*American Educational Research Journal*
*Month XXXX, Vol. XX, No. X, pp. 1–23*
*DOI: 10.3102/0002831218760213*


# Inequity and Excellence in Academic Performance: Evidence From 27 Countries

**Philip D. Parker**
**Herbert W. Marsh**
*Australian Catholic University*
**John P. Jerrim**
*University College London*
**Jiesi Guo**
**Theresa Dicke**
*Australian Catholic University*

*Research suggests that a country does not need inequity to have high performance. However, such research has potentially suffered from confounders*

---

Philip D. Parker is an associate professor and program lead at the Institute for Positive Psychology and Education at the Australian Catholic University, Lvl 9, 33 Berry St, North Sydney, NSW 2060, Australia; e-mail: *philip.parker@acu.edu.au*. He received his doctorate in educational psychology from the University of Sydney. His major research interests include educational inequality, developmental transitions, and educational attainment.

Herbert W. Marsh is professor at the Institute for Positive Psychology and Education and Emeritus Professor at Oxford University. Using sophisticated quantitative methods, he is a leading researcher in self-concept and motivation, evaluation of teaching, developmental psychology, sports psychology, peer review and peer support, gender issues, and anti-bullying interventions.

John P. Jerrim is a professor in educational and social statistics at the UCL Institute of Education, UK. He received his doctorate in social statistics from the University of Southampton. His research focuses on international comparisons of educational attainment and the evaluation of educational programs and policies.

Jiesi Guo is a post-doc research fellow at the Institute for Positive Psychology and Education. His areas of interest include educational and developmental psychology with a particular focus on how multiple systems on the cultural, social, and motivational development of youth shape individual and gender difference in achievement choice.

Theresa Dicke is a post-doc research fellow at the Institute for Positive Psychology and Education at the Australian Catholic University. Her research interests lie in the realm of organizational, health, and educational psychology. Applying complex quantitative methods, she has extensively studied the role of (occupational) well-being and achievement in school, thus, focusing on students', teachers', and principals' perspectives alike.

*present in between-country comparative research (e.g., latent cultural differences). Likewise, relatively little consideration has been given to whether the situation may be different for high- or low-performing students. Using five cycles of the Programme for International Student Assessment (PISA) database, the current research explores within-country trajectories in achievement and inequality measures to test the hypothesis of an excellence/equity tradeoff in academic performance. We found negative relations between performance and inequality that are robust and of statistical and practical significance. Follow-up analysis suggests a focus on low and average performers may be critical to successful policy interventions.*

Keywords: educational inequality, achievement stratification, PISA, cross-cultural comparisons

## Introduction

Educational policy aims to maximize educational excellence and reduce inequity. The need to balance these demands is an ongoing concern in social mobility (Burger, 2016), educational attainment (Goldthorpe, 2007), and to a lesser degree, concerns about performance in standardized tests (Checchi, 2006). Our paper is primarily concerned with issues relating to the association between national performance in standardized tests (educational excellence) and the degree of variation in performance within a nation (our measure of educational inequity). It is our hypothesis that greater variance in test scores—greater inequality—will be negatively associated with higher average educational achievement—or higher excellence. We seek to directly challenge views that a country's educational policies must incorporate some inequality to produce higher average test scores. To test this hypothesis, we consider a range of inequality measures. Unlike previous research, we focus on (a) changes that occur within countries over time and (b) where in the academic achievement distribution the changes occur. In the following sections, we first position our research within the broader domain of educational inequity before outlining competing positions on the excellent/equity tradeoff in educational ability. Finally, we consider what empirical research currently suggests about this debate and the limitations with the existing evidence base that we seek to overcome.

### Excellence Versus Equality

Debates over excellence in education often suggest that educational systems produce the highest average performance if schools can tailor offerings to children of different academic ability (for a review, see Hoxby, 2003; Van de Werfhorst & Mijs, 2010; for an applied introduction, see Walberg, 2000). Checchi (2006; see also Hoxby, 2003) provides a detailed treatment of this line of reasoning, but the argument states that in the absence of government interference, families will choose a level and type of education for their children that will maximize the child's achievement and should this occur for most children, maximize the achievement of the nation as a whole (Friedman, 2002; Hoxby, 2003). At the core of this idea is that differentiated, stratified, decentralized, and/or private or privatized education (for a review, see Bol & Van de Werfhorst, 2013; Kerckhoff, 1995; Parker, Jerrim, Schoon, & Marsh, 2016) provides a context that prepares children with different ability with appropriate skills. This may mean less talented children are provided with educational content specifically focused on vocational skills (for an overview, see Brunello & Checchi, 2007). For talented children no longer hampered by the need for teachers to limit the scope and speed of content for the benefit of less talented children, increased education system variance will maximize their learning gains (see Van de Werfhorst & Mijs, 2010).

Under this model, increased academic *excellence* for a country will tend to be associated with greater variance in achievement due to selection effects, signaling, and different educational content (Jakubowski, Patrinos, Porta, & Wisniewski, 2010; Parker et al., 2016; Pfeffer, 2015). Thus, inequality is a necessary condition for excellence. Because of this, there is a potential conflict in policy between maximizing excellence (i.e., maximizing average levels of achievement by allowing children to match their education to their potential) and limiting inequity (i.e., minimizing the variability in outcomes between children; Gans & King, 2014). According to the tradeoff position, excellence (i.e., high performance) comes with the cost of inequality. But does the empirical evidence support this?

### *Excellence/Equality Tradeoff*

Underlying the excellence/equity tradeoff position is the belief that policymakers need to balance the competing demands of promoting excellence and reducing inequality. Inequality may come about via several mechanisms. First, educational differentiation or school choice means that different children receive different levels or types of education. Second, it may be that excessive variance in academic achievement occurs due to mechanisms unrelated to government policy or at least unrelated to government *education* policy. For example, increased variance may come about due to wider social stratification by race, ethnicity, or social class (e.g., Rowe & Lubienski, 2017). Third, there may be barriers that prevent children from disadvantaged backgrounds from gaining access to the type of education best suited to their underlying talent. Indeed, due to limited access to economic or other resources, risk adversity, or poor decision making, parents might choose a type of education that is inappropriate for the child and policies are required that provide such children with educational chances more in keeping with their ability (Friedman, 2002; Gans & King, 2014).

Suggestions for policy interventions that emerge from the tradeoff position often indicate that apart from ensuring that talented children are not misplaced, policy should not try to minimize variance in academic ability within a country (Walberg, 2000). Under this belief, achievement differentiation, decentralization, privatization, and stratification should be encouraged as they increase the options available to parents and improve overall performance. Yet government intervention should still focus on reducing the risks of student misplacement within this system (Friedman & Friedman, 1980).

*Empirical Evidence*

The underlying theory of the excellence/equity tradeoff is elegant. Yet, it has been increasingly disputed by empirical evidence derived mainly from studies using large-scale international student assessments (for a review, see Van de Werfhorst & Mijs, 2010). Anecdotally, criticism of the excellence/equity tradeoff comes from the observation that high-performing countries like Finland appear to combine low levels of inequality (including both low barriers to entry and relatively undifferentiated education) with high levels of performance in international tests (Simola, 2005). Empirically, evidence questioning the benefits of inequality comes from two strands of evidence: (a) empirical results that show that high variance leads to considerable inequality in educational outcomes and (b) empirical research that shows that educational systems with high academic ability variance may have poorer average performance. In relation to the former, Brunello and Checchi (2007) found that tracking is related to disadvantages for poorer children in both educational attainment and labor market outcomes and that these effects are larger the earlier the tracking. Jerrim, Chmielewski, and Parker (2015) found that private schooling in Australia, the UK, and the United States was associated with advantages in both education and labor market outcomes. Finally, Parker et al. (2016) found that ability stratification was associated with lower expectations of university attainment for poorer children controlling for academic achievement.

In relation to the second stream of evidence, Hanushek and Wößmann (2005) found that early tracking increased educational inequality and that it was associated with lower mean performance. Micklewright and Schnepf (2007) showed that the distance between the 95th and 5th percentiles in achievement and the median performance within a country was negatively correlated. Likewise, Checchi, van de Werfhorst, Braga, and Meschi (2014) found no or negative relationships between various forms of variance and stratification and average achievement. In addition to data on educational performance is research on educational attainment. Thomas, Wang, and Fan (2001) found a negative relationship between a Gini index (a relative measure of inequality) of years of education and the average years of education within a country for countries within the developed world. Pfeffer

(2015) found that there was no relationship between performance in international adult skills assessment and inequity of opportunities. Overall, this suggests that there is little evidence that inequality is needed for excellence in academic performance, at least within rich countries.

Significantly, almost all of the research to date has focused on between- (cross-sectional) rather than within-country (multicohort) relationships. It is also important to note that changes in stratification may or may not occur evenly across the achievement distribution, with changes in variance at the top or bottom half potentially being of most importance. Where changes in variance occur could potentially have different implications. For example, Micklewright and Schnepf (2007) suggest that inequality tends to be largest in the bottom half of the achievement distribution. Thus, *increases* in polarity (movement from the median of the distribution to the tails) at the bottom end of the distribution may be most important. Indeed, Poland has had particular success at improving performance by introducing policy targeting such students (Breakspear, 2012). Alternatively, Ryan (2013), focusing only on Australia, suggests that declines in the top half of the distribution account for that country's decline in math performance. This indicates that reduction in polarity at the top end of the distribution (i.e., the highest performers becoming more similar to the median performer) is of most concern.

## Current Research

The current research makes use of over a decade of the Programme for International Student Assessment (PISA) data to explore the association between changes in country inequality and changes in country average achievement. As such, we advance the following hypotheses:

*Hypothesis 1:* Trends in inequality from 2000 to 2012 will be zero or negatively related to trends in performance over the same period.

*Hypothesis 2:* Changes in inequality from one PISA round to the next will be zero or negatively related to changes in performance for the same rounds. Both Hypotheses 1 and 2 are founded on the hypothesis that inequality is not a necessary requirement for academic excellence.

*Hypothesis 3:* When large changes in inequality occur, changes at the top or bottom of the achievement distribution will be differentially associated with changes of average achievement.

We use the term *trend* when considering movements in inequality or average achievement as a linear line through all the PISA cycles under consideration. The term *change* is reserved for when we are averaging the changes from one PISA cycle to the next. We do not consider the effect of a previous wave in inequality on changes in achievement (or vice versa). Rather, in all models, we are focused on simultaneous change in achievement and inequality.

### Measures of Inequality

We note that a number of different measures of inequality have been used in the literature. These include measures focused on how children of different levels of ability are sorted into schools such as the between-school achievement variance or intraclass correlation coefficient (ICC) (Marks, 2006; Parker et al., 2016; Salchegger, 2016). These measures provide an index of the degree to which a country's education system segregates children of different levels of academic ability into different schools. This measure incorporates both formal (e.g., tracking) and informal (e.g., social segregation) differentiation (see Parker et al., 2016). Other measures focus on the degree of variance in academic performance between children within the same country. These include absolute or relative (i.e., scale invariant) measures (for a review, see Handcock & Morris, 1999). We use a selection of all of these indexes, including (a) ICC as a measure of the amount of between-school ability stratification, (b) the distance between the 95th and 5th percentiles in achievement as a measure of absolute variance in achievement (see Micklewright & Schnepf, 2007), (c) a constructed Gini index of achievement, and (d) where possible, relative polarity as relative indexes of variance (Handcock & Morris, 1999).

There are few criteria for what indicates large or small variation in these measures, and this is particularly the case in the context in which we use them, where we rely on trends or change over time. In the absence of criteria then, we undertake extensive sensitivity analyses using multiple measures across multiple academic domains with multiple statistical methods. Thus, our focus is on results that show consistency across these approaches.

## Method

### Participants

All analyses were done at the country level using participant-level indicators of math, science, and reading achievement from all five PISA rounds. PISA provides data on a representative sample of 15-year-olds. We focused on OECD countries (based on membership as of 2000) with the exception of Mexico and Turkey.[1] The data are collected in a two-stage procedure with schools selected proportional to size and a random sample of 15-year-olds selected from within each school (OECD, 2004). A set of weights is provided so that the sample is representative of the target population. In total, participants came from 27 countries for a total sample size of 1,026,173 for analysis related to reading achievement and 957,735 for analysis related to math and science achievement. The reason for the difference in participant numbers is that all participants received reading scores in PISA 2000 but only a subsample received either math or the science scores. In all other PISA rounds, participants received estimated performance scores for all domains.

## Measures

### *Academic Performance*

Children's academic achievement was measured via performance on a standardized test in math, reading, and science. The achievement tests used in PISA are designed specifically to enable cross-national comparisons in academic achievement. PISA differs from other international measures of academic performance as it focuses on functional ability rather than knowledge or mastery of a curriculum. Answers from the achievement tests were summarized by the survey organizers into a single score for each of the three domains using an item-response model, the intention being that true skill in each subject is unobserved and must be estimated from the answers to the test (for further details, see OECD, 2004). Five plausible values were generated for each pupil, estimating their true proficiency in each subject. These scores were scaled by the survey organizers to have a mean of 500 points and standard deviation of 100 points across OECD countries in the first PISA round. Country average performance, Gini, and ICC were all estimated with the five plausible values separately and then averaged to provide country-specific point estimates.

### *Gini Index*

The Gini index was calculated separately for each academic domain, country, and PISA round. As with all measures used in the present research, the Gini was calculated using the population weight via the reldist package in R (Handcock & Aldrich, 2002). This index varies between 0 (indicating a uniform distribution of achievement) and 1 (indicating that only a single individual had a non-zero achievement score). We multiplied the Gini index by 100.

### *Intraclass Correlation*

ICCs estimate the degree to which students within a school resemble each other and differ on average from those in other schools in terms of academic achievement. Thus, higher estimates of ICCs reflect the degree to which schools were homogenous in academic achievement. ICCs were estimated after applying population weights. We also multiplied these by 100 so that they varied from 0 to 100 (see Marks, 2006).

### *95th Percentile – 5th Percentile*

The distance between the 95th percentile (P95) and 5th percentile (P5) of achievement was likewise calculated after applying the population weight. There was evidence of change in achievement and all inequality indexes across the PISA cycles; however, this differed in size by country (see supplementary material in the online version of the journal).

## Statistical Analysis

### *Modeling Approach*

Hypotheses 1 and 2 relied on exploring the relationship between estimates derived for each country. We focus here on estimates derived using a series of multilevel models with PISA cycle estimates nested within country. As such, all analyses were done at the country level, and no individual-level data were modeled in the analysis reported in the results. There are debates about how appropriate the use of multilevel models is in the context of country comparisons. In particular, there are concerns that random effects models remain common despite the fact that (a) countries are rarely sampled randomly from a population (or in our case include all or almost all countries in a relevant population; i.e., the OECD) and (b) country-specific estimates can be biased (due to shrinkage) when there are few countries (e.g., Byran & Jenkins, 2015). As such, we also tested the robustness of the results using country fixed effects models. Detailed consideration of model development is provided in the supplementary material in the online version of the journal.

For trends, multilevel growth curve models were estimated (Hypothesis 1). In each case, both the intercept (i.e., initial level at year 2000) and slope (i.e., slope of the linear interpolated trajectories from 2000 to 2012) were estimated as country random effects. Such models were run separately for academic achievement and inequality measures. Country-specific slope estimates were drawn from the resulting parameter estimates. We also calculated the simple difference between PISA 2000 and 2012 achievement and inequality measures and looked at the relationship between these. Growth curve models treat PISA cycles as an ordinal variable and thus summarize the change across PISA cycles in relation to, for example, achievement as a linear trend. The benefit of this is that it provides a simple summary measure that reduces the influence of noise around this trajectory, thus reducing the impact of outlier cycles (e.g., where a country experiences a notable increase in only one PISA cycle before returning to baseline levels).

However, it is possible that these results may be biased as they impose a linear trajectory from PISA 2000 to PISA 2012. We aimed to account for this by using change score models (Hypothesis 2). In this case, achievement at round $k + 1$ was regressed on achievement at round $k$ with the regression estimate fixed to 1 (i.e., a simple difference score) and the change score of inequality from round $k$ to $k + 1$. The result of this specification was that change in achievement was predicted by change in inequality over the same lag. Random effects for country were included.

### *Variance Location*

Hypothesis 3 focused on where changes in inequality occurred in the achievement distribution. Using the reldist package in R (Handcock &

Morris, 1999), we isolated changes in the achievement distribution from 2000 to 2012 in relation to shape (e.g., changes in skewness) and location (e.g., movement of the population as a whole up or down the achievement distribution). We took two approaches to this. First, we explored the relationship between relative polarity (RP; i.e., degree of movement from the median to the tails of the distribution from one PISA cycle to the next) and changes in achievement for all countries. Second, we selected several countries that displayed considerable change in achievement from 2000 to 2012 for a more detailed analysis. We use both RP measures as well as plots of changes in the achievement distribution, decomposed into location and shape changes. All RP indexes vary from −1 to 1, with negative values indicating decreased polarity or a movement of values toward the median. The median relative polarity (RPM) index provides an overall estimate. This can be decomposed to explore the upper (RPU) and lower (RPL) portions of the distribution.

## Results

### Hypothesis 1: Associations in Trends

We first looked at whether linear trends in achievement from 2000 to 2012 were related to linear trends in inequality. For this, we extracted country-level trends from (a) a series of random intercepts and slopes models, (b) a series of country fixed effect models, and (c) the simple difference between achievement and inequality measures from PISA 2000 to PISA 2012 (hereafter *simple*). As shown in Table 1, the relationship between the trend in achievement and the trend in inequality was negative in all cases. In support of Hypothesis 1, countries that increased in achievement from 2000 to 2012 tended to also decline in inequality measures. Relationships were strongest for Gini and ICC indexes (Gini: mean $r = -.658$; ICC: mean $r = -.524$), with correlations routinely around –.50 and frequently above –.70. The relationships were more moderate for P95 – P5 (mean $r = -.313$) and typically only significant for science (note that average correlations are Pearson correlations). The correlations were similar for all achievement domains, with Figure 1, derived from the multilevel models, showing the relationship between the linear trajectory of math achievement and inequality. The supplementary material in the online version of the journal provides figures for reading and science.

### Hypothesis 2: Association in Change Scores

The analysis presented previously focused on linear trends in achievement and inequality from 2000 to 2012. However, it is possible that these results do not give an accurate reflection of the relationship between associations in simultaneous changes in achievement and inequality (see methodology for a discussion). To account for this, we looked at the relationship

*Table 1*
**Correlation Between Achievement Trajectory and Inequality Trajectory**

| | Random Effect (Pearson/Spearman) | Country Fixed Effect (Pearson/Spearman) | Simple (Pearson/Spearman) |
|---|---|---|---|
| Math | | | |
| Gini | −.722***/−.727*** | −.681***/−.756*** | −.689***/−.725*** |
| ICC | −.536**/−.519** | −.571**/−.563** | −.503**/−.457* |
| P95 – P5 | −.347/−.374 | −.313/−.369 | −.323/−.439* |
| Reading | | | |
| Gini | −.613***/−.523** | −.567**/−.524** | −.542**/−.489*** |
| ICC | −.495**/−.485** | −.558**/−.552** | −.670***/−.694*** |
| P95 – P5 | −.255/−.180 | −.230/−.227 | −.188/−.149 |
| Science | | | |
| Gini | −.704***/−.727*** | −.695***/−.662*** | −.706***/−.767*** |
| ICC | −.440*/−.318 | −.500**/−.414* | −.443*/−.281 |
| P95 – P5 | −.371*/−.464* | −.395*/−.392* | −.391*/−.518** |

*Note.* Random effect = correlation of slope with achievement slope from a multilevel growth curve model; country fixed effect = correlation of slope with achievement slope from a country fixed effect model; simple = correlation of difference from PISA 2000 to 2012 in achievement and inequality measures; ICC = intraclass correlation; P95 = 95th percentile; P5 = 5th percentile; PISA = Programme for International Student Assessment.
**p* < .05. ***p* < .01. ****p* < .001.

between changes in achievement and changes in inequality from one PISA wave to the next (Table 2). For all academic domains, a change in the Gini index from one PISA round to the next was associated with a significant counteracting change across the same cycles in average achievement ($\beta$ range = –.369 to –.413). On their original metrics, a 1-point increase in Gini (inequality) was associated with a 6-point (for science) to 10-point (for math) decline in achievement for changes across the same PISA cycles. Significant associations were likewise found for reading and science for ICCs and for reading for P95 – P5. Effect sizes were moderate for the Gini index (mean $\beta$ = –.387) and ICCs (mean $\beta$ = –.361) and small for P95 – P5 (mean $\beta$ = –.132).

### Hypothesis 3: Where Does Inequality Change?

A focus on change scores also allowed us to consider changes in relative polarity from one PISA wave to the next. In all cases, the estimates were negative, suggesting that inequality is not a requirement for excellence (see Table 2). Supporting Hypothesis 3, the effects for RPM and RPL were only significant in one case. Overall, the relationships were strongest for the upper half of the achievement distribution and significant or marginally significant for all domains (mean $\beta$ = –.264). This indicates that declines in

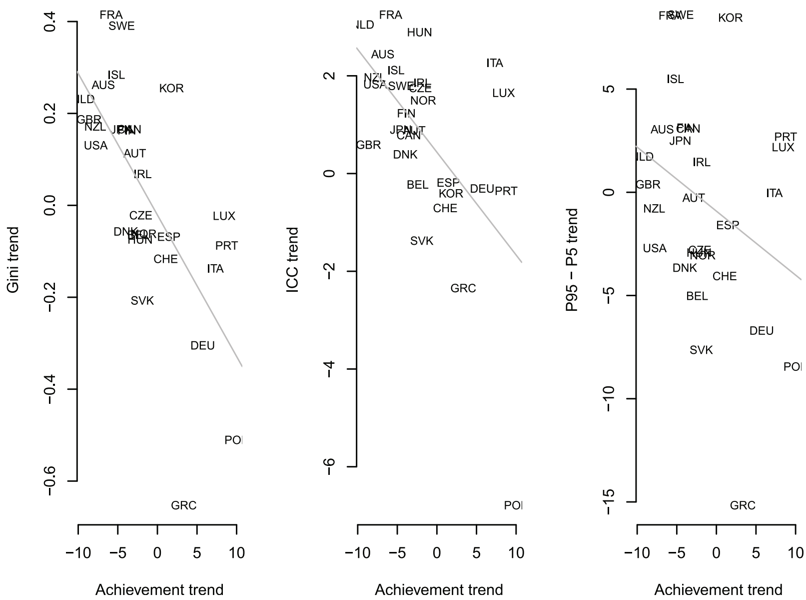


*Figure 1.* **Math trends.**

*Note.* Country given using the ISO three-letter code. Regression line represented in gray.

achievement may be more strongly weighted toward increases in inequality in the upper portion of the achievement distribution. Put simply, declining PISA scores tended to be associated with average-performing students falling further behind the highest performing students such that the right tail of the distribution became increasingly elongated (i.e., the highest performing students tend to be protected against declines in achievement). However, the difference between RPL and RPU were relatively small though nevertheless sufficient enough to suggest a more in-depth consideration would be beneficial.

We finally considered where in the achievement distribution changes in inequality tended to occur for countries that experienced notable changes. Given space constraints, we focused on Germany, Poland, Sweden, and Iceland as these were the countries in which the largest changes in achievement and inequality occurred. Germany and Poland were the only two countries to improve by over 20 achievement points and decreased in Gini by over 1 point for each domain between 2000 and 2012. Sweden declined by almost 30 points in each domain and increased in Gini by over 1 point in both reading and science (and over half a point in math). Likewise, Iceland increased in Gini by over 1 point in each domain and declined in achievement by over 20 points in math and reading (and over 17 points in science).

*Table 2*
**Lagged Results**

| | Est | *SE* | β | *p* |
|---|---|---|---|---|
| Math | | | | |
| Gini | −10.556 | 1.908 | −.369 | *** |
| ICC | −0.342 | 0.357 | −.245 | |
| P95 – P5 | −0.138 | 0.075 | −.104 | ^ |
| RPM | −0.454 | 0.314 | −.151 | |
| RPL | −0.178 | 0.275 | −.060 | |
| RPU | −0.562 | 0.284 | −.223 | ^ |
| Reading | | | | |
| Gini | −7.917 | 1.320 | −.413 | *** |
| ICC | −0.531 | 0.178 | −.458 | ** |
| P95 – P5 | −0.138 | 0.059 | −.165 | * |
| RPM | −0.454 | 0.314 | −.151 | |
| RPL | −0.614 | 0.242 | −.250 | * |
| RPU | −0.419 | 0.252 | −.172 | ^ |
| Science | | | | |
| Gini | −9.252 | 1.422 | −.380 | *** |
| ICC | −0.499 | 0.183 | −.380 | ** |
| P95 – P5 | −0.123 | 0.072 | −.127 | ^ |
| RPM | −0.706 | 0.292 | −.295 | * |
| RPL | −0.234 | 0.278 | −.100 | |
| RPU | −0.824 | 0.242 | −.397 | *** |

*Note.* Est = estimated effect; β = estimates taken from a model in which achievement and stratification are standardized around the grand mean; Gini = Gini estimates of achievement; ICC = intraclass correlation of achievement; P95 – P5 = distance in achievement between the 95th and 5th percentiles; RPM = relative polarity median of achievement; RPL = relative polarity lower of achievement; RPU = relative polarity upper of achievement.
^$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

The results indicated significant changes in polarity for each country in at least two of the three achievement domains (see Table 3). Germany and Poland declined in polarity (see plots in supplementary material in the online version of the journal). Germany predominantly declined in the upper portion of the distribution, with Poland displaying most change in the lower portion. However, for reading in Germany and reading and science in Poland, significant declines in polarity occurred in both RPL and RPU. This shape change resulted in fewer individuals in the lower and upper deciles than would have been the case if changes in achievement from 2000 to 2012 were due to location changes alone (i.e., mean rather than distribution shape). Sweden and Iceland both significantly increased in relative polarity. In both cases, changes were predominantly located in the upper portion of the distribution. What this means is that as Sweden and Iceland

*Table 3*
**Relative Changes in Polarity**

| | Germany | | Poland | | Sweden | | Iceland | |
|---|---|---|---|---|---|---|---|---|
| | Est | *p* | Est | *p* | Est | *p* | Est | *p* |
| Math | | | | | | | | |
| RPM | −.014 | | −.059 | *** | .017 | | .072 | *** |
| RPL | −.036 | | −.117 | *** | −.003 | | .039 | |
| RPU | .009 | | −.002 | | .037 | | .107 | *** |
| Read | | | | | | | | |
| RPM | −.093 | *** | −.085 | *** | .075 | *** | .021 | |
| RPL | −.087 | *** | −.115 | *** | .063 | ** | .017 | |
| RPU | −.099 | *** | −.054 | * | .086 | *** | .026 | |
| Science | | | | | | | | |
| RPM | −.066 | *** | −.092 | *** | .026 | * | .086 | *** |
| RPL | −.030 | | −.095 | *** | −.004 | | .061 | * |
| RPU | −.101 | *** | −.088 | ** | .055 | * | .111 | *** |

*Note*. Est = estimated effect; RPM = relative polarity median of achievement; RPL = relative polarity lower of achievement; RPU = relative polarity upper of achievement.
*$p$ < .05. **$p$ < .01. ***$p$ < .001.

declined in average achievement, the most talented students were partially protected. Thus, there were frequently 20% to 30% more students in the top decile than would be expected if achievement declines were consistent across the whole distribution. Indeed, for science achievement in Iceland, there were approximately equal numbers of students in the top decile of the reference distribution at both PISA 2000 and 2012 when there should have been only 60% as many individuals in 2012 if there was no change in relative polarity (see Figure 2).

## Discussion

Consistent with growing evidence (e.g., Checchi et al., 2014; Micklewright & Schnepf, 2007), our results suggest that inequality, indexed by stratification or variance in achievement, is negatively associated with average achievement at the country level. Importantly, effect sizes were routinely of a similar size for both relative Gini (variance) and ICC (stratification) indexes of inequality. Relationships were smaller but still negative and often significant for absolute measures of variance (see the following). We extended previous research by focusing on within-country changes in inequality and its association with within-country changes in average achievement. Not only were results consistent with previous research in showing that inequality is not necessary to produce excellence, they also

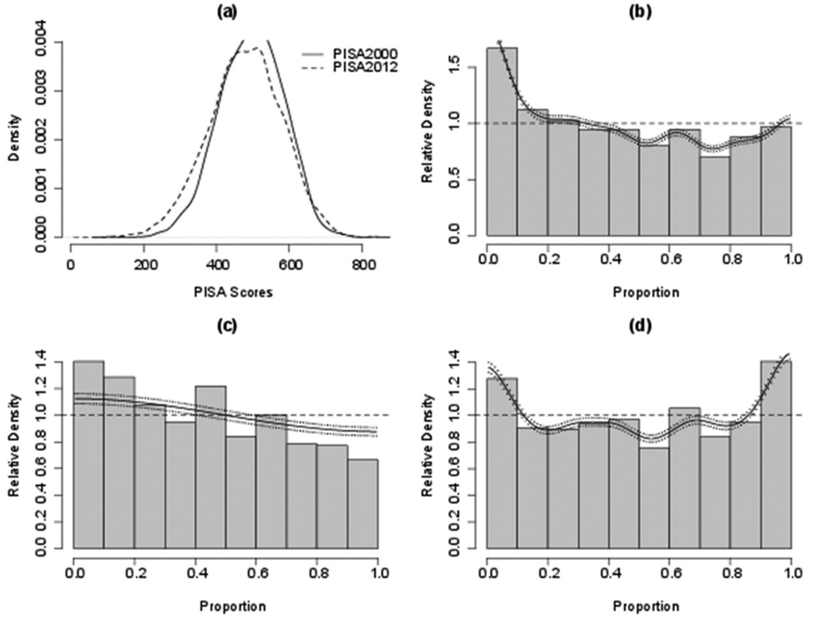


*Figure 2.* **Distribution change for science achievement in Iceland.**

*Note.* Panel (a) represents the achievement distributions for PISA 2000 and 2012. Panel (b) represents changes in achievement from 2000 to 2012 using the 2000 distribution as a reference. Panel (c) indicates what the change in the achievement distribution from 2000 to 2012 is if the change was due to location change alone. Panel (d) indicates the changes in variance at different points in the distribution. The bar plots indicate calculations of change within each decile of the achievement distribution. The solid line indicates estimates changes in the distribution using a Gaussian kernel density estimator. The dotted lines indicate 95% confidence intervals. For both the bar graph and estimated change line, a value above 1 indicates values in 2012 above those observed in 2000. Values below 1 indicate values in 2012 that are below those observed in 2000. Thus, when the confidence intervals do not cross the horizontal line at 1 on the y-axis, significant differences exist between 2000 and 2012. PISA = Programme for International Student Assessment.

suggested that increases in inequality within a country may be associated with declines in performance.

When considered from a within-country perspective, traditional dividing lines between educational systems evaporated. Nordic countries have often been shown to be among the most equal in between-country studies (e.g., Parker et al., 2016, 2017). When considering within-country estimates however, Iceland and Sweden had some of the most evident declines in achievement and increases in inequality of all countries considered. Alternatively, while Germanic countries have been shown to be some of the most unequal due to early and extensive tracking, Germany has shown considerable improvement in academic achievement, which has been associated with notable decreases in inequality. Taken together, while between-country differences continue to follow traditional demarcations in inequality—Nordic $<$ Anglophone $<$ Germanic—(see Dupriez & Dumay, 2006), within-country analysis shows a shifting landscape where these monikers hold less relevance. This could be taken to suggest that overall, the inter-country landscape is becoming more equal. However, there were notable increases in intraclass correlations PISA 2000 to 2012 (see supplementary material in the online version of the journal). Thus, the trend for OECD countries is actually toward greater inequality.

It may be that changes unrelated to direct educational policy are driving these results. As such, we ran further sensitivity analysis on the country fixed effects presented in Table 1. In this case, we calculated the partial correlation coefficients between academic excellence and our inequity measures controlling for trajectories across the same period (2000–2012) in Gross Domestic Product (GDP, in US dollars), average disposable income, and the percentage of GDP spent on social welfare. As Table 4 shows, the results were similar to those reported in Table 1.

### Why Is Excellence Not Positively Related to Higher Variance?

A major question that emerges from the current research is why there is so little evidence that inequity is a requirement for excellence. To some degree, this is answered by proponents of the tradeoff argument themselves, namely, that decisions relating to the amount and type of education that a child should invest in is a decision not made by the child themselves but rather by parents or guardians. Such parents may not make decisions that lead to the best possible school placement (Friedman, 2002). Widespread and systematic inefficiencies in child assignment could account for the results noted here (see Pfeffer, 2015). Indeed, PISA data suggest that misplacement occurs across the socioeconomic ladder (Parker et al., 2017). For example, Maaz, Trautwein, Lüdtke, and Baumert (2008) note that in the Germanic system, parents from well-off families often ensure that their children are located in university track systems even when teacher recommendations are for lower track placements. Further, Parker et al. (2017) show that children of richer parents pay for poor placement with decreased academic self-concept. Conversely, children of poorer parents would likely gain in self-concept by inaccurate school placement but pay in terms of more difficult pathways to university.

This would suggest that the problem is not with the idea that a school system should tailor offerings to different levels of the achievement distribution but rather with its application in context. However, inherent problems

*Table 4*

**Country Fixed Effects Controlling for Country-Level Covariates**

| | No Controls (Pearson/Spearman) | Social Welfare (Pearson/Spearman) | GDP (Pearson/Spearman) | Disposable Income (Pearson/Spearman)[a] |
|---|---|---|---|---|
| Math | | | | |
| Gini | −.681***/−.756*** | −.680***/−.752*** | −.694***/−.754*** | −.710***/.775*** |
| ICC | −.571**/−.563** | −.574**/−.557** | −.600***/−.562** | −.676***/−.656*** |
| P95 – P5 | −.313/−.369 | −.313/−.381* | −.324/−.379 | −.368/−.445* |
| Reading | | | | |
| Gini | −.567**/−.524** | −.574**/−.536** | −.558**/−.551** | −.609***/−.519** |
| ICC | −.558**/−.552** | −.560**/−.549** | −.586**/−.540** | −.591**/−.544** |
| P95 – P5 | −.210/−.264 | −.210/−.271 | −.226/−.281 | −.327/−.311 |
| Science | | | | |
| Gini | −.695***/−.662*** | −.696***/−.654*** | −.753***/−.688*** | −.785***/−.739*** |
| ICC | −.500**/−.414* | −.503**/−.397* | −.536**/−.427* | −.553**/−.453** |
| P95 – P5 | −.395*/−.392* | −.394*/−.379 | −.496**/−.427* | −.558**/−.539** |

*Note*. All covariates were taken from the OECD (https://data.oecd.org/). GDP = Gross Domestic Product; ICC = intraclass correlation; P95 = 95th percentile; P5 = 5th percentile.

[a]These results exclude Luxemburg for whom disposable income data were not available.

*$p < .05$. **$p < .01$. ***$p < .001$.

with school inequality suggest issues may continue to occur even with perfect placement. Evidence from educational psychology points to a natural bias in the way young people form expectations. Dicke et al. (in press) argue that children in more selective schools have lower academic self-concepts than they would have had they gone to more comprehensive schools—a so-called big-fish-little-pond effect. An important extension of this is that lower self-concept leads to lower performance in a reciprocal spiral (a reciprocal effects model [REM]; Dicke et al., in press). PISA data show that this effect is larger in countries with more tracking or higher ICCs (Salchegger, 2016). It is possible that this bias in self-perceptions may account for some of the reason why more stratified systems do worse than expected if inequality led to more efficient teaching and thus great academic excellence.

Alternatively, nonlinear peer effects in learning quality likely provide equally compelling explanation of these results for the low end of the achievement distribution. Nonlinear effects suggest that high-performing students tend to lose very little from association with poorer performing students but poorer performing students gain considerable benefits in terms of motivation and quality of peer interaction (Checchi, 2006; Hanushek, Kain, Markman, & Rivkin, 2003; Hanushek & Wößmann, 2005). Countries should consider the findings here and determine whether students across the achievement distribution may actually benefit from more integrated classrooms—though always with an eye to the local policy context.

**Changes in Inequality**

The current research suggested that increases in stratification measures of inequality are associated with decreases in average achievement. We considered average change in variance for all countries but also the form of this change. Declines in achievement were mostly associated with protection of high-performing students and fall in average- and low-performing students. Taken as a whole, there was evidence of an effective hollowing out of the middle of the achievement distribution where there was increasing polarization between the most talented students and the rest. In-depth analysis of countries that changed considerably from PISA 2000 to PISA 2012 (i.e., 20 PISA points and 1 Gini point) provided a more nuanced perspective on this issue.

Ringarp and Rothland (2010) note that Sweden has moved from one of the most to one of the least centralized educational systems, with increased school choice and privatization in the past few decades. Iceland has long had a decentralized school system with considerable school choice. However, decentralization was strengthened by policy in 2008, and the implication of this policy likely increased after the global financial crisis, where local communities responded to a reduction in educational funding in a diverse number of ways (Ministry of Education, Science and Culture, 2014). Importantly, this led to considerable regional differences in declines in PISA performance. In contrast, the "PISA shock" of 2000 in Germany led to a national conversation on education, an increase in centralization, and a focus on lower performers and immigrants (Breakspear, 2012). In Poland, there was a strong focus on the poorest performing students in response to PISA results (Breakspear, 2012). Our findings suggest that for Germany, increases in performance mostly centered on the middle of achievement distribution. For Poland, our results show the success of their focus on the bottom of the achievement distribution. Taking all the results together, a hypothesis emerges that a country's educational policy that mainly serves talented students will be associated with lower average performance. Alternatively, a focus on the lower and middle portions of the achievement distribution will be associated with higher average ability. Overall, there is a need for future research that focuses not just on changes in inequality overall but on where changes occur and what implications this has for how a given country should determine its educational policies when its own unique context is considered.

### Measures of Inequality

There were modest differences in the results depending on the measure of inequality used. However, before discussing these differences, we want to emphasize the broad consistencies. First, the direction of the relationship between inequality and performance was always negative regardless of the measure used or the model used to test the relationship. Second, each measure of inequality was significantly negative for at least one achievement domain in each model. Nevertheless, we did observe modest differences. Primary among them was that the relative measures of variance (Gini) and stratification (ICC) were similar in size and routinely larger than the absolute measure (95th – 5th percentile). This may be due to the relative measures having proportional scale invariance while the absolute measures do not (Handcock & Morris, 1999). Given this property, it may be that the relative measures are more clearly comparable across time and context than the absolute measures.

### Education Policy Consideration and Limitations Given the Current Evidence

Our research findings are consistent with a broader set of research (e.g., Checchi et al., 2014; Hanushek & Wößmann, 2005; Micklewright & Schnepf, 2007; Van de Werfhorst & Mijs, 2010) that has questioned the value of educational policies, at both a state or nation level, that promote school differentiation, and thus there is a continued need to consider aspects of government policy related to decentralization, private or privatized schooling, and tracking. All these policies promote stratification by ability and as such do not appear to lead to higher average academic ability. As noted previously, countries such as Sweden and Iceland have increased decentralization and school choice and have seen notable declines in performance, while Germany has moved toward increased centralization and seen an increase in performance. However, readers should consider three caveats when interpreting what our results suggest for policy in a given country.

First, average PISA achievement is only one measure of an education system's performance, and the achievement tests on which our results are based are low stakes. Speaking against this concern is modeling that implies that improvements in PISA scores are linked with real-world outcomes such as economic growth (see Hanushek & Wößmann, 2010). Nevertheless, future researchers might want to consider a wider range of outcomes. For example, in Germany, though tracking is associated with poorer average achievement, retention through the full program of study is high (Checchi et al., 2014).

Likewise, readers should consider if policies and social change at other levels of society may require an increase in decentralization and school choice or at least make such policies more appealing. As Friedman (2002) notes, school choice may be one of the only, or at least one of the most effective, means of reducing educational inequality in the face of increasing geographic segregation by income. School choice could do this by providing children in very poor regions' access to high-quality schools in other districts. Indeed, countries like the United States have seen exceptional increase in geographic segregation in recent years (Owens, Reardon, & Jencks, 2016), and thus, there is good opportunity to test Friedman's hypothesis. However, readers should be aware that initial empirical evidence suggests that school choice in the context of residential segregation may actually exacerbate inequality for disadvantaged children (Saporito, 2003).

While the multicohort evidence presented here is a step forward over previous cross-sectional evidence, the results should not be taken as indicative of causation. In particular, the causal direction is unclear because we concern ourselves only with associations between trends in achievement and inequality for the same cycles. Our results show that a country can combine both excellence and low inequality (see also Simola, 2005). However, it is not certain that inequality leads to poorer performance or whether poorer performance limits the scope for countries to focus more closely on issues of inequality. Likewise, the correlation between excellence and inequality may be a proxy for other factors. In particular, social structure, not school structure, could drive these results—although previous research suggests this is unlikely (Dupriez & Dumay, 2006). More probable is that changes in funding between schools or between regions within countries could account for these results (Owens et al., 2016). Likewise, changes in school-to-school or regional differences in school quality could account for our findings.

### Age of First Selection and Other Challenges to Our Conclusions

A notable challenge to our interpretation of the results presented here is that they compare systems with different ages of first selection (Pfeffer, 2015), that is, the age at which students are streamed into different tracks. Thus, for example, PISA tests students at age 15, and yet a number of OECD countries begin tracking students at age 16 (Bol & Van de Werfhorst, 2013; Pfeffer, 2015). This has implications for our interpretation of the results for Poland that, as part of the reform of the education system, lifted the age of first selection from 15 to 16 years of age (Jakubowski et al., 2010). Thus, a criticism of our work is that systems that do not track before age 15 are merely delaying the inevitable. Indeed, Jakubowski et al. (2010) showed that achievement differences between vocational and academic track Polish students increased at age 16, after the change in policy, in the same way they did before the policy change. There are several points to be made here. First, Pfeffer's results are similar in conclusion to ours despite focusing on the adult population. Namely, there appears to be little evidence that inequality is a necessary condition for excellence (or quality in Pfeffer's terminology; see literature review) when measured after schooling. Second, even if it is the case that inequality observed in differentiated systems eventually emerges in late tracking systems (Jakubowski et al., 2010), it is certainly not clear that the achievement advantage that late tracking countries have over early tracking countries disappears when considering older samples. Again, the consistency between ours and Pfeffer's results would seem to indicate that this fear is unfounded.

## Conclusion and Future Directions

This paper, in combination with a growing amount of cross-sectional empirical research, provides compelling evidence that a negative relationship exists between average academic excellence and inequality. This is a problem for policies that promote decentralization, school choice, privatization, and segregation. However, future research and theory needs to explain why this negative relationship exists and under what social conditions it holds. Furthermore, there is clearly a need for research that further evaluates how changes in variance at different points in the achievement distribution affect average achievement. Put simply, research needs to determine whether and when policies directed toward those in the bottom half of the distribution are most effective. There is a need for researchers to consider what forces are behind changes in variance over relatively short periods of time in some countries (the current study covers only a single decade). In particular, further in-depth analysis of countries that have shown large change in achievement and inequality are needed to unpack the various structures and policies that lead to increases or decreases in inequality.

Finally, there is a need for longitudinal versions of large-scale assessments such as PISA to determine long-term outcomes of inequity and excellence. As PIAAC (the adult skills assessment version of PISA) develops, linking PISA and PIAAC in a synthetic panel design may have advantages.

Alternatively, assessments that incorporate a larger number of age groups and at different points in their schooling careers will be important to overcome difficulties associated with country differences in the age of first selection. In particular, as Pfeffer (2015) argues, large-scale assessment that includes the final year of compulsory schooling are needed. Nevertheless, utilizing multiple cycles of PISA, as we do here, provides a means of focusing attention on within-country changes (where policy contexts tend to be less variant).

## ORCID iD

Philip D. Parker https://orcid.org/0000-0002-4604-8566

## Notes

We wish to acknowledge the generous support of the Australian Research Council (DE140100080).

[1]The use of OECD countries excluding Mexico and Turkey is relatively common (e.g., Micklewright & Schnepf, 2007; Parker, Jerrim, Schoon, & Marsh, 2016). The reason for this is (a) considerable differences between Mexico and Turkey and the rest of the OECD in GDP and human development indexes and (b) a large number of not at school youth in these countries at the age of interest leading to potential systematic bias in estimates.

## References

Bol, T., & Van de Werfhorst, H. G. (2013). Educational systems and the trade-off between labor market allocation and equality of educational opportunity. *Comparative Education Review*, *57*, 285–308.

Breakspear, S. (2012). *The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance* (OECD Education Working Papers, No. 71). Paris: OECD Publishing.

Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, *22*, 782–861.

Bryan, M. L., & Jenkins, S. P. (2015). Multilevel modelling of country effects: A cautionary tale. *European Sociological Review*, *32*, 3–22.

Burger, K. (2016). Intergenerational transmission of education in Europe: Do more comprehensive education systems reduce social gradients in student achievement? *Research in Social Stratification and Mobility*, *44*, 54–67.

Checchi, D. (2006). *The economics of education: Human capital, family background, and inequality*. Cambridge, UK: Cambridge University Press.

Checchi, D., van de Werfhorst, H., Braga, M., & Meschi, E. (2014). The policy response: Education. In W. Salverda, B. Nolan, I. Marx, A. McKnight, I. G. Tóth, & H. van de Werfhorst (Eds.), *Changing inequalities and societal impacts in rich countries: Analytical and comparative perspectives* (pp. 294–327). Oxford, UK: Oxford University Press.

Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (in press). Effects of school-average achievement on individual self-concept and

achievement: Unmasking phantom effects masquerading as true compositional effects. *Journal of Educational Psychology*.
Dupriez, V., & Dumay, X. (2006). Inequalities in school systems: Effect of school structure or of society structure?. *Comparative Education*, *42*, 243–260.
Friedman, M. (2002). *Capitalism and freedom*. Chicago, IL: University of Chicago Press.
Friedman, M., & Friedman, R. (1980). *Free to choose*. New York, NY: Harcourt.
Gans, J., & King, S. (2014). *Finishing the job: Real world policy solutions in health, education, and transport*. Melbourne, Australia: Melbourne University Press.
Goldthorpe, J. H. (2007). *On sociology* (2nd ed.). Stanford, CA: Stanford University Press.
Handcock, M., & Aldrich, E. M. (2002). *Applying relative distribution methods in R*. Seattle, WA: Center for Statistics and Social Sciences, University of Washington. Retrieved from https://www.csss.washington.edu/Papers/wp27.pdf
Handcock, M. S., & Morris, M. (1999). *Relative distribution methods in the social sciences*. New York, NY: Springer.
Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, *18*, 527–544.
Hanushek, E. A., & Wößmann, L. (2005). Does educational tracking affect performance and inequity? Differences-in-differences evidence across countries. *The Economics Journal*, *116*, 63–76.
Hanushek, E. A., & Wößmann, L. (2010). *The high cost of low educational performance: The long-run economic impact of improving PISA outcomes*. Paris: OECD Publishing.
Hoxby, C. M. (2003). School choice and school productivity. Could school choice be a tide that lifts all boats? In C. M. Hoxby (Ed.), *The economics of school choice* (pp. 287–342). Chicago, IL: University of Chicago Press.
Jakubowski, M., Patrinos, H. A., Porta, E. E., & Wisniewski, J. (2010). *The impact of the 1999 education reform in Poland*. Washington, DC: World Bank.
Jerrim, J., Chmielewski, A. K., & Parker, P. (2015). Socioeconomic inequality in access to high-status colleges: A cross-country comparison. *Research in Social Stratification and Mobility*, *42*, 20–32.
Kerckhoff, A. C. (1995). Institutional arrangements and stratification processes in industrial societies. *Annual Review of Sociology*, *21*(1), 323–347.
Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives*, *2*, 99–106.
Marks, G. N. (2006). Are between-and within-school differences in student performance largely due to socio-economic background? Evidence from 30 countries. *Educational Research*, *48*, 21–24.
Micklewright, J., & Schnepf, S. (2007). Inequalities in industrialised countries. In S. P. Jenkins & J. Micklewright (Eds.), *Inequality and poverty re-examined* (pp. 129–145). Oxford, UK: Oxford University Press.
Ministry of Education, Science and Culture. (2014). *Review of policies to improve the effectiveness of resource use in schools: Country background report Iceland*. Retrieved from https://tinyurl.com/y95pk96d
Organisation for Economic Co-operation and Development. (2004). *PISA 2003 technical report*. Paris: OECD Publishing.
Owens, A., Reardon, S. F., & Jencks, C. (2016). Income segregation between schools and school districts. *American Educational Research Journal*, *53*, 1159–1197.

Parker, P. D., Jerrim, J., Schoon, I., & Marsh, H. W. (2016). A multination study of socioeconomic inequality in expectations for progression to higher education: The role of between-school tracking and ability stratification. *American Educational Research Journal*, *53*, 6–32.

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. doi:10.1037/edu0000215

Pfeffer, F. T. (2015). Equality and quality in education. A comparative study of 19 countries. *Social Science Research*, *51*, 350–368.

Ringarp, J., & Rothland, M. (2010). Is the grass always greener? The effect of the PISA results on education debates in Sweden and Germany. *European Educational Research Journal*, *9*, 422–430.

Rowe, E. E., & Lubienski, C. (2017). Shopping for schools or shopping for peers: Public schools and catchment area segregation. *Journal of Education Policy*, *32*, 340–356.

Ryan, C. (2013). What is behind the decline in student achievement in Australia? *Economics of Education Review*, *37*, 226–339.

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*, 405–423.

Saporito, S. (2003). Private choices, public consequences: Magnet school choice and segregation by race and poverty. *Social Problems*, *50*(2), 181–203.

Simola, H. (2005). The Finnish miracle of PISA: Historical and sociological remarks on teaching and teacher education. *Comparative Education*, *41*, 455–470.

Thomas, V., Wang, Y., & Fan, X. (2001). *Measuring educational inequality: Gini coefficients of education* (Policy Research Paper 2525). Washington, DC: World Bank.

Van de Werfhorst, H. G., & Mijs, J. J. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology*, *36*, 407–428.

Walberg, H. J. (2000). Market theory of school choice. *Education Week*, *19*, 46–49.

Manuscript received November 13, 2016
Final revision received January 27, 2018
Accepted January 28, 2018