The Negative Year in School Effect: Extending Scope and Strengthening Causal Claims

By: Philip D. Parker
*Institute for Positive **Psychology** and Education, Australian Catholic University*;
Herbert W. Marsh
*Institute for Positive **Psychology** and Education, Australian Catholic University*
Felix Thoemmes
College of Human Ecology, Cornell University
Nicholas Biddle
Centre for Social Research and Methods, Australian National University
Acknowledgement: We acknowledge the generous support of the Australian Research Council (DP130102713, DE140100080).

The Negative Year in School Effect: Extending Scope and Strengthening Causal Claims

Self-beliefs are a driving force in young people's academic decision-making processes (Marsh, 2006). Numerous studies show that such self-beliefs predict educational aspirations, even when controlling for performance in high-quality standardized tests (Guo, Marsh, Morin, Parker, & Kaur, 2015; Guo, Marsh, Parker, Morin, & Yeung, 2015; Guo, Parker, Marsh, & Morin, 2015; Parker et al., 2012; Parker, Marsh, Ciarrochi, Marshall, & Abduljabbar, 2014; Parker, Nagy, Trautwein, & Lüdtke, 2014). Youth are thus not perfectly rational agents. Indeed, recent research highlights how public policy, historical change, and institutions feed into the biases young people use to construct a view of their academic selves (Parker, Marsh, et al., 2017; Parker, van Zanden, et al., 2017). It is no surprise that macroforces influence children's sense of self and such thinking has been prominent in political philosophy since at least Jean-Jacques Rousseau. What modern psychological research is showing is that macrostructures can influences a child's sense of self by manipulating the experiences a child has and, more importantly, manipulating who her most frequent interactions are with. This influence need not be direct nor does it have to be intentional. For this reason, educational research must pay close attention to how policy, institutional setting, and historical change may influence a child's social context and illuminate the effect this has on how they come to see themselves. One such area of interest is school enrolment policies. Indeed, an important international concern is how enrolment age policy (both its content and its enforcement) can ensure equity, while respective parents' rights to make important educational choices for their children.

Marsh and colleagues (Marsh, 2016; Marsh, Pekrun, et al., 2017) proposed the Negative Year in School Effect (NYiSE) as a critical predictor of academic self-beliefs. The NYiSE suggests that, holding achievement constant, grade-relative-to-age influences self-concept (i.e., perceptions of competence within a given domain). Thus, if most 15-year-olds are in Grade 9, then being a 15-year-old in Grade 8 (older than classmates) will be associated with higher self-concept, whereas being a 15-year-old in Grade 10 will result in lower self-concepts. Issues related to self-concept and grade-relative-to-age is not an uncommon concern for academics, policymakers, or parents given the potential short and long-term impacts on cognitive and psychosocial outcomes (Jimerson & Brown, 2013). Such concerns, however, have often been muddled because of confounding the process by which a child is

enrolled in a given grade (i.e., delayed enrolment, retention, or acceleration) with the fact of that placement. Thus, there are contradictory claims that retaining a child has a negative effect on students' self-beliefs, whereas some parents engage in red-shirting (holding their children back from entering school for as long as possible) under the assumption that it will provide their child with cognitive and noncognitive advantages over their younger same year grade peers (Edwards, Taylor, & Fiorini, 2011). Indeed, five-percent of the U.S. starting school population now delays school entry by a year beyond when they are eligible (Bedard & Dhuey, 2006), most likely because of parents' belief that their child will benefit from their relatively older age (Lincove & Painter, 2006). Evidence from across the OECD suggests such parents may be right to do so as being relatively old for your grade is beneficial in terms of achievement (Bedard & Dhuey, 2006; cf. Lincove & Painter, 2006). Likewise, Marsh (2016) found that despite large country-to-country variation in school age enrolment and progression policies, being young for your year grade was negatively related to academic self-concept in all OECD countries.

In the current study, we focus on social comparison theory, as expressed in NYiSE, as a foundation for our research and use modern approaches to control for selection effects to assess the effect of the NYiSE on self-belief and educational attainment outcomes.

Competing Self-Concept Theories

Self-concept consists of at least three components: a temporal component in which people evaluate current performance against past performance, a dimensional component in which people compare their performance in one domain against performance in another domain, and a social component in which individuals compare their performance against others (van Zanden, Marsh, Seaton, & Parker, 2015). We now have considerable research on the relative importance of these components, and it appears that social component is by far the strongest (Müller-Kalthoff, Helm, & Möller, 2017). Indeed, social comparison and self-concept are extremely highly correlated ($r = .85$; correcting for unreliability $r = .93$; based on Marsh, Kuyper, Morin, Parker, & Seaton, 2014).

When considering academic self-concept as social comparison, three theories exist: (a) an economic rational actor model, (b) an assimilation to group norms based model, and (c) an identity distortion model (Parker, van Zanden, et al., 2017) based on frame-of-reference theory (Festinger, 1954) and local dominance effect theory (Zell & Alicke, 2010). The rational actor model is the most parsimonious. It states that individuals have a stable

preference to form self-beliefs that are as accurate as possible and that they will do so by pursuing the highest quality objectively information they can. They will do so up to the point at which the costs of pursuit are higher than the benefits gained. In a world of high quality standardized testing there is, according to this approach, no reason why self-concept should not be entirely explainable by academic tests scores and certainly no reason why self-concepts should vary systematically when high quality achievement tests are in plentiful supply (Goldthorpe, 2007). Assimilation models assume individuals adjust their self-concepts to align with the expectations of the group they are most closely related to (Akerlof & Kranton, 2000). The defining feature of these models is that self-concepts conform to group stereotypes or norms and are thus somewhat immune to individual achievement feedback. Finally, information distortion models assume bounded rational actors. In these models, young people are said to want accurate self-concepts but hold a preference for information that is local (i.e., information based on performance relative to class mates rather than information about relative standing in a country as a whole) and salient (i.e., information from individuals who are like them). Thus, although self-concepts are still a function of academic performance, a preference is given to performance feedback that position individuals within their local sphere of influence (Parker, Marsh, et al., 2017; Parker, van Zanden, et al., 2017). The NYiSE assume such a bounded rational actor model and hypothesizes that being relatively young for your year grade will have a negative effect on self-concept. Competing models hypothesize either no effect or a positive effect of being younger for rational actor and assimilation models respectively.

The NYiSE

Frames-of-reference are at the heart of Festinger's (1954) social comparison theory and are the conceptual basis of the NYiSE. Simply put, in an academic setting, students compare their own academic accomplishments with those of their peers. Importantly, recent research has noted that it is proximate or direct classmates and not more distal or generalized conceptions of peers (e.g., those in the region or state or ideas of an average other) that are most powerful in determining academic self-beliefs (Zell & Alicke, 2010). The peers that a child regularly interacts with, as a function of the school and the grade a child is enrolled in, have a strong determining influence on their academic self-beliefs. Indeed, the Big-Fish-Little-Pond-Effect (BFLPE), which is based on frames-of-reference, is one of the most

consistently replicated findings in educational psychology (for reviews, see Marsh et al., 2008; Marsh & Seaton, 2015). The BFLPE is the negative effect of school- or class-average achievement on academic self-concept controlling for individual level achievement. It thus suggests that it is this relative position within a given local frame-of-reference, rather than just objective ability, that forms the basis of academic self-belief.

Marsh (2016) proposed that the NYiSE as an extension to the BFLPE and contended that it was a unifying framework for different research literatures traditions on the negative effects associated with starting school at a younger age and accelerating grades, versus the positive effects for starting school at an older age (i.e., academic red shirting) and repeating a grade. This NYiSE focuses attention on the fact of placement (i.e., the child's repeated exposure to their year grade peers) as the driving force of this effect, rather than the initiating event of, or reason for, placement (e.g., red-shirting, retention, or acceleration; Marsh, 2016). Although considerable research in this area exists (e.g., Alexander, Entwisle, & Dauber, 2003; Jimerson, 2001; Reynolds, 1992; Roderick, 1994; Roderick & Engel, 2001), the NYiSE provides a comprehensive theoretical model that suggests a common set of processes regardless of how a child ends up outside their modal year. The NYiSE relies on the following four premises:

1. Age relative to same year-grade peers influences students sense of self as it relates to their academic prowess,
2. This effect is driven by social comparison of in-grade peers rather than same-age peers,
3. The effect of social comparison is enduring such that it quickly overrides more transitory contrasting effects associated with particular enrolment events, and
4. As a consequence of (4), over the long-term, the NYiSE is a more significant influence on self-beliefs and associated outcomes rather than the type of initiating event for how a child ends up outside the modal year for their age.

Building from these premises, early empirical results have supported tenants of NYiSE theory. Marsh (2016) found that grade was a significant negative predictor of self-concept, holding achievement and demographics constant, in 38 of 41 countries (and negative in all countries) in the 2003 cycle of the program for international student assessment (PISA). Marsh et al. (2017) used longitudinal data and NYiSE theory to show that retention has a

positive or neutral association with a suite of cognitive and psychosocial outcomes and this predictive effect was persistent across early to-middle adolescence in a large German sample.

Potential Issues With NYiSE Research

Despite these findings there are several limitations in the existing research. First, grade-relative-to-age is not distributed at random and thus the NYiSE may reflect selection effects linked to preexisting differences in demographics or other factors (see Edwards et al., 2011). Previous research on grade in school effects has noted that results from OLS regression can give rise to highly biased findings where such preexisting differences are not taken into account (Cliffordson & Gustafsson, 2010). NYiSE research to date has attempted to account for this by including an extensive set of covariates (Marsh et al., 2017) and demonstrating its cross-national generalizability (Marsh, 2016). However, these efforts cannot rule out potential confounding variables that are not observed in the dataset. In defense of the robustness of the negative effects of school- or class-average achievement, Marsh and colleagues (2017) noted that many potential confounders would work against the effect, such that better controls should lead to even more negative frame of reference effects. The logic is captured in Figure 1. Here we assume that most research on effects like the BFLPE contain unmeasured variables U. U is a hidden variable which affects both selection into a particular class, school, or year grade and has a direct effect on self-concept. The argument of Marsh and colleagues is variables that tend to positively effect self-concept (raw intelligence, social advantage, etc.) tend to increase the chances that an individual will enroll in a more competitive environment (the negative effect present in both the BFLPE and the NYiSE). Thus, unless these confounds are controlled for, the path from selection to self-concept will be downwardly biased. Controlling for concurrent academic achievement goes some way to correcting for this but is unlikely to completely resolve this issue (this is because achievement confounds cognitive maturity and school based skill development and will thus likely overcorrect the estimates; see below).

Nevertheless, Marsh and colleagues (2017) noted that this rationale might not extend to NYiSEs and called for further research to more explicitly test this hypothesis. In this paper, we opt for using month of birth as an instrumental variable to define the effect of negative year in school effect (NYiSE). Such a method is likely to overcome most concerns about the biasing effect of unmeasured covariates (though this requires confirmation that birth month is unrelated to other sources of exogenous variance). However, reviewers of this article did

point out that birth month does not resolve all potential issues. Namely, unless participants fill out self-concept questionnaires on their birthday, the NYiSE, even corrected by a birth month instrument, may be biased. This is because those students who are born in January 1988 are both more likely to be a year ahead of the modal year for their age and more cognitively developed than a similar child born in December 1988. This is a common finding in economic research that looks at the effect of birth month on achievement (Lincove & Painter, 2006). The argument is that young people have a biological clock that clicks and a particular rate regardless of environmental context. Thus, effects of being older for a grade may reflect this biological clock rather than enrolment practices (Lincove & Painter, 2006). Put simply, the NYiSE may reflect age-at-testing effects on self-concept rather than enrolment effects (see Black, Devereux, & Salvanes, 2011).

There are, however, two factors in our favor. First, Crawford et al.'s (2014) careful analysis suggests that while achievement differences by birth month can be completely explained by age at test effects, differences in self-concept cannot. This would be consistent with a NYiSE theory that, where achievement differences are driven by IQ, knowledge, and acquired skills, self-concept differences are driven primarily by frame-of-reference effects. Second, being ahead for your age (i.e., positive grade-relative-to-age) is hypothesized to have a negative effect on self-concept, yet the effect of natural cognitive maturation would be expected to have a positive effect on self-concept. As such an instrumental variable approach that controlled for all exogenous variation apart from cognitive maturation effects (i.e., absolute age differences at the time of testing) would result in smaller estimates of the NYiSE (see Figure 2).

Additional limitations of research on NYiSE are that it has only explicitly been tested with reference to math self-concept. We extend the research by not only using two measures of math self-beliefs (self-concept and social comparison), but also two additional measures related to general academic and English social comparison. Finally, when NYiSE is integrated into broader social comparison approaches to self-concept a natural implication is that age relative to peers should have implications for educational outcomes.

Education Transition Choices: A Corollary of the NYiSE

A further aspect of academic self-concept theory is that children's positive self-beliefs are not only desirable outcomes, but worthy of study as significant influences on educational, developmental, and occupational outcomes. This is particularly the case for educational

attainment (rather than just achievement) in which there is significant empirical evidence that academic self-concept influences academic aspirations, attainment, and decision making that effects significant post high-school choices (Guo, Marsh, Morin, et al., 2015; Guo, Marsh, Parker, et al., 2015; Guo, Parker, Marsh, & Morin, 2015; Parker et al., 2012; Parker, Marsh, et al., 2014; Parker, Nagy, et al., 2014). Not only does the importance of self-concept pass empirical evidence tests but it also bridges theories between psychology and sociology. For example, the sociological framework of Breen and Goldthorpe (1997) proposes academic self-beliefs as one cornerstone of their theory on the way in which young people navigate the many sequential branching points of educational transitions. From this we can develop a corollary of the NYiSE. Namely that the NYiSE should be significant and negative for educational attainment, and that the effect should be explained by self-concept. Put simply, for children of equal academic achievement and age, those who are ahead of the modal year grade will be less likely to attend university resulting from their overall lower academic self-beliefs.

Causality and the NYiSE

NYiSE research to date has been explicit that findings do not provide unproblematic evidence of causation. Primarily this is a result of issues related to research design. OLS regression models that aim to estimate the effect of relative age are often confounded and thus estimates may be biased (Angrist & Keueger, 1991). In our case, there may exist an unobserved confounding variable that predicts both self-concept and school starting age. For example, the practice of redshirting by parents who are particularly engaged in the educational process and are strategic in the educational decisions they make for their child may lead to systematic bias in relative age in year group and self-concept. This may thus bias the strength and direction in the NYiSE.

Clearly a randomized control trial (RCT) approach to NYiSE would provide one mechanism by which selection effects can be mitigated and a robust counterfactual constructed. However, to the authors' knowledge, no such trials exist, and were they to take place it would be a number of years (more than a decade in the case of the age group we are studying) for those trials to yield causal inference on outcomes in early adolescence and young adulthood. From some perspectives, this would suggest no causal claims can be advanced at all, regardless of whether such claims are advanced with due caution. Indeed, we have noticed a problematic trend in education and psychology, to treat RCTs as the only

means of providing any confidence in causal claims. This trend treats causal claims as binary such that standards for causality are met if, and only if, claims come from an RCT and not otherwise. This has severely curtailed the introduction of new design and statistical methodological approaches that, while often resting on stronger assumptions than RCTs, nevertheless aim to increase confidence in causal claims (see Morgan & Winship, 2007; Pearl, 2009; Pearl, Glymour, & Jewell, 2016). Providing minor improvements in confidence in causal claims, the explosion of probability based longitudinal data has provided a means for researchers to consider temporal ordering of variables and thus rule out reverse causation (i.e., where a fitted regression model suggests y is a function of x but in nature x is actually a function of y). It is upon this basis that we test the mediation hypothesis of the NYiSE on university entry via academic self-beliefs.

Providing stronger evidence, more sophisticated approaches have aimed to replicate RCTs as closely as possible by matching similar participants in different groups (e.g., Propensity score matching; see Stuart, 2010; Thoemmes & Kim, 2011) or take advantage of natural or quasi-experimental approaches (difference-in-differences, regression of discontinuity, or instrumental variable regression; see Angrist & Pischke, 2009). In the current research, we take advantage of instrumental variable regression to provide stronger evidence of the causal effect of the NYiSE (though see section above).

Instrumental variable (IV) regression takes advantage of cases in which an external variable is related to the outcome variable of interest in no manner other than through the proposed explanatory variable (the 'through and only through' assumption). A basic causal directed acyclic graph (DAG) for such a model is given in Figure 1 where z is the instrument. This figure shows that IV regression has two strong assumptions or what Westhoff (2013) calls 'good' instrument conditions. First the instrument variable z must be associated with the main explanatory variable x. Second, the exclusion restriction states that z cannot be associated with u. IV regression is then estimated in two steps.

In step one variable x is regressed on z. In step two, the predicted x variable, ($\hat{x}$), from the previous step is used as a surrogate for x and is used to predict the outcome variable y, while also controlling for a vector of other observable exogenous characteristics used as control variables (X). When the two conditions mentioned above are met, resulting estimates provide a causal estimate of the effect of x on y (Westhoff, 2013). Mathematically, the process is as follows. The standard OLS equation for individuals i = 1, . . . , n is:

(insert equation images)

A detailed treatment of instrumental variables with examples of economics can be found in Westhoff (2013) and Baltagi (2011). For an introduction based in educational research see Murnane and Willett (2011; particularly Ch. 10).

In our case, we assume that month of birth is not associated with the outcome of interest apart from through the child's age (though see above), but plays a vital role in determining a child's grade-relative-to-age (here we take an age cohort all those who qualify for the program for international student assessment [PISA] in the 2003 cycle). The use of birth month as an instrument to explore educational outcomes is used relatively frequently in educational economics to define causal effects (e.g., Angrist & Keueger, 1991). Yet to our knowledge birth month has rarely been used in educational psychology research to determine the effect of school context on individual difference variables (cf. Crawford et al., 2014).

Our assumptions then are that (a) birth month is associated with year grade in school, (b) birth month is independent of potential confounders (or independent of all confounders but the potential suppressing effect of test taking age) of the effect of year grade on self-beliefs and other outcomes, and (c) birth month only has an effect on self-beliefs and other outcomes of interest via its relationship with year grade.

Birth month is a particularly useful instrument as it is likely measured with little or no error and for the fact that it is so often a strong candidate for meeting the 'through and only through' assumption of IV regression. This has made it particularly popular as an instrumental variable (see Buckles & Hungerman, 2013) starting with the research on the influence of season of birth on earnings via education (Angrist & Keueger, 1991). However, there have been claims that season of birth is not independent of potential confounding variables. Buckles and Hungerman (2013) provided some demographic evidence that high socioeconomic (SES) women were more likely to give birth outside of winter months (cf. Solli, 2017 who found no evidence that families time births differently across a range of background variables in Norway). Currie and Schwandt (2013) review existing literature and suggest birth month may vary by a range of parental background variables. In this research, we test whether birth month is associated with SES, parental education, parental occupational prestige, jurisdiction (i.e., state of residence), migrant status (English speaking, Indigenous, non-English speaking), and number of siblings.

Current Study

We outline a set of hypothesis derived directly from the NYiSE theory and its corollaries derived from integrating it with social comparison and self-concept theory more broadly. The hypotheses are as follows:

Hypothesis 1. OLS regression will show a zero or positive relationship between grade and outcome variables. This is because OLS regression estimates provide biased estimation of the NYiSE because of the presence of confounding variables and do not account for selection effects.

Hypothesis 2. More rigorous causal models will reveal a negative effect between grade and outcome variables, providing a lower-bound estimate of the true causal effect of the NYiSE. In particular, the use of instrumental regression provides a means of accounting for selection effects, such as parents' strategic enrolment of their child in a particular year group, that likely downwardly bias estimates of the NYiSE (see Potential Issues with NYiSE Research section above).

Hypothesis 3. Using the standardized achievement tests from the Program for International Student Assessment (PISA) as a proxy for cognitive development, instrumental variable regression will reveal even stronger negative effects of grade on outcome variables, providing a potential upper-bound estimate of the true effect of the NYiSE (under the assumption that students gain significant self-concept advantages, independently from the NYiSE, from cognitive maturation; see Figure 2).

Hypothesis 4. Testing the corollary of the NYiSE, we hypothesize that grade will have a negative effect on university entry and that this effect will be entirely mediated by self-concept. The rationale is that self-concept is a major predictor of educational choices and thus, any mechanisms that effect self-concept will in turn effect educational choices (all else being equal).

Method

Participants

The Longitudinal Study of Australia Youth (LSAY) 2003 is an extension of the Programme for International Student Assessment (PISA) 2003. It includes most of the PISA 2003 cohort, consisting of 10,370 fifteen-year-old Australians surveyed over 10 years. Participants were born between May 1987 and April 1988. At wave 1 (approximate age 15), achievement, demographics, and math self-concept were measured as part of PISA 2003

cycle. Wave 2 (approximate age 16) provided the measures of math, English, and general academic social comparison (see below). Data relating to university entry from waves 5–10 were also used to measure academic achievement and attainment. The sample had approximately equal numbers of females (49.7%) and males and consisted largely of children born to native-born Australians (78%), with smaller populations of first (11%) and second (9%) generation Australian immigrants (the remaining participants did not report sufficient information on this topic). Three percent of the sample identified as being of Aboriginal or Torres Islander descent. Using international classifications, 40% of the participants had at least one parent with a university level of education, and 43% had at least one parent with either short cycle or postsecondary nontertiary level of education. The remaining participants had at least one parent with some high school (13%) or lower level of education. The average socioeconomic index of the participants' parents on the International Socioeconomic Index was 52.84 (SD = 15.93), which is considerably higher than the OECD average (OECD, 2011).

Strietholt, Rosén, and Bos (2013) suggest that retention and acceleration can bias instrumental regression results. As such, those who had repeated a grade before the age of 15 were excluded from the analysis. Likewise, following Marsh (2016) we used information on school starting age and current grade to identify and exclude accelerators.

Measures

Math Self-concept (MSC) is measured with five Likert response items (e.g., I learn mathematics quickly). We use the continuous scale score (SCMAT) provided by the PISA survey organizers. The full scale and response points are provided in supplementary materials. In the current sample the greatest lower bound and Cronbach's alpha estimates of reliability were both .88. A one-factor congeneric model provided a good fit to the data (TLI = .986, RMSEA = .059).

Social comparison

Wave 2 of the LSAY data contained a general academic social comparison item: "Compared with most of the students in your year level at school, how well are you doing in your school subjects overall?" This was assessed on a 5-point Likert scale of very well, better than average, about average, not very well, and very poorly. In addition, similar questions for

both math ("Compared with most of the students in your year level at school, how well are you doing in Mathematics?") and English ("Compared with most of the students in your year level at school, how well are you doing in English?") were asked. These were reverse scored so that high values reflected more positive perceptions of the individual's position in the class. It is critical to note that these items are strict measures of social comparison that is very similar to but not identical to self-concept measures (Marsh et al., 2014).

Year grade

PISA provides a grade variable that is adjusted such that zero is the modal grade for the country, negative integers represent being behind the modal grade (i.e., old for grade), and positive integers represent being ahead of the modal grade (i.e., young for grade). Thus, the NYiSE would be consistent with a negative effect of grade on a given outcome. This variable is scored for the country as a whole. However, there are some jurisdictional differences by state. In the case of LSAY 2003, however, only one state had a modal year grade that was different than that of the country; Western Australia. In our analysis, we correct the grade variable for Western Australia to bring it into alignment with the other states. Year grade was thus coded as 0 for the modal year, −1 for a year behind, and 1 for a year ahead.

Instrumental variable

We use self-reported birth month, as 11 dummy variables, as the instrumental variables.

University entry

Each year of the LSAY survey, participants undertake an extensive education and occupation interview. The data from this are incorporated with data from previous years to form derived variables reflecting the participants' educational and occupational attainment and status so far. As outlined below using a longitudinal approach we use data from all waves from 2007 (age 19) to 2013 (age 25). We exclude data collected in 2003 to 2006 as in those years only those participants who were ahead of the modal year or at the modal year would have had the opportunity to qualify to attend university (as they had not finished school yet) thus biasing the results. Waiting to the 2007 wave provided sufficient time for the sample to have completed schooling. We restricted analysis to those who indicated that they had

completed year 12 in any year 2006 to 2008 as year 12 graduation is typically a requirement for university entry.

Statistical Analysis

NYiSE hypotheses

We performed an OLS regression using the PISA population weights. We used both cluster robust standard errors and school fixed effects to account for the complex design (see supplementary materials). The survey organizers provided these weights. All outcome variables were in the main analysis were z-scored and thus estimates of the NYiSE are in effect size units that reflect the change in the outcome variable in standard deviation units for an increase in one year grade. Instrumental variable regression was conducted using two staged least squares (2SLS) using the ivpack package, which allows for the use of weights and provides cluster robust standard errors (Jiang & Small, 2014).

Three diagnostic tests were run for each model. The first was the Sargan test. This tests whether the set of instrumental variables and exogenous covariates is identified (i.e., uncorrelated with the residuals; see Baltagi, 2011). This test should be nonsignificant. Second, a Wald Test was used to ensure that birth month was not a 'weak' instrument. This test should be significant (i.e., the inclusion of birth month should significant improve the prediction of year grade). Finally, we use the Wu-Hausman test. This compares the estimated effect of year grade on our response variables estimated from an OLS regression model with that obtained from an equivalent 2SLS (i.e., where both models include all covariates) using a chi-square test. If this test is not significant, it means that results from 2SLS models do not significantly differ from OLS regression results. In such cases, we prefer the OLS results as this model is both simpler and has less uncertainty (i.e., has smaller standard errors). Indeed, IV regression always has larger standard errors than comparable OLS regression models (Baltagi, 2011).

We ran two models for each outcome. The first included exogenous covariates (state/jurisdiction, SES, ethnicity, and number of older siblings). Second, we included the three PISA standardized achievement tests as proxies for intellectual development. These models aimed to provide an approximation of an upper-bound of the NYiSE if the underlying causal model was consistent with that in Figure 2. All models contained a set of state/jurisdiction dummy variables.

Sensitivity tests

IV regression using birth month can be biased if families time birth differently and this difference is systematically related to important covariates. We investigated whether birth month was associated with SES as measured by the PISA ESCS scale, $F(11, 302) = .87$, $p = .57$, parents highest education level, $F(11, 302) = .95$, $p = .49$, parents highest occupational prestige, $F(11, 302) = .75$, $p = .69$, number of older siblings, $F(11, 302) = 1.00$, $p = .44$, ethnicity, $F(22, 292) = 1.52$, $p = .07$, and state/jurisdiction, $F(77, 237) = 1.02$, $p = .45$. These covariates were nevertheless included in all models.

We ran a series of models for each outcome variable, each with slight variations in the estimator used, the use of weights, or the way in which the clustering of students within schools was accounted for. The aim of this approach was to ensure that the results were robust to model specification. The model set-ups consisted of the following:

1. IV reg with cluster robust standard errors, no weights, using the Limited Information Maximum Likelihood, rather than the 2SLS estimator;
2. IV reg with cluster robust standard errors, no weights, using 2SLS;
3. IV reg with cluster robust standard errors, weights, using 2SLS;
4. IV reg with cluster robust standard errors, no weights, including achievement in both stage 1 and stage 2, using 2SLS;
5. IV reg with cluster robust standard errors, no weights, including achievement and grade as outcomes in Stage I (i.e., both year grade and achievement were predicted by the birth month instrument);
6. IV reg with cluster robust standard errors, weights, including achievement in both stage 1 and stage 2, using 2SLS;
7. IV reg with cluster robust standard errors, weights, including achievement and grade as outcomes in Stage I;
8. Models 2–7 including school fixed effects in both stage 1 and stage 2.

Results from all models were consistent and were never statistically significantly different. However, for some model's diagnostic tests revealed problems with models that included weights or school fixed effects. In particular the Sargan test was significant for one, or in one case, two response variable (see Supplementary Material). For this reason, we focus our reporting on Models 2 and 4 above. All other model results, and the full set of exact p

values for the diagnostic tests, can be found in the supplemental material. The interpretation of model results was the same regardless of which model set-up was used.

NYiSE corollary hypothesis

Testing the effect of the NYiSE on educational attainment required longitudinal data. Participants were asked every year from age 16 if they attended university. However, we focused on the ages in which youth typically qualify for university entry (ages 19 to 25; see above). At each wave of LSAY, approximately 10% of the sample from the previous wave did not complete the questionnaire. As such, attrition is a notable concern. There are two options for dealing with this. The first is to choose a particular wave (say age 19) and impute the missing data (we use multiple imputations for some of the self-belief; see below). The problem with this is that the variables for university entry are derived variables, which use responses from multiple questions and multiple time waves making imputation at a given point in time dubious. The other option is to use all available data over the time period of interest and summarize estimates for each individual. When data are represented in long format (i.e., each observation for each individual is a separate row) estimates modeling such data necessarily incorporate all available information. As all available data are used, this avoids the problem present in balanced sample designs in which the researcher must balance the desire to use early waves to have the largest possible sample versus the desire to use later waves when an individuals' entry into university is most likely to be revealed. We chose the second option and analyzed the data using Generalized Estimating Equations (GEE) in the yags package (Carey, 2012). An autoregressive correlation structure of lag 1 was used, though it should be noted that GEE models tend to be robust under mild departures from the specified correlation structure (Park & Shin, 1999). GEE models provide estimates at the population averaged effects and are useful in instances, such as the present, where within person covariates are not of interest, outcomes are qualitative in measurement, and where maintenance of the largest possible effective sample size in the presence of attrition is of concern.

Missing Data

There were very few missing data present for the PISA wave of the LSAY sample (i.e., Wave 1) or subsequent wave (Wave 2) at less than 1% for all outcomes and less that 2% for all exogenous covariates. Nevertheless, it was deemed prudent to impute missing data. We used the bootstrapped EM procedure for imputation from the Amelia package (Honaker,

King, & Blackwell, 2011). For university entry, there were no missing data from unit nonresponse. As noted above, the LSAY database has notable attrition including up to 35% in year 2007 (wave 5) to 64% by year 2013 (wave 10). This attrition is a known issue in LSAY, and to account for it the survey organizers provide well developed sample plus attrition weights. The aim of these weights is to ensure that the sample remains representative of the original population of interest (Lim, 2011). For details on how these weights are derived please see Lim (2011).

Results

Weak Instrument Test

A first step in IV regression is to ensure that the proposed IV is not a weak instrument (i.e., poorly related to the proposed independent variable). We do this with a Wald test between a model predicting the grade with all covariates and a second model that also includes the instrumental variable. Staiger and Stock (1997) suggest that this first stage regression should have an F value of at least 10 or more, and that smaller F values indicate a weak instrumental variable. In our case, the ratio was 440 ($p < .001$). Table 1 indicates the mean age of first data collection for each month and the number of participants in the modal year (Year 10), a year behind (Year 9), or a year ahead (Year 11). The small number of participants (<1%) who were outside these grades were excluded from analysis. Table S1 in the supplemental material shows the first stage results from the 2SLS models.

NYiSE Effect on Outcomes

Results from the OLS models with either cluster robust standard errors or school-fixed effects suggest that grade in school is not related to math self-concept, or math, English, and general academic social comparison (see Table 2 and Table S3b). The use of instrumental variable regression resulted in consistent negative effects which were significant for all variables with the exception of English social comparison (though this effect was significant at $p < .10$ and was not statistically significantly different from the other effects; p values ranged from $p = .271$ for math social comparison to $p = .381$ for math self-concept). Wu-Hausman tests were significant for every variable. This indicated that the results from the IV models were significantly different than their counterpart estimates from OLS regression. The effect size estimates were generally around .10 of a standard deviation decrease in self-concept for every year grade in school, such that it was estimated that, at a similar age, year 9 had approximately .20 higher math self-concept and math social comparison than year 11

students. Results using a limited information maximum likelihood estimator (LIML) had almost identical results to the results presented in Table 2 (see Table S2).

Using the PISA achievement scores as a proxy for differences in cognitive maturity that may be present in the sample, we reestimated the IV reg models (see Table 3). In all these cases, the Wu-Hausman test was not significant when achievement was included in both the 2SLS and OLS regression models. This indicated that an OLS model returned similar results to the 2SLS. Although, the OLS effect sizes are smaller than the IV reg models, they are considerably more efficient than the 2SLS and thus we interpret these results.

One interpretation of these results is that they reflect an upper-bound on the effect size of the NYiSE; consistent with the strategy we outlined in the methodology. That is that while the models without achievement likely reflect as close as possible causal effects for the NYiSE, if cognitive maturation has a positive effect on achievement, which in turn positively predicts self-concept and no other confounders and/or collider bias is present, then the NYiSE may be better represented by controlling for achievement. In this case, the estimated NYiSE was approximately double the size of the models without achievement and were significant at traditional levels for all variables. This indicated that, for example, a 15-year-old in year 9 would have almost half a standard deviation higher math self-concept than an equally capable 15-year-old in year 11.

NYiSE Corollary

Utilizing the advantage of longitudinal data to rule out reverse causation, we also considered a corollary of the NYiSE. Namely, that year grade should have a negative effect on educational attainment that is explained by self-beliefs (see Marsh, 1991). We used generalized estimating equation (GEE) models to predict whether individuals entered university at any time from 2007 to 2013. We chose not to use instrumental variable regression in this case as standard IV regression models produce inconsistent results when used to estimate binary outcome variables (Wooldridge, 2010). Although there are means around this for standard cases, estimating such models with panel data such as the present becomes difficult. For this reason, we estimated a GEE model with an extensive set of controls (state, number of older siblings, school, SES, ethnicity, achievement in math, reading, and science). All results are reported using robust standard errors. When achievement is not controlled for, the effect of grade on university entry is positive (log-odds

= .296 [.255, .366]). Upon controlling for achievement, however, we see that NYiSE had a significant negative effect on university entry (log-odds = −.115 [−.197, −.032]). Table 4 provides the predicted probabilities of university entry at different levels of achievement for those at, ahead, and behind the modal year grade (evaluated at the average of socioeconomic status, the modal values for number of siblings and ethnicity and for the most populous state, NSW). For reference, a moderately achieving individual 1-year behind the modal grade was nearly six-percentage points more likely to enter university than a peer 1 year ahead of the modal year.

In the next step, we controlled for all self-beliefs (math self-concept, math, English, and academic social comparison) in reestimating the model. The significant NYiSE was found to be nonsignificant (log-odds = −.001 [−.091, .072]). This indicates that the NYiSE was completely explained by self-beliefs. The predicted probabilities for this model (holding all self-beliefs at the sample average) are presented in Table 5 and show the disappearance of the NYiSE.

Discussion

The current research aimed to (a) provide more robust evidence to help evaluate the causal claims of the NYiSE, (b) determine the extent to which the effect was also evident in constructs beyond math self-concept, and (c) evaluate a corollary of the NYiSE that grade, for similar aged peers, has a negative effect on ambitious educational choices. We found support in all three cases. First, the NYiSE was significant under IV regression. Furthermore, when controlling for academic achievement, the effects sizes were approximately twice as large. This may provide an upper-bound to the effect size of the NYiSE should the assumptions present in Figure 2 represent the true data-generating model in nature. Importantly, not only did these results provide increased confidence in the causal effect of the NYiSE for math self-concept, findings suggested it extended to a range of social comparison variables measured a year later. Finally, we showed that the NYiSE, when integrated with social comparison and sociological theories of educational choice theory, predicts a negative effect on university entry and this effect can be entirely explained by self-concept. We found evidence consistent with this perspective.

Self-concept theory—the reciprocal effects model (see Marsh, 2006 for a review)—can be used to help explain the effect of the NYiSE on university entry beyond the direct effect of self-concept on the choice processes of young people. This theory states that

academic self-beliefs and academic achievement mutually reinforce each other such that anything that leads to a rise in self-beliefs will have on corresponding effect on academic achievement. As such, a fully articulated process model may suggest that the NYiSE, via social comparison processes, influences self-beliefs, which in turn influence subsequent academic achievement and that both these self-beliefs and resulting achievement are major inputs into the choice process that students go through at educational transition points. In this way, the social psychology of Festinger's (1954) is linked to the sociological focus of Boudon (1974) on the need to consider both achievement and nonachievement (including self-belief) influences on children's negotiation of the various stages of education.

Non-Linear Effects

An anonymous reviewer noted that the NYiSE may not be linear. Thus, those behind the modal year grade for their age may experience the effect in a fundamentally different way to those who are above the modal year grade. We did not originally hypothesize this and as such do not provide a test of this in the main results. However, the possibility of nonlinear effects is both possible and intriguing. As such we categorized year grade into modal year (reference group), behind the modal year, and above the modal year. The NYiSE would assume that the behind group would have higher self-beliefs (self-concept and social comparison) than the modal group and hence positive coefficient, while the ahead group would have lower self-beliefs and hence negative coefficients. As this approach required the categorization of year grade, these models (particularly the 2SLS models) had lower power. Nevertheless, interesting results were obtained (see Table S5 in supplementary material). In particular for math self-concept and social comparison the behind group did indeed have significantly higher self-beliefs but the effects for the ahead group were nonsignificant (and positive). A similar pattern emerged for academic social comparison though the behind group was not significantly lower. As with the main results English social comparison was an outlier in which the behind group were almost identical to the modal group but the ahead group had lower (though nonsignificant) social comparison scores. More research on nonlinear NYiSE is clearly warranted but a tentative hypothesis can be derived from these results. That is, at least for math, being behind the modal grade may be particularly beneficial whereas being ahead may not represent a particular disadvantage and may have some benefits. It maybe that school staff and parents are particularly sensitive to children who are youngest in their grade and provide preventative intervention. Alternatively, being young for your year grade may trigger assimilation mechanisms that increase motivation and effort in

the student in order not to be left behind. Given the large standard errors noted here and the lack of a priori hypothesizing on our part, all such claims should be regarded as tentative until research specifically on this issue can be conducted.

Are the Effect Sizes Meaningful?

The current research reflects some of the most robust evidence on the effect size of the NYiSE. Overall, our results suggest self-concept and self-beliefs declines by between .10 and .20 standard deviation units for each year above the modal age grade. This is quite consistent with Marsh (2016) multicountry research in which the average NYiSE, put onto relatively similar metric to ours, was −.20, with countries ranging from −.012 to −.426 (10/90th percentile = −.064/−.294). Australia, the context of the current research, had an effect size of about −.20 in the Marsh study. By traditional standards (Cohen, 1992) these would reflect a small effect. However, these traditional metrics are largely decontextualized from a given research context. More critical is a consideration of what the policy implications may be, what cost-benefits from intervening might there be, and how large the population a given effect relates to is. Our research suggests that the NYiSE we found was sufficient to account for approximately 6 percentage-point difference in the university entry rates of moderate achievers behind the modal year compared with those ahead. Such an effect size should be of interest to policymakers, particularly given the number of new students enrolled in school in a given year. But careful consideration must be given to the relative cost-benefit trade-off of intervening.

Limitations

Although the IV approach is a considerable improvement over standard OLS models, it is still reliant on a set of assumptions that are not empirically testable. Namely, that the instruments (month of birth) are correlated with the outcome variable (self-concept) only through the effect on the main explanatory variable (age). Although this assumption is theoretically robust, a complimentary set of research on NYiSE would seek to identify situations where school starting age is randomly determined (e.g., the cost of strict enforcement of enrolment ages or the provision of additional resources to those youngest in a school grade).

The other limitation was that for some of the outcomes, in some of the sensitivity models, the Sargan test was significant. This seemed to be driven mostly by the inclusion of school fixed effects, but was also occasionally significant with the inclusion of weights. In all

cases, however, results from supplementary models were very similar to models reported in text where Sargan tests were not significant for any outcomes. Likewise, the significant Sargan tests were typically only for math social comparison, however, in such cases math self-concept gave consistent results with math social comparison and typically did not have a significant Sargan test. It may well be that the significant effect is driven by a correlation that is introduced by the two-stage sampling of the LSAY database (schools are the primary sampling unit) or by oversampling of particular groups (e.g., Indigenous students). The sampling strategy is very complex for LSAY, including sampling of schools proportional to size within strata based on a number of factors (Lim, 2011). Given the complexity of the sampling procedure, it is hard to narrow down where issues might be present, if indeed they are present at all. Dearing, Zachrisson, & Nærde (2015) suggest that IV models exist on a continuum from less to more plausible. As such the few significant Sargan tests under some model specifications should be balanced against the consistency and robustness of the results and the theoretical plausibility of the IV (along with the theoretical justification of the IV's independence from potential covariates) when interpreting the results. We believe this tips the balance in favor of our instrument being more plausible rather than less. Nevertheless, readers should take into account the few significant Sargan tests when interpreting the results.

Implications for Theory and Practice

A critical implication of this research, and NYiSE theory more generally is that too much consideration is given to the events which lead children to be out of their modal year grade and insufficient attention is given the ongoing implications of being outside the modal year irrespective of their pathway to it. Initiating events like retention or promotion often garner considerable attention in the minds of parents, policymakers, and researchers because they have defined temporal boundaries and have immediate and recognizable effects on children; at least in the short term. Yet, soon over, these events pale in significance to the accumulation of children's daily interactions with their peers that seem to be most important to the formation of their self-beliefs. Similar arguments have been made in relation to enrolment into selective schools (Marsh, 2006). Further, Marsh (2016) showed that the NYiSE had the same association for both retention and red-shirting. From our perspective, the joy or shame that comes from acceleration or retention is likely to erode over time,

whereas the ever-present comparison with same class peers becomes an indomitable influence in the way in which children come to view themselves academically.

This research is consistent with a broader perspective of the centrally important role that cultural traditions, practices, or theories about school selection, educational policy and bureaucracy play in the inner lives of children. The central role that social comparison plays in the way in which children form their self-beliefs ensures that any decision that effects who a child's peers will be, will play a significant determining role in the way they feel about themselves. This acknowledgment of macroforces on the intrapsychic constitution of individual has long been a clear feature of developmental psychology including the theoretical models of Bronfenbrenner (1979) and Glen Elder (1985). However, relatively more attention has been given to the theoretical possibility of macrocontextual effects on individual identity than on articulating the processes by which this might happen. As such, there is an increasing need for research, such as the present, that clearly articulates the mechanisms by which government and bureaucracy at macrocontextual level, and mediating structures at the meso-contextual level, influences children's growing sense of self and identity in the various fields of their life.

It is also worth noting that the NYiSE has implications for self-belief theory more generally. In the literature review we stated that there are three major competing models of self-concept. An economic rational actor model indicates that where high-quality objective information exists, individuals will form their self-beliefs in relation to it (e.g., Goldthorpe, 2007). From this model, there is rarely a reason to expect systematic variation in self-beliefs provided high-quality information is available. An assimilation models suggests that humans have a powerful motive to conform the normative self-beliefs and identity of their local group (e.g., Akerlof & Kranton, 2000). This theory implies that groups will exert a significant positive effect. Finally, an information distortion model, agrees with economic models that people desire accurate self-concept, but that they tend to weight information more heavily that is local and salient (Parker, Marsh, et al., 2017; Parker, van Zanden, et al., 2017). The negative NYiSE effects shown here strongly support this latter model. When considered along with the consistent evidence in favor of the BFLPE it can be suggested that classroom exhibit much more comparative rather than normative power in the inner lives of children.

From a practical perspective, we suggest that both parents and policymakers not be misled by the immediacy of the impact on children when it comes to decisions about school

placement or enrolment age. Rather, consideration should also be given to the formative role that daily interactions with peers will have on the child over the longer term. However, such recommendations should be considered in light of the effect sizes presented here and the fact that we estimate average effects. As such, results like these will always be more useful in relation to macropolicy and should be considered with much greater caution when considering the specific lives of individual children.

Footnotes

1 Assumes reliability alpha of .85 for social comparison.

2 This is quite similar to Festinger's (1954) evolved drive to accurately assess one's abilities against others.

3 Queensland also has had a slightly different policy regarding school starting age. However, in these data, the modal grade is consistent with the rest of the country.

4 This effect was just significant when an un-adjusted Wald test (unadjusted for the complex sampling strategy) was used ($F[22, 292] = 1.63$, $p = .04$).

5 Marsh (2016) reports effect sizes in standard deviation units changes in self-concept given a 1-standard deviation unit change in year grade. Using information from this paper we convert these reported effect sizes into approximately similar effect sizes to ours where changes in self-concept are given for a change in 1-year grade.

References

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. The Quarterly Journal of Economics, 115, 715–753. 10.1162/003355300554881

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (2003). On the success of failure: A reassessment of the effects of retention in the primary school grades. New York, NY: Cambridge University Press.

Angrist, J., & Keueger, A. (1991). Does compulsory school attendance affect schooling and earnings? The Quarterly Journal of Economics, 106, 979–1014. 10.2307/2937954

Angrist, J. D., & Pischke, J. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton, NJ: Princeton University Press.

Baltagi, B. H. (2011). Econometrics. Amsterdam, the Netherlands: Springer Texts in Business and Economics. 10.1007/978-3-642-20059-5

Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. The Quarterly Journal of Economics, 121, 1437–1472.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Too young to leave the nest? The effects of school starting age. The Review of Economics and Statistics, 93, 455–467. 10.1162/REST_a_00081

Boudon, R. (1974). Education, opportunity, and social inequality. New York, NY: Wiley.

Breen, R., & Goldthorpe, J. H. (1997). Explaining educational differentials towards a formal rational action theory. Rationality and Society, 9, 275–305. 10.1177/104346397009003002

Bronfenbrenner, U. (1979). The ecology of human development. Cambridge, MA: Harvard University Press.

Buckles, K. S., & Hungerman, D. M. (2013). Season of birth and later outcomes: Old questions, new answers. The Review of Economics and Statistics, 95, 711–724. 10.1162/REST_a_00314

Carey, V. (2012). yags: Yet another GEE solver. (R package version 6.1–13/r50). Retrieved from https://R-Forge.R-project.org/projects/yags

Cliffordson, C., & Gustafsson, J. E. (2010). Effects of schooling and age on performance in mathematics and science: A between-grade regression discontinuity design with instrumental variables applied to Swedish TIMSS 1995 data. In 4th IEA International Research Conference (IRC), Gothenburg, Sweden.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155–159. 10.1037/0033-2909.112.1.155

Crawford, C., Dearden, L., & Greaves, E. (2014). The drivers of month-of-birth differences in children's cognitive and non-cognitive skills. Journal of the Royal Statistical Society, Series A: Statistics in Society, 177, 829–860. 10.1111/rssa.12071

Currie, J., & Schwandt, H. (2013). Within-mother analysis of seasonal patterns in health at birth. Proceedings of the National Academy of Sciences of the United States of America, 110, 12265–12270. 10.1073/pnas.1307582110

Dearing, E., Zachrisson, H. D., & Nærde, A. (2015). Age of entry into early childhood education and care as a predictor of aggression: Faint and fading associations for young Norwegian children. Psychological Science, 26, 1595–1607. 10.1177/0956797615595011

Edwards, B., Taylor, M., & Fiorini, M. (2011). Who gets the "gift of time" in Australia? Exploring delayed primary school entry. Australian Review of Public Affairs, 10, 41–60.

Elder, G. H., Jr. (1985). Perspectives on the life course. In G. H.Elder, Jr., (Ed.), Life course dynamics: Trajectories and transitions (pp. 23–49). Ithaca, NY: Cornell University Press.

Festinger, L. (1954). A theory of social comparison processes. Human Relations, 7, 117–140. 10.1177/001872675400700202

Goldthorpe, J. H. (2007). On sociology. Stanford, CA: Stanford University Press.

Guo, J., Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2015). Directionality of the associations of high school expectancy-value, aspirations, and attainment: A longitudinal study. American Educational Research Journal, 52, 371–402. 10.3102/0002831214565786

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Yeung, A. S. (2015). Expectancy-value in mathematics, gender and socioeconomic background as predictors of achievement and aspirations: A multi-cohort study. Learning and Individual Differences, 37, 161–168. 10.1016/j.lindif.2015.01.008

Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. S. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. Developmental Psychology, 51, 1163–1176. 10.1037/a0039440

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. Journal of Statistical Software, 45, 1–47. 10.18637/jss.v045.i07

Jiang, Y., & Small, D. (2014). ivpack: Instrumental variable estimation. (R package version 1.2.) Retrieved from https://CRAN.R-project.org/package=ivpack

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. School Psychology Review, 30, 420–437.

Jimerson, S. R., & Brown, J. A. (2013). Grade retention. In J. A.Hattie & E. M.Anderman (Eds.), International guide to student achievement (pp. 42–44). New York, NY: Routledge.

Lim, P. (2011). Weighting the LSAY programme of international student assessment cohorts. Adelaide, South Australia: NCVER. Retrieved from https://www.lsay.edu.au/__data/assets/pdf_file/0024/181437/LSAY_Technical_report_61_a.pdf

Lincove, J. A., & Painter, G. (2006). Does the age that children start kindergarten matter? Evidence of long-term educational and social outcomes. Educational Evaluation and Policy Analysis, 28, 153–179. 10.3102/01623737028002153

Marsh, H. W. (1991). Failure of high-ability high schools to deliver academic benefits commensurate with their students' ability levels. American Educational Research Journal, 28, 445–480. 10.3102/00028312028002445

Marsh, H. W. (2006). Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology. Leicester, England: British Psychological Society.

Marsh, H. W. (2016). Cross-cultural generalizability of year in school effects: Negative effects of acceleration and positive effects of retention on academic self-concept. Journal of Educational Psychology, 108, 256–273. 10.1037/edu0000059

Marsh, H. W., Kuyper, H., Morin, A. J., Parker, P. D., & Seaton, M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. Learning and Instruction, 33, 50–66. 10.1016/j.learninstruc.2014.04.002

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Lichtenfeld, S. (2017). Long-term positive effects of repeating a year in school: Six-year

longitudinal study of self-beliefs, anxiety, social relations, school grades, and test scores. Journal of Educational Psychology, 109, 425–438. 10.1037/edu0000144

Marsh, H. W., & Seaton, M. (2015). The Big-Fish–Little-Pond Effect, competence self-perceptions, and relativity: Substantive advances and methodological innovation. In A. J.Elliott (Ed.), Advances in motivation science (Vol. 2, pp. 127–184). Amsterdam, the Netherlands: Elsevier. 10.1016/bs.adms.2015.05.002

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish-little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. Educational Psychology Review, 20, 319–350. 10.1007/s10648-008-9075-6

Morgan, S., & Winship, C. (2007). Counterfactuals and causal inference: Methods and principles for social research. New York, NY: Cambridge University Press. 10.1017/CBO9780511804564

Müller-Kalthoff, H., Helm, F., & Möller, J. (2017). The big three of comparative judgment: On the effects of social, temporal, and dimensional comparisons on academic self-concept. Social Psychology of Education, 20, 849–873. 10.1007/s11218-017-9395-9

Murnane, R. J., & Willett, J. B. (2011). Methods matter. Oxford, UK: Oxford University Press.

OECD. (2011). Education at a glance. Paris, France: OECD.

Park, T., & Shin, D. Y. (1999). On the use of working correlation matrices in the GEE approach for longitudinal data. Communications in Statistics Simulation and Computation, 28, 1011–1029. 10.1080/03610919908813590

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. Educational Psychology, 34, 29–48. 10.1080/01443410.2013.797339

Parker, P. D., Marsh, H. W., Guo, J., Anders, J., Shure, N., & Dicke, T. (2017). An information distortion model of social class differences in math self-concept, intrinsic value and utility value. Journal of Educational Psychology. Advance online publication. 10.1037/edu0000215

Parker, P. D., Nagy, G., Trautwein, U., & Lüdtke, O. (2014). The internal/external frame of reference as predictors of career aspirations and university majors. In J.Eccles & I.Schoon (Eds.), Gender differences in aspirations and attainment: A life course perspective. Cambridge, UK: Cambridge University Press. 10.1017/CBO9781139128933.015

Parker, P. D., Schoon, I., Tsai, Y. M., Nagy, G., Trautwein, U., & Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: A multicontext study. Developmental Psychology, 48, 1629–1642. 10.1037/a0029167

Parker, P. D., van Zanden, B., & Parker, R. B. (2017). Girls get smart, boys get smug: Historical changes in gender differences math, English, and academic social comparison and achievement. Learning and Instruction. Advance online publication. 10.1016/j.learninstruc.2017.09.002

Pearl, J. (2009). Causality. New York, NY: Cambridge University Press. 10.1017/CBO9780511803161

Pearl, J., Glymour, M., & Jewell, N. P. (2016). Causal inference in statistics: A primer. West Sussex, UK: Wiley.

Reynolds, A. J. (1992). Grade retention and school adjustment: An explanatory analysis. Educational Evaluation and Policy Analysis, 14, 101–121. 10.3102/01623737014002101

Roderick, M. (1994). Grade retention and school dropout: Investigating the association. American Educational Research Journal, 31, 729–759. 10.3102/00028312031004729

Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. Educational Evaluation and Policy Analysis, 23, 197–227. 10.3102/01623737023003197

Solli, I. F. (2017). Left behind by birth month. Education Economics, 25, 323–346. 10.1080/09645292.2017.1287881

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. Econometrica, 65, 557–586. 10.2307/2171753

Strietholt, R., Rosén, M., & Bos, W. (2013). A correction model for differences in the sample compositions: The degree of comparability as a function of age and schooling. Large-scale Assessments in Education. Advance online publication. 10.1186/2196-0739-1-1

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. Statistical Science, 25, 1–21. 10.1214/09-STS313

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. Multivariate Behavioral Research, 46, 90–118. 10.1080/00273171.2011.540475

van Zanden, B., Marsh, H. W., Seaton, M., & Parker, P. D. (2015). Self-concept: From unidimensional to multidimensional and beyond. In J. D.Wright (Ed.), International encyclopedia of the social and behavioral sciences (pp. 460–468). Amsterdam, the Netherlands: Springer. 10.1016/B978-0-08-097086-8.25089-7

Westhoff, F. (2013). An introduction to econometrics: A self-contained approach. Cambridge, MA: MIT Press.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. Cambridge, MA: MIT Press.

Zell, E., & Alicke, M. D. (2010). The local dominance effect in self-evaluation: Evidence and explanations. Personality and Social Psychology Review, 14, 368–384. 10.1177/1088868310366144