46167). We are grateful to the Twins Early Development Study families for their participation and support for more than a decade. This study was also supported by a British Academy Grant (SG 42038) to the first author.

References

- Ackerman PL, Wolman S. How accurate are self-estimates of intelligence? *Journal of Experimental Psychology: Applied*. 2007; 13:57–78. [PubMed: 17535132]
- Akaike H. Factor analyses and AIC. *Psychometrika*. 1987; 52:317–332.
- Alicke MD. Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*. 1985; 49:1621–1630.
- Arbuckle J. Full information estimation in the presence of incomplete data. In: Marcoulides, GA.; Schumacker, RE., editors. *Advanced structural equation modelling issues and techniques*. Mahwah: NJ: Erlbaum; 1996.
- Bandura A. The psychology of chance encounters and life paths. *American Psychologist*. 1982; 37:747–755.
- Bandura, A. *Self-efficacy: The exercise of control*. New York: W.H. Freeman; 1997.
- Baumeister R, Campbell J, Krueger J, Vohs K. Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest*. 2003 May;4(1)

- Byrne BM. The general/academic self-concept nomological network: A review of construct validation research. *Review of Educational Research*. 1984; 54:427–456.
- Calsyn R, Kenny D. Self-concept of ability and perceived evaluations by others: Cause or effect of academic achievement? *Journal of Educational Psychology*. 1977; 69:136–145. [PubMed: 853147]
- Chamorro-Premuzic, T. *Personality and individual differences*. Oxford: Blackwell; 2007.
- Chamorro-Premuzic T, Arteche A. Intellectual competence and academic performance: Preliminary validation of a model. *Intelligence*. 2008; 36:564–573.
- Chamorro-Premuzic T, Furnham A. A possible model to understand the personality-intelligence interface. *British Journal of Psychology*. 2004; 95:249–264. [PubMed: 15142305]
- Chamorro-Premuzic, T.; Furnham, A. *Personality and intellectual competence*. Mahwah: Lawrence Erlbaum Associates; 2005.
- Chamorro-Premuzic T, Furnham A. Intellectual Competence and the intelligent personality: A third way in differential psychology. *Review of General Psychology*. 2006; 10:251–267.
- Cogan LC, Conklin AM, Hollingworth HL. An experimental study of self-analysis, estimates of associates, and the results of tests. *School & Society*. 1915; 2:171–179.
- Cohen, J. *Statistical power analysis for the behavioral sciences*. 2. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
- Dale PS, Harlaar N, Plomin R. Correspondence between telephone and teacher assessments of reading: I. Substantial correspondence for a sample of 5808 children and for extremes. *Reading & Writing: An Interdisciplinary Journal*. 2005; 18:385–400.
- Deary IJ, Whiteman MC, Starr JM, Whalley LJ, Fox HC. The impact of childhood intelligence in later life: Following up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology*. 2004; 86:130–147. [PubMed: 14717632]
- Denissen JA, Zaretzky NR, Eccles JS. I like to do it, I'm able, and I know I am: Longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Development*. 2007; 78:430–447. [PubMed: 17381782]
- Department for Education and Employment. *QCA Key Stage 1: Assessment and reporting arrangements*. London: HMSO; 2000.
- The National Curriculum. *The National Curriculum. Handbook for secondary teachers in England*. London: Department for Education and Skills, Qualifications and Curriculum Authority; 2004.
- Dweck CS. Motivation processes affecting learning. *American Psychologist*. 1986; 41:1040–1048.
- Dweck CS, Leggett EL. A social-cognitive approach to motivation and personality. *Psychological Review*. 1988; 95:256–273.
- Eccles (Parsons), J.; Adler, TF.; Futterman, R.; Goff, SB.; Kaczala, CM.; Meece, JL.; Midgley, C. Expectancies, values, and academic behaviours. In: Spence, JT., editor. *Achievement and achievement motives*. San Francisco: Freeman; 1983. p. 75-146.
- Eccles, JS.; Wigfield, A.; Schiefele, U. Motivation to succeed. In: Eisenberg, N., editor. *Handbook of child psychology*. Vol. 3: Social, emotional and personality development. New York: Wiley; 1998. p. 1017-1095.
- Enders CK, Bandalos DL. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*. 2001; 8:430–457.
- Gose A, Wooden S, Muller D. The relative potential of self-concept and intelligence as predictors of achievement. *Journal of Psychology*. 1980; 104:279–287.
- Gottfredson L. Where and why *g* matters: Not a mystery. *Human Performance*. 2002; 15:25–46.
- Greven CU, Harlaar N, Kovas Y, Chamorro-Premuzic T, Plomin R. More than just IQ: School achievement is predicted by self-perceived abilities – but for genetic rather than environmental reasons. *Psychological Science*. 2009; 20:753–762. [PubMed: 19470122]
- Gustafsson, JE.; Undheim, JO. Individual differences in cognitive functions. In: Berliner, DC.; Calfee, RC., editors. *Handbook of educational psychology*. New York: Prentice Hall International; 1996. p. 186-242.

- Haworth CM, Harlaar N, Kovas Y, Davis OP, Oliver BR, Hayiou-Thomas ME, Frances J, Busfield P, McMillan A, Dale PS, Plomin R. Internet cognitive testing of large samples needed in genetic research. *Twin Research and Human Genetics*. 2007; 10:554–63. [PubMed: 17708696]
- Hedges, LV.; Olkin, I. *Statistical methods for meta-analysis*. Orlando: Academic Press; 1985.
- Judge TA, Jackson CL, Shaw JC, Scott BA, Rich BL. Self-Efficacy and Work-Related Performance: The Integral Role of Individual Differences. *Journal of Applied Psychology*. 2007; 92:107–127. [PubMed: 17227155]
- Kanfer R, Ackerman PL. Individual differences in work motivation: Further explorations of a trait framework. *Applied Psychology: An International Review*. 2000; 49:470–482.
- Kanfer, R.; Heggstad, ED. Motivational traits and skills: A person-centered approach to work motivation. In: Cummings, LL.; Staw, BM., editors. *Research in Organizational Behavior*. Vol. 19. Greenwich, CT: JAI Press, Inc; 1997. p. 1-56.
- Kaplan, E.; Fein, D.; Kramer, J.; Delis, D.; Morris, R. *WISCIII as a process instrument (WISC-III-PI)*. New York: The Psychological Corporation; 1999.
- Kovas Y, Haworth CMA, Dale PS, Plomin R. The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development*. 2007; 72(3 Serial No 188):1–144. [PubMed: 17661895]
- Loehlin, JC. *Latent variable models*. 4. Lawrence Erlbaum; Mahwah, NJ: 2004.
- Mabe PA III, West SA. Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*. 1982; 67:280–296.
- McCarthy, D. *McCarthy Scales of Children's Abilities*. New York: Psychological Corporation; 1972.
- Marsh HW. The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*. 1990a; 82:623–636.
- Marsh, HW. *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. Leicester, UK: British Psychological Society; 2007.
- Marsh HW, Trautwein U, Ludtke O, Koller O, Baumert J. Academic self-concept, interest, grades, and standardized test scores: Reciprocal effect models of causal ordering. *Child Development*. 2005; 76:397–416. [PubMed: 15784090]
- Middleton MJ, Midgley C. Avoiding the demonstration of lack of ability: An under explored aspect of goal theory. *Journal of Educational Psychology*. 1997; 89:710–718.
- Muthén, LK.; Muthén, BO. *Mplus User's Guide*. 5. Los Angeles, CA: Muthén & Muthén; 2007.
- Oliver B, Harlaar N, Thomas ME, Hayiou, Kovas Y, Walker SO, Petrill SA. A twin study of teacher-reported mathematics performance and low performance in 7-year-olds. *Journal of Educational Psychology*. 2004; 96:504–517.
- Oliver B, Plomin R. *Twins' Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behaviour problems from childhood through adolescence*. *Twin Research and Human Genetics*. 2007; 10:96–105. [PubMed: 17539369]
- Raven, J.; Raven, JC.; Court, JH. *Raven's Advanced Progressive Matrices*. Oxford: Oxford Psychologists Press Ltd; 1998.
- Reeves DJ, Boyle WF, Christie T. The relationship between teacher assessments and pupil attainments in standard test tasks at Key Stage 2, 1996–98. *British Educational Research Journal*. 2001; 27:141–160.
- Roeser RW, Midgley C, Urdan T. Perceptions of the school psychological environment and early adolescents' self-appraisals and academic engagement: The mediating role of goals and belonging. *Journal of Educational Psychology*. 1996; 88:408–422.
- Schutte TH. Students' estimates of their ability and achievement. *Journal of Educational Research*. 1929; 20:394–396.
- Shen E. The validity of self-estimate. *Journal of Educational Psychology*. 1925; 16:104–107.
- Shicke MC, Fagan TK. Contributions of self-concept and intelligence to the prediction of academic achievement among grade 4, 6, and 8 students. *Canadian Journal of School Psychology*. 1994; 1:62–69.
- Smith, P.; Fernandes, C.; Strand, S. *Technical Manual*. 3. Windsor: nferNelson; 2001. *Cognitive Abilities Test*.

- Spinath B, Spinath FM, Harlaar N, Plomin R. Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence*. 2006; 34:363–374.
- Spinath FM, Spinath B, Plomin R. The nature and nurture of intelligence and motivation in the origins of sex differences in elementary school achievement. *European Journal of Personality*. 2008; 22:211–229.
- Wigfield A, Karpachian M. Who am I and what can I do: Children's self-concepts and motivation in achievement situations. *Educational Psychologist*. 1991; 26:233–261.
- Valentine JC, Dubois DL, Cooper H. The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*. 2004; 39:111–133.
- Wechsler, D. Wechsler Intelligence Scale for Children. 3. San Antonio, Tex: Psychological Corporation; 1992. (WISC-III)

APPENDIX. Perceived Ability in School Scale

How good do you think you are at:	Very good	Quite good	Doing OK	Not so good	Not at all good
1. Reading?	☐	☐	☐	☐	☐
2. Writing?	☐	☐	☐	☐	☐
3. Spelling?	☐	☐	☐	☐	☐
4. Solving number and money problems?	☐	☐	☐	☐	☐
5. Doing maths in your head?	☐	☐	☐	☐	☐
6. Multiplying and dividing?	☐	☐	☐	☐	☐
7. Learning about nature and living things?	☐	☐	☐	☐	☐
8. Tasting things out to see what they can do? (e.g. magnets)	☐	☐	☐	☐	☐
9. Finding out how things work? (e.g. the human body)	☐	☐	☐	☐	☐

Figure 1. Cross-lag model for achievement (Ach) and self-perceived ability (SPA) at ages 9 and 12. Suffix of 1 indicates time 1 (age 9) and suffix of 2 indicates time 2 (age 12) testing.

Figure 2.

Standardized parameter estimates from confirmatory factor analysis of achievement (Ach) and self-perceived ability (SPA) at ages 9 and 12. Suffix of 1 indicates time 1 (age 9) and suffix of 2 indicates time 2 (age 12) testing. The numbers separated by commas that are presented below the estimates refer to upper and lower bounds of 95% confidence intervals. All path coefficients are significant at $p < .01$.

Figure 3.

Standardized parameter estimates from cross-lag model of achievement (Ach) and self-perceived ability (SPA) at ages 9 and 12. Suffix of 1 indicates time 1 (age 9) and suffix of 2 indicates time 2 (age 12) testing. The numbers separated by commas that are presented below the estimates refer to upper and lower bounds of 95% confidence intervals. All path coefficients are significant at $p < .01$.

Table 1

Means and standard deviations for females and males on achievement (Ach), self-perceived ability (SPA), and cognitive ability measures

	Girls			Boys			<i>d</i>
	M	SD	N	M	SD	N	
Age 9							
Ach: Math	-.06	.95	2294	.07	1.05	1927	-.13
Ach: Science	-.04	.95	2276	.05	1.06	1917	-.09
Ach: English	.11	.95	2310	-.13	1.04	1944	.24
SPA: Math	-.18	1.00	3068	.22	.95	2584	-.41
SPA: Science	-.03	1.00	3066	.03	.99	2579	-.06
SPA: English	.08	.97	3071	-.09	1.02	2584	.17
Cognitive ability	-.02	.99	2939	.03	1.01	2448	-.05
Age 12							
Ach: Math	-.02	.99	1474	.03	1.01	1233	-.05
Ach: Science	-.03	.98	1458	.04	1.03	1241	-.07
Ach: English	.06	.97	1527	-.07	1.04	1259	.13
SPA: Math	-.15	.99	3016	.18	.99	2537	-.33
SPA: Science	-.02	1.00	3015	.03	1.00	2536	-.05
SPA: English	.12	.95	3017	-.14	1.04	2536	.26
Cognitive ability	-.08	1.00	2532	.12	.98	1945	-.20

Note: *d* = Cohen's *d* (unbiased estimate, following Hedges and Olkin, 1985). Negative estimate indicates greater effect in boys. Scores were standardized on the whole sample to a mean of zero and standard deviation of 1.

Table 2

Model-fitting comparisons.

Model	χ^2	df	BIC	Sample-size adjusted BIC	SRMR
1 Null model	2101.918	63	152361.258	152199.195	0.090
2 Full cross-lagged effects (SPA ₁ → Ach ₂ , Ach ₁ → SPA ₂ , Ach ₁ → Ach ₂ , SPA ₁ → SPA ₂)	1760.743	63	151988.446	151826.383	0.064
3 SPA ₁ → Ach ₂ cross-lag effects (SPA ₁ → Ach ₂ , plus Ach ₁ → Ach ₂ and SPA ₁ → SPA ₂)	1777.711	64	151996.079	151837.194	0.065
4 Ach ₁ → SPA ₂ cross-lag effects (Ach ₁ → SPA ₂ , plus Ach ₁ → Ach ₂ and SPA ₁ → SPA ₂)	1876.075	64	152112.211	151953.325	0.072
5 Stability only (Ach ₁ → Ach ₂ and SPA ₁ → SPA ₂)	1895.487	65	152123.214	151967.506	0.074

Note: BIC = Bayesian Information Criterion; SRMR = Standardized Root Mean Square Residual.