Learning and Instruction 54 (2018) 125–137

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Girls get smart, boys get smug: Historical changes in gender differences in math, literacy, and academic social comparison and achievement

Philip D. Parker [a,*], Brooke Van Zanden [a], Rhiannon B. Parker [b]

[a] Institute for Positive Psychology and Education, Australian Catholic University, Australia
[b] University of Wollongong, Australia

ARTICLE INFO

Article history:
Received 7 September 2016
Received in revised form
21 August 2017
Accepted 2 September 2017
Available online 22 September 2017

Keywords:
Gender differences
Self-concept
Social comparison
Assimilation
Contrast

ABSTRACT

Girls' lack of self-belief has frequently been cited as a major barrier to advancement in both empirical research and in the popular imagination. With girls now outcompeting boys at almost every educational level, this paper considers if girls still have lower self-concept than boys, if this changes when controlling for academic ability, and what mechanisms explain gender differences. We compare and contrast rational choice, contrast, and assimilation approaches to self-concept and juxtapose historical trajectories in gender differences in self-concept and achievement to distinguish between them. We do this in five age cohorts born between 1981 and 1993 (N = 66, 522) for math, literacy, and general academic domains. Results suggest that there are still significant differences in self-concept between equally able boys and girls and that a mix of assimilation and contrast mechanisms likely explains the size and direction of these effects.

1. Introduction

Gender differences in human capital have largely been eliminated from the labor market (Goldin, 2014). In education, females now outperform males at most levels of education and are better represented in universities (OECD, 2015). Yet gender gaps persist in average income, and employment in prestigious occupations and leadership roles (CEDA, 2013; Goldin, 2014). There are a variety of reasons why this may be the case including both structural issues and differences in non-cognitive factors (Chevalier & Arnaud, 2007; Goldin, 2014). Both academic research (Hyde, 2014; Phelan, Moss-Racusin, & Rudman, 2008; Rudman, 1998) and the media (Duberman, 2014; League, 2011) have highlighted a reason of particular relevance – self-concept and self-promotion. Indeed, cross-cultural and developmental research has demonstrated a relatively stable moderate self-esteem advantage for men of about 0.25 of a standard deviation (Bleidorn et al., 2015) while also suggesting that women not only have more negative general self-concept but can also be socially penalised for overt displays of confidence (Phelan et al., 2008).

It is natural to ask then, does this difference in self-concept have its origin in schooling and, if so, how has this difference responded to historical increases in human capital and the educational attainment of women from the preceding decades? This is certainly not the first research to consider self-concept in education as a central explanatory variable in gender differences in long-term outcomes. Indeed, the most recent Organisation for Economic Co-operation and Development (OECD, 2015) report on gender differences discusses such beliefs as the central non-cognitive factor in explaining gendered outcomes in math and science domains internationally. However, our focus is on historical trends in differences in self-concept conditioned on achievement (i.e., of equally able boys and girls) and how these trends are related to trends in gender gaps in achievement. As such, we aim to: a) describe the historical trends in gender gaps for adolescents over a historical period of more than a decade, b) identify particular mechanisms that may be relevant to these differences, c) address how such mechanisms may work together, and c) consider whether historical trends support a particular mechanism or combination of mechanisms. Below, we outline the advantages of using historical data, present competing theoretical mechanisms for how gender differences may emerge in self-concept, and provide a review of the literature.

* Corresponding author.
E-mail address: philip.parker@acu.edu.au (P.D. Parker).

https://doi.org/10.1016/j.learninstruc.2017.09.002
0959-4752/

1.1. The use of historical data

There is a considerable literature on gender differences in self-concept (see Hyde, 2014 for a review). However, little research has specifically focused on gender differences in self-concept controlling for academic achievement (i.e., the portion of self-concept differences that would seem incompatible with objective reality). Although some research on this exists (e.g., Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002), this mostly represents single cohort longitudinal or cross-sectional research that is ill placed to determine the mechanisms behind such gender differences. This is partly due to a lack of explanation of the various competing theoretical arguments, which we specify in detail here. Even so, showing a gender difference from a single sample in a survey design rarely provides sufficient information to choose between competing mechanisms even if various mechanisms are specified. In order to do this, several options are available to researchers, each with their own strengths and weaknesses.

The first method of comparing mechanisms is experimental with the random assignment of participants. This method is often seen as the gold standard due to its unrivalled ability to provide evidence of causation. However, experimental research in this area often suffers from low power and a lack of external validity (Flore & Wicherts, 2015; Ganley, Mingle, Ryan, Ryan & Vasilyeva, 2013). A second approach is the use of comparative data from multiple countries (e.g., Charles, Harr, Cech, & Hendley, 2014; Else-Quest, Hyde, & Linn, 2010; Mann & Di Prete, 2016; Skaalvik, 1990; Stoet, Bailey, Moore, & Geary, 2016). Variation in achievement or gender inequality indexes or other factors at the country level can be correlated with gender differences in self-concept and these results compared against what would be expected on the basis of different theoretical mechanisms. However, countries often differ from each other in a vast number of ways (e.g., response set differences, latent cultural differences, etc.) that make it difficult to determine what factors are driving variation in results.

Finally, historical data provides evidence for or against competing mechanisms (e.g., Schoon, 2006). While not so common in educational psychology, this approach is often undertaken in life course studies (e.g., Bynner, 2016; *cf.*; Cimpian, Lubienski, Timmer, Makowski, & Miller, 2016). Its advantage is that cohorts are inherently ordered in time allowing for the construction of trajectories of historical change in multiple relevant variables. This is particularly useful where several counteracting mechanisms may be at work. In particular, where a single set of results may seem to favour only one mechanism, careful attention to historical changes can reveal that such results may actually be due to a mix of competing mechanisms of differing strength. This is particularly the case where historical trends include notable changes in context that can be used as a natural experiment (Bronfenbrenner, 1979). As noted above, one of the most notable changes in the educational context has been the increasing performance of women. Thus research, such as the present, can provide an indication of how self-concept is likely to change in response to an intervention that targets closing any remaining gender gaps in achievement.

Furthermore, the presentation of historical trajectories alone often represents a useful scientific endeavour above and beyond what it may reveal about different mechanisms (see Goldthorpe, 2016). Historical data over a moderate time frame also holds constant a number of often-unmeasured variables that can reduce confidence in multi-country studies. The disadvantage is that researchers must rely on data collected by others and thus have little control over the measures, populations, or time periods covered (see Elder, Pavalko, & Clipp, 1993). Additionally, variation across short historical periods is likely to be smaller than variation across countries.

All three approaches (experimental, multi-country, and historical) have complementary strengths and weaknesses, and all are certainly stronger than one-shot cross-section studies. We contribute to the advancement of research in this area by focusing on historical data. This is of particular relevance given that one of the defining features of recent history has been the dramatic rise of female academic achievement (Goldin, 2014; OECD, 2015). Yet little research has mapped the rise in achievement with change in self-concept, and certainly no research has done so with comparable databases explored over a decade. We aim to present evidence of trajectories in gender differences in academic achievement, self-concept, and self-concept controlling for achievement in math, English, and general academic domains. Below, we outline several theories on what such trajectories might look like if particular mechanisms were in operation.

1.2. Self-concept theories

Self-concepts are of interest in multiple fields of the social sciences, each with different approaches, and each, based on the number and strength of the assumptions that they hold, more or less likely to be true a-priori. We discuss these theories from simplest (fewest assumptions) to most complex.

1.2.1. Rational action theories of self-concept

The simplest models are those that suppose that individuals are rational actors (Little, 2012). This economic based approach has only three assumptions: 1) individuals have stable preferences, 2) individuals aim to maximise their utility with respect to those preferences, and 3) individuals do so under resource and/or information constraints (Becker, 1974). From these three assumptions a compelling model of self-concept can be built. We start with the assumption that people have a stable preference for accurately knowing their position relative to others (i.e., ability self-concept) and seek to maximise the accuracy of their self-concept by forming them with the best available information. However, accurate and objective information can be difficult and costly to obtain. As such, individuals will only seek to maximise the objectivity of their self-concept as long as the cost of gaining access to more objective information outweighs the benefits received from the increased accuracy of their self-concept.

The clearest application of this approach to self-concept is by the sociologist John Goldthorpe (2007; Breen & Goldthorpe, 1997) who saw self-concept as a function of achievement and thus group differences by social class or gender as purely a function of differences in underlying achievement distributions. Any group differences beyond achievement were thought to be ephemeral in nature and would quickly be resolved by students continuing to receive feedback by way of additional test scores and other information over the course of their school careers (Goldthorpe, 2007). Such rational choice theories are often seen sceptically within psychology, however, it must be noted that they do have the benefit of being self-contained explanations and of having relatively few assumptions (i.e., that people do what is best for them with the resources that they have to hand is its own explanation) (see Becker, 1974; Goldthorpe, 2007; Little, 2012). For our purposes, this theoretical approach would hypothesise that, after controlling for academic ability, there would be no difference in self-concept by gender, or at least no systematic difference, and that this would remain the case regardless of how large gender differences in achievement became over time, or which gender such achievement differences favoured.

1.2.2. Contrast theories of self-concept

A more modest proposal (which decreases the requirement for humans to act as pure rational agents but nevertheless increases the complexity of relevant models) is that individuals act according to bounded rationality (Simon, 1972) or operate according to what Gigerenzer (2008) calls rationality for mortals (i.e., allow for cognitive shortcuts). In line with this perspective are early psychological models that use surprisingly similar definitions of self-concept to the rational choice theory we explored above. Like the above definition, which highlights a stable preference for self-concepts that match objective reality, Festinger (1954, p. 117) notes that "there exists, in the human organism, a drive to evaluate his opinions and his abilities", and that the drive to evaluate abilities causes individuals to engage in behaviour that attempts to maximise the accuracy of their self-concept. Individuals do this because the costs of inaccuracy can be exceedingly high. While there are striking similarities in self-concept definitions from economics and psychology, they differ in one crucial aspect. While Becker (1974) explicitly denies systematic biases in the way individuals seek information and process it, Festinger (1954) hypothesises that, in a world of imperfect information, people assess their ability by comparing themselves to others. Individuals do not, however, tend to compare themselves to just anyone, or even to those that would give them the best information. Rather, they have certain biases in who they choose based on location (the closer to them the better) and salience (the more similar to them the better). This leads to a focus on contrast processes within local salient frames-of-reference (Festinger, 1954). Contrast processes, like the big-fish-little-pond effects, appear to be surprisingly resilient to individual level psychological mechanisms, which suggests that they are likely universal (Marsh, 2006). Nevertheless, new research shows that contrast effects do appear to respond to macro-contexts (see Parker et al., 2017; Salchegger, 2016).

This focus on macro-context has lead Parker et al. (2017) to integrate both the economic and psychological models into a single model referred to as the Information Distortion Model. This model focuses on how the use of local and salient frames-of-reference can distort the information about ability that individuals use to develop their self-concept. The primary hypothesis of this theory is that group differences in self-concept controlling for ability will be opposite in direction to group differences in achievement. The rationale for this is that the group with higher achievement has a more competitive frame-of-reference and thus receives distorted information on their underlying abilities that places *downward* pressure on self-concept. The group with lower achievement has a less competitive frame-of-reference and therefore receives distorted information on their ability that puts *upward* pressure on their self-concept. The information distortion model was developed in relation to social class, however, it is very similar to a key tenant of the Self-Protective Properties of Stigma theory by Crocker and Major (1989) that was developed in relation to gender, race, and other minority or marginalised groups. This theory aims to explain why stigmatised groups do not tend to suffer from lower self-concept than non-stigmatised groups and may even have more positive views of themselves. Crocker and Major (1989) suggest that individuals in stigmatised groups are likely to compare their abilities and attributes with in-group rather than out-group peers. Mapping this to historical trends, we would expect that whenever gender differences in achievement favour one gender, a corresponding self-concept difference controlling for ability would favour the other gender. Furthermore, as gender differences in achievement grow or shrink over time, we should observe counteracting changes in the strength of such differences in self-concept.

1.2.3. Assimilation theories of self-concept

A more complex model, but one which may provide a better account of the extant literature (see below), is based on assimilation theories of self-concept. Whenever research refers to stereotypes that are internalised by a given group of individuals, such references usually explicitly or implicitly reflect assimilation assumptions about self-concept. The best-formulated version of this theory is Akerlof and Kranton's (2010) Identity Economics. This theory states that individuals are constantly monitoring their relative distance from the core of a given set of groups. Large distances from the core of at least one group is emotionally costly and thus individuals aim to reduce that distance wherever possible. Thus, individuals will seek to alter their beliefs and behaviours to match stereotypical perceptions about the group they perceive to be closest to or to which they perceive they are most likely to belong. Mapping this theory to historical trends, we would expect gender differences in self-concept to follow prevailing stereotypes. Further, given that stereotypes are surprisingly resistant to change (Haines, Deaux, & Loaro, 2016), gender differences in self-concept should be largely uncorrelated with changes in gender differences in achievement. Thus while women have rapidly closed the gap to men in academic achievement and human capital, assimilation models would hypothesise that changes in self-concept would be much slower. The following section provides an account of the literature and the degree to which it favours these different theoretical mechanisms.

1.3. Literature review

It must be stated upfront that gender differences in achievement and self-concept tend to be modest. While such differences often persist across studies, are useful in exploring the role of various underlying psychological processes, and, when accumulated over time, have significant effects on attainment outcomes, the defining feature of research in this area is that boys and girls are more similar than they are different (Hyde, 2014). Nevertheless, we now review the relevant literature and its implications for various elements of the theories reviewed above. Firstly, there is now extensive research showing gender differences in various math self-concept variables favouring boys (Hyde, 2014; Van Zanden, Marsh, Seaton, & Parker, 2015) and literacy self-concept favouring girls (Marsh & Yeung, 1998; Van Zanden et al., 2015). Importantly, these differences are in the same direction as gender differences in achievement (Falch, Torberg, & Naper, 2013; Hyde & Mertz, 2009; Lietz & Petra, 2006; Machin & McNally, 2005; Machin & Pekkarinen, 2008; Marsh & Yeung, 1998; Reilly, Neumann, & Andrews, 2015; Voyer & Voyer, 2014) as would be hypothesised by rational choice and assimilation theories. Interestingly, past research showing these patterns has explicitly found support for rational choice contentions and denied the likelihood that gender assimilation processes were in operation. Skaalvik (1990) found in two countries that general academic, math, and English self-concept gender differences were only as large as what would be expected on the basis of academic ability. It should be noted that research has also suggested that gender differences are not equally distributed across the achievement distribution, with several sets of results suggesting that gender differences are larger for higher performing students and that this has been somewhat resistant to change over successive cohorts (see Cimpian et al., 2016; Reilly et al., 2015).

However, the bulk of the literature shows that these differences in self-concept by gender remain when controlling for academic achievement; albeit they narrow considerably (Jacobs et al., 2002; Parker, Nargy, Trautwein, & Lüdtke, 2014; Wang & Degol, 2013). Furthermore, there is also some evidence that girls have lower general academic self-concept than boys, though such differences tend to be small (Marsh, 1993). This is despite girls out-performing boys in general academic achievement (OECD, 2015). This indicates that pure rational choice models of self-concept are unlikely.

As noted above, research in favour of contrast models is strongest in relation to general academic self-concept in which boys have higher self-concept than girls despite having lower achievement. In relation to the assumptions of contrast theories, previous research had supported the supposition that people have a preference for comparing themselves to those closest to them. Referred to as the local dominance effect, experimental research has shown that social comparison processes are stronger for frames-of-reference that are most proximal (Zell & Alicke, 2010). Likewise, large-scale survey research has found that relative position within class has a larger effect on self-concept than relative position within school or country (Marsh, Kuyper, Morin, Parker, & Seaton, 2014). Moving from locality to salience, in one large study when students were asked to select a reference target to compare their grades against, over 80% selected a member of the same gender (Huguet et al., 2009). Similar findings have been observed in workplace evaluations (Crocker & Major, 1989; McKinsey & Company, 2015). Further, individuals have a strong preference for using same-sex comparisons, even when their gender is unrelated to outcomes in the domain of interest (Miller, 1984). This suggests that the necessary conditions are present for processes related to contrast to be in operation. Research on self-concept has persistently shown the power of contrast processes to shape self-concept as evidenced by decades of research on frog-pond (Davis, 1966) and big-fish-little-pond effects (Marsh, 2006). Furthermore, previous research comparing assimilation and contrast mechanisms in predicting self-concept suggests that assimilation effects are smaller in size and duration than contrast effects (Marsh, Köller, & Baumert, 2001). However, little research has considered whether contrast processes are in operation to explain gender differences in self-concept.

Is there any evidence that such processes are indeed in operation? A particularly striking example are school systems that educate boys and girls separately, thereby ensuring gender-specific frames-of-reference. Marsh, Abduljabbar, et al. (2014) explored such a system in Saudi Arabia finding that, while girls significantly outperformed their male peers in math, they did not differ on self-concept. Using the statistical information from the Marsh, Abduljabbar, et al. (2014) paper, we estimated the gender difference in residual self-concept (i.e., controlling for achievement) and found results consistent with contrast effects. Boys' math self-concept was a tenth of a SD higher than girls despite performing objectively worse (Cohen's $d = 0.108$, SE $= 0.017$, $p < 0.001$; see the R script in supplementary material). However, using data from the same paper, we also calculated the residual self-concept advantage in boys from the United States who did not significantly differ from girls in math achievement. In this case boys still had higher self-concept when controlling for achievement (Cohen's $d = 0.118$, SE $= 0.014$, $p < 0.001$). This contradiction suggests that contrast theories alone are likely insufficient and that assimilation processes must also be considered.

Thus, the question is whether the necessary conditions are present for assimilation to take place. As noted above, assimilation theories have the most assumptions. This is because they require individuals not only to know their relative position within their own group but also the relative position of their group compared to other groups. For stereotypes the assumptions become even stronger requiring individuals to be aware of general, often abstract, and often in contradiction to objective reality, opinions of their group. If stereotypes are to influence self-concept, individuals must also internalise these opinions and to some extent endorse them as true. Thus, there needs to be evidence that a strong persistent socialisation environment acculturating these views is present.

Girls and boys are repeatedly told from a young age how girls and boys should act and think and parental attitudes toward their children's interests and abilities have been shown to fall along gendered lines (Eccles, Jacobs, & Harold, 1990). The stereotypical idea that boys are better at math and girls are better at literacy is ingrained in children at an extremely young age (Lummis, Max & Stevenson, 1990; Tiedemann, 2000). This suggests the normative forces are present in sufficient force for assimilation mechanisms to explain gender differences in self-concept. Further, the size and direction of gender differences noted above, particularly for the math domain, suggest that gender assimilation/stereotype processes do affect self-concept (e.g. Jacobs et al., 2002; Wang & Degol, 2013).

Nevertheless, assimilation and contrast processes are not mutually exclusive and both could be in operation with the direction of gender differences indicating the relative strength of the two processes. Whether one or both processes are in operation is impossible to determine from single population cross-sectional studies. However, by closely monitoring change in gender differences in self-concept with respect to change in gender difference in achievement, it may be possible to determine if both mechanisms are in operation. For example, suppose we found, as so much previous research previous has, that girls have lower math self-concept controlling for achievement and lower math achievement than boys. This would indicate the presence of assimilation effects. However, suppose that, from one birth cohort to the next, girls narrowed the achievement gap to boys in math and this corresponded with an increase in the gender gap in self-concept favouring boys. This would be compelling evidence that, in addition to assimilation effects, contrast effects were also in operation, albeit to a weaker extent.

2. Current research

Using historical data from multiple age cohorts evaluated with similar measures of achievement and self-concept, we seek to overcome this limitation in a number of ways. First, the research aims to show historical changes in gender differences in achievement, self-concept, self-concept controlling for achievement in math, literacy, and general academic domains. Given that previous research (e.g., Reilly et al., 2015) has suggested that gender differences in achievement may differ in strength at different points in the achievement distributions, we also consider gender differences in these variables for poor, moderate, and excellent performing students. Second, by paying careful attention to the association in changes in these variables over time, we consider whether rational choice, contrast, assimilation processes, or some combination of these are in operation to explain gender differences in self-concept. As noted above, existing research exploring gender differences in self-concept is often ill suited to consider such a question, particularly when both assimilation and contrast mechanisms may simultaneously occur. Table 1 summarises how different patterns in gender difference trajectories for self-concept and achievement provide evidence for different mechanisms or set of mechanisms.

3. Methodology

3.1. Data sources and participants

The current research uses all cohorts of the Longitudinal Study of Australian Youth (LSAY; $N = 66{,}522$). The LSAY consists of five distinct databases, each focused on a particular age cohort. The databases in LSAY are complex and consist of two different designs. The early cohorts (modal birth years 1981 and 1984) were

Table 1
Hypothesised differences in self-concepts.

Case Number	Trends	Supports
1	All gender differences in self-concept are zero (or non-systematic) once accounting for achievement.	Rational Choice Theory
2	When achievement differences favour one gender, self-concept differences will favour the other gender. As the former gets larger, so too will the latter.	Contrast
3	Gender differences in self-concept controlling for achievement follow stereotypical directions and are largely unresponsive to changes in gender differences in achievement.	Assimilation/Stereotype
4	Similar to case 3 but, where achievement gender gaps get smaller, self-concept gender gaps controlling for achievement get larger.	Assimilation/Stereotype with Contrast

nationally representative samples of year 9 students, the vast majority of which were aged 14 at the time of testing. The latter cohorts (modal birth years 1987, 1990, and 1993) represented extensions of the Programme for International Student Assessment (PISA) in 2003, 2006, and 2009 and, as such, were nationally representative samples of 15 year olds. These cohorts were thus slightly older than the average sample in the previous cohorts at 14. However, the pre-PISA cohorts had much greater variation in age. As such, it was critical that we controlled for age in the analysis. Descriptives of the different cohorts can be found in Table 2.

All cohorts utilised a complex design rather than a purely random sample. In all cases the data were nested (students nested within schools) and sampling by strata was undertaken with the primary sampling unit being the school. In addition, oversampling of underrepresented groups (either by state or by particular demographic) was utilised. To correct for this, sample weights and cluster robust standard errors were used for all analyses (see below for details).

3.2. Materials

3.2.1. Academic achievement

Both the PISA based and pre-PISA LSAY cohorts included standardised numeracy and literacy tests. The pre-PISA cohorts of the LSAY included a 20-item tests of each domain with scores given as counts of correct responses. Thus, possible scores ranged from 0 to 20. The final LSAY cohorts took achievement scores from the PISA achievement tests. The PISA tests were matrix sampled and thus true score performance was taken from an item response theory model scaled to have an international average of 500 and a standard deviation of 100. Each individual received five plausible values drawn from their estimated underlying achievement distribution. Analysis was undertaken with each of the plausible values and were integrated (we use the survey package in R for this process; Lumley, 2011).

In the current research all performance measures were within cohort standardised to have a cohort mean of 0 and a standard deviation of 1. For all cohorts we used math and literacy achievement scores when considering math and literacy social comparison respectively and the unweighted average of the individual's math and literacy scores when considering general academic social comparison. The later LSAY cohorts had measures of science achievement; however, in order to harmonise data across all cohorts, only math and literacy scores were used.

3.2.2. Self-concept

All cohorts of both studies had 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 five-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 literacy (“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 the social comparison component of self-concept (see Marsh et al., 2014). This is a particular advantage in the current case as the aspect of self-concept we were interested in was this very component. This allowed us to focus our attention on this process while being largely unaffected by the dimensional or temporal comparison that are encapsulated in broader self-concept measures.

3.2.3. Age

The PISA cohorts of LSAY were age cohorts and thus all children were 15 at some point during the year of testing. However, the pre-PISA cohorts of LSAY were representative samples of year 9 students where the vast majority of participants were 14 at the time of testing, but ages ranged from as low as 13 to as high as 18 (see Table 2). It was thus critical to control for age. In all models, age was entered as a standardised variable based on a weighted (by sample size) mean of the weighted (utilizing each cohorts sample weights) means and standard deviations of age in each cohort. This corresponded to a mean of 15.22 years and a standard deviation of 0.36.

Table 2
Basic demographics.

	LSAY Cohort Name				
	1995	1998	2003	2006	2009
Modal Birth Year	1981	1984	1987	1990	1993
N	13613	14118	10370	14170	14251
Age (SD)	14.49(0.47)	14.62(0.44)	15.69(0.29)	15.70(0.28)	15.69(0.29)
Age Range^	13–17	13–18	15–16	15–16	15–16
% Male	48.87	49.27	50.82	51.14	48.91

Notes. LSAY = Longitudinal Study of Australian Youth. SD = Standard Deviation. The PISA LSAY cohorts are representative of individuals who were aged 15 for at least part of the calendar year. The pre-PISA cohorts were representative of Australian school grade 9 and thus have a much wider age range.

3.3. Analysis

As noted above, all cohorts had a complex design. Further, the pre-PISA and PISA cohorts took a different approach to accounting for this complex design. For the pre-PISA cohorts, weights were provided to account for oversampling of particular groups and stratified sampling. In this case we used these samples plus strata weights along with cluster robust standard errors for school. For the PISA wave population, sample weights were provided. In keeping with the pre-PISA waves, we used sample plus attrition weights (see below) to account for oversampling of some groups. Strata and school nesting were accounted for by the provided Fay weights with a rho of 0.5. As such, all standard errors accounted for the dominant clustering levels. In all cases analysis was undertaken using the survey package in R (Lumley, 2011). Missing data was negligible for the pre-PISA waves (never more that 3.5% for a given item). Not all of the PISA participants completed the social comparison items in the PISA cohorts of LSAY. To account for this we utilised the attrition weights provided by the LSAY organisers.

The major outcome variable for the current paper was a single item measuring social comparison on a five-point Likert scale. Thus, these items were not on a continuous or near continuous scale. As such, we treated these variables as ordinal and used proportional odds logistic regression. Results are largely reported in log-odds with their associated standard errors. In addition to log-odds, we also graphed the predicted probabilities of responding *very well* when asked about how a participant compared themselves to the rest of their class. All predicted probabilities were evaluated at the mean age across all cohorts and, where applicable, average levels of achievement. In addition, we also considered gender differences in academic achievement. In these cases ordinary least squares were used and results were reported as gender differences in standard deviation units of achievement.

Taken together, the models predicting social comparison take the form:

$$y^* = \beta_{male} + \beta_{age} \quad (1)$$

or

$$y^* = \beta_{male} + \beta_{age} + \beta_{achievement} \quad (2)$$

For equations (1) and (2), a series of thresholds μ_k, where $k = 4$, were used to capture the probability of transitioning from level k−1 (e.g., better than average) to level k (very well) in the outcome variable such that:

$$y = \begin{cases} 0 & if \quad y^* \le \mu_1, \\ 1 & if \quad mu_1 < y^* < mu_2 \\ 2 & if \quad mu_2 < y^* < mu_3 \\ 3 & if \quad mu_3 < y^* < mu_4 \\ 4 & if \quad \mu_4 < y^* \end{cases} \quad (3)$$

In addition to the proportional odds models predicting social comparison, we also explored gender differences in academic achievement in Australian adolescents based on:

$$y_{achievement} = \alpha + \beta_{male} + \beta_{age} \quad (4)$$

There have been claims in the literature that gender gaps in achievement (and particularly math achievement) differ at different points along the achievement distribution. Thus, to provide a more nuanced exploration of gender differences in achievement, we applied quantile regression in order to consider gender gaps at the 0.25, 0.50 (median), and 0.75 quantiles representing poor, average, and excellent performers respectively. This required only a small change to equation (4):

$$y_{achievement}(\tau|x) = \alpha + \beta_{male} + \beta_{age} \quad (5)$$

Here we looked at the τ-quantile of achievement conditioned on our predictors where τ is either the 0.25, 0.50, or 0.75 quantile (see Kleiber & Zeileis, 2008). Again, we incorporated sample weights and used cluster robust standard errors.

4. Results

Given the different weighting procedures, all models (i.e., equations (1), (2), (4), and (5)) were run for each cohort and for general academic, math, and literacy separately. We focused mainly on a graphical display of the resulting parameters of interest and associated confidence intervals in order to focus attention on historical patterns. Full model results in table form can be found in the Appendix.

4.1. General academic

We first considered how gender differences in general academic ability have changed over the decade and a half period under investigation. Our results suggested that for the years under study gender differences for adolescents in Australia increasingly favoured females; moving from almost zero in effect size and non-significant to significant and equalling over a tenth of a SD. Given that previous research has stated that gender differences may be different at different points in the achievement distribution, we also used quantile regression to explore historical trends in gender differences at the 0.25, 0.50, and 0.75 quantiles (see Table 3a). The general trend was replicated here; however, there was relatively few gender differences for brighter children (though point estimates moved from favouring boys to favouring girls). For children in the left tail of the distribution there was an upward trend in gender differences favouring girls. Finally, gender differences at the median went from non-significant to significant favouring girls.

For general academic self-concept, gender differences fluctuated somewhat but were always higher for boys than for girls. We reran these models controlling for achievement and there was a small upward trend favouring boys over girls. When the trends for self-concept and achievement were considered together, results were consistent with what would be expected on the basis of self-concept theories associated with gender specific contrast mechanisms (see Fig. 1). The predicted probabilities of an individual of average age and average achievement reporting that they perceived

Table 3a
Quantile regression for general Academic achievement.

	Birth Cohort				
	1981	1984	1987	1991	1993
0.25					
Intercept	−0.80(0.05)	−0.69(0.10)	−0.71(0.05)	−0.62(0.04)	−0.61(0.04)
Male	−0.12(0.05)	−0.16(0.08)	−0.27(0.05)	−0.21(0.04)	−0.22(0.04)
Age	−0.08(0.02)	−0.07(0.02)	0.10(0.03)	0.09(0.02)	0.10(0.02)
Median					
Intercept	−0.06 (0.07)	0.04(0.10)	−0.04(0.04)	0.02(0.04)	0.04(0.04)
Male	−0.01(0.05)	−0.10(0.09)	−0.17(0.04)	−0.12(0.04)	−0.13(0.04)
Age	−0.07(0.02)	−0.06(0.03)	0.10(0.02)	0.09(0.02)	0.10(0.02)
0.75					
Intercept	0.63(0.05)	0.64(0.08)	0.57(0.04)	0.61(0.04)	0.65(0.04)
Male	0.09(0.06)	0.12(0.11)	−0.08(0.05)	−0.04(0.05)	−0.07(0.04)
Age	−0.03(0.02)	−0.05(0.02)	0.10(0.02)	0.10(0.02)	0.10(0.02)

Notes. Standard errors are presented in brackets.

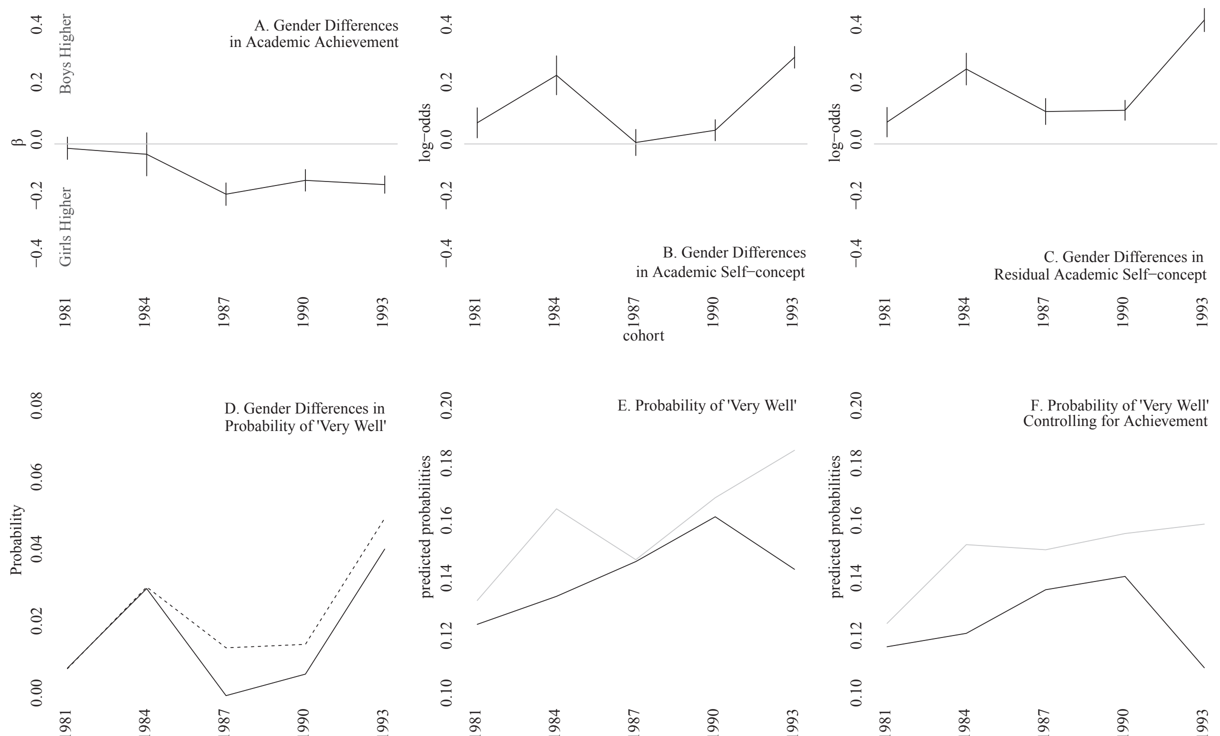

Fig. 1. Gender difference in general academic achievement and self-concept.

Notes. Panels from left to right, top to bottom, A. gives the standardised beta coefficients for gender difference, B. gives the log-odds gender difference controlling for age, C. gives the log-odds gender difference conditioned on age and achievement, D. represents the difference in predicted probabilities controlling for age (solid line) and controlling for age and achievement (broken line), E. represents the predicted probabilities for girls (in black) and boys (in grey) at the mean of age, and F. provides the same information as E but at the mean of age and achievement. The 95% confidence intervals are provided for the beta and log-odds coefficients.

themselves as performing *very well* in comparison to their peers is presented in the second row of Fig. 1. This shows a gap opening up in favour of boys over the five cohorts despite the increasing performance advantage of girls. The 1984 and 1993 cohorts represent a particular social comparison advantage for boys and these correspond with the cohorts in which girls have the greatest advantage in terms of achievement. It is worth emphasising that the predicted probabilities in panel *f* of Fig. 1 are presented for boys and girls with the same overall level of achievement. As will be discussed below, the 1984 cohort represents a particular outlier in the overall linear trend.

4.2. Math

Gender differences in math achievement were relatively stable across the period of study, though a slight trend toward the gender gap closing can be observed. Consistent with previous literature, quantile regression indicated that gender differences in achievement were absent in the left tail of the distribution, were small and occasionally significant at the median, and were larger for better performing children (see Table 3b). In terms of trends, the downward trend observed in OLS regression was most obvious for the highly able children as it halved in size.

Math self-concept tended to be relatively stable across the period of study with no notable trend. When controlling for achievement there was likewise little evidence of change in self-concept differences once controlling for achievement. Gender gaps still favoured boys across cohorts to relatively the same degree. Mapping this against the achievement differences we can see that, as achievement gaps declined over time, self-concept differences tended to remain fairly stable (see Fig. 2). This suggests that a stereotype only model was a relatively good fit to this data.

Table 3b
Quantile regression for math achievement.

	Birth Cohort				
	1981	1984	1987	1991	1993
0.25					
Intercept	−0.90(0.10)	−0.81(0.10)	−0.86(0.05)	−0.78(0.04)	−0.76(0.04)
Male	0.11(0.07)	0.02(0.07)	−0.00(0.04)	0.10(0.04)	0.08(0.05)
Age	−0.06(0.03)	−0.07(0.03)	0.11(0.03)	0.08(0.02)	0.10(0.02)
Median					
Intercept	−0.11(0.04)	−0.04(0.10)	−0.20(0.05)	−0.16(0.04)	−0.10(0.04)
Male	0.28(0.10)	0.06(0.11)	0.08(0.05)	0.16(0.04)	0.12(0.04)
Age	0.00(0.03)	0.00(0.03)	0.11(0.03)	0.10(0.02)	0.10(0.02)
0.75					
Intercept	0.41(0.09)	0.51(0.13)	0.46(0.05)	0.45(0.04)	0.54(0.05)
Male	0.29(0.12)	0.25(0.13)	0.13(0.05)	0.23(0.05)	0.16(0.03)
Age	−0.07(0.03)	−0.05(0.03)	0.11(0.02)	0.12(0.02)	0.09(0.02)

Notes. Standard errors are presented in brackets.

4.3. Literacy

Literacy ability showed the biggest increase in gender gap of all three domains increasing from girls having a 0.20 of an SD to almost 0.40 of an SD advantage over boys. Quantile regression suggested that there was no trend at the median with a persistent 0.30 of an SD difference favouring girls (see Table 3c). At the left tail of the achievement distribution the gender gap increased by about half from the 1981 cohort to the 1993 cohort. The largest increase was

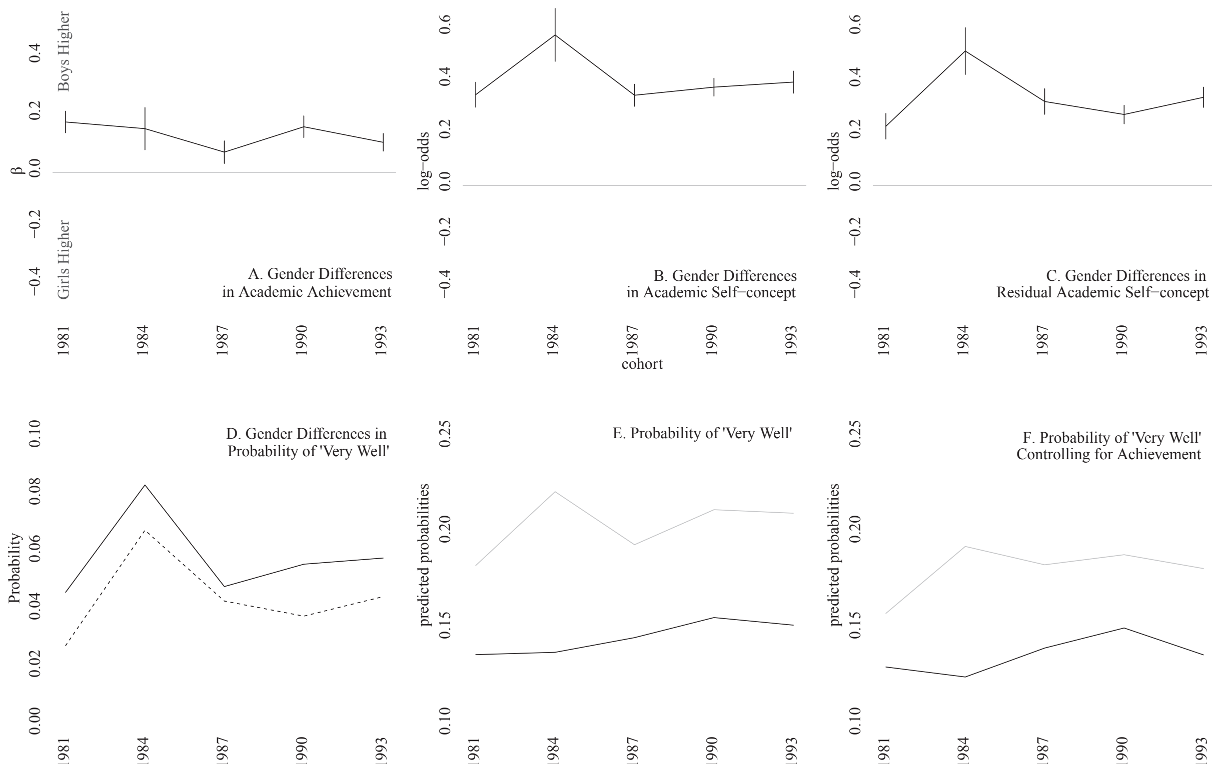

Fig. 2. Gender difference in math achievement and self-concept.
Notes. From left to right, top to bottom, panel A. gives the standardised beta coefficients for gender difference, B. gives the log-odds gender difference controlling for age, C. gives the log-odds gender difference conditioned on age and achievement, D. represents the difference in predicted probabilities controlling for age (solid line) and controlling for age and achievement (broken line), E. represents the predicted probabilities for girls (in black) and boys (in grey) at the mean of age, and F. provides the same information as E but at the mean of age and achievement. Standard error bars are provided for the beta and log-odds coefficients.

Table 3c
Quantile regression for literacy achievement.

	Birth Cohort				
	1981	1984	1987	1991	1993
0.25					
Intercept	−0.75(0.07)	−0.61(0.10)	−0.54(0.04)	−0.45(0.04)	−0.46(0.04)
Male	−0.29(0.05)	−0.34(0.07)	−0.48(0.05)	−0.46(0.04)	−0.45(0.04)
Age	−0.10(0.03)	−0.09(0.03)	0.08(0.02)	0.08(0.02)	0.09(0.02)
Median					
Intercept	0.21(0.05)	0.35(0.12)	0.11(0.04)	0.17(0.03)	0.16(0.04)
Male	−0.26(0.09)	−0.27(0.08)	−0.39(0.05)	−0.38(0.040)	−0.35(0.04)
Age	0.00(0.03)	0.00(0.03)	0.08(0.02)	0.090(0.02)	0.09(0.02)
0.75					
Intercept	0.73(0.08)	0.89(0.05)	0.70(0.04)	0.74(0.04)	0.76(0.04)
Male	0.00(0.09)	0.00(0.12)	−0.30(0.04)	−0.29(0.04)	−0.29(0.04)
Age	0.00(0.00)	0.00(0.02)	0.07(0.02)	0.09(0.02)	0.09(0.02)

Notes. Standard errors are presented in brackets.

for highly able students where there were no gender differences observed for early cohorts but this difference jumped to 0.30 of an SD for the later cohorts. Gender differences in literacy self-concept showed little evidence of trends from earlier to later cohorts. However, controlling for achievement, there was a general downward trend in the self-concept advantage for girls across the five cohorts. Interestingly, this decreasing in self-concept advantage for girls contrasted with girls increasing advantage in literacy achievement (see Fig. 3). Taken together, the results for math social comparison support dominant assimilation plus weaker contrast mechanisms. Again, the 1984 cohort represented a notable outlier.

5. Discussion

The current research used standardised achievement data and self-concept from representative samples of five Australian cohorts to explore historical trends in gender differences in these variables and what mechanisms might explain them. We considered rational choice theories, which suggested there will be no relationship between gender and self-concept after accounting for achievement. We also considered a contrast hypothesis, which stated that gender differences in self-concept controlling for achievement should favour the opposite gender to that which had higher achievement.

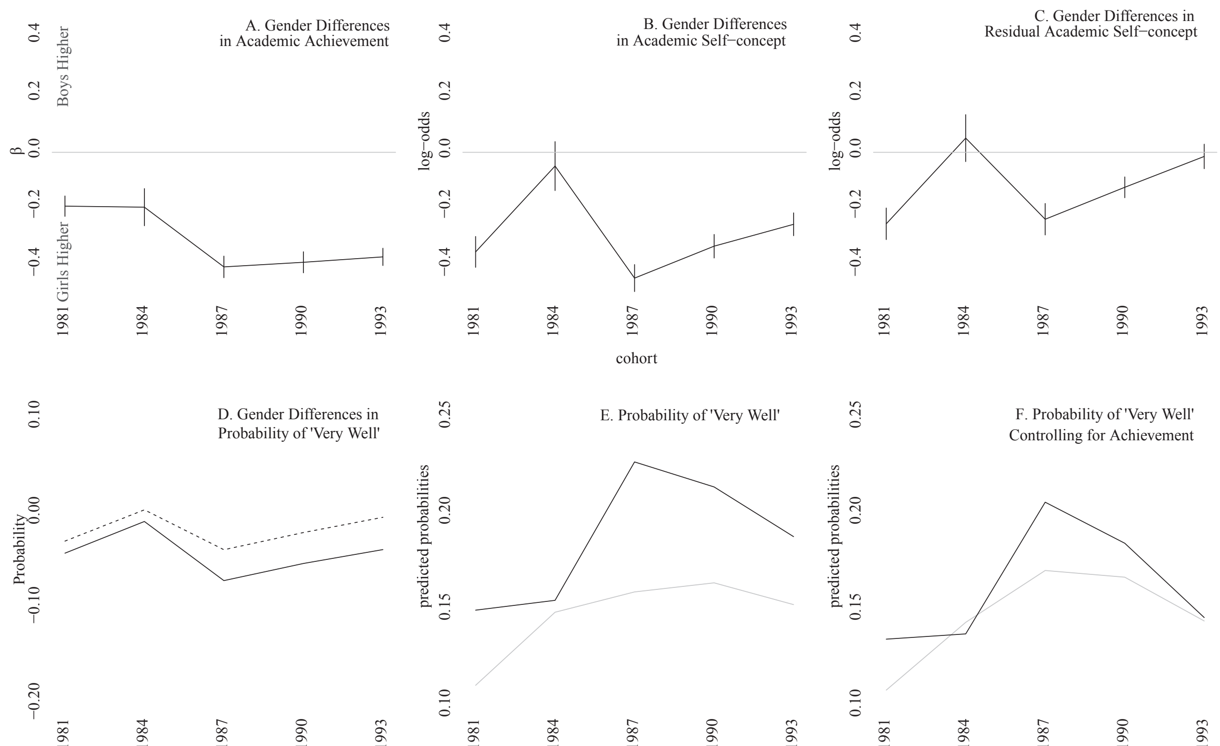

Fig. 3. Gender difference in literacy achievement and self-concept.
Notes. From left to right, top to bottom, panel A. gives the standardised beta coefficients for gender difference, B. gives the log-odds gender difference controlling for age, C. gives the log-odds gender difference conditioned on age and achievement, D. represents the difference in predicted probabilities controlling for age (solid line) and controlling for age and achievement (broken line), E. represents the predicted probabilities for girls (in black) and boys (in grey) at the mean of age, and F. provides the same information as E but at the mean of age and achievement. Standard error bars are provided for the beta and log-odds coefficients.

Assimilation hypotheses suggest that gender differences in self-concept controlling for achievement would be in stereotypical directions and would be relatively immune to change in achievement gender differences. Finally, we considered that contrast and assimilation processes might both be in operation. Below, we discuss the historical trends we observed before considering each theory in turn and the degree to which the evidence we provide is inline with these perspectives.

5.1. Historical trends

Goldthorpe (2016) suggests that one of the key goals of the social sciences is to accurately describe social phenomena. This includes not only determining how large particular effects are but also the contexts under which they change. As noted above, recent history has seen a dramatic increase in the achievement of women (OECD, 2015). We also observed an increase in the performance of girls, relative to boys. For literacy, gender differences doubled in size from the first to the last cohort. For the general academic domain, gender difference went from non-significant to significant. Gender differences in math achievement have also declined in size. Of particular interest was when these effects were explored at difference points in the achievement distribution. For general academic achievement, increases in effects favouring girls tended to mainly occur for low and moderate performing students. For literacy achievement, gender differences particularly increased for better performing students. For math, girls notably tended to close the gap in achievement at the top rather than at the bottom of the distribution, most likely as gender differences at the bottom were already weak. This does not contradict previous research, which has shown that gender differences in math achievement are largest for high performing students (Reilly et al., 2015). Indeed, despite the decline, gender differences were consistently larger for better performing students than at any other point in the distribution. However, this does contradict research by Cimpian et al. (2016) who looked at younger children in the US. This may suggest that societal and/or developmental differences in achievement gaps are present at different positions across the distribution.

In comparison, gender differences in self-concept showed much more modest declines. This may be due to the smaller variability in these measures as self-concept was estimated using single 5-point Likert scale items. However, as we note below, the relatively slower pace of change in self-concept compared to achievement suggests that assimilation mechanisms are likely in operation. The exception to this was general academic self-concept in which gender differences notably increased favouring boys. This was in direct contrast to changes in achievement which increasingly favoured girls. As we argue below, this suggests the presence of contrast mechanisms. No trend supported a strict rational choice theory.

5.2. Rational choice theories

Across five cohorts for three self-concept variables we found evidence of significant gender differences in self-concept controlling for achievement in 13 of 15 cases. Further, there was no evidence of a consistent historical trend downward in these results. This is inconsistent with a strict rational choice perspective. For example, Goldthorpe (2007) suggested that as children receive more information about their ability their self-concept should become increasingly objective. However, the period under investigation is associated with a considerable increase in the degree of testing against national standards, particularly with the introduction of a comprehensive national testing scheme launched in 2008 (Lingard, 2010; Lobascher, 2011). This arguably means that children in the last cohort under investigation (1993 cohort) had access to higher quality information about their objective standing in the nation than any previous cohort as they were among the first to sit these tests. Further, as this information was provided to all students, it was easy to obtain. Rational choice theories would suggest that gender differences in self-concept, once achievement had been accounted for, should certainly diminish, if not evaporate completely, with the introduction of such high-quality information. Yet in both general academic and math self-concept the trend was toward increased differences. That we do not see a trend downward despite increased access to cheaper high-quality information about childrens' objective ability (at least at the national level) suggests that students must use some cognitive shortcuts when evaluating themselves. This indicates that assimilation and contrast processes in relation to local conditions may be in operation.

5.3. Assimilation theories

Evidence was much stronger for assimilation theories of self-concept. In particular, boys had higher math self-concept than equally able girls, while girls had significantly higher literacy self-concept than equally able boys. Both of these results are in the same direction as gender differences in achievement. These results point toward the important formative role of assimilation to stereotypes or perceived group norms as suggested in identity economics (Akerlof & Kranton, 2010). We have previously presented Goldthorpe's (2007) theory as not providing allowances for such perspectives. However, we should note that Goldthorpe primarily wrote in relation to social class and his argument was not that assimilation to group norms was not possible but that social classes tend to hold little normative power. In contrast, gendered norms are likely sufficiently crystallised and consistent within society to impact an individual's ways of viewing themselves and their behaviours (e.g., Mann, Legewie, & DiPrete, 2015; Parker et al., 2012). This would seem to at least be the case for math and literacy domains.

As noted in the literature review, cross-cultural research has shown that parents impart, and children take on, the view that boys are good at math and girls are good at literacy from a very young age (Lummis et al., 1990). As such, research in this area may be improved by considering identity economics in concert with traditional comparison processes when exploring how gendered self-concept emerges. Indeed, Akerlof and Kranton (2010) have explored how assimilation processes operate within an educational context, although they have largely focused on race and social class. Application of this theory to the study of gender in the classroom and to its intersection with other elements of identity will likely be beneficial given its level of sophistication, its systematic nature, and careful articulation of its premises. However, we should note one caveat; assimilation theory would suggest that gender differences in self-concept controlling for achievement should be relatively independent of the size of gender differences in achievement. In our findings, gender differences in self-concept tend to grow in the opposite direction as gender differences in achievement. This suggests the need to also consider contrast mechanisms.

5.4. Contrast models

The final theory under consideration was contrast models of self-concept. For general academic self-concept the results tended to support such a mechanism. As contrast theories would hypothesise, there was a negative correlation between the size of the achievement advantage for girls and the size of the self-concept gender differences when controlling for achievement. With only five data points one cannot read too much into a correlation; however, the result were consistent with contrast models ($r = -0.25$) and the relationship was much stronger when the outlier 1984 cohort was excluded ($r = -0.44$). More importantly, when there were no achievement gaps, self-concept gaps controlling for achievement were also not significant. Put simply, differences in self-concept in favour of boys only emerged when girls significantly outcompeted them. It is this counterintuitive hypothesis that stands at the heart of contrast theory.

The results for math and literacy self-concept were less compelling. Indeed, the differences were almost always in the same direction as gender differences in achievement, which is the exact opposite of what contrast models would hypothesise. This would suggest that gender differences in self-concept in these domains wouldn't be solely due to contrast mechanisms. Nevertheless, some of the historical trends in how these self-concepts changed over time may suggest that a more nuanced theory is needed.

5.5. Combined contrast and assimilation processes

As noted in the literature review, no research to our knowledge has considered whether assimilation and contrast mechanisms may both be in operation to explain gender differences in self-concept. This may be due to most research being poorly placed to differentiate assimilation only effects from a strong assimilation effect with a weaker contrast effect. As we argue above, careful attention to historical trends in the size of self-concept differences provides an avenue to distinguish between these competing views. Our results for math and literacy self-concept tended to support assimilation mechanisms from a global perspective. However, we note that for both math ($r = -0.13$; without 1984 cohort $r = -0.93$) and literacy, when excluding the 1984 cohort ($r = -0.52$), the correlation between gender differences in achievement and gender differences in self-concept controlling for achievement were negative. Again, we stress that these correlations should be considered with scepticism given the few data points. They nonetheless make the direction of differences that is apparent in the figures clear. Such results tend to lend credence to the view that assimilation processes primarily drive gender differences in these domains, but weaker yet notable contrast processes may also be in operation.

5.6. Implications

The findings in this research suggest that gender differences in self-concept might not be as easily explained by gender stereotypes as they may appear to be on the surface. It is likely that both assimilation and contrast mechanisms are in operation in many areas of academic self-concept and there is a need to develop better methods to distinguish the relative strengths of the two and the conditions under which one dominates the other. This is particularly the case where big-fish-little-pond effect research suggests that, in relation to school compositional effects, assimilation effects tend to be much weaker and short lived than contrast effects (Marsh, 2006). It was interesting to note that assimilation evidence was strongest in math and literacy in which we found a wealth of previous literature relating to stereotypes. For general academic self-concept we found relatively less literature, at least not for modern Western children, discussing gender stereotypes. Interestingly, this was the domain in which contrast mechanisms seemed to be most consistent with our results. Thus, a speculative hypothesis is that assimilation is strongest wherever clear, consistent, and crystalised normative processes are pronounced and, where they are relatively absent, contrast mechanisms dominate.

Much attention has been directed towards the problem of gender stereotypes and their potential negative impacts on later attainment (e.g., Eccles et al., 1990). However, it should also be noted that contrast mechanisms are likewise potentially problematic for children and society. As Covington and Beery (1976) note, for self-concept to be optimal and functional in directing future behaviour and choices, they should be as realistic as possible. Both assimilation and contrast mechanisms represent processes that result in students holding self-concepts that are inconsistent with high-quality objective information about them. Thus, continued consideration of how such processes work and under what conditions they can be mitigated, or where relevant, exploited, is an ongoing research concern.

5.7. Limitations

The historical method that we used holds many advantages over considering only one cohort alone. Nevertheless, there were limitations. Indeed, there are four caveats that should be placed on an interpretation of these results. First, the data we explore here covers both a long period of time, and uses consistent measures and multiple representative cohorts. Nevertheless, the main conclusions are based on results across five cohorts. Thus, results should be interpreted with caution until they can be replicated with other methods. Second, as noted in the literature review, an experimental approach that is able to directly manipulate aspects of the environment or a study comparing multiple countries or multiple contexts on comparable measures would also be beneficial. Given that our results suggest a mix of contrast and assimilation processes, experimental research that holds one of these mechanisms constant and manipulates the other would be useful. Third, it should be noted that, consistent with the meta-analyses of Hyde and colleagues (1988; 1990a), gender differences in achievement were either small or, at best, moderate. Likewise, differences in residual social comparison also tended to be small. This is not to suggest the results are unimportant. Indeed, Hyde and colleagues (1990b) note that even small differences can have large cumulative effects in guiding choices and decisions over time. However, the results should be interpreted in light of the extremely large overlap between boys and girls. Finally, in all cases the 1984 cohort represented a clear outlier in relation to social comparison, consistently favouring males. Careful examination of the data, documentation, and discussion with the individuals associated with the original LSAY data collection and management suggested that this was not a clerical error and, as such, we retain the results here. Given the existing information, the size of the anomalous findings for this cohort is difficult to explain. It may be that this effect reflects a period effect (i.e., some specific event that occurred in 1998 – the year in which this data was collected) rather than a cohort effect of which we are unaware.

6. Conclusion

The current study revealed that girls had lower general academic and math social comparison and that their advantage in literacy social comparison diminished over time. Such patterns remained, and in some cases strengthened, when considering equally able boys and girls. Paradoxically, as girls fell further behind boys in their self-concept, their relative advantage in achievement tended to increase. This inverse relationship between achievement gains and self-concept declines can be explained by contrast social comparison mechanisms. However, the results for math and literacy clearly indicated that assimilation social comparison processes play a larger role in these domains. Stereotypes about relative performance in math and literacy exist. However, they are unlikely to be the only mechanisms driving gender differences in self-concept. Taken together, models of gender differences in self-concept should incorporate both contrast and assimilation social comparison processes.

Acknowledgments

This research was funded by the Australian Research Council (DE140100080, DP130102713). The authors would like to thank Dr. Baljinder Sahdra for comments on the manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.learninstruc.2017.09.002.

Appendixes. Full model results

Table A1
Model 1 Results: Gender Differences in Achievement.

	Birth Cohort				
	1981	1984	1987	1991	1993
General Academic					
Intercept	−0.133(0.043)	−0.037(0.073)	−0.084(0.032)	−0.019(0.031)	0.012(0.034)
Male	−0.015(0.039)	−0.036(0.075)	−0.175(0.039)	−0.126(0.038)	−0.141(0.030)
Age	−0.06(0.013)	−0.065(0.016)	0.093(0.015)	0.094(0.012)	0.092(0.013)
Math					
Intercept	−0.199(0.041)	−0.131(0.067)	−0.209(0.032)	−0.171(0.034)	−0.112(0.036)
Male	0.175(0.038)	0.152(0.073)	0.070(0.039)	0.159(0.038)	0.105(0.031)
Age	−0.045(0.012)	−0.052(0.015)	0.100(0.014)	0.096(0.012)	0.092(0.013)
Literacy					
Intercept	−0.041(0.041)	0.051(0.068)	0.049(0.031)	0.123(0.029)	0.129(0.031)
Male	−0.187(0.035)	−0.190(0.064)	−0.397(0.038)	−0.381(0.037)	−0.362(0.030)
Age	−0.061(0.012)	−0.060(0.016)	0.075(0.017)	0.085(0.012)	0.086(0.013)

Notes. Estimates given in cohort-specific standard deviation units of achievement. Age is standardised to the global mean and standard deviation.

Table A2
Model 2 Results: Gender Differences in Social Comparison.

	Birth Cohort				
	1981	1984	1987	1991	1993
General Academic					
VP > NVW	−5.215(0.157)	−4.799(0.216)	−6.260(0.279)	−4.900(0.109)	−3.328(0.073)
NVW > AA	−3.011(0.078)	−3.095(0.107)	−3.877(0.084)	−2.910(0.056)	−1.560(0.053)
AA > BTA	0.153(0.060)	0.187(0.058)	−0.120(0.053)	−0.0529(0.042)	0.496(0.047)
BTA > VW	1.941(0.068)	1.855(0.085)	1.755(0.054)	1.636(0.045)	1.78(0.045)
Male	0.073(0.052)	0.239(0.068)	0.005(0.046)	0.048(0.037)	0.301(0.038)
Age	−0.046(0.019)	−0.073(0.022)	−0.022(0.029)	−0.040(0.025)	−0.004(0.022)
Math					
VP > NVW	−3.715(0.098)	−3.578(0.148)	−3.499(0.077)	−3.216(0.068)	−3.183(0.072)
NVW > AA	−1.472(0.053)	−1.502(0.094)	−1.585(0.054)	−1.496(0.049)	−1.389(0.054)
AA > BTA	0.443(0.053)	0.505(0.095)	0.370(0.052)	0.342(0.045)	0.453(0.046)
BTA > VW	1.838(0.062)	1.828(0.098)	1.766(0.049)	1.684(0.049)	1.714(0.052)
Male	0.347(0.048)	0.576(0.101)	0.345(0.042)	0.376(0.035)	0.396(0.042)
Age	−0.028(0.018)	−0.059(0.020)	−0.008(0.025)	−0.054(0.022)	0.013(0.022)
Literacy					
VP > NVW	−4.505(0.129)	−4.826(0.163)	−4.926(165)	−4.732(0.106)	−4.139(0.096)
NVW > AA	−2.338(0.074)	−2.524(0.085)	−2.965(0.059)	−2.662(0.051)	−2.361(0.052)
AA > BTA	0.259(0.063)	0.133(0.071)	−00.342(0.056)	−0.157(0.042)	0.0823(0.048)
BTA > VW	1.728(0.070)	1.689(0.079)	1.220(057)	1.296(0.048)	1.456(0.052)
Male	−0.345(0.054)	−0.048(0.085)	−0.436(0.047)	−0.325(0.041)	−0.249(0.040)
Age	−0.074(0.019)	−0.046(0.021)	−0.028(0.029)	−0.019(0.026)	0.002(0.026)

Notes. Response to comparison with class peers: VP = very poorly, NVW = not very well, AA = about average, BTA = better than average, VW = very well. Estimates reported in log-odds with associated standard errors.

Table A3
Model 3 Results: Gender Differences in Social Comparison Controlling for Achievement

	Birth Cohort				
	1981	1984	1987	1991	1993
General Academic					
VP > NVW	−5.471(0.156)	−4.973(0.221)	−6.501(0.273)	−5.183(0.106)	−3.585(0.072)
NVW > AA	−3.249(0.079)	−3.253(0.106)	−4.108(0.082)	−3.159(0.056)	−1.720(0.051)
AA > BTA	0.102(0.063)	0.188(0.060)	−0.255(0.050)	−0.066(0.041)	0.628(0.042)
BTA > VW	2.014(0.076)	1.969(0.079)	1.835(0.054)	1.797(0.046)	2.086(0.042)
Male	0.076(0.051)	0.260(0.055)	0.113(0.045)	0.117(0.035)	0.431(0.040)
Age	−0.017(0.019)	−0.045(0.019)	−0.083(0.028)	−0.123(0.025)	−0.109(0.022)
Achievement	0.636(0.034)	0.599(0.043)	0.616(0.024)	0.785(0.027)	0.914(0.032)
Math					
VP > NVW	−4.144(0.094)	−3.936(0.134)	−3.939(0.079)	−3.546(0.069)	−4.823(0.104)
NVW > AA	−1.810(0.054)	−1.758(0.078)	−1.952(0.055)	−1.770(0.047)	2.724(0.048)
AA > BTA	0.327(0.054)	0.458(0.082)	0.238(0.051)	0.243(0.044)	−0.072(0.042)
BTA > VW	1.894(0.066)	1.941(0.088)	1.810(0.048)	1.726(0.046)	1.479(0.049)
Male	0.227(0.049)	0.514(0.090)	0.321(0.049)	0.272(0.037)	0.338 (0.038)
Age	0.005(0.019)	−0.027(0.020)	−0.093(0.025)	−0.133(0.024)	0.073(0.026)
Achievement	0.811(0.028)	0.777(0.046)	0.797(0.029)	0.739(0.028)	0.631 (0.026)
Literacy					
VP > NVW	−4.767(0.134)	−4.929(0.174)	−5.076(0.168)	−4.823(0.104)	−4.270(0.103)
NVW > AA	−2.512(0.078)	−2.590(0.089)	−3.095(0.060)	−2.724(0.048)	−2.440(0.054)
AA > BTA	0.283(0.067)	0.201(0.074)	−0.315(0.057)	−0.072(0.042)	0.242(0.046)
BTA > VW	1.851(0.076)	1.828(0.093)	1.343(0.063)	1.479(0.049)	1.759(0.053)
Male	−0.248(055)	0.049(0.081)	−0.232(0.055)	−0.121(0.036)	−0.014(0.042)
Age	−0.051(0.018)	−0.026(0.022)	−0.072(0.028)	−0.073(0.026)	−0.067(0.027)
Achievement	0.624(0.031)	0.530(0.045)	0.573(0.034)	0.631(0.026)	0.804(0.032)

Notes. Response to comparison with class peers: VP = very poorly, NVW = not very well, AA = about average, BTA = better than average, VW = very well. Estimates reported in log-odds with associated standard errors.

References

Akerlof, G. A., & Kranton, R. E. (2010). *Identity economics: How our identities shape our work, wages, and well-being*. Princeton University Press.
Becker, G. S. (1974). *The economic approach to human behavior*. University of Chicago press.
Bleidorn, W., Arslan, R. C., Denissen, J. J. A., Rentfrow, P. J., Gebauer, J. E., Potter, J., et al. (2015). Age and gender differences in self-esteem: A cross-cultural window. *Journal of Personality and Social Psychology, 111*, 396–410.
Breen, R., & Goldthorpe, J. H. (1997). Explaining educational differentials towards a formal rational action theory. *Rationality and Society, 9*, 275–305.
Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.
Bynner, J. (2016). Institutionalization of life course studies. In M. J. Shanahan, J. T. Mortimer, & M. K. Johnson (Eds.), *Handbook of the life course* (pp. 27–58). Heidelberg: Springer.
Charles, M., Harr, B., Cech, E., & Hendley, A. (2014). Who likes math where? Gender differences in eighth-graders' attitudes around the world. *International Studies in Sociology of Education, 24*(1), 85–112.
Chevalier, A., & Arnaud, C. (2007). Education, occupation and career expectations: Determinants of the gender pay gap for UK graduates. *Oxford Bulletin of Economics and Statistics, 69*, 819–842.
Cimpian, J. R., Lubienski, S. T., Timmer, J. D., Makowski, M. B., & Miller, E. K. (2016).

Have gender gaps in math Closed? Achievement, teacher perceptions, and learning behaviors across two ECLS-K cohorts. *AERA Open, 2*(4), 2332858416673617.
Committee for Economic Development of Australia [CEDA]. (2013). *Women in leadership: Understanding the gender gap*. Retrieved from CEDA: http://adminpanel.ceda.com.au/folders/Service/Files/Documents/15355~cedawiljune%202013final.pdf.
Covington, M. V., & Beery, R. G. (1976). *Self-worth theory and school learning*. New York, NY: Holt, Rinehart and Winston.
Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96*, 608–630.
Davis, J. A. (1966). The campus as a frog pond: An application of the theory of relative deprivation to career decisions of college men. *American Journal of Sociology, 72*(1), 17–31.
Duberman, A. (2014, March 20). *Why it's harder for women to 'brag' about themselves at work – and why we really need to*. The Huffington Post. Retrieved from Huffington Post: http://www.huffingtonpost.com.au.
Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues, 46*(2), 183–201.
Elder, G. H., Pavalko, E. K., & Clipp, E. C. (1993). *Working with archival data: Studying lives* (Vol. 88). Sage.
Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103.
Falch, T., Torberg, F., & Naper, L. R. (2013). Educational evaluation schemes and gender gaps in student achievement. *Economics of Education Review, 36*, 12–25.
Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117–140.
Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology, 53*, 25–44.
Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effects on girls' mathematics performance. *Developmental Psychology, 49*(10), 1886.
Gigerenzer, G. (2008). *Rationality for mortals: How people cope with uncertainty*. Oxford, UK: Oxford University Press.
Goldin, C. (2014). A grand gender convergence: Its last chapter. *The American Economic Review, 104*, 1091–1119.
Goldthorpe, J. H. (2007). *On sociology*. Stanford, CA: Stanford University Press.
Goldthorpe, J. H. (2016). *Sociology as a population science*. Cambridge, U.K.: Cambridge University Press.
Haines, E. L., Deaux, K., & Lofaro, N. (2016). The times they are a-changing… or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly, 40*, 353–363.
Huguet, P., Pascal, H., Florence, D., Herbert, M., Isabelle, R., Ladd, W., … John, N. (2009). Clarifying the role of social comparison in the big-fish–little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology, 97*, 156–170.
Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65*, 373–398.
Hyde, J. S., Fennema, E., & Lamon, S. J. (1990a). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139–155.
Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990b). Gender comparisons of mathematics attitudes and affect. *Psychology of Women Quarterly, 14*, 299–324.
Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*, 53–69.
Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences, 106*, 8801–8807.
Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*, 509–527.
Kleiber, C., & Zeileis, A. (2008). *Applied econometrics with R*. New York, NY: Springer.
League, L. (2011, December 2). *Why is self-promotion so hard for women?* Forbes. Retrieved from Forbes: www.forbes.com.
Lietz, P., & Petra, L. (2006). A meta-analysis of gender differences in reading achievement at the secondary school level. *Studies in Educational Evaluation, 32*, 317–344.
Lingard, B. (2010). Policy borrowing, policy learning: Testing times in Australian schooling. *Critical Studies in Education, 51*, 129–147.
Little, D. (2012). *Varieties of social explanation: An introduction to the philosophy of social science*. Kindle Edition.
Lobascher, S. (2011). What are the potential impacts of high-stakes testing on literacy education in Australia? *Literacy Learning: The Middle Years, 19*(2), 9.
Lumley, T. (2011). *Complex surveys: A guide to analysis using R*. Hoboken, NJ: John Wiley & Sons.
Lummis, M., Max, L., & Stevenson, H. W. (1990). Gender differences in beliefs and achievement: A cross-cultural study. *Developmental Psychology, 26*, 254–263.
Machin, S., & McNally, S. (2005). Gender and student achievement in English schools. *Oxford Review of Economic Policy, 21*, 357–372.
Machin, S., & Pekkarinen, T. (2008). Global sex differences in test score variability. *Science, 322*, 1331–1332.
Mann, A., & DiPrete, T. A. (2016). The consequences of the national math and science performance environment for gender differences in STEM aspiration. *Sociological Science, 3*, 568.
Mann, A., Legewie, J., & DiPrete, T. A. (2015). The role of school performance in narrowing gender gaps in the formation of STEM aspirations: A cross-national study. *Frontiers in Psychology, 6*, 171.
Marsh, H. W. (1993). The multidimensional structure of academic self-concept: Invariance over gender and age. *American Educational Research Journal, 30*, 841–860.
Marsh, H. W. (2006). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. London, UK: British Psychological Society.
Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J., Abdelfattah, F., & Nagengast, B. (2014a). The big-fish-little-pond effect in mathematics: A cross-cultural comparison of US and Saudi Arabian TIMSS responses. *Journal of Cross-Cultural Psychology, 45*(5), 777–804.
Marsh, H. W., Köller, O., & Baumert, J. (2001). Reunification of East and West German school systems: Longitudinal multilevel modeling study of the big-fish-little-pond effect on academic self-concept. *American Educational Research Journal, 38*(2), 321–350.
Marsh, H. W., Kuyper, H., Morin, A. J., Parker, P. D., & Seaton, M. (2014b). 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.
Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and English constructs. *American Educational Research Journal, 35*, 705–738.
McKinsey & Company. (2015). *Women in the workplace*. Retrieved from LeanIn.Org and McKinsey & Company: http://womenintheworkplace.com.
Miller, C. T. (1984). Self-schemas, gender, and social comparison: A clarification of the related attributes hypothesis. *Journal of Personality and Social Psychology, 46*, 1222–1229.
OECD. (2015). *PISA the ABC of gender equality in education aptitude, behaviour, confidence: Aptitude, behaviour, confidence*. OECD Publishing.
Parker, P. D., Marsh, H. W., Guo, J., Anders, J. D., Shure, D., & Dicke, T. (2017). An information distortion model of social class differences in math self-concept, intrinsic value and utility value. *Journal of Educational Psychology*. Online First.
Parker, P., Nagy, G., Trautwein, U., & Ludtke, O. (2014). Predicting career aspirations and university majors from academic ability and self-concept: A longitudinal applications of the internal-external frame of reference model. In I. Schoon, & J. S. Eccles (Eds.), *Gender differences in aspirations and attainment: A life course perspective* (pp. 247–266). Cambridge, United Kingdom: Cambridge University Press.
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*(6), 1629–1642.
Phelan, J. E., Moss-Racusin, C. A., & Rudman, L. A. (2008). Competent yet out in the cold: Shifting criteria for hiring reflect backlash toward agentic women. *Psychology of Women Quarterly, 32*, 406–413.
Reilly, D., Neumann, D. L., & Andrews, G. (2015). Sex differences in mathematics and science achievement: A meta-analysis of national assessment of educational progress assessments. *Journal of Educational Psychology, 107*, 645–662.
Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counter-stereotypical impression management. *Journal of Personality and Social Psychology, 74*, 629–645.
Salchegger, S. (2016). Selective school systems and academic self-concept: How explicit and implicit school-level tracking relate to the big-fish-–little-pond effect across cultures. *Journal of Educational Psychology, 108*(3), 405–423.
Schoon, I. (2006). *Risk and resilience: Adaptations in changing times*. Cambridge: Cambridge University Press.
Simon, H. A. (1972). Theories of bounded rationality. *Decision and Organization, 1*(1), 161–176.
Skaalvik, E. M. (1990). Gender differences in general academic self-esteem and in success expectations on defined academic problems. *Journal of Educational Psychology, 82*(3), 593.
Stoet, G., Bailey, D. H., Moore, A. M., & Geary, D. C. (2016). Countries with higher levels of gender equality show larger national sex differences in mathematics anxiety and relatively lower parental mathematics valuation for girls. *PloS One, 11*(4), e0153857.
Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology, 92*(1), 144–151.
Van Zanden, B., Marsh, H. W., Seaton, M., & Parker, P. D. (2015). Self-concept: From unidimensional to multidimensional and beyond. In *International encyclopedia of the social and behavioral sciences* (pp. 460–468). Springer.
Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*, 1174–1204.
Wang, M. T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy–value perspective to understand individual and gender differences in STEM fields. *Developmental Review, 33*, 304–340.
Zell, E., & Alicke, M. D. (2010). The local dominance effect in self-evaluation: Evidence and explanations. *Personality and Social Psychology Review, 14*, 368–384.