Highlights

Non-cognitive peer effects in secondary education

Nikki Shure

- Having more conscientious classmates improves individual math and language performance
- The effect of conscientious peers works through their average level
- Peers' non-cognitive traits have a large impact on individuals' learning outcomes
- The concept of 'peer effects' is extended to include 'Big Five' personality traits

Non-cognitive peer effects in secondary education

Nikki Shure[1]

Abstract

The peer effects literature has established that peers impact each other in the classroom through academic achievement and cognitive ability, but has not explored many alternative channels. This paper examines how the non-cognitive traits of peers in the classroom impact an individual's learning outcomes. I estimate a linear-in-means model and alternative models of peer effects with additional peer effects terms accounting for "Big Five" personality traits. Controlling for selection into schools, cognitive and non-cognitive ability, and family background, there is a significant, positive relationship between average peer conscientiousness and individual academic performance of the order of a 0.15 standard deviation increase in math scores and a 0.12 standard deviation increase in language scores. This is the first evidence relating non-cognitive traits to peer effects in secondary school and lends support for programs in schools targeting the development of non-cognitive skills.

Keywords: educational attainment, non-cognitive skills, peer effects, personality

JEL: I20, J20, J24

Email address: `nikki.shure@ucl.ac.uk` (Nikki Shure)

[1]UCL Social Research Institute and IZA

[2]I would like to thank conference participants at the IWAEE 2016, Sandra McNally, Patricia Rice, Margaret Stevens, Brian Bell, Ulf Zölitz, John Jerrim, Katarzyna Kubacka, Hedvig Horvath, and Anna Adamecz-Völgyi for their helpful comments and Jan Van Damme, Bieke De Fraine, and the LOSO Project Team for giving me permission to use the data.

1. Introduction

From smoking, teen pregnancy, and other risky behaviors, to grades and labor market outcomes, friends and acquaintances impact each other, for better or for worse. Much of the economics literature on peer effects, however, has focused on how the past academic performance of peers impacts an individual's learning outcomes. The standard peer effects literature has used mostly administrative data sets from the United States and found a range of results from positive, non-linear effects (Vigdor and Nechyba, 2007) to positive effects at the class, but not grade level (Betts and Zau, 2004; Burke and Sass, 2013) to positive effects with heterogeneity (Hoxby, 2000) of past peer academic performance on later individual outcomes.[3] The results of non-linearity and heterogeneity in peer effects models indicate that how peers impact each other is a complex process, yet there has been almost no exploration of how peers impact each other in the classroom beyond their past academic performance.

In this paper, I use longitudinal data from Flanders, Belgium to look at the relationship between non-cognitive peer effects and academic performance in mathematics and language.[4] This focus on non-cognitive peer effects fits within Sacerdote's (2011) "broad definition of peer effects to encompass nearly any externality in which peers' backgrounds, current behavior, or outcomes affect an outcome" and it is the first paper to explicitly look at how non-cognitive traits of peers impact individual learning outcomes in secondary school. This is important since secondary school is a key formative period in a young person's life and the results produced in this paper can speak to a broad audience of educators. I build on the existing peer effects literature by explicitly using measures of non-cognitive peer and individual level traits in my analysis. I use teacher assessments of personality to construct measures of conscientiousness, agreeableness, and extroversion based on the "Big Five taxonomy," first coined by Goldberg (1981). As noted by Golsteyn et al. (2021), this is the first paper to explicitly consider non-cognitive peer effects.

In this paper, non-cognitive ability is closely related to personality and "soft skills," which

[3] Ammermueller and Pischke (2009), Feld and Zölitz (2019), Gibbons and Telhaj (2015), Lavy, Silva, and Weinhardt (2015) find similar positive results across Europe.

[4] It should be noted that "language" refers to a test in Dutch, which is also the language of instruction.

Table 1: The Big Five Domains

Factor	Definition of Factor
I. Openness to Experience (Intellect)	The tendency to be open to new aesthetic, cultural, or intellectual experiences.
II. Conscientiousness	The tendency to be organized, responsible, and hardworking.
III. Extraversion	An orientation of one's interests and energies toward the outer world of people and things rather than the inner world of subjective experience; characterized by positive affect and sociability.
IV. Agreeableness	The tendency to act in a cooperative, unselfish manner.
V. Neuroticism (Emotional Stability)	Neuroticism is a chronic level of emotional instability and proneness to psychological distress. Emotional stability is predictability and consistency in emotional reactions, with absence of rapid mood changes.

Source: American Psychological Association Dictionary (2007) in Almlund et al. (2011)

Heckman and Kautz (2012) define as "personality traits, goals, motivations, and preferences that are valued in the labor market, in school, and in many other domains." To provide further information on the Big Five, I reproduce Table 1 from Almlund et al. (2011) as Table 1. As Almlund et al. (2011) point out, this is a widely used and accepted taxonomy of personality.

In order to conceptualize how these non-cognitive peer effects might impact learning outcomes, I draw on a range of psychological literature and the standard educational production function. As Almlund et al. (2011) discuss, attainment is a function of IQ, effort, and environment. We can think of non-cognitive peer effects entering the education production function through either effort (e.g. in terms of whether or not peers motivate each other to succeed) or environment (e.g. if peers' behavior causes disruption to learning or if teachers are particularly motivated by engaged pupils). Golsteyn et al. (2017) helpfully note that when the environment is affected, the efficiency of learning is affected. In this context, peers' Big Five personality traits may impact the environment (learning efficiency) and individual effort, which I discuss further in the Empirical Strategy section.

I use these non-cognitive measures to estimate three sets of models, beginning with the linear-in-means model (LIM), as well as several robustness checks. The assumption underlying the LIM is that having higher ability peers on average impacts individual level outcomes. Second, I estimate models that include a mean peer effect term and variance peer effect term in order to disentangle the effect of the tightness of the ability distribution from the mean on outcomes. There are differing schools of thought on how variance in ability can impact individual outcomes. Sacerdote (2011) highlights the "rainbow" and the "boutique" models. The rainbow model posits that high levels of variance, i.e. diversity, is beneficial, while the boutique model posits that a similar level of ability is beneficial because it allows teachers to teach at one level. Policies of tracking or streaming pupils based on ability are common in secondary school (Slavin, 1990) and reflect the boutique model. They work on the assumption that a class with a lower variance, i.e. more similar pupils, positively affects a pupil's outcome. Including the variance terms allows me to test this.

Third I estimate models of peer effects that focus on the impact of having pupils with very high or very low non-cognitive skills in a classroom. The "shining light" model hypothesizes that having very good pupils in the classroom serves as an example for the others and thereby pulls them all up (as they increase effort), while the "bad apple" model hypothesizes that having very poor performing or ill-behaved pupils disturbs learning in such a way to drag her peers down (decreasing efficiency) (Sacerdote, 2011). The bad apple model of peer effects has received considerable attention in the literature (e.g. Lazear, 2001). In order to test the validity of these two hypotheses, I estimate a model that includes the percent of pupils in each class in the bottom and top quintiles of the entire cohort's ability distribution. This is important since there may be differential returns to very high or very low values of these traits.

Golsteyn et al. (2021) is the only other paper to examine the importance of peer personality on individual learning outcomes. They exploit random allocation to university seminar groups at Maastricht University to examine how peers' persistence, self-confidence, anxiety, and risk attitude impact test outcomes. Their findings show that being exposed to more persistent peers positively affects performance while being exposed to more risk taking peers negatively affects performance. My paper is different from Golsteyn et al. in several key ways. First, I use data from compulsory schooling, which is arguably more representative of how peer effects might work for a broad swath of the population. Secondary school is a key point in a young person's life before transitioning into further education or the labor market, so understanding peer effects in this context is important. Second, I use school fixed effects as the identification strategy. Most secondary schools do not use random allocation to classrooms. The evidence generated by this paper is therefore useful to a broad range of policymakers and educators who do not plan on randomly allocating students to classes. Third, I use Big Five personality traits. These measures have been used in range of papers across disciplines, which allows for comparability of results. They are also easy to collect given recent advancements in 10 question versions, which only take one minute to administer (Rammstedt and John (2007) and Hahn et al. (2012)).

Neidell and Waldfogel (2010) also look at non-cognitive peer effects in their paper examining the impact of peer group kindergarten attendance on first grade learning outcomes. Unlike this paper, they use measures of peer behavior as predictors of individual learning outcomes. Their findings show that high peer externalizing problems, behaviors associated with "defiance, impulsivity, disruptiveness, aggression, antisocial features, and overactivity" (Hinshaw, 1992), negatively affect learning outcomes. Apart from their paper, most other peer effects studies looking beyond the traditional channel of past academic performance have focused on gender (e.g. Hoxby (2000) and Lavy and Schlosser (2011)). Hoxby (2000) shows that having a higher proportion of girls in a classroom positively impacts learning outcomes. The mechanism through which this works, however, has yet to be fully explained. One potential explanation is that girls and boys have on average different strengths and weaknesses in terms of non-cognitive skills. Bertrand and Pan (2011) review the medical and psychological evidence that points to this, e.g. the fact that boys are statistically more likely to be diagnosed with attention deficit hyperactivity disorder (ADHD). Aizer (2009) explicitly looks at ADHD and exploits exogeneity in access to treatment as a way to change peer behavior. She finds that improvements in peer behavior improve academic outcomes, which indicates that peer behavior is perhaps just as important as peer cognitive ability in determining outcomes, supporting the mechanism of environment affecting learning efficiency. This paper contributes to the peer effects literature by being the first to explicitly model how personality traits affect learning outcomes in secondary school over and above traditional academic peer effects in a context relevant to most young people.

There is a related literature on the linkage between non-cognitive traits and learning and labor market outcomes. The growing evidence out of this literature shows that non-cognitive ability may matter just as much or even more than cognitive ability for labor market outcomes. Borghans et al. (2008) and Almlund et al. (2011) provide extensive overviews of this literature.[5] Heckman and Rubenstein (2001) provide evidence on the importance of non-cognitive skills on "General Educational Development" (GED) test outcomes, a high school diploma equivalency exam, in the United States. They do not identify any one non-cognitive skill in their analysis, but find that GED recipients earn less, have lower hourly wages, and obtain a lower level of schooling than high school dropouts, which they attribute to lower non-cognitive skills. Not only are non-cognitive traits important for future outcomes, but they can be developed, which makes them of interest to policy makers. Heckman et al. (2013) show that non-cognitive traits are malleable in a long-term way using data from the Perry Preschool Project.

The data used in this paper allows me to account for many of the pitfalls commonly associated with the peer effects literature. I account for the endogeneity of peers in part through a school fixed effects approach and in part through the construction of the data set. All of the pupils in this data set are starting secondary school at a new school with mostly new peers, which enables me to avoid many of the problems that arise from administrative data including the "reflection problem" (Manski, 1993). The reflection problem arises because peers influence each other over time, which means that in most cases, an individual will have helped to shape the peer effect term used in analysis. Because my peer effects are calculated using individual level measures collected from the end of primary school, most of the new peers in their secondary school will not have influenced each other over time. I show that most peers in math and language classes did not attend primary school together and that my results are robust to excluding those who did know many of the peers from primary school.

[5] Borghans et al. (2008) make an important distinction regarding non-cognitive traits as compared to much of the economics literature in this area. They purposefully avoid the distinction between "cognitive" and "non-cognitive" traits because ultimately every aspect of human behavior relies on some degree of cognition. Instead, they focus on personality with a significant amount of focus placed on the Big Five. Since much of the economics literature, especially work by Heckman, uses the division between cognitive and non-cognitive, I choose to do the same. I am aware, however, of the discussion in the literature as to the validity of terms.

The data set also includes a rich array of socio-economic status (SES) measures, personality measures, and an intelligent quotient (IQ) test, allowing me to control for much of the unobserved heterogeneity between pupils and account for most, if not all, of the selection into classrooms. I run a series of robustness checks around the issue of selection into classrooms. They show that individuals are not selected into classes on the basis of their non-cognitive traits. Importantly, while peer non-cognitive traits impact learning outcomes at the end of the school year, they do not impact tests taken at the very beginning of the school year.

I find that having more conscientious peers in a classroom is positively related to individual math and language performance. I also find that having more extraverted peers in a classroom is negatively associated with individual math performance. I find no support for peer IQ or past subject performance positively impacting results in the standard linear-in-means framework. I also find limited support for alternative models of non-cognitive peer effects. These peer effects seem to work primarily through the class average. Overall, the findings in this paper show that peers influence each other's learning outcomes in ways beyond the traditional channels of IQ and past subject performance.

The rest of this paper is structured as follows. In section 2, I discuss the data used and some key stylized facts. I then present the models estimated and the identification strategy in section 3, which I follow with a discussion of the estimates obtained in section 4. In section 5, I discuss the conclusions.

2. Data and descriptive statistics

The data used in this paper comes from an educational study carried out in Flanders, Belgium in the 1990s, known as the "Longitudinaal Onderzoek in het Secundair Onderwijs"[6] or in English, "Longitudinal Research in Secondary Education" Project (henceforth referred to as the "LOSO Project"). Van Damme et al. (2002) provide an overview of the project and its aim to assess the effectiveness of secondary schooling in Flanders, Belgium. The project began in 1990 with data collection on more than 6,000 pupils, who were all 12 years old at

[6]The LOSO data were collected by Professor Jan Van Damme (KU Leuven) and financed by the Flemish Ministry of Education and Training as part of the OBPWO program, on the initiative of the Flemish Minister of Education.

the time and about to begin secondary school. These pupils comprise a cohort beginning secondary school in 1990, and attending 57 different secondary schools[7] from three regions of Flanders (Van Damme et al., 2002). For each of the 57 schools sampled, the entire cohort of first year pupils is surveyed; I only look at the first year of secondary school.

Table 2 provides descriptive statistics on the sample of pupils used in this paper. Just under half of the sample are boys (47.5 percent), one in five has a father with a university degree, and less than ten percent have a least one foreign born parent. My variables of interest are the teacher reported non-cognitive measures for each pupil and the corresponding non-cognitive peer effect terms of agreeableness, conscientiousness, and extraversion. Each pupil's teacher from their primary school was asked to assess the non-cognitive ability of the pupil before they started secondary school. I use this assessment and an exploratory factor analysis to construct three non-cognitive measures based on the Big Five taxonomy and then validate them.[8] These individual measures as well as their peer effects terms have been standardized to mean zero, standard deviation one for the estimation sample, so that all regression results may be interpreted in terms of effect sizes.

Approximately half of the individuals in the LOSO study have the measures of personality provided by their primary school teacher, which gives me an estimation sample of 3,174. This has to do with the inability of the LOSO researchers to reach all primary school teachers for each pupil. Since the primary schools did not participate in the LOSO project, there was less scope to track down every primary school teacher. These missing values are most likely not completely missing at random and pose a significant challenge to identification in the context of peer effects (Sojourner, 2013). To check this, I explore the distribution of missing data and the characteristics of individuals with missing data in Appendix C. I run balance

[7]It should be noted that not every secondary school in Flanders is part of this study. As the investigators of the study state: "the sample of students was taken from almost all schools of three regions in Flanders. The set of schools is to a *certain extent* representative of the Flemish secondary schools in general" [emphasis added] (Van Damme et al., 2002). Schools were selected based on the following criteria: the size of the school, the type of school, the curriculum offered, and the religious affiliation of the school (Van de gaer et al., 2009). Once schools were selected, consent was sought from pupils and their parents; less than one percent declined to participate in the study (Van de gaer et al., 2009).

[8]The actual measures used in this analysis were developed in conjunction with Dr. Katarzyna Kubacka from the Organization for Economic Co-operation and Development as part of their "Education and Social Progress" project. See Appendix A for a more detailed description of how these measures were calculated and validated.

tests on the individuals who have the non-cognitive measures and those who do not. The results of these tests are shown in Table C.12. They show that the students with missing data tend to be lower ability and from more disadvantaged backgrounds, which should be kept in mind when interpreting the complete case results, which are based on a complete case analysis, often referred to as the individual deletion procedure (IDP) in the peer effect literature (Sojourner, 2013). Importantly, the distribution of individuals with missing data across schools is relatively even. There is no school with less than 15 percent of individuals with missing data and the distribution of pupils with missing data across classes is roughly normal (see Figure C.2 in Appendix C).

In order to address this missing data issue more robustly, I take three main approaches. First, I control for the share of peers with missing data. Second, I impute the missing non-cognitive traits using multiple imputation (Rubin, 1976; 1987) and re-run all of the main models. Third, I re-estimate all models using inverse probability weights to account for the likelihood of missing data. These results may be found in Appendix C and show that the main results in this paper hold.

At the beginning of the LOSO Project, all pupils took the Getlov battery for intelligence, an IQ test (Lancksweerdt, 1991). Again, the descriptive statistics for this measure are available in Table 2. In this paper, I use the IQ test to measure underlying cognitive ability. While there is still some debate as to how accurately IQ measures cognitive ability amongst researchers who focus on latent factor models, it is a better measure than previous grades or subject specific test scores. In Table 3, I present the correlations between the three non-cognitive measures, conscientiousness, agreeableness, and extraversion, and the overall IQ test score. This table shows that the three non-cognitive measures are relatively weakly correlated with one another and that conscientiousness is most correlated with IQ. This correlation between conscientiousness and IQ has been found in other studies (Borghans et al., 2008).

The outcome measures I use in this paper are scores on math and language (Dutch) achievement tests created by the LOSO Project team and thus standardized across all schools, unlike teacher assessed grades used in other studies. These tests were administered at the beginning and at the end of the first year of secondary school, which allows me to see how a

Table 2: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Pupil level					
Conscientiousness	0	1	-2.42	1.223	3174
Agreeableness	0	1	-3.139	0.943	3174
Extraversion	0	1	-3.147	1.224	3174
Math score at start of school year	0	1	-7.131	5.023	3174
Math score at end of school year	0	1	-6.563	2.351	3174
Language score at start of school year	0	1	-8.076	5.278	3174
Language score at end of school year	0	1	-6.782	2.998	3174
IQ	0	1	-4.172	2.911	3174
Male	0.475	0.499	0	1	3174
Father has tertiary education	0.218	0.413	0	1	3174
At least one foreign parent	0.073	0.261	0	1	3174
Income category 1	0.01	0.098	0	1	3174
Income category 2	0.14	0.347	0	1	3174
Income category 3	0.273	0.445	0	1	3174
Income category 4	0.266	0.442	0	1	3174
Income category 5	0.168	0.374	0	1	3174
Income category 6	0.143	0.35	0	1	3174
School and class level					
Grade size	118.86	76.868	17	363	57
Number of math classes	6.263	3.368	2	18	57
Number of language classes	6.211	3.39	2	18	57
Math class size	19.108	5.221	5	29	334
Language class size	19.253	5.155	5	29	332

NB: Cognitive and non-cognitive measures have been standardized to be mean 0, standard deviation 1 using the full sample. The variable for income is coded into six categories where 1 means the family income is less than 25,000 Francs per month; 2 between 25,000-40,000 Francs; 3 between 40,000-60,000 Francs; 4 between 60,000-80,000 Francs; 5 between 80,000-100,000 Francs; and 6 more than 100,000 Francs per month. This is the net family income and includes pensions, unemployment and other benefits, deductions for taxes, and an allowance for children. When this data was collected in 1990, the Belgian Franc was still in use. Statistics from the St. Louis Federal Reserve indicate that in September 1990, one U.S. Dollar was worth 32.28 Francs (Federal Reserve of St. Louis, 2006). This means a net income of 100,000 Francs per month in 1990 was equivalent to 3,097.89 USD.

Table 3: Correlations between Non-cognitive Measures and IQ

	Conscientiousness	**Agreeableness**	**Extraversion**	**IQ**
Conscientiousness	1			
Agreeableness	0.455***	1		
Extraversion	0.532***	0.358***	1	
IQ	0.627***	0.218***	0.265***	1

NB: N = 3174. *** p<0.01, ** p<0.05, * p<0.1

new group of peers impact the end of school year test outcomes while controlling for the initial test score in a value added framework. Since both tests were created by the LOSO Project team and "Item Response Theory" (IRT) scores[9] were computed using the raw scores, the assessments are comparable across schools and time.

The parents of the pupils in the study also filled out a background questionnaire at the beginning of the secondary school year, which covered parental education and other family characteristics. There is a large literature on which variables best represent SES and how these are then related to educational outcomes. Sirin (2005) conducts a meta-analysis of all papers that have been published between 1990-2000 to examine the strength of various measures in explaining educational achievement. He acknowledges the "tripartite nature of SES that incorporates parental income, parental education, and parental occupation as the three main indicators of SES" (Sirin, 2005), which is how I determine which variables to include in my analysis. I include a binary variable for whether or not the father of the pupil attended tertiary education and information on family income.

Because the peer effects I am interested in occur at the school and class level, I also present some key descriptive statistics for these levels. Table 2 presents information on the size of each cohort beginning secondary school at each of the 57 schools (119 pupils on average), along with the average number of pupils in each math and language class at each school (approximately 19 for both). This table also includes information on the number of math and language classes at each school. It is important for my identification strategy that I have at least two classes at every school since I am including school fixed effects; any school with just one language or math class will otherwise be dropped from my estimation. As these descriptives show, however, all of my schools have at least two language and math classes, which means I am able to include all schools from the LOSO Project in my analysis. Unfortunately I do not have information on the teacher, which limits the ability to verify if there is selection of teachers to certain classes or include controls for teacher characteristics

[9]IRT scores are preferred in the educational assessments literature due to the fact that they take the difficulty and discrimination of the test question into account (DeMars, 2010). IRT scores for the LOSO Project were calculated by their research team from raw test scores using a one parameter logistic model commonly known as the Rasch Model. The scores are standardized in such a way that the mean is zero and the standard deviation is one, which implies that having a negative score is below average and a positive score above average.

in any of the models.

Table 4 includes information on the class averages for the key cognitive and non-cognitive measures of interest. This table presents the standard deviations of these peer effects terms at the class level, which is important because of my inclusion of school fixed effects, there needs to be enough variation in these variables within schools. I further explore the variation in the peer effect terms in the bottom panel of Table 4 by presenting the raw variation in each of the non-cognitive and cognitive peer effects terms as well as the variation net of school fixed-effects. This is calculated across all individuals in the estimation sample. Since the peer effect terms have been standardized to mean zero, standard deviation one, the variance of each term is one. While the second column clearly shows that school fixed-effects reduce the variation in the peer effects terms by approximately half, they do not completely eliminate it, which supports the use of this empirical strategy.

3. Empirical strategy

3.1. Framework

As previously mentioned, we can think about non-cognitive peer effects using the standard education production function while drawing on insights from the psychology literature. In the standard education production function, attainment is produced via three inputs: IQ, effort, and environment (Almlund et al., 2011). Following similar notation to Golsteyn et al. (2017), I define the education production function for individual i in class c as:

$$Achievement_{ic} = f(IQ_{ic}, effort_{ic}, environment_{ic}) \tag{1}$$

Peers' Big Five personality traits may impact the environment (learning efficiency) and/or individual effort via the environment in the classroom. A classroom full of well-prepared and attentive classmates may create a positive learning environment where an individual is motivated to do her best. Similarly a classroom full of chatty and disruptive peers may create an environment where it is difficult for an individual to focus and succeed. This implies that the environment is a function of the vector of peers' (non-cognitive) traits:

$$environment_{ic} = f(peer\, traits'_{\text{-}ic}) \tag{2}$$

Table 4: Variance in peer effect terms

Variable	Mean	Std. Dev.	Min.	Max.	N
Math class peer effects					
Mean peer math score at t=0	-0.159	0.863	-3.356	1.376	334
Mean peer IQ	97.872	12.578	60.203	123.551	334
Mean peer conscientiousness	-0.178	0.803	-2.042	1.148	334
Mean peer agreeableness	-0.064	0.486	-1.84	0.878	334
Mean peer extraversion	-0.065	0.53	-2.397	1.245	334
Mean male classmates	0.504	0.338	0	1	334
Variance math score at t=0	0.528	0.734	0.027	5.727	334
Variance IQ	98.809	47.811	17.911	396.868	334
Variance conscientiousness	0.541	0.358	0	2.851	334
Variance agreeableness	0.939	0.528	0.028	3.21	334
Variance extraversion	0.875	0.445	0	2.909	334
Language class peer effects					
Mean peer language score at t=0	-0.154	0.876	-3.475	1.531	332
Mean peer IQ	97.853	12.594	60.203	123.551	332
Mean peer conscientiousness	-0.179	0.802	-2.042	1.148	332
Mean peer agreeableness	-0.065	0.487	-1.84	0.878	332
Mean peer extraversion	-0.065	0.531	-2.397	1.245	332
Mean male classmates	0.503	0.338	0	1	332
Variance language score at t=0	0.454	0.644	0.01	6.069	332
Variance IQ	99.25	47.745	17.911	396.868	332
Variance conscientiousness	0.544	0.359	0	2.851	332
Variance agreeableness	0.939	0.523	0.028	3.21	332
Variance extraversion	0.878	0.444	0	2.909	332
Variable	**Variance**	**Variance net of school FE**			**N**
Math class peer conscientiousness	1	0.524			3174
Math class peer agreeableness	1	0.585			3174
Math class peer extraversion	1	0.604			3174
Math class peer math performance t=0	1	0.461			3174
Math class peer IQ	1	0.461			3174
Language class peer conscientiousness	1	0.521			3174
Language class peer agreeableness	1	0.583			3174
Language class peer extraversion	1	0.602			3174
Language class peer language performance t=0	1	0.441			3174
Language class peer IQ	1	0.459			3174

Effort is defined as a function of individual traits, including IQ and non-cognitive traits, and the learning environment:

$$effort_{ic} = f(\textit{non-cog}_{ic}, IQ_{ic}, environment_{ic}) \tag{3}$$

Equations 2 and 3 can be substituted into the production function and then the production function can be partially differentiated with respect to peer traits using chain rule. This produces:

$$\frac{\partial A_{ic}}{\partial peer\,traits'_{\text{-}ic}} = \frac{\partial environment_{ic}}{\partial peer\,traits'_{\text{-}ic}}\left[\frac{\partial A_{ic}}{\partial environment_{ic}} + \frac{\partial effort_{ic}}{\partial environment_{ic}}\frac{\partial A_{ic}}{\partial effort_{ic}}\right] \tag{4}$$

This set-up allows for peers' traits to impact the learning environment and therefore attainment. The partial derivative shows both a direct effect, $\partial A_{ic}/\partial environment_{ic}$, and an indirect effect, $(\partial effort_{ic}/\partial environment_{ic})(\partial A_{ic}/\partial effort_{ic})$, of this on attainment. The sign of this indirect effect, however, is ambiguous. Peer traits may either create an environment that stimulates effort or encourages free riding, which makes the direction of the overall peer effect on attainment unclear and will depend on the trait in question (Golsteyn et al., 2017).

This framework can help us predict how individual and peer conscientiousness, agreeableness, and extraversion work in the production function. Drawing on Table 1, conscientiousness manifests in organization, discipline, and achievement-orientation; agreeableness manifests in helpfulness, sympathy towards others, and cooperation; and extraversion manifests in higher sociability, assertiveness, and talkativeness. At the individual level, personality traits enter the production function in a non-monotonic manner (Almlund et al., 2011; Heckman et al., 2006). There is evidence from the psychology literature that Big Five personality traits are related to achievement, work ethic, and motivation (Komarraju et al., 2011; Mammadov et al., 2018). Conscientiousness and agreeableness have been shown to be positively related to four learning styles (synthesis analysis, methodical study, fact retention, and elaborative processing), all of which positively predict attainment (Komarraju et al., 2011) as well as being direct predictors of GPA (Vedel, 2014). This is one mechanism through which conscientiousness and agreeableness could increase individual effort. Conscientiousness has also been shown to be strongly related to persistence, motivation, and grit (Duckworth et al., 2007), which would also imply increased effort. Extraversion, on the other hand, has been found to be less predictive of learning outcomes. In some studies, it negatively predicts certain learning styles associated with attainment, but that this may be dependent on the type of task at hand, i.e. extroverts exhibit worse performance on concrete tasks vs. abstract tasks (Riding and Dyer, 1980). More recent work has highlighted that extraversion may be context specific, limiting the ability to make predictions about the direction of the relationship (Komarraju et al., 2011). At a certain point, however, it is possible that too much of any trait has a diminishing (or even negative) effect on attainment (see Figure 19 in Heckman et al., 2006).

At the peer level, much of the same literature can inform our predictions as to how conscientiousness, agreeableness, and extraversion will affect the environment and therefore an individual's effort. Having classmates with high levels of conscientiousness and agreeableness who employ the aforementioned learning strategies helps to create a positive learning environment, which is collaborative and achievement-oriented (Komarraju et al., 2011). This should increase the efficiency of learning for the individual (direct effect) and also increase her own motivation and effort (indirect effect), making the peer effect positive. As in the case of individual extraversion, the direction of the relationship between peer extraversion and individual attainment is unclear. Having more extraverted peers can create a social environment where individuals feel welcome, but may also be disruptive since higher extraversion leads to increased talkativeness. This may decrease the efficiency of learning strategies (direct effect) and cause the individual to decrease her own effort due to distractions (indirect effect). As previously highlighted, this may vary by context and subject.

3.2. Challenges to identification

In this paper I estimate the "linear-in-means model," the most ubiquitous model in the peer effects literature, and then move on to several alternative models of peer effects. The linear-in-means model includes average peer group ability and average peer group characteristics in addition to individual level variables. The drawback of this model is that it forces all peer effects to enter the model linearly and have the same homogenous impact on all students.[10] Regardless of the type of model used, there are several key empirical challenges that all researchers looking at peer effects must address. These are: simultaneity, selection, and measurement error. Each of these issues poses a challenge to the identification of peer effects and must somehow be taken into consideration.

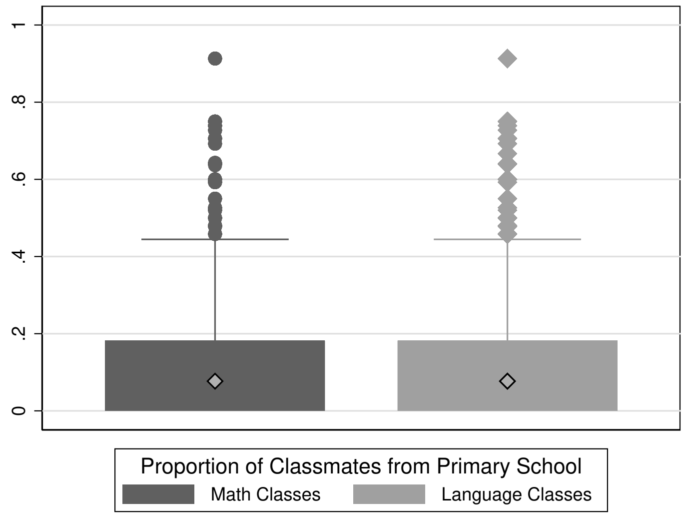
Manski (1993) was the first to raise the issue of simultaneity, which many studies do not fully address, as something he termed the "reflection problem." He uses this term to describe the issue that since peers form a group, any group level outcome will be impacted by all members. This means that using peers' smoking habits to estimate the probability that an individual will smoke is inherently biased by the issue that that individual may also impact her peers' smoking habits. Since peers affect each other, their habits and outcomes evolve in a dependent way over time, which leads to a simultaneity bias. In this sense, Manski points out that most peer effects are actually endogenous.

This is an important concern to keep in mind when estimating peer effects in this context, and something I take into account in my empirical strategy. All of the "peer effect" terms that I include are averages across all pupils in a class, excluding a given individual. This means that every pupil will have a slightly different peer effect term since she is not included in that average. Manski points out that it might not be enough to simply exclude the individual from the peer effect term since this person still influenced her peers. I attempt to avoid this issue by using peer effects measured at the end of primary school or at the beginning of the school year, at a new school before pupils have had time to influence each other. This is possible because of the point in time at which the LOSO data was collected. Children in Flanders attend compulsory education from age six until age 18. Secondary school, the focus of this paper, begins at age 12, when pupils transition to a new school.

In order to explore how many peers a pupil already knew in her class from primary school, I look at the relationship between primary school attendance and secondary school class composition. Since I also know which primary school all of the pupils in the LOSO project attended, I am able to construct a measure of how many peers in a given class went

[10] Angrist (2014) has leveled some more severe criticisms of the linear-in-means model, and the peer effects literature in general, which I attempt to partially address by using more precise measurements of cognitive and non-cognitive ability and by dealing with the reflection problem.

Figure 1: Boxplot of Proportion of Primary School Peers in Secondary School Class



Notes: N=3174

to the same primary school as an individual. This measure tells me what proportion of her peers in a language or math class a pupil knows when starting secondary school. Because of issues associated with the reflection problem, a low proportion of peers from primary school makes my identification strategy cleaner. Figure 1 shows that the median pupil, marked in the graph with a red diamond, knows less than 10 percent of her classroom peers from primary school. This figure also shows that even the pupil at the 75th percentile of the distribution knows less than half of her peers in either language or math from primary school.

Figure 1 provides descriptive evidence that the reflection problem does not pose a major challenge to identification; however, as a robustness check I run the main linear-in-means models on a subsample of individuals who know less than 10 percent of their secondary school classmates from primary school. These results in Appendix E are very similar to the main results presented in Section 4, which should provide reassurance that the reflection problem is not driving the main findings of this paper.

The issue of selection still proves challenging for identification. Here the problem arises because students are not randomly assigned to schools (or potentially even classes) and this leads to correlated effects. This means that students in a particular school may have some other common characteristics, such as socio-economic status or parental education levels, that are actually driving the direction of the peer effects. In general, most research on peer effects may be divided into studies that exploit exogenous variation in assignment to schools or classrooms and studies that attempt to control for selection through the introduction of school, class, or individual fixed effects. Some studies have also relied on variation in gender or ethnic composition of a particular school in consecutive years (e.g. Hoxby, 2000).

In this paper, I rely on school fixed effects to account for selection into schools. These fixed effects will absorb any time invariant characteristics of the schools, e.g. facilities, teaching style, etcetera. One characteristic of the secondary schooling landscape in Flanders is the prevalence of private schools. Many of these private schools are religious, and of the 57 schools surveyed in the LOSO Project, 38 are private and 19 are public (Van Damme et al., 2002). Freedom of school choice is guaranteed by the Belgian Constitution. I take this into account in the empirical analysis through the inclusion of school fixed effects. In practice this means that my identification relies on within school variation. If all pupils at a given school are very similar, which may be the case when pupils select into schools, this limits the strength of my identification strategy. As shown in Table 4, however, there seems to be enough variation in the variation of the key variables of interest within the 57 schools that this does not pose a major challenge to identification.

How classes at secondary school are formed will impact my identification strategy. If classes are formed randomly, which is not the case at most schools, then we do not have to consider what selection mechanisms might be driving formation and thereby affect peer effects. Since classes are generally formed with some criteria in mind, however, we need to be concerned with selection. One possible explanation for how classes are formed is that pupils who sign up for similar elective courses (e.g. Latin) are then placed in the same math or language class. This elective choice will be correlated with cognitive ability since pupils who want to go to university will sign up to take Latin. Another possibility is that pupils with high past academic performance are grouped together. Without the random allocation of students to classes, there is very little that can be done to fully address this; however, I am able to control for elective choice. I am also able to control for a broader range of characteristics than most studies, including IQ, socio-economic status, and non-cognitive traits, which should reduce some concern about selection on unobservables.

To probe the endogenous group formation concern more formally, I run several tests, the results of which are presented in Appendix D. The first test is to examine whether peer non-cognitive traits are related to individual test scores from the beginning of the school year. This is an alternative test for the exogeneity of peer group formation and follows Neidell and Waldfogel (2010). Here the math and language tests taken at the beginning of the school year are regressed on all of the same controls as in Equation 5 (outlined below). None of the coefficients on the non-cognitive peer effects in Table D.17 are statistically significant, indicating that selection into classrooms is not occurring on this basis.

The second test is to compare individuals in above and below peer median conscientiousness classes and examine how the fixed effects approach addresses their potential differences along a range of characteristics. This is based on Neidell and Waldfogel (2010), who also use a fixed effects estimation strategy. Table D.18 shows that although individuals in these two types of classes differ along every dimension, the inclusion of school fixed effects eliminates many of these differences. For the differences that remain, they are greatly reduced in magnitude, highlighting the strength of this identification strategy.

The third tests is to regress each individual non-cognitive trait on the peer average of that trait while also controlling for the school level take out mean of that trait as proposed by Guryan et al. (2009). The idea is to reduce bias since an individual cannot be her own peer and therefore a high and low ability individual at a small school could end up with very different peer effect terms than the same individuals at a larger school. The results of these tests, presented in Table D.19, show that assignment to classrooms is exogenous for agreeableness and extraversion. The coefficients on the peer effects terms for these two traits in Columns (2) and (3) are small, positive, but not statistically significant. The results in Column (1) show a positive and statistically significant relationship between individual level conscientiousness and peer conscientiousness; however, the strength of this relationship is weak (0.2). Taken together, the results of these three tests provide some evidence that endogenous peer group formation based on non-cognitive traits is not a major concern.

The third major challenge to this type of empirical work lies in measurement error. These issues arise because many of the variables used to capture cognitive ability are not very precise. Grades from the previous school year are often used as a measure of ability, but this ignores the fact that grades actually measure something much more complicated than cognitive ability since they often have a subjective component, such as classroom participation, which is a combination of cognitive and non-cognitive ability. Grades also often take into account the teacher's perception of the individual. If peer ability is measured with measurement error, the direction of the bias is unclear (Angrist, 2014; Feld and Zölitz, 2017). The IRT scores I use as outcome and control measures should suffer less from measurement error since they have been standardized by the LOSO researchers to take into account the difficulty and discrimination of test items.

In addition to these three challenges, the previous discussion raised issues of ambiguity in the direction of effects. Cooley Fruehwirth (2014) highlights how to interpret peer effects, even when models are well specified. She acknowledges that peer achievement proxies for a variety of unobservable peer characteristics in many peer effects studies. I am able to address some of these concerns by introducing the non-cognitive peer effects on top of the traditional achievement peer effects. This should reduce some of the issues introduced when the achievement peer effects work alone and thereby proxy unobservables. I probe the sensitivity of my analyses to the addition of proxies in Appendix F. In this appendix, I take two approaches. First, instead of controlling for IQ and performance in that subject at the beginning of the school year, I only control for IQ. These results show that average peer IQ has a positive and statistically significant effect on individual math scores when it is included without peer math performance. This effect decreases in magnitude and loses significance, however, once average peer conscientiousness is introduced.

Second, I attempt to further disentangle how variables may be proxying for each other by "building up" the models in a different order than in the main analysis of the paper. These models begin with just the non-cognitive peer effect terms and the results show, for example, that the magnitude of the peer effect term for conscientious is significantly decreased (approximately by half) once the traditional peer effects for cognitive ability are included. This implies that conscientiousness is also a proxy for ability, but still has a positive effect over and above it. Taken together, these analyses confirm that the findings on peer conscientiousness are robust to the order of inclusion and that average peer cognitive ability, which is found to be significant in other papers, may be proxying for peer non-cognitive traits.

3.3. Models estimated

With all of these considerations in mind, I begin the analysis in this paper with the linear-in-means (LIM) peer effects model:

$$\begin{aligned} Y_{ics,1} &= \beta_1 IQ_{ics,0} + \beta_2 \overline{IQ}_{-ics,0} + \beta_3 NC_{ics,0} + \beta_4 \overline{NC}_{-ics,0} + \beta_5 X_{ics,0} \\ &\quad + \beta_6 \overline{X}_{-ics,0} + \alpha_s + \varepsilon_{icst,1} \end{aligned} \tag{5}$$

In this model, $Y_{ics,1}$, is the outcome measure: grades in either language or math at the end of the first year of secondary school. Here $t = 0$ indicates the beginning of the first year of secondary school and $t = 1$ represents the end of the school year. I include both individual IQ, $IQ_{ics,0}$, measured at the beginning of secondary school, and the class average IQ score, excluding the individual, $\overline{IQ}_{-ics,0}$. The subscript i denotes the individual, while the peer effect terms have the subscript $-i$ to indicate that they are calculated across all individuals except individual i. This means that every individual will have a slightly different peer effect term since they are "leave out means." The vector $NC_{ics,0}$ includes measures of non-cognitive skills as reported by the primary school teacher at the end of primary school and the vector $\overline{NC}_{-ics,0}$ includes the averages of these measures for all the peers in the new secondary school class. Similarly, the vector $X_{ics,0}$ includes measures of background characteristics including father's education, family income, and gender as well as previous math or language performance and the vector $\overline{X}_{-ics,0}$ includes the averages of these measures for all the peers in a class, excluding individual, i. I include school fixed effects, α_s, in order to account for unobserved heterogeneity at the school level. It should be noted that by including school fixed effects, my identification is being driven by the differences across classes. This is why having two or more classes at a school is fundamental to this identification strategy. The descriptive statistics show that there are on average six classes per school and at least two at every school, which allows me to obtain identification.

I will first estimate a series of basic linear-in-means models with school fixed effects. Following this analysis, I will explore alternative peer effects models. These peer effects models allow for variance in ability to enter into the model and also allow for peers in different portions of the ability distribution to have different effects on the individual. The variance model I estimate takes the following form:

$$\begin{aligned} Y_{ics,1} &= \beta_1 IQ_{ics,0} + \beta_2 \overline{IQ}_{-ics,0} + \beta_3 Var(IQ)_{-ics,0} + \beta_4 NC_{ics,0} + \beta_5 \overline{NC}_{-ics,0} \\ &+\beta_6 Var(NC)_{-ics,0} + \beta_7 X_{ics,0} + \beta_8 \overline{X}_{-ics,0} + \alpha_s + \epsilon_{icst,1} \end{aligned} \tag{6}$$

The inclusion of both the mean peer effect term and variance peer effect term for the measures of IQ, non-cognitive ability, and previous math or language performance will allow me to disentangle the effect of the tightness of the ability distribution from the mean on outcomes. This allows me to the test the rainbow and the boutique models.

In order to test the validity of the shining light and bad apple hypotheses, I estimate a model that includes the percent of pupils in each class in the bottom and top quintiles of the entire cohort's ability distribution. I include these quintile variables for both cognitive and non-cognitive measures in the following model:

$$\begin{aligned} Y_{ics,1} &= \beta_1 IQ_{ics,0} + \beta_2 \overline{IQ}_{-ics,0} + \beta_3 PercentTopQuintile(IQ)_{-ics,0} \\ &+\beta_4 PercentBottomQuintile(IQ)_{-ics,0} + \beta_5 NC_{ics,0} + \beta_6 \overline{NC}_{-ics,0} \\ &+\beta_7 PercentTopQuintile(NC)_{-ics,0} + \beta_8 PercentBottomQuintile(NC)_{-ics,0} \\ &+\beta_9 X_{ics,0} + \beta_{10} \overline{X}_{-ics,0} + \alpha_s + \xi_{icst,1} \end{aligned} \tag{7}$$

I cluster the standard errors in all models at the school level to account for the sampling frame (Abadie et al., 2017). In the following Results section, I present the results from the aforementioned models in the same order discussed here.

4. Results

In this section of the paper, I present the results from the models discussed in the Empirical Strategy section. All of the tables in this section report effect sizes due to the standardization of the variables. I first present the tables for the linear-in-means models, followed by the results from the alternative models of peer effects. This includes a table presenting the variance model discussed in the Empirical Strategy section as well as the top and bottom quintile models.

4.1. Linear-in-means models

In each table presented in this section for the linear-in-means model results there are five columns. Each of these five columns represents a slightly different specification of a linear model[11] and the order is consistent across tables for the sake of clarity. Appendix F includes several sensitivity checks to probe the ordering of introducing these variables into the models.

I group the peer effect terms together and the individual level controls together in each table. In Column (1) of each table I run a basic OLS regression that represents a "value added model." This is a standard model within the literature on academic performance. Here math or language achievement at the end of the school year is regressed on past math or language achievement, IQ, gender, SES, and class size. This value added specification gives us an idea of what the baseline model without any inclusion of peer effects looks like and serves as a reference point for comparison.

In Column (2) of each table I introduce individual level non-cognitive measures to the baseline value added model. This extends the traditional value added model because generally such measures of personality traits are not available. I include these measures so that we may differentiate the effect of individual level non-cognitive traits from the effect of non-cognitive peer effects once they are added to the baseline.

Column (3) presents the results of the traditional linear-in-means peer effects model common in the literature. This model includes all of the variables from the baseline in Column (1), but also includes the peer effect averages for IQ, previous math or language performance, income, and proportion of males in the classroom. I do not include any non-cognitive measures in this specification because I want to be able to compare my model with non-cognitive peer effects to the traditional peer effects model. In Column (4) I add individual level non-cognitive measures to the traditional linear-in-means peer effects model in order to differentiate between individual and peer level non-cognitive traits. I am also interested to see if the inclusion of individual level non-cognitive traits changes the results of

[11]This logic does not apply to the tables containing the results for variance model and the top/bottom quintile peer effects model.

the traditional peer effects model.

Finally, in Column (5) of each table I present the results from the full model. This includes all of the baseline variables and traditional peer effects terms plus individual non-cognitive measures and the average of these measures across peers. This is the model of interest as it fully extends the baseline to include both cognitive and non-cognitive peer and individual effects.

The results of the complete linear-in-means model of the math scores in Column (5) show that the only non-cognitive peer effects to have a statistically significant relationship with individual level learning outcomes are conscientiousness and extraversion. The peer effect term for conscientiousness positively and significantly impacts individual level learning outcomes. The effect of a one standard deviation increase in peer conscientiousness is 0.15 of a standard deviation on the math score in the full model, Column (5) of Table 5. This is a moderate effect size within the education literature (Kraft, 2020).

The peer effect for extraversion is statistically significant and negative in Table 5, meaning a one standard deviation increase in average peer extraversion is associated with a -0.07 standard deviation decrease in math scores. If we interpret the peer effect on extraversion to be highly correlated with a disruptive learning environment due to the talkative nature of the pupils, then we would expect peer level extraversion to be negatively related to individual learning outcomes. This result is similar to Neidell and Waldfogel's finding that peer externalizing problems, which they describe as "most likely capture[ing] classroom disturbance", negatively affect learning outcomes. The peer effect of agreeableness is not statistically significant; however, it is positive, which is in accordance with our beliefs about peer agreeableness fostering a positive learning environment.

The peer effect term for IQ is consistently small and not statistically significant. This does not mean that having higher ability peers does not improve learning outcomes, but may be the result of including proxies for IQ in the model (Cooley Fruehwirth, 2013). To probe this finding further, I re-run all models and only include either IQ or prior performance in language or math. I also run a series of models where I build up to the full models presented in Column (5) by adding the various peer effects terms in a different order. These

Table 5: Linear-in-Means Models: Math Scores

VARIABLES	(1) math score	(2) math score	(3) math score	(4) math score	(5) math score
Average peer effect terms					
Conscientiousness					0.1458***
					(0.0515)
Agreeableness					0.0214
					(0.0253)
Extraversion					-0.0679*
					(0.0355)
Math score at t=0			0.0365	0.0176	-0.0089
			(0.0702)	(0.0680)	(0.0684)
IQ			0.0624	0.0387	-0.0232
			(0.0724)	(0.0693)	(0.0698)
Male			-0.1299	-0.0735	-0.0067
			(0.1733)	(0.1711)	(0.1655)
Father tertiary education			0.4760**	0.3759**	0.3123*
			(0.1878)	(0.1711)	(0.1739)
Pupil variables					
Conscientiousness		0.1870***		0.1657***	0.1597***
		(0.0269)		(0.0268)	(0.0258)
Agreeableness		0.0063		0.0067	0.0100
		(0.0160)		(0.0157)	(0.0155)
Extraversion		-0.0270*		-0.0258*	-0.0253
		(0.0153)		(0.0154)	(0.0152)
Math score at t=0	0.3568***	0.3156***	0.3273***	0.3024***	0.3018***
	(0.0265)	(0.0217)	(0.0228)	(0.0201)	(0.0205)
IQ	0.3310***	0.2646***	0.2952***	0.2489***	0.2472***
	(0.0329)	(0.0312)	(0.0346)	(0.0328)	(0.0328)
Male	-0.0847*	-0.0502	-0.0650*	-0.0424	-0.0385
	(0.0448)	(0.0404)	(0.0338)	(0.0331)	(0.0327)
Father tertiary education	0.0534*	0.0246	0.0353	0.0160	0.0160
	(0.0317)	(0.0321)	(0.0319)	(0.0321)	(0.0321)
Math class size	-0.0113	-0.0140**	-0.0185**	-0.0177**	-0.0172**
	(0.0073)	(0.0067)	(0.0091)	(0.0087)	(0.0083)
Constant	-0.2033	-0.2377*	-0.0301	-0.1509	-0.2483
	(0.1381)	(0.1272)	(0.2054)	(0.1994)	(0.1886)
School FE	Yes	Yes	Yes	Yes	Yes
Observations	3,174	3,174	3,174	3,174	3,174
R-squared	0.5338	0.5494	0.5414	0.5528	0.5553

Cluster robust standard errors in parentheses. All coefficients presented here represent effect sizes. These regressions also include control variables for peer and individual family income not presented in this table. *** p<0.01, ** p<0.05, * p<0.1

results (presented in Appendix F) show that including just the IQ peer effects term without any other peer effects term produces a statistically significant and positive coefficient on the IQ peer effects term while the non-cognitive peer effect results remain the same. Since most other peer effects studies do not control for two rich measures of prior attainment as well as non-cognitive traits, some of which are positively correlated with IQ, it is possible that the significant findings on cognitive ability in other studies are actually proxying for conscientiousness.

The peer effect term for having peers with educated fathers is also statistically significant and positive across specifications, unlike the peer effect terms for income (not reported in the tables), which are never statistically significant in any specification. Having peers from more educated families is positively related to math outcomes.

These results also show the importance of individual level non-cognitive measures on math scores. Across all specifications, conscientiousness at the individual level is statistically significant and positive. In the literature, conscientiousness has been found to be positively correlated with educational and labor market outcomes (Borghans et al., 2008). Poropat (2009) conducts a meta-analysis of studies using the Big Five model of personality and academic outcomes and finds that the correlation between an individual's conscientiousness and academic performance is independent of IQ and that conscientiousness predicts as much of tertiary education performance as IQ once past schooling performance is included in the model. If the indirect effect of peer conscientiousness on attainment via the classroom learning environment is positive and large, perhaps the individual level finding can be extrapolated to the classroom level. This would support my findings on the importance of peer conscientiousness.

The introduction of the individual level non-cognitive measures in Column (2) of Table 5 is interesting because the general results from the baseline do not change. The effect of individual level IQ is approximately 0.3 of a standard deviation, and also remains statistically significant and roughly the same size throughout all specifications. Compared to the non-cognitive peer effects, individual IQ has a much stronger relationship with math scores, which is to be expected. Individual past performance in math positively predicts math performance with an effect size of 0.3 of a standard deviation, which also remains relatively constant in terms of size and statistical significance across specifications.

The role of gender in the math score regressions reveals an interesting result also shown in the literature on non-cognitive traits. Overall, boys perform worse in math than girls; however, this result only holds in models without non-cognitive measures. As discussed, gender may often proxy for non-cognitive traits (Bertrand and Pan, 2013), which seems to be the case here. Once non-cognitive traits are included, the negative relationship between being a boy and math performance disappears.

The results for the regressions using language scores have some key differences and some similarities to the math results. At the individual level and the peer level, the non-cognitive traits that matter differ from math to language. The results of the full model in Column (5) of Table 6 show that peer level conscientiousness is also positively related to individual performance, as was the case with math. A one standard deviation increase in average peer conscientiousness is associated with a 0.12 standard deviation increase in language scores. This supports the idea that having diligent peers benefits an individual's learning outcomes. None of the other non-cognitive peer effects have a statistically significant relationship with individual language outcomes.

The peer effect term for having peers whose fathers attended tertiary education is large, positive, and statistically significant in Columns (3) and (4) of Table 6. Once I add in peer level non-cognitive measures, this relationship gets weaker and loses statistical significance at traditional levels. This indicates that much of the benefit individuals get from having peers from better educated families may actually be transmitted through their non-cognitive traits when it comes to subjects that focus on language.

Unlike in math, the peer effect term for extraversion does not have a statistically significant association with language performance. The potentially disruptive nature of more extraverted peers is not related to language learning outcomes, again highlighting potential differences between learning quantitative and qualitative subjects.

At the individual level, there is a statistically significant and positive relationship between conscientiousness and language performance of approximately 0.15 standard deviations. Again, this is in line with the literature on conscientiousness being highly related

Table 6: Linear-in-Means Models: Language Scores

VARIABLES	(1) language score	(2) language score	(3) language score	(4) language score	(5) language score
Average peer effect terms					
Conscientiousness					0.1209**
					(0.0502)
Agreeableness					-0.0087
					(0.0216)
Extraversion					-0.0135
					(0.0267)
Language score at t=0			0.0061	-0.0047	-0.0394
			(0.0718)	(0.0653)	(0.0652)
IQ			0.0736	0.0528	0.0058
			(0.0682)	(0.0651)	(0.0641)
Male			-0.1983	-0.1493	-0.1040
			(0.1370)	(0.1302)	(0.1309)
Father tertiary education			0.4473**	0.3647*	0.2651
			(0.1913)	(0.1867)	(0.1892)
Pupil variables					
Conscientiousness		0.1770***		0.1589***	0.1503***
		(0.0182)		(0.0193)	(0.0173)
Agreeableness		-0.0019		-0.0020	0.0008
		(0.0125)		(0.0129)	(0.0132)
Extraversion		-0.0340**		-0.0335**	-0.0328**
		(0.0145)		(0.0140)	(0.0132)
Language score at t=0	0.4700***	0.4160***	0.4418***	0.4034***	0.4060***
	(0.0474)	(0.0450)	(0.0457)	(0.0439)	(0.0430)
IQ	0.2941***	0.2449***	0.2663***	0.2313***	0.2294***
	(0.0304)	(0.0270)	(0.0285)	(0.0253)	(0.0254)
Male	-0.2345***	-0.2186***	-0.2190***	-0.2095***	-0.2053***
	(0.0346)	(0.0336)	(0.0322)	(0.0337)	(0.0327)
Father tertiary education	0.0803**	0.0580*	0.0664**	0.0511*	0.0472
	(0.0305)	(0.0292)	(0.0302)	(0.0294)	(0.0289)
Language class size	-0.0099	-0.0125**	-0.0160**	-0.0158**	-0.0156**
	(0.0061)	(0.0054)	(0.0065)	(0.0063)	(0.0061)
Constant	0.0292	0.0113	0.2284*	0.1357	0.0988
	(0.1156)	(0.1010)	(0.1282)	(0.1214)	(0.1129)
School FE	Yes	Yes	Yes	Yes	Yes
Observations	3,174	3,174	3,174	3,174	3,174
R-squared	0.6948	0.7071	0.7012	0.7105	0.7120

Cluster robust standard errors in parentheses. All coefficients presented here represent effect sizes. These regressions also include control variables for peer and individual family income not presented in this table. *** p<0.01, ** p<0.05, * p<0.1

to labor market and academic success (Borghans et al., 2008). The results for individual level extraversion are also similar to math: in the language specifications, there is a statistically significant and negative relationship between being more extraverted and language performance. In the language regressions, agreeableness does not predict performance in a statistically significant way.

4.2. Variance and top and bottom quintile models

I now turn my attention to the issue of alternative models of peer effects. Until now, all of the models I have estimated have been linear-in-means models, which restrict the interpretation of peer effects to the average of peer characteristics. These restrictions might not be realistic if we believe that other characteristics of the ability distribution within a class matter. Perhaps having a low variance in ability is better for learning outcomes because teachers are able to teach to one level? Alternatively, if we believe the shining light or bad apple model, then having a higher percentage of very high and very low-performing pupils might also affect learning outcomes in a way the traditional linear-in-means model is unable to capture.

In order to examine these alternative models of peer effects, I first estimate the same linear-in-means model as before, but include a variance term for each of the cognitive and non-cognitive peer effects terms. This is the model presented in Equation (6). This variance term measures the variance of all peers' ability, excluding the individual pupil.

The results of this estimation on the math scores do not reveal any major differences. None of the variance terms in Column (1) of Table 7 are statistically significant; however, the mean peer effect terms for conscientiousness and extraversion still have the same signs they had in the linear-in-means models presented in Table 5. The average of peer level conscientiousness has a statistically significant and positive relationship with math scores and the average peer level measure of extraversion has a negative and statistically significant relationship with math scores. The magnitude of these effects has increased as a result of accounting for the variance peer effect terms.

The results of the variance model for language in Table 7 are similar. Again none of the variance peer effects terms are statistically significant. The mean peer effect term of

Table 7: Variance Models

VARIABLES	(1) math score	(2) language score
Variance peer effects terms		
Conscientiousness	-0.0008	-0.0475
	(0.0532)	(0.0480)
Agreeableness	-0.0577	-0.0042
	(0.0417)	(0.0288)
Extraversion	-0.0424	0.0167
	(0.0355)	(0.0373)
Math/language score at start of school year	0.0208	-0.0184
	(0.0404)	(0.0156)
IQ	-0.0066	-0.0036
	(0.0435)	(0.0640)
Average peer effect terms		
Conscientiousness	0.1486**	0.1056*
	(0.0583)	(0.0606)
Agreeableness	0.0075	-0.0133
	(0.0286)	(0.0246)
Extraversion	-0.0760*	0.0003
	(0.0398)	(0.0326)
Math/language score at start of school year	-0.0016	-0.0404
	(0.0637)	(0.0762)
IQ	-0.0376	0.0067
	(0.0705)	(0.0614)
Proportion father tertiary education	0.2640	0.2085
	(0.1744)	(0.1921)
Pupil variables		
Conscientiousness	0.1639***	0.1495***
	(0.0262)	(0.0162)
Agreeableness	0.0127	-0.0015
	(0.0155)	(0.0131)
Extraversion	-0.0255*	-0.0306**
	(0.0151)	(0.0135)
Math/language score at start of school year	0.3027***	0.4217***
	(0.0209)	(0.0441)
IQ	0.2416***	0.2224***
	(0.0329)	(0.0255)
Father tertiary education	0.0195	0.0522*
	(0.0315)	(0.0289)
Class size	-0.0163*	-0.0163**
	(0.0085)	(0.0061)
Constant	-0.2495	0.2210*
	(0.6071)	(0.1168)
School FE	Yes	Yes
Observations	3,174	3,174
R-squared	0.5574	0.7146

Cluster robust standard errors in parentheses. All coefficients presented here represent effect sizes. These regressions also include control variables for peer and individual family income and gender not presented in this table. *** p<0.01, ** p<0.05, * p<0.1

conscientiousness is still positive and statistically significant, with a one standard deviation increase in average peer conscientiousness being associated with a 0.14 standard deviation increase in language scores. This indicates that having more conscientious peers on average is beneficial, but that there is no relationship between the dispersion of conscientiousness in a classroom and individual learning outcomes. This result supports the idea of implementing interventions to improve pupil conscientiousness across the distribution. In both the language and the math regressions, having a higher variance of cognitive ability, both in terms of past performance and IQ, does not affect learning outcomes in a statistically significant way.

These results do not rule out the possibility of alternative models of peer effects. I now turn my attention to the effect of having more very high or very low ability peers in a classroom. In order to test this, I take the distribution of all pupils in the data set for a variety of measures and divide it into five quintiles. Using these quintiles, I then calculate the percentage of classmates a given pupil has from each quintile. This excludes the pupil's own place in the distribution from these measures and gives me a relative measure of how many peers fall into each quintile that I can compare across pupils. In Column (1) of each of the quintile model tables I present the full linear-in-means model, in Column (2) I include the top and bottom quintile measures, but do not control for the average peer effect term of each variable, and in Column (3) I include the top and bottom quintile measures as well as the average peer effect term.

The results in Column (3) of Table 8 indicate that there is a relationship between having more peers in the bottom quintile of agreeableness and in the top quintile of IQ. A one percent increase in the share of bottom quintile agreeable peers is associated with a 0.85 standard deviation increase in math scores. It could be the case that having more peers at the bottom of the agreeableness distribution makes the learning environment more focused and less social and thereby improves learning outcomes. A one percent increase in the share of peers at the top of the IQ distribution leads to a 0.32 standard deviation increase in individual math scores, potentially lending support for the shining light hypothesis.

The results for the language scores in Table 9 are different and show support for the bad apple model of peer effects. The results in Column (2) show that having more peers in the bottom quintile of conscientiousness is negatively related to individual language scores and that having more peers in top quintile of extraversion is beneficial. The impact of very peers with very low conscientiousness does not remain when the average peer level is introduced in Column (3). This supports previous findings on bad apples disrupting the learning environment (e.g. Lazear, 2001), which may be especially detrimental when peers have extremely low levels of conscientiousness characterized by low levels of organization, discipline, and achievement-orientation.

As in the linear-in-means framework reproduced in Column (1), there is a statistically significant relationship between peer extraversion and language outcomes in Columns (2) and (3) of Table 9. While on average having more extroverted peers negatively impacts an individual's learning outcomes, having peers in the top quintile of the extraversion distribution proves helpful. Here the shining light model of having more very extraverted peers to potentially lead the class with their participation might explain these results.

5. Conclusion

The analysis in this paper shows that there is a relationship between peers' non-cognitive characteristics and individual learning outcomes. Estimating a standard linear-in-means model, peer level conscientiousness positively predicts math scores by 0.15 standard deviations, while higher peer level extraversion hurts them by -0.07 standard deviations. In the case of language scores, peer level conscientiousness positively predicts language scores by 0.12 standard deviations. These are moderate effect sizes in the education context (Kraft, 2020). The peer level conscientiousness results are robust to a range of checks and are in line with theory and empirical evidence from the psychology literature that relates personality to learning.

I find limited evidence to support other models of non-cognitive peer effects. For both math and language, I find no support for the hypothesis that greater variance in ability within the classroom hurts (i.e. the boutique model) or benefits learning outcomes (i.e. the rainbow model) (Sacerdote, 2011). In the math results, having more peers who score very low on agreeableness positively impacts scores. In the same model, average peer agreeableness positively predicts higher individual attainment. This average effect is in line with evidence from the

Table 8: Top and Bottom Quintile Model: Math Scores

VARIABLES	(1) math score	(2) math score	(3) math score
Top and bottom quintile peer effect terms			
Percent of peers in bottom quintile of conscientiousness		-0.2280	0.1164
		(0.2979)	(0.3600)
Percent of peers in top quintile of conscientiousness		0.2634	0.1140
		(0.3043)	(0.2621)
Percent of peers in bottom quintile of agreeableness		0.3487	0.8467**
		(0.2999)	(0.3389)
Percent of peers in top quintile of agreeableness		0.1930	0.0585
		(0.1897)	(0.2028)
Percent of peers in bottom quintile of extraversion		0.3804	0.0342
		(0.3225)	(0.3786)
Percent of peers in top quintile of extraversion		0.0594	0.2235
		(0.1974)	(0.2479)
Percent of peers in bottom quintile of math scores at start of school year		-0.0293	-0.1908
		(0.1742)	(0.2441)
Percent of peers in top quintile of math scores at start of school year		0.1139	0.1140
		(0.1755)	(0.2087)
Percent of peers in bottom quintile of IQ		-0.0578	-0.1219
		(0.2202)	(0.2453)
Percent of peers in top quintile of IQ		0.3399**	0.3235*
		(0.1654)	(0.1907)
Average peer effect terms			
Conscientiousness	0.1458***		0.1081
	(0.0515)		(0.0712)
Agreeableness	0.0214		0.0918***
	(0.0253)		(0.0308)
Extraversion	-0.0679*		-0.0793
	(0.0355)		(0.0480)
Math score at start of school year	-0.0089		-0.0176
	(0.0684)		(0.0720)
IQ	-0.0232		-0.0943
	(0.0698)		(0.0825)
Pupil variables			
Conscientiousness	0.1597***	0.1565***	0.1541***
	(0.0258)	(0.0256)	(0.0259)
Agreeableness	0.0100	0.0067	0.0105
	(0.0155)	(0.0158)	(0.0158)
Extraversion	-0.0253	-0.0234	-0.0247
	(0.0152)	(0.0153)	(0.0155)
Math score at start of school year	0.3018***	0.3024***	0.3053***
	(0.0205)	(0.0206)	(0.0216)
IQ	0.2472***	0.2463***	0.2467***
	(0.0328)	(0.0338)	(0.0327)
Constant	-0.2483	-0.4542**	-0.4514**
	(0.1886)	(0.1985)	(0.2214)
School FE	Yes	Yes	Yes
Observations	3,174	3,174	3,174
R-squared	0.5553	0.5554	0.5579

Cluster robust standard errors in parentheses. All coefficients presented here represent effect sizes. These regressions also include control variables for peer and individual family income, gender, father's education, and class size not presented in this table. *** p<0.01, ** p<0.05, * p<0.1

Table 9: Top and Bottom Quintile Model: Language Scores

VARIABLES	(1) language score	(2) language score	(3) language score
Top and bottom quintile peer effect terms			
Percent of peers in bottom quintile of conscientiousness		-0.3023**	-0.0033
		(0.1464)	(0.1983)
Percent of peers in top quintile of conscientiousness		0.0392	-0.1431
		(0.1770)	(0.1930)
Percent of peers in bottom quintile of agreeableness		0.2467	0.3047
		(0.1565)	(0.2634)
Percent of peers in top quintile of agreeableness		0.2214	0.1410
		(0.4022)	(0.3885)
Percent of peers in bottom quintile of extraversion		0.1975	-0.0283
		(0.2120)	(0.3050)
Percent of peers in top quintile of extraversion		0.3600**	0.4577**
		(0.1758)	(0.2100)
Percent of peers in bottom quintile of language scores at start of school year		0.2476*	0.2788
		(0.1420)	(0.1938)
Percent of peers in top quintile of language scores at start of school year		0.3742***	0.3426**
		(0.1058)	(0.1386)
Percent of peers in bottom quintile of IQ		-0.0085	-0.0639
		(0.1626)	(0.1852)
Percent of peers in top quintile of IQ		0.2145	0.1649
		(0.1586)	(0.2104)
Average peer effect terms			
Conscientiousness	0.1209**		0.1087*
	(0.0502)		(0.0619)
Agreeableness	-0.0087		0.0136
	(0.0216)		(0.0372)
Extraversion	-0.0135		-0.0462
	(0.0267)		(0.0435)
Language score at start of school year	-0.0394		-0.0171
	(0.0652)		(0.0752)
IQ	0.0058		-0.0011
	(0.0641)		(0.0881)
Pupil variables			
Conscientiousness	0.1503***	0.1502***	0.1481***
	(0.0173)	(0.0173)	(0.0172)
Agreeableness	0.0008	-0.0009	0.0008
	(0.0132)	(0.0129)	(0.0136)
Extraversion	-0.0328**	-0.0328**	-0.0331**
	(0.0132)	(0.0131)	(0.0130)
Language score at start of school year	0.4060***	0.4091***	0.4083***
	(0.0430)	(0.0423)	(0.0423)
IQ	0.2294***	0.2301***	0.2295***
	(0.0254)	(0.0260)	(0.0252)
Constant	0.0988	-0.2341	-0.1872
	(0.1129)	(0.1622)	(0.1708)
School FE	Yes	Yes	Yes
Observations	3,174	3,174	3,174
R-squared	0.7120	0.7147	0.7153

Cluster robust standard errors in parentheses. All coefficients presented here represent effect sizes. These regressions also include control variables for peer and individual family income, gender, father's education, and class size not presented in this table. *** p<0.01, ** p<0.05, * p<0.1

psychology literature that shows agreeableness is strongly related to academic motivations, including persistence and self-improvement, which lead to positive attainment (Komarraju et al., 2011). In the language results I find that having more peers who have low conscientiousness hurts scores. This supports the bad apple model of peer effects where a student with poor discipline or someone who requires extra attention from the teacher detracts from the learning outcomes of others (Lazear, 2001; Sacerdote, 2011). This result is not robust to the inclusion of the average effect of conscientiousness. Taken together, these results provide limited support for other models of non-cognitive peer effects.

I cannot completely rule out the possibility that these results are potentially limited by the issue of selection into classrooms within schools. As with many papers on school peer effects, the context explored in this paper does not benefit from random assignment to groups. I am, however, able to control for many traditionally unobserved aspects of pupil heterogeneity, including IQ and non-cognitive traits, which means issues of unobserved heterogeneity and measurement error are less of a problem here. I also conduct three tests to further probe potential endogeneity issues in this context. Taken together, the results of these tests support the idea that peer groups are not selected based on non-cognitive traits.

There is also a high proportion of missing data in this sample, something traditionally ignored in the peer effects literature (Sojourner, 2013). I explore the distribution of the missing data and find it is normally distributed across schools; however, individuals with missing data are more likely to be boys and be from disadvantaged backgrounds. When I account for missing data in my estimation strategies, the results are broadly similar to the main results obtained in this paper. If anything, it seems like the estimates presented here may be slightly downward biased.

It could also be the case that individuals switch classes throughout the school year, which would also affect my results since peer groups would change. I only observe an individual's classroom at the beginning of the school year, so am unable to check this. Anecdotal evidence suggests that pupils do not regularly change classes throughout the school year, but in the case that they did, I would essentially be estimating intention to treat effects. It would mean I am underestimating the impact of the non-cognitive peer effects, although this seems unlikely based on discussions with the LOSO research team.

Despite these limitations, the results in this paper are very similar to those obtained by Golsteyn et al. (2021), who exploit random assignment to peer groups in university. They find that exposure to persistent peers leads to long term learning gains for the individual, which is in line with the LIM conscientiousness results obtained in this study since conscientiousness and persistence have been shown to be related (Duckworth et al., 2007). Golsteyn et al. also find that exposure to risk-loving peers decreases academic attainment, which is in line with the LIM extraversion results from this paper since extraversion has been found to be negatively related to risk aversion (Oehler and Wedlich, 2018). On balance this gives my results validity in informing the peer effects debate.

I have shown that peers influence each other's learning outcomes in ways beyond the traditional channels of IQ and past subject performance. Aspects of personality, both at the individual and perhaps more interestingly, at the peer group level, have a meaningful relationship with learning outcomes. Interestingly, they seem to matter more than peer measures of cognitive ability. The finding that higher average peer conscientiousness is positively related to math and language outcomes means that schools should target interventions to improve pupil conscientiousness. Other studies have confirmed that non-cognitive skills are malleable, e.g. self-concept, which strengthens the case for interventions in schools (O'Mara et al., 2006).

Policymakers in some contexts are already aware of the importance of non-cognitive skills for educational outcomes. In the United Kingdom, for example, the Office for Standards in Education, Children's Services and Skills (Ofsted), which rates schools, is particularly interested in classroom behavior and how it is related to learning outcomes (Ofsted, 2019). Ofsted has not only placed a focus on "low level disruption" in the classroom, but also included key behavioral and non-cognitive terms (e.g. whether pupils are "confident", "self-disciplined", and "self-assured") in their grade descriptors used to rate schools (Ofsted, 2019). Interventions to specifically target and develop non-cognitive skills may not only benefit the individual pupil in her life, but have a knock-on effect in terms of positively impacting her peers through peer effects. Non-cognitive skills can and should be developed in order to improve both learning and labor market outcomes.

References

Abadie, A., Athey, S., Imbens, G., Wooldridge, J., 2017. When Should You Adjust Standard Errors for Clustering? Technical Report w24003. National Bureau of Economic Research. Cambridge, MA. URL: `http://www.nber.org/papers/w24003.pdf`, doi:`10.3386/w24003`.

Aizer, A., 2008. Peer Effects and Human Capital Accumulation: the Externalities of ADD. Technical Report w14354. National Bureau of Economic Research. Cambridge, MA. URL: `http://www.nber.org/papers/w14354.pdf`, doi:`10.3386/w14354`.

Almlund, M., Duckworth, A.L., Heckman, J., Kautz, T., 2011. Personality Psychology and Economics, in: Hanushek, E.A., Machin, S., Woessmann, L. (Eds.), Handbook of the Economics of Education. Elsevier. volume 4 of *Handbook of The Economics of Education*, pp. 1–181. URL: `http://www.sciencedirect.com/science/article/pii/B9780444534446000018`, doi:`10.1016/B978-0-444-53444-6.00001-8`.

Ammermueller, A., Pischke, J.S., 2009. Peer effects in European primary schools: Evidence from the progress in international reading literacy study. Journal of Labor Economics 27, 315–348. ISBN: 0734-306X Publisher: The University of Chicago Press.

Angrist, J.D., 2014. The perils of peer effects. Labour Economics 30, 98–108. ISBN: 0927-5371 Publisher: Elsevier.

Bertrand, M., Pan, J., 2013. The trouble with boys: Social influences and the gender gap in disruptive behavior. American economic journal: applied economics 5, 32–64. ISBN: 1945-7782.

Betts, J., Zau, A., 2004. Peer Groups and Academic Achievement: Panel Evidence from Administrative Data. mimeo .

Borghans, L., Duckworth, A.L., Heckman, J.J., Weel, B.t., 2008. The Economics and Psychology of Personality Traits. Journal of Human Resources 43, 972–1059. URL: `http://jhr.uwpress.org/content/43/4/972`, doi:`10.3368/jhr.43.4.972`.

Burke, M.A., Sass, T.R., 2013. Classroom peer effects and student achievement. Journal of Labor Economics 31, 51–82. ISBN: 0734-306X Publisher: University of Chicago Press Chicago, IL.

Cooley Fruehwirth, J., 2013. Identifying peer achievement spillovers: Implications for desegregation and the achievement gap. Quantitative economics 4, 85–124. ISBN: 1759-7323 Publisher: Wiley Online Library.

Cooley Fruehwirth, J., 2014. Can Achievement Peer Effect Estimates Inform Policy? A View from Inside the Black Box. The Review of Economics and Statistics 96, 514–523. URL: https://doi.org/10.1162/REST_a_00385, doi:10.1162/REST_a_00385. publisher: MIT Press.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. psychometrika 16, 297–334. ISBN: 0033-3123 Publisher: Springer.

Damme, J.V., Fraine, B.D., Landeghem, G.V., Opdenakker, M.C., Onghena, P., 2002. A New Study on Educational Effectiveness in Secondary Schools in Flanders: An Introduction. School Effectiveness and School Improvement 13, 383–397. URL: https://www.tandfonline.com/doi/abs/10.1076/sesi.13.4.383.10285, doi:10.1076/sesi.13.4.383.10285. publisher: Routledge _eprint: https://www.tandfonline.com/doi/pdf/10.1076/sesi.13.4.383.10285.

DeMars, C., 2010. Item response theory. Oxford University Press.

Duckworth, A.L., Peterson, C., Matthews, M.D., Kelly, D.R., 2007. Grit: perseverance and passion for long-term goals. Journal of personality and social psychology 92, 1087. ISBN: 1939-1315 Publisher: American Psychological Association.

Feld, J., Zölitz, U., 2017. Understanding peer effects: On the nature, estimation, and channels of peer effects. Journal of Labor Economics 35, 387–428. ISBN: 0734-306X Publisher: University of Chicago Press Chicago, IL.

Van de gaer, E., Pustjens, H., Van Damme, J., De Munter, A., 2009. School Engagement and Language Achievement: A Longitudinal Study of Gender Differences across Secondary

School. Merrill-Palmer Quarterly 55, 373–405. URL: https://www.jstor.org/stable/23096232. publisher: Wayne State University Press.

Gibbons, S., Telhaj, S., 2016. Peer effects: Evidence from secondary school transition in England. Oxford Bulletin of Economics and Statistics 78, 548–575. ISBN: 0305-9049 Publisher: Wiley Online Library.

Ginkel, J.R.v., Linting, M., Rippe, R.C.A., Voort, A.v.d., 2020. Rebutting Existing Misconceptions About Multiple Imputation as a Method for Handling Missing Data. Journal of Personality Assessment 102, 297–308. URL: https://doi.org/10.1080/00223891.2018.1530680, doi:10.1080/00223891.2018.1530680. publisher: Routledge _eprint: https://doi.org/10.1080/00223891.2018.1530680.

Goldberg, L.R., 1981. Language and individual differences: The search for universals in personality lexicons. Review of personality and social psychology 2, 141–165.

Golsteyn, B., Non, A., Zölitz, U., 2017. The impact of peer personality on academic achievement. University of Zurich, Department of Economics, Working Paper .

Golsteyn, B.H., Non, A., Zölitz, U., 2021. The impact of peer personality on academic achievement. Journal of Political Economy 129, 1052–1099. ISBN: 0022-3808 Publisher: The University of Chicago Press Chicago, IL.

Guryan, J., Kroft, K., Notowidigdo, M.J., 2009. Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. American Economic Journal: Applied Economics 1, 34–68. ISBN: 1945-7782.

Hahn, E., Gottschling, J., Spinath, F.M., 2012. Short measurements of personality – Validity and reliability of the GSOEP Big Five Inventory (BFI-S). Journal of Research in Personality 46, 355–359. URL: http://www.sciencedirect.com/science/article/pii/S0092656612000578, doi:10.1016/j.jrp.2012.03.008.

Heckman, J., Pinto, R., Savelyev, P., 2013. Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. American Economic Review 103, 2052–2086. URL: https://pubs.aeaweb.org/doi/10.1257/aer.103.6.2052, doi:10.1257/aer.103.6.2052.

Heckman, J.J., Kautz, T., 2012. Hard evidence on soft skills. Labour economics 19, 451–464. ISBN: 0927-5371 Publisher: Elsevier.

Heckman, J.J., Rubinstein, Y., 2001. The Importance of Noncognitive Skills: Lessons from the GED Testing Program. American Economic Review 91, 145–149. URL: https://pubs.aeaweb.org/doi/10.1257/aer.91.2.145, doi:10.1257/aer.91.2.145.

Heckman, J.J., Stixrud, J., Urzua, S., 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. Journal of Labor economics 24, 411–482. ISBN: 0734-306X Publisher: The University of Chicago Press.

Hinshaw, S.P., 1992. Externalizing behavior problems and academic underachievement in childhood and adolescence: causal relationships and underlying mechanisms. Psychological bulletin 111, 127. ISBN: 1939-1455 Publisher: American Psychological Association.

Hoxby, C., 2000. Peer Effects in the Classroom: Learning from Gender and Race Variation. Technical Report w7867. National Bureau of Economic Research. Cambridge, MA. URL: http://www.nber.org/papers/w7867.pdf, doi:10.3386/w7867.

Komarraju, M., Karau, S.J., Schmeck, R.R., Avdic, A., 2011. The Big Five personality traits, learning styles, and academic achievement. Personality and Individual Differences 51, 472–477. URL: http://www.sciencedirect.com/science/article/pii/S0191886911002194, doi:10.1016/j.paid.2011.04.019.

Kraft, M.A., 2020. Interpreting Effect Sizes of Education Interventions. Educational Researcher 49, 241–253. URL: https://doi.org/10.3102/0013189X20912798, doi:10.3102/0013189X20912798. publisher: American Educational Research Association.

Lancksweerdt, P., 1991. Getlov (Gemeenschappelijke Testbatterij Lager Onderwijs). Onderzoeksverslag over de gegevens van de leerlingen van het 6de leerjaar lager onderwijs uit de vrije P.M.S.-centra in Oost-Vlaanderen schooljaar 1989–1990 [Getlov (Test battery for primary education). Research report about the data of six-grade students in primary education fromprivate school guidance centres of East Flanders]. Technical Report. Vrij P.M.S.-centrum 1. Deinze.

Lavy, V., Schlosser, A., 2011. Mechanisms and impacts of gender peer effects at school. American Economic Journal: Applied Economics 3, 1–33. ISBN: 1945-7782.

Lavy, V., Silva, O., Weinhardt, F., 2012. The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools. Journal of Labor Economics 30, 367–414. URL: https://www.journals.uchicago.edu/doi/10.1086/663592, doi:10.1086/663592.

Lazear, E.P., 2001. Educational production. The Quarterly Journal of Economics 116, 777–803. ISBN: 1531-4650 Publisher: MIT Press.

Mammadov, S., Cross, T.L., Ward, T.J., 2018. The Big Five personality predictors of academic achievement in gifted students: Mediation by self-regulatory efficacy and academic motivation. High Ability Studies 29, 111–133. URL: https://doi.org/10.1080/13598139.2018.1489222, doi:10.1080/13598139.2018.1489222. publisher: Routledge _eprint: https://doi.org/10.1080/13598139.2018.1489222.

Manski, C.F., 1993. Identification of Endogenous Social Effects: The Reflection Problem. The Review of Economic Studies 60, 531. URL: https://academic.oup.com/restud/article-lookup/doi/10.2307/2298123, doi:10.2307/2298123.

Neidell, M., Waldfogel, J., 2010. Cognitive and noncognitive peer effects in early education. The Review of Economics and Statistics 92, 562–576. ISBN: 0034-6535 Publisher: MIT Press.

OECD, 2010. PISA 2009 at a Glance. OECD. URL: https://www.oecd-ilibrary.org/education/pisa-at-a-glance-2010_9789264095298-en, doi:10.1787/9789264095298-en.

Oehler, A., Wedlich, F., 2018. The relationship of extraversion and neuroticism with risk attitude, risk perception, and return expectations. Journal of Neuroscience, Psychology, and Economics 11, 63–92. doi:10.1037/npe0000088. place: US Publisher: Educational Publishing Foundation.

Ofsted, 2019. Ofsted School Inspection Handbook. Technical Report 190017. Office for Standards in Education, Children's Services and Skills (Ofsted). URL: https://www.gov.uk/ofsted.

O'Mara, A.J., Marsh, H.W., Craven, R.G., Debus, R.L., 2006. Do Self-Concept Interventions Make a Difference? A Synergistic Blend of Construct Validation and Meta-Analysis. Educational Psychologist 41, 181–206. doi:10.1207/s15326985ep4103_4.

Poropat, A.E., 2009. A meta-analysis of the five-factor model of personality and academic performance. Psychological Bulletin 135, 322–338. doi:10.1037/a0014996.

Rammstedt, B., John, O.P., 2007. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality 41, 203–212. URL: http://www.sciencedirect.com/science/article/pii/S0092656606000195, doi:10.1016/j.jrp.2006.02.001.

Riding, R.J., Dyer, V.A., 1980. The relationship between extraversion and verbal-imagery learning style in twelve-year-old children. Personality and Individual Differences 1, 273–279. URL: http://www.sciencedirect.com/science/article/pii/0191886980900598, doi:10.1016/0191-8869(80)90059-8.

Rubin, D.B., 1976. Inference and missing data. Biometrika 63, 581–592. ISBN: 1464-3510 Publisher: Oxford University Press.

Rubin, D.B. (Ed.), 1987. Multiple Imputation for Nonresponse in Surveys. Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ, USA. URL: http://doi.wiley.com/10.1002/9780470316696, doi:10.1002/9780470316696.

Sacerdote, B., 2011. Peer effects in education: How might they work, how big are they and how much do we know thus far?, in: Handbook of the Economics of Education. Elsevier. volume 3, pp. 249–277.

Shewbridge, C., Hulshof, M., Nusche, D., Stoll, L., 2011. OECD Reviews of Evaluation and Assessment in Education: School Evaluation in the Flemish Community of Belgium 2011. OECD Reviews of Evaluation and Assessment in Education, OECD. URL: https://www.oecd-ilibrary.org/education/oecd-reviews-of-evaluation-and-assessment-in-education-belgium-flemish-community-2011_9789264116726-en, doi:10.1787/9789264116726-en.

Sirin, S.R., 2005. Socioeconomic status and academic achievement: A meta-analytic review of research. Review of educational research 75, 417–453. ISBN: 0034-6543 Publisher: Sage Publications Sage CA: Thousand Oaks, CA.

Slavin, R.E., 1990. Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. Review of educational research 60, 471–499. ISBN: 0034-6543 Publisher: Sage Publications Sage CA: Thousand Oaks, CA.

Sojourner, A., 2013. Identification of peer effects with missing peer data: Evidence from Project STAR. The Economic Journal 123, 574–605. ISBN: 0013-0133 Publisher: Oxford University Press Oxford, UK.

of St. Louis, F.R.B., 2006. Belgium / U.S. Foreign Exchange Rate (DISCONTINUED SERIES). Technical Report. Federal Reserve Bank of St. Louis.

Van Damme, J., De Fraine, B., Van Landeghem, G., Opdenakker, M.C., Onghena, P., 2002. A new study on educational effectiveness in secondary schools in Flanders: An introduction. School Effectiveness and School Improvement 13, 383–397. ISBN: 0924-3453 Publisher: Taylor & Francis.

Vedel, A., 2014. The Big Five and tertiary academic performance: A systematic review and meta-analysis. Personality and Individual Differences 71, 66–76. URL: http://www.sciencedirect.com/science/article/pii/S0191886914004073, doi:10.1016/j.paid.2014.07.011.

Vigdor, J., Nechyba, T., 2007. Peer effects in North Carolina public schools. Schools and the equal opportunity problem , 73–101Publisher: MIT Press Cambridge, MA.

Appendix A. Non-cognitive Measures

Table A.10 presents the questions from the teacher questionnaire used to construct the non-cognitive measures used in this paper. On this questionnaire, the primary school teacher had to answer how much the statement applied to the pupil using a Likert Scale. These questions were selected from the questionnaire using principal component analysis and then reviewed to assure quality. Each scale was assessed using Cronbach's Alpha to ensure scale reliability (Cronbach, 1951). All measures exceed the rule of thumb value of 0.7 established by Cronbach for a reliable scale.

Appendix B. Education in Flanders

Children in Flanders attend compulsory education from age six until age 18. Secondary school, the focus of this paper, begins at age 12. At this point, children pursue one of two streams, known as the A or B stream; in 2010, 84 percent of children were enrolled in the A stream, which has been relatively stable since the school reform of 1989 (Shrewbridge et al., 2011). The A stream is the standard stream, while the B stream is focused on children with developmental or learning difficulties. There is less differentiation between schools at this point, as pupils still follow a more standardized curriculum with fewer electives and specialization; however, there is a high degree of school autonomy in Flanders (Shrewbridge et al., 2011). This means that selection into a specific school is more likely since schools can differentiate themselves and thereby attract prospective pupils. After completing the A or B stream, which takes two years, pupils then move to one of four types of upper secondary schools. These four types of upper secondary schools are: general secondary education (ASO, Algemeen secundair onderwijs); technical secondary education (TSO, Technisch secundair onderwijs); secondary arts education (KSO, Kunst secundair onderwijs); and vocational secondary education (BSO, Beroepssecundair onderwijs) (Shrewbridge et al., 2011). As indicated by their names, these schools range from academically focused to vocationally focused. Pupils can only attend university if they receive a diploma from an ASO or TSO secondary school. In this paper I only look at the first year of secondary school because of the relative homogeneity between content at this level.

Table A.10: Non-cognitive measures

Original Question Number	Dutch Question	English Translation	Scale	Cronbach's Alpha
Q1	*Kon goed volgen in de klas, heeft voldoende intellectuele mogelijkheden; is verstandig*	Has sufficient intellectual capabilities to follow well in the classroom; is wise	CON	
Q2	*Was gemotiveerd voor het schoolwerk; wilde het echt goed doen; werkte zonder tegenzin*	Was motivated for school work; wanted to do it really well; worked without reluctance	CON	
Q12	*Kon een samenhangend verhaal vertellen; een onderwerp uitdiepen; bij het onderwerp blijven*	Could tell a coherent story; explore a topic; stay on the subject	CON	
				0.865
Q11	*Stelde zich gemakkelijk open voor de onderwijzer(es); was spontaan; niet defensief*	Was open to the teacher(es); was spontaneous; not defensive	EXTRA	
Q18	*Maakte een energieke en vitale indruk; was vrolijk en zag er gelukkig uit*	Made an energetic and vital impression; was smiling and looked happy	EXTRA	
Q21	*Zocht contact met de medeleerlingen; was open en aanspreekbaar*	Made contact with fellow students; was open and approachable	EXTRA	
				0.793
Q5	*Stoorde de les niet opzettelijk; was niet gericht op het boycotten van het lesverloop*	Did not disturb the lesson intensionally; did not aim to boycott learning	AGREE	
Q9	*Hield zich goed (dit is uit zichzelf) aan de klasregels; wachtte zijn/haar beurt af; voortdurend tot de orde roepen was niet nodig*	Held herself to the class rules; waited for her turn; it was not necessary to constantly call her to order	AGREE	
Q15	*Was afkerig van vijandelijkheden; wilde vriendelijk en aardig zijn voor anderen; beleefde geen genoegen aan het plagen en pesten van anderen*	Was averse to hostilities; was friendly and kind to others; experienced no pleasure in teasing and bullying of others	AGREE	
				0.843

Source: LOSO Data Set ("Beoordeling van der leerling door de leerkracht Basisonderwijs")

One main feature of the education system in Flanders is the prevalence of Catholic schools, which contributes to the high degree of school autonomy. The Belgian Constitution includes a freedom to education clause, which gives any individual the right to open up a school (Shrewbridge et al., 2011). In order for schools to receive public funding and issue diplomas, however, they must follow a core curriculum set out by the Flemish authorities and they must agree to be inspected by the Flemish authorities (Shrewbridge et al., 2011). This same freedom to education clause in the constitution gives parents the right to school choice, which naturally complicates the identification strategy in this context.

The quality of education in the Flemish context has been deemed relatively high. On every Program for International Student Assessment (PISA) test including the first round in 2000, pupils in Flanders have consistently performed well above average (Shrewbridge et al., 2011). Over 80 percent of Flemish adults aged 25-34 years old have completed upper secondary education and in 2008, 42 percent of 25-34 years olds held a tertiary degree, which is higher than the OECD average (Shrewbridge et al., 2011). Results from the 2009 PISA for all of Belgium showed that are still vast differences in performance based on socio-economic status of pupils and schools (OECD, 2011). These same results also pointed to significantly lower performance on the part of immigrant children as opposed to native Dutch speakers; the proportion of low-performing immigrant pupils was three times as large as the proportion of low-performing native Dutch speakers (OECD, 2011). While these stylized facts come from data collected significantly after the period I look at in this paper, immigration is not new to Flanders and neither are issues of inequality.

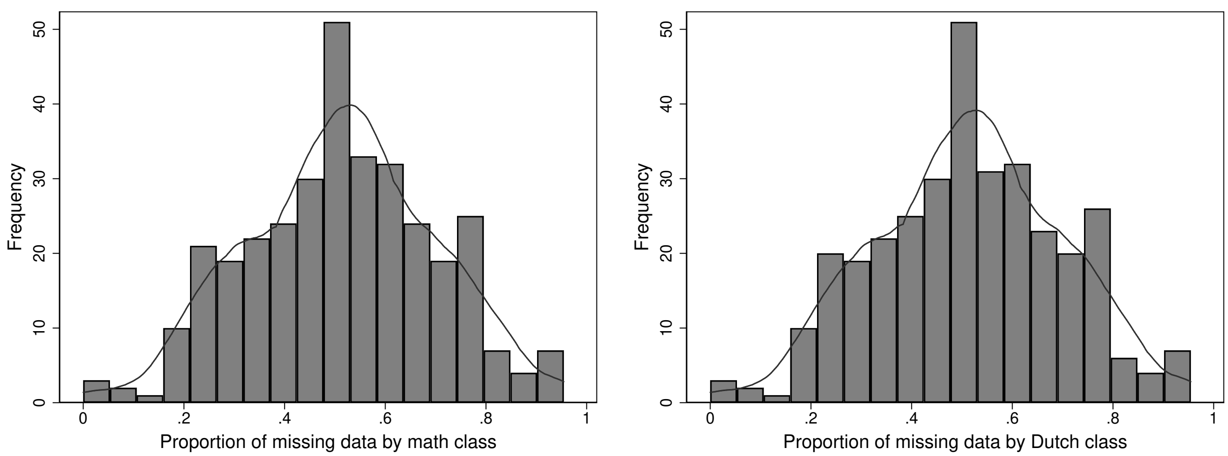
Appendix C. Missing data

As outlined in the main text, there is a large proportion of the original LOSO sample with missing data on the non-cognitive measures as assessed by their primary school teachers (roughly 50%). This is because primary schools were not part of the sampling frame for the LOSO project, so response rates of these teachers tended to be low. In Table C.11, I present an overview of the pattern of observability. This shows that the problem of missing data is driven primarily by lack of data on the non-cognitive traits as well as missing data on demographics. The number of non-observed cognitive skills or outcome measures tends to be

Table C.11: Pattern of Observability

Outcome	**Individual Non-Cog.**	**Individual Cog.**	**Demographics**	**N**
Yes	Yes	Yes	Yes	3,174
Yes	No	Yes	Yes	1,592
Yes	Yes	No	Yes	89
Yes	Yes	Yes	No	849
No	Yes	Yes	Yes	184
No	No	Yes	Yes	97

Figure C.2: Histograms of Missing Data in Secondary School Classes



(a) N=334 math classes

(b) N=332 Dutch classes

much smaller. When computing the peer effect terms, I use all available data. This means that even though the complete case sample is 3,174, the peer effect terms are calculated using a larger sample (i.e. any pupils with available data on that variable).

Figure C.2 shows the distribution of missing data across classes. The seemingly identical figures show a roughly normal distribution of the proportion of missing non-cognitive measures across the 334 and 332 math and Dutch classes respectively. Importantly, this missingness is relatively evenly distributed across secondary schools in the sample. There is no school of the 57 with less than 15 percent missing data in these variables. This lends support to the usage of school fixed effects as a means to deal with this issue.

In Table C.12 I report the results of a balance test for those individuals who have non-cognitive measures and those who do not. The results show that individuals without the non-cognitive measures are from less advantaged backgrounds, more likely to be boys, and

Table C.12: Missing Data Balance Tests

	Missing			Non-missing			
	N	Mean	SD	N	Mean	SD	Diff.
Male	3601	0.55	0.50	3174	0.48	0.50	-0.074*
Father has tertiary education	3601	0.18	0.38	3174	0.22	0.41	0.039**
Low income	2109	0.19	0.39	3174	0.15	0.36	-0.037***
High income	2109	0.12	0.32	3174	0.14	0.35	0.024*
IQ	3361	-0.08	1.04	3174	0.08	0.95	0.157**
Math score at end of school year	3263	-0.08	1.04	3174	0.08	0.95	0.164***
Language score at end of school year	3325	-0.11	1.06	3174	0.12	0.92	0.230***
Math score at start of school year	3400	-0.10	1.07	3174	0.11	0.91	0.208***
Language score at start of school year	3380	-0.08	1.06	3174	0.09	0.93	0.168**

NB: *** p<0.01, ** p<0.05, * p<0.1

Table C.13: The Relationship Between Missing Data and Peer Effect Terms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	missing	missing	missing	missing	missing	missing	missing	missing
Peer effect term								
Conscientiousness	-0.0492***	-0.0492***					-0.0660***	-0.0684***
	(0.0057)	(0.0077)					(0.0100)	(0.0120)
Agreeableness			-0.0251***	-0.0334***			-0.0103	-0.0189**
			(0.0058)	(0.0074)			(0.0083)	(0.0096)
Extraversion					-0.0156***	-0.0257***	0.0372***	0.0168
					(0.0058)	(0.0069)	(0.0097)	(0.0106)
Constant	0.3471***	0.1784**	0.3482***	0.1654*	0.3482***	0.1675*	0.5308***	0.4048***
	(0.0058)	(0.0884)	(0.0058)	(0.0887)	(0.0058)	(0.0888)	(0.0060)	(0.0928)
School FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	6,775	6,775	6,775	6,775	6,775	6,775	6,775	6,775
R-squared	0.0134	0.1139	0.0028	0.1100	0.0011	0.1091	0.0114	0.1169

Cluster robust standard errors in parentheses. The outcome variable is a binary indicator for whether or not an individual has missing data. The results presented here are for math class peer effect terms. *** p<0.01, ** p<0.05, * p<0.1

lower attaining. This means that the data are not missing completely at random (MCAR), but does not tell us whether the data are missing at random (MAR) or missing not at random (MNAR). MAR occurs when the probability of data missing is the same within a group defined by the observed data and MNAR occurs when the probability of data missing occurs for reasons unknown (Rubin, 1976).

I also explore whether individuals with missing data are more likely to be in classrooms with higher or lower peer level non-cognitive traits. The results in Table C.13 show that individuals with missing data are more likely to be in classrooms with lower peer means.

The analysis in this paper is done on a complete case sample (also referred to as the individual-deletion procedure or IDP (Sojourner, 2013)), but the results in Tables C.12 and

C.13 indicate that a more robust approach could be taken. In order to address this, I take several approaches. First, I control for the share of missing peers in a class in my models. Second, I re-estimate all of my models on a multiply imputed sample. And third, I re-estimate my models using inverse probability weights.

Table C.14 presents the analogous model in Column (5) of Tables 5 and 6 with an additional control variable for the share of peers with missing data in each class. The results from this robustness check show very similar results to the main linear-in-means specifications: the point estimates on the non-cognitive peer effect terms are very similar in magnitude and statistical significance. The coefficient on the proportion of missing peers variable is weakly significant in the language scores regression, indicating that having a higher proportion of missing peers is associated with lower performance in language.

To get a better idea of what the bias introduced by individuals with missing data might look like, I follow Rubin (1976; 1987) and perform multiple imputation by chained equation (MICE) in Stata 15. I impute values for conscientiousness, agreeableness, extraversion, math scores, language scores, IQ, and family income using gender, parental education, class size, and elective choice in the imputation model. I impute 10 datasets, which are combined following Rubin (1987). This allows me to account for the uncertainty around the possible values of the missing variables. While many social scientists are skeptical of multiple imputation, there is a growing movement to assuage concerns and increase knowledge around this methodology (e.g. Ginkel et al., 2020).

Multiple imputation allows me to run the same models on the entire sample of 6,775 individuals. For the sake of brevity, I only include results for the linear-in-means models for math, but the results are broadly consistent across all models when comparing the complete case analysis with the multiply imputed sample. The results in Table C.15 are very similar to the main results in Table 5. The coefficient on average peer conscientiousness in Column (5) is still the most robust peer effect and is statistically significant and positive. Its magnitude is increased here as compared to the complete case analysis. The main difference is that the effect of average peer extraversion is no longer statistically significant, but the magnitude of the coefficient is nearly the same. If we believe that the missing data are MAR, which is

Table C.14: Linear-in-Means Models Controlling for Share of Missing Peers

VARIABLES	(1) math score	(2) language score
Average peer effect terms		
Conscientiousness	0.1417***	0.1161**
	(0.0504)	(0.0501)
Agreeableness	0.0187	-0.0112
	(0.0260)	(0.0212)
Extraversion	-0.0676*	-0.0128
	(0.0347)	(0.0263)
Math/language score at start of school year	-0.0236	-0.0419
	(0.0681)	(0.0628)
IQ	-0.0240	-0.0033
	(0.0670)	(0.0586)
Proportion father tertiary education	0.3220*	0.2799
	(0.1763)	(0.1900)
Pupil variables		
Conscientiousness	0.1600***	0.1502***
	(0.0257)	(0.0173)
Agreeableness	0.0094	0.0003
	(0.0156)	(0.0133)
Extraversion	-0.0256*	-0.0329**
	(0.0151)	(0.0131)
Math/language score at start of school year	0.3005***	0.4050***
	(0.0210)	(0.0435)
IQ	0.2465***	0.2285***
	(0.0325)	(0.0255)
Father tertiary education	0.0175	0.0486*
	(0.0315)	(0.0289)
Class size	-0.0155*	-0.0143**
	(0.0083)	(0.0062)
Proportion missing peers	-0.2470	-0.2167*
	(0.1670)	(0.1285)
Constant	-0.1975	0.1506
	(0.1888)	(0.1168)
School FE	Yes	Yes
Observations	3,174	3,174
R-squared	0.5562	0.7127

Cluster robust standard errors in parentheses. All coefficients presented here represent effect sizes. These regressions also include control variables for peer and individual family income and gender not presented in this table. *** p<0.01, ** p<0.05, * p<0.1

possible given the results in Table C.12, and that the MI model has been correctly specified, this provides some indication that the peer effects estimated in this paper may be downward biased due to the missing data on pupils.

Finally, I follow Golsteyn et al. (2021) and generate inverse probability weights (IPW) to account for the probability of being in the complete case sample. I then re-run my models using these weights. In the interest of brevity, I only present the preferred specification found in Column (5) in of Tables 5 and 6 in Table C.16. These results show that weighting the sample using the IPW, generated using the variables discussed in the balance table, does not change the overall story. The point estimates remain nearly the same as in the unweighted specifications. This indicates that missing data, or the probability of being in the complete case sample, is not driving these results.

Taken together this analysis shows that the missing data in this sample is not MCAR. The robustness checks undertaken to probe the validity of the results, however, indicate that the results generated using IDP hold up to methods to account for missing data. This should assuage concerns that the results generated in this paper are driven by missing peers.

Appendix D. Exogeneity of peer groups

This appendix contains the results of three tests to probe potential endogeneity issues around the formation of peer groups in this context. The first test is to examine whether peer non-cognitive traits are related to individual test scores from the beginning of the school year. This is a test for the exogeneity of peer group formation. The second test is to compare individuals in above and below peer median conscientiousness classes and examine how the fixed effects approach addresses their potential differences. The third test is to examine the relationship between peer non-cognitive traits and individual traits to see if they are related. This is a test for the exogeneity of peer group formation.

As a first check on the endogeneity of peer group formation, I replicate part of Table 2 from Neidell and Waldfogel (2010). Table D.17 shows the results of regressing math and language test scores from the beginning of the school year (instead of the end, which are the

Table C.15: Linear-in-Means Models with Multiple Imputation: Math Scores

VARIABLES	(1) math score	(2) math score	(3) math score	(4) math score	(5) math score
Average peer effect terms					
Conscientiousness					0.2334***
					(0.0555)
Agreeableness					0.0318
					(0.0282)
Extraversion					-0.0507
					(0.0347)
Math score at t=0			0.1285***	0.0674*	-0.0909*
			(0.0414)	(0.0385)	(0.0462)
IQ			-0.0323	-0.0309	-0.0301
			(0.0241)	(0.0228)	(0.0226)
Male			-0.1837	-0.1301	-0.0060
			(0.1222)	(0.1170)	(0.1181)
Father tertiary education			0.8208***	0.6748***	0.4653***
			(0.1794)	(0.1580)	(0.1592)
Pupil variables					
Conscientiousness		0.2537***		0.2233***	0.2116***
		(0.0255)		(0.0252)	(0.0241)
Agreeableness		0.0141		0.0160	0.0183
		(0.0134)		(0.0133)	(0.0131)
Extraversion		-0.0337***		-0.0320**	-0.0327***
		(0.0121)		(0.0121)	(0.0121)
Math score at t=0	0.5490***	0.4358***	0.4773***	0.4055***	0.4011***
	(0.0227)	(0.0220)	(0.0213)	(0.0215)	(0.0216)
IQ	0.0113	0.0076	0.0164	0.0119	0.0124
	(0.0103)	(0.0109)	(0.0105)	(0.0109)	(0.0109)
Male	-0.0403	-0.0033	-0.0177	0.0083	0.0083
	(0.0312)	(0.0280)	(0.0254)	(0.0253)	(0.0250)
Father tertiary education	0.1053***	0.0549**	0.0762***	0.0424*	0.0351
	(0.0254)	(0.0236)	(0.0229)	(0.0221)	(0.0222)
Math class size	-0.0015	-0.0087*	-0.0161**	-0.0161**	-0.0159***
	(0.0056)	(0.0049)	(0.0065)	(0.0061)	(0.0057)
Constant	-0.4510***	-0.3062***	-0.0619	-0.1007	-0.1885
	(0.1056)	(0.0917)	(0.1576)	(0.1431)	(0.1328)
School FE	Yes	Yes	Yes	Yes	Yes
Observations	6,775	6,775	6,775	6,775	6,775

Cluster robust standard errors in parentheses. All coefficients presented here represent effect sizes. These regressions also include control variables for peer and individual family income not presented in this table. Regressions performed on 10 multiply imputed datasets and combined following Rubin (1987). *** p<0.01, ** p<0.05, * p<0.1

Table C.16: Linear-in-Means Models with Inverse Probability Weights

VARIABLES	(1) math score	(2) language score
Average peer effect terms		
Conscientiousness	0.1501***	0.1229**
	(0.0530)	(0.0508)
Agreeableness	0.0246	-0.0079
	(0.0253)	(0.0221)
Extraversion	-0.0727*	-0.0158
	(0.0363)	(0.0272)
Math/language score at start of school year	-0.0039	-0.0390
	(0.0690)	(0.0657)
IQ	-0.0334	0.0042
	(0.0709)	(0.0651)
Proportion father tertiary education	0.3486*	0.2770
	(0.1797)	(0.1913)
Pupil variables		
Conscientiousness	0.1628***	0.1517***
	(0.0263)	(0.0181)
Agreeableness	0.0087	0.0015
	(0.0157)	(0.0136)
Extraversion	-0.0262*	-0.0327**
	(0.0153)	(0.0136)
Math/language score at start of school year	0.2921***	0.3960***
	(0.0207)	(0.0472)
IQ	0.2523***	0.2291***
	(0.0338)	(0.0266)
Father tertiary education	0.0149	0.0460
	(0.0326)	(0.0302)
Class size	-0.0180**	-0.0158**
	(0.0087)	(0.0062)
Constant	-0.2383	0.0961
	(0.1941)	(0.1152)
School FE	Yes	Yes
Observations	3,174	3,174
R-squared	0.5556	0.7100

Cluster robust standard errors in parentheses. All coefficients presented here represent effect sizes. These regressions also include control variables for peer and individual family income and gender not presented in this table. *** p<0.01, ** p<0.05, * p<0.1

Table D.17: Exogeneity of peer groups I: Neidell and Waldfogel (2010)

VARIABLES	(1) math score t=0	(2) language score t=0
Mean peer conscientiousness	0.0314	-0.0718
	(0.0379)	(0.0509)
Mean peer agreeableness	0.0038	0.0224
	(0.0202)	(0.0234)
Mean peer extraversion	-0.0128	-0.0046
	(0.0208)	(0.0214)
Conscientiousness	0.0947***	0.1853***
	(0.0199)	(0.0172)
Agreeableness	0.0296**	-0.0043
	(0.0127)	(0.0124)
Extraversion	-0.0088	-0.0332***
	(0.0114)	(0.0120)
Constant	-0.3230***	-0.2548**
	(0.1128)	(0.1204)
Observations	3,174	3,174
R-squared	0.6749	0.7028
School fixed effects	Yes	Yes
Individual controls	Yes	Yes
Peer controls	Yes	Yes

This table provides the results of the same tests as Columns (1) and (2) of Table 2 in Neidell and Waldfogel (2010). Individual controls include IQ, gender, father's education, and family income. Peer controls include average peer IQ, proportion male peers, proportion father with tertiary degree, and proportion of high and low income peers. *** p<0.01, ** p<0.05, * p<0.1

standard outcome variables throughout this paper) on peer non-cognitive traits. As Neidell and Waldfogel argue, if selection into classes is based on peer non-cognitive traits, then they will have a positive effect on test scores from the beginning of the year. Table D.17 shows this is not the case. None of the coefficients on the non-cognitive peer effects terms are statistically significant.

Table D.18 presents the results of the second test by replicating Table 1 from Neidell and Waldfogel (2010). They divide their sample based on peer pre-school enrollment (their peer effect of interest) into above and below median levels. Analogously, I divide the sample into above and below peer median conscientiousness (this has been done for each non-cognitive trait, but only the results of conscientiousness are reported here in the interest of brevity).

Table D.18: Exogeneity of peer groups II: Neidell and Waldfogel (2010)

	(1) Mean	(2) Std. Dev	(3) Diff.	(4) p-value	(5) Diff. (with FE)	(6) p-value (with FE)
Math score at t=1	0.5059	0.6814	-1.0119***	0.0000	0.2600***	0.0000
Language score at t=1	0.5634	0.7570	-1.1268***	0.0000	0.2493***	0.0000
Math score at t=0	0.5039	0.6862	-1.0077***	0.0000	0.1959***	0.0000
Language score at t=0	0.5282	0.7653	-1.0564***	0.0000	0.1918***	0.0000
Conscientiousness	0.5342	0.7316	-1.0684***	0.0000	0.1728***	0.0000
Agreeableness	0.2065	0.8621	-0.4130***	0.0000	0.0850	0.1233
Extraversion	0.2478	0.9001	-0.4956***	0.0000	0.0694	0.2043
IQ	0.5181	0.7588	-1.0362***	0.0000	0.1615***	0.0001
Male	0.4398	0.4965	0.0712***	0.0001	-0.0356	0.1142
Low income	0.0800	0.2714	0.1393***	0.0000	-0.0221	0.2695
High income	0.2199	0.4143	-0.1537***	0.0000	0.0081	0.6759
Father tertiary education	0.3321	0.4711	-0.2275***	0.0000	0.0480**	0.0311
At least one foreign parent	0.0246	0.1549	0.0977***	0.0000	-0.0187	0.1368
Math class size	22.1676	3.3280	-2.7171***	0.0000	0.2493*	0.0663
Language class size	22.2325	3.2148	-2.6862***	0.0000	0.1522	0.2445

Following Table 1 in Neidell and Waldfogel (2010), 'Mean' is the average for individuals in classes above the median conscientiousness, 'Diff.' is the difference in means of the variables for individuals in classes above vs. below the median class conscientiousness. 'p-value' is from t-test of variables below/above median that cluster on class. 'FE' adjusts variables for school fixed effect. N=3,174 *** $p<0.01$, ** $p<0.05$, * $p<0.1$

The results in the Column (3) show that individuals in classes above and below the median level of conscientiousness are different along many other dimensions and that all of these differences are statistically significant. This highlights the need to further address the endogeneity of peer group formation. The results in Column (5) highlight the strength of the fixed effect approach: now only half of the differences are statistically significant and their magnitudes are substantially reduced in each case. Importantly, these groups do not differ in terms of agreeableness and extraversion once I control for school fixed effects. Of course this does not fully resolve concerns about selection into classrooms, which is why I conduct three tests.

The third test is to regress each individual non-cognitive trait on the peer effect term for that trait while also controlling for the school level take out mean of that trait (Guryan et al., 2009). The results from this test are presented in Table D.19. They show that conscientiousness is the only non-cognitive trait where there is a positive relationship between individual and peer conscientiousness, although this relationship is weak. There is no relationship between peer agreeableness or extraversion with the individual level measure of that trait. As an additional check, I run the same models in Table D.19, but include peer cognitive effects and individual controls. This further reduces the strength of the relationship between peer conscientiousness and individual conscientiousness to 0.1.

Taken together, these three tests provide evidence that selection into classes on the basis of non-cognitive traits is not a major concern. Nevertheless, I remain cautious in my interpretation of the results and return to this point in the Conclusion of the paper.

Appendix E. Excluding primary school peers

In this section I re-estimate the main LIM specifications for math and language scores using only individuals who knew less than 10 percent of their secondary school classmates from primary school. The goal here is to show that the main results of this paper are not driven by the reflection problem. The results in Tables E.20 and E.21 show that this is not the case. Although the sample is reduced to 1,899 individuals, the results remain broadly the same as in the main LIM specifications. In the case of math, there is a positive coefficient on the peer effect for conscientiousness and a negative coefficient on the peer effect of extraversion. In the case of language, there is a positive coefficient on the peer effect for conscientiousness. The magnitude is broadly similar in both cases.

Appendix F. Building up models

In this appendix, I explore the potential issues around including two measures of cognitive ability in the main models, IQ and math and language skills from the beginning of the school year, as well the order in which peer effects are introduced into the model. This allows me to address issues of non-cognitive traits potentially proxying for measures of cognitive traits or vice versa (Cooley Fruehwirth, 2014).

Tables F.22 and F.23 present the LIM models for math and language, but instead of controlling for IQ and performance in that subject at the beginning of the school year, I only

Table D.19: Exogeneity of peer groups III: Guryan et al. (2009)

VARIABLES	(1) CON	(2) AGREE	(3) EXTRA
Mean peer conscientiousness math	0.1666***		
	(0.0485)		
Mean peer agreeableness math		0.0336	
		(0.0271)	
Mean peer extraversion math			0.0350
			(0.0375)
Constant	-8.4603***	-14.7455***	-21.6532***
	(1.7158)	(2.5420)	(3.4258)
Observations	3,174	3,174	3,174
R-squared	0.7142	0.5485	0.5791
School leave out means	Yes	Yes	Yes
Elective dummies	Yes	Yes	Yes
Mean peer conscientiousness language	0.1686***		
	(0.0490)		
Mean peer agreeableness language		0.0337	
		(0.0270)	
Mean peer extraversion language			0.0346
			(0.0374)
Constant	-8.4675***	-14.7450***	-21.6540***
	(1.7150)	(2.5421)	(3.4261)
Observations	3,174	3,174	3,174
R-squared	0.7144	0.5485	0.5791
School leave out means	Yes	Yes	Yes
Elective dummies	Yes	Yes	Yes

This table presents regression results from a model similar to model 2 in Guryan et al. (2009). The top panel contains regressions, for math peers and the bottom panel for language peers. Each regression also includes the school level take out mean of that trait as well as the elective choices. *** p<0.01, ** p<0.05, * p<0.1

Table E.20: Linear-in-Means Math Models without Primary School Peers

VARIABLES	(1) math score	(2) math score	(3) math score	(4) math score	(5) math score
Average peer effect terms					
Conscientiousness					0.1359**
					(0.0545)
Agreeableness					0.0625*
					(0.0328)
Extraversion					-0.0733*
					(0.0430)
Math score at t=0			0.1139	0.0808	0.0468
			(0.0841)	(0.0825)	(0.0820)
IQ			0.0111	-0.0037	-0.0855
			(0.0696)	(0.0669)	(0.0751)
Male			-0.0908	-0.0571	0.0172
			(0.2958)	(0.2994)	(0.2885)
Father tertiary education			0.2086	0.1209	0.0859
			(0.1959)	(0.1852)	(0.1903)
Pupil variables					
Conscientiousness		0.1709***		0.1494***	0.1500***
		(0.0279)		(0.0288)	(0.0292)
Agreeableness		0.0346		0.0352*	0.0409*
		(0.0208)		(0.0208)	(0.0210)
Extraversion		-0.0333*		-0.0294	-0.0313
		(0.0198)		(0.0202)	(0.0198)
Math score at t=0	0.3578***	0.3246***	0.3275***	0.3088***	0.3057***
	(0.0389)	(0.0351)	(0.0357)	(0.0334)	(0.0328)
IQ	0.3631***	0.2966***	0.3307***	0.2848***	0.2831***
	(0.0443)	(0.0400)	(0.0479)	(0.0432)	(0.0429)
Male	-0.0381	0.0074	-0.0280	0.0110	0.0150
	(0.0446)	(0.0442)	(0.0441)	(0.0464)	(0.0455)
Father tertiary education	0.0516	0.0290	0.0302	0.0179	0.0204
	(0.0436)	(0.0428)	(0.0437)	(0.0429)	(0.0433)
Math class size	-0.0174*	-0.0196**	-0.0274**	-0.0255**	-0.0228**
	(0.0094)	(0.0088)	(0.0120)	(0.0114)	(0.0111)
Constant	-0.3081***	-0.3696***	-0.1719	-0.2853	0.0018
	(0.1088)	(0.1084)	(0.2145)	(0.2190)	(0.2359)
School FE	Yes	Yes	Yes	Yes	Yes
Observations	1,899	1,899	1,899	1,899	1,899
R-squared	0.5438	0.5579	0.5497	0.5603	0.5636

Cluster robust standard errors in parentheses. All coefficients presented here represent effect sizes. These regressions also include control variables for peer and individual family income not presented in this table. The sample only includes individuals who knew less than 10 percent of their secondary school math classmates from primary school. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table E.21: Linear-in-Means Language Models without Primary School Peers

VARIABLES	(1) language score	(2) language score	(3) language score	(4) language score	(5) language score
Average peer effect terms					
Conscientiousness					0.1361**
					(0.0538)
Agreeableness					-0.0044
					(0.0236)
Extraversion					-0.0099
					(0.0330)
Language score at t=0			0.0258	0.0085	-0.0332
			(0.0860)	(0.0821)	(0.0792)
IQ			0.0457	0.0289	-0.0309
			(0.0914)	(0.0900)	(0.0898)
Male			-0.2414	-0.2024	-0.1476
			(0.1639)	(0.1593)	(0.1578)
Father tertiary education			0.5209***	0.4452**	0.3220*
			(0.1759)	(0.1768)	(0.1827)
Pupil variables					
Conscientiousness		0.1748***		0.1537***	0.1471***
		(0.0228)		(0.0230)	(0.0220)
Agreeableness		-0.0190		-0.0195	-0.0160
		(0.0169)		(0.0175)	(0.0176)
Extraversion		-0.0303		-0.0269	-0.0266
		(0.0184)		(0.0171)	(0.0162)
Language score at t=0	0.4616***	0.4162***	0.4351***	0.4049***	0.4070***
	(0.0674)	(0.0644)	(0.0657)	(0.0637)	(0.0626)
IQ	0.2885***	0.2373***	0.2623***	0.2259***	0.2229***
	(0.0398)	(0.0333)	(0.0381)	(0.0325)	(0.0323)
Male	-0.1803***	-0.1734***	-0.1717***	-0.1691***	-0.1633***
	(0.0377)	(0.0392)	(0.0369)	(0.0400)	(0.0396)
Father tertiary education	0.0718*	0.0515	0.0570	0.0440	0.0371
	(0.0384)	(0.0371)	(0.0373)	(0.0368)	(0.0364)
Language class size	-0.0133*	-0.0154**	-0.0186**	-0.0178**	-0.0168**
	(0.0072)	(0.0063)	(0.0070)	(0.0067)	(0.0065)
Constant	-0.2487***	-0.3518***	-0.1123	-0.2652**	-0.0738
	(0.0847)	(0.0763)	(0.1312)	(0.1296)	(0.1512)
School FE	Yes	Yes	Yes	Yes	Yes
Observations	1,897	1,897	1,897	1,897	1,897
R-squared	0.7105	0.7216	0.7180	0.7261	0.7281

Cluster robust standard errors in parentheses. All coefficients presented here represent effect sizes. These regressions also include control variables for peer and individual family income not presented in this table. The sample only includes individuals who knew less than 10 percent of their secondary school language classmates from primary school. *** p<0.01, ** p<0.05, * p<0.1

control for IQ. These results show that average peer IQ has a positive and statistically significant effect on individual math scores when it is included without peer math performance. This effect decreases in magnitude and loses significance, however, once average peer conscientiousness is introduced in Column (5) of Tables F.22 and F.23. This implies that average peer cognitive ability, which is found to be significant in other papers, may be proxying for peer non-cognitive traits which are not included in those models.

In Tables F.24 and F.25 I attempt to further disentangle how variables may be proxying for each other by building up the models in a different order than in the main analysis of the paper. In these models, I first include only the non-cognitive peer effects in Column (1). In Column (2), I add the other peer effects terms. In Column (3), I add the individual non-cognitive measures and in Column (4) the individual cognitive ability measures. Finally, in Column (5) I add all measures, which is makes it the same Column (5) as in Tables 5 and 6. These tables show that the magnitude of the peer effect term for conscientious is significantly decreased (approximately by half) once the traditional peer effects for cognitive ability are included. It is further reduced with the inclusion of individual level variables. This implies that conscientiousness is also a proxy for ability, but still has a positive effect over and above it.

Table F.22: Linear-in-Means Math Models with Just IQ

VARIABLES	(1) math score	(2) math score	(3) math score	(4) math score	(5) math score
Average peer effect terms					
Conscientiousness					0.1600***
					(0.0474)
Agreeableness					0.0204
					(0.0281)
Extraversion					-0.0687*
					(0.0356)
IQ			0.1816***	0.1237**	0.0299
			(0.0507)	(0.0473)	(0.0560)
Male			-0.1998	-0.1240	-0.0473
			(0.1819)	(0.1781)	(0.1683)
Father tertiary education			0.4297**	0.3131*	0.2398
			(0.1929)	(0.1735)	(0.1699)
Pupil variables					
Conscientiousness		0.2352***		0.1992***	0.1912***
		(0.0262)		(0.0251)	(0.0243)
Agreeableness		0.0158		0.0147	0.0185
		(0.0165)		(0.0160)	(0.0159)
Extraversion		-0.0304*		-0.0280*	-0.0272*
		(0.0158)		(0.0157)	(0.0154)
IQ	0.5470***	0.4300***	0.4629***	0.3905***	0.3881***
	(0.0305)	(0.0307)	(0.0322)	(0.0323)	(0.0323)
Male	-0.0987**	-0.0486	-0.0655*	-0.0343	-0.0298
	(0.0481)	(0.0417)	(0.0351)	(0.0342)	(0.0341)
Father tertiary education	0.0870**	0.0453	0.0553	0.0301	0.0298
	(0.0349)	(0.0345)	(0.0348)	(0.0344)	(0.0346)
Math class size	0.0014	-0.0039	-0.0127	-0.0126	-0.0128
	(0.0073)	(0.0064)	(0.0089)	(0.0084)	(0.0080)
Constant	-0.5526***	-0.5428***	-0.2049	-0.3214*	-0.4030**
	(0.1395)	(0.1227)	(0.1919)	(0.1848)	(0.1726)
School FE	Yes	Yes	Yes	Yes	Yes
Observations	3,174	3,174	3,174	3,174	3,174
R-squared	0.4878	0.5149	0.5044	0.5221	0.5251

Cluster robust standard errors in parentheses. All coefficients presented here represent effect sizes. These regressions also include control variables for peer and individual family income not presented in this table. *** p<0.01, ** p<0.05, * p<0.1

Table F.23: Linear-in-Means Language Models with Just IQ

VARIABLES	(1) language score	(2) language score	(3) language score	(4) language score	(5) language score
Average peer effect terms					
Conscientiousness					0.1024**
					(0.0444)
Agreeableness					-0.0012
					(0.0263)
Extraversion					-0.0139
					(0.0248)
IQ			0.1948***	0.1344***	0.0679
			(0.0438)	(0.0410)	(0.0519)
Male			-0.2635*	-0.1756	-0.1237
			(0.1353)	(0.1286)	(0.1321)
Father tertiary education			0.4868***	0.3529*	0.2661
			(0.1803)	(0.1777)	(0.1830)
Pupil variables					
Conscientiousness		0.2761***		0.2387***	0.2316***
		(0.0158)		(0.0166)	(0.0154)
Agreeableness		-0.0028		-0.0041	-0.0016
		(0.0140)		(0.0144)	(0.0146)
Extraversion		-0.0487***		-0.0469***	-0.0464***
		(0.0134)		(0.0129)	(0.0124)
IQ	0.5886***	0.4582***	0.5000***	0.4162***	0.4149***
	(0.0213)	(0.0207)	(0.0230)	(0.0200)	(0.0202)
Male	-0.3459***	-0.3002***	-0.3096***	-0.2836***	-0.2809***
	(0.0354)	(0.0354)	(0.0332)	(0.0365)	(0.0357)
Father tertiary education	0.1461***	0.0989***	0.1123***	0.0829**	0.0803**
	(0.0354)	(0.0324)	(0.0355)	(0.0333)	(0.0327)
Language class size	0.0036	-0.0030	-0.0129*	-0.0133**	-0.0134**
	(0.0065)	(0.0052)	(0.0068)	(0.0063)	(0.0060)
Constant	-0.1877	-0.1748*	0.2515*	0.1130	0.0868
	(0.1245)	(0.0989)	(0.1387)	(0.1275)	(0.1209)
School FE	Yes	Yes	Yes	Yes	Yes
Observations	3,174	3,174	3,174	3,174	3,174
R-squared	0.6200	0.6529	0.6392	0.6615	0.6626

Cluster robust standard errors in parentheses. All coefficients presented here represent effect sizes. These regressions also include control variables for peer and individual family income not presented in this table. *** p<0.01, ** p<0.05, * p<0.1

Table F.24: Building up Linear-in-Means Math Models

VARIABLES	(1) math score	(2) math score	(3) math score	(4) math score	(5) math score
Average peer effect terms					
Conscientiousness	0.5260***	0.2643***	0.1730***	0.1494***	0.1458***
	(0.0448)	(0.0509)	(0.0458)	(0.0509)	(0.0515)
Agreeableness	0.0268	0.0041	0.0291	0.0252	0.0214
	(0.0289)	(0.0293)	(0.0282)	(0.0271)	(0.0253)
Extraversion	-0.0792**	-0.0629*	-0.0704**	-0.0680*	-0.0679*
	(0.0344)	(0.0348)	(0.0323)	(0.0359)	(0.0355)
Math score at t=0		0.1241**	0.0702	-0.0465	-0.0089
		(0.0590)	(0.0596)	(0.0615)	(0.0684)
IQ		0.1599**	0.0975	-0.0333	-0.0232
		(0.0692)	(0.0686)	(0.0679)	(0.0698)
Male		-0.1551	-0.0266	-0.0089	-0.0067
		(0.1819)	(0.1724)	(0.1719)	(0.1655)
Father tertiary education		0.5108***	0.3523**	0.3748**	0.3123*
		(0.1794)	(0.1595)	(0.1667)	(0.1739)
Pupil variables					
Conscientiousness			0.3392***	0.1603***	0.1597***
			(0.0242)	(0.0251)	(0.0258)
Agreeableness			-0.0025	0.0140	0.0100
			(0.0177)	(0.0156)	(0.0155)
Extraversion			-0.0480***	-0.0244	-0.0253
			(0.0154)	(0.0153)	(0.0152)
Math score at t=0				0.2990***	0.3018***
				(0.0204)	(0.0205)
IQ				0.2453***	0.2472***
				(0.0328)	(0.0328)
Male					-0.0385
					(0.0327)
Father tertiary education					0.0160
					(0.0321)
Math class size					-0.0172**
Constant	-0.5296***	-0.4773***	-0.6826***	-0.6003***	-0.2483
	(0.0359)	(0.1078)	(0.1089)	(0.1138)	(0.1886)
School FE	Yes	Yes	Yes	Yes	Yes
Observations	3,174	3,174	3,174	3,174	3,174
R-squared	0.3946	0.4084	0.4613	0.5531	0.5553

Cluster robust standard errors in parentheses. All coefficients presented here represent effect sizes. These regressions also include control variables for peer and individual family income not presented in this table. *** p<0.01, ** p<0.05, * p<0.1

Table F.25: Building up Linear-in-Means Language Models

VARIABLES	(1) language score	(2) language score	(3) language score	(4) language score	(5) language score
Average peer effect terms					
Conscientiousness	0.5527***	0.2318***	0.1276***	0.1362**	0.1209**
	(0.0350)	(0.0491)	(0.0466)	(0.0512)	(0.0502)
Agreeableness	0.0022	-0.0263	0.0008	-0.0104	-0.0087
	(0.0335)	(0.0270)	(0.0270)	(0.0210)	(0.0216)
Extraversion	-0.0322	-0.0183	-0.0236	-0.0158	-0.0135
	(0.0248)	(0.0223)	(0.0242)	(0.0271)	(0.0267)
Language score at t=0		0.1447**	0.0914	-0.0759	-0.0394
		(0.0699)	(0.0587)	(0.0645)	(0.0652)
IQ		0.2030***	0.1276*	-0.0057	0.0058
		(0.0737)	(0.0737)	(0.0656)	(0.0641)
Male		-0.3732***	-0.2460*	-0.2299*	-0.1040
		(0.1317)	(0.1243)	(0.1325)	(0.1309)
Father tertiary education		0.5189***	0.3603**	0.3513*	0.2651
		(0.1802)	(0.1790)	(0.1794)	(0.1892)
Pupil variables					
Conscientiousness			0.3888***	0.1478***	0.1503***
			(0.0203)	(0.0175)	(0.0173)
Agreeableness			-0.0059	0.0159	0.0008
			(0.0183)	(0.0147)	(0.0132)
Extraversion			-0.0663***	-0.0292**	-0.0328**
			(0.0135)	(0.0127)	(0.0132)
Language score at t=0				0.4251***	0.4060***
				(0.0427)	(0.0430)
IQ				0.2095***	0.2294***
				(0.0247)	(0.0254)
Male					-0.2053***
					(0.0327)
Father tertiary education					0.0472
					(0.0289)
Language class size					-0.0156**
					(0.0061)
Constant	-0.1967***	-0.0232	-0.2528***	-0.2309***	0.0988
	(0.0383)	(0.0706)	(0.0713)	(0.0711)	(0.1129)
School FE	Yes	Yes	Yes	Yes	Yes
Observations	3,174	3,174	3,174	3,174	3,174
R-squared	0.4956	0.5143	0.5825	0.7043	0.7120

Cluster robust standard errors in parentheses. All coefficients presented here represent effect sizes. These regressions also include control variables for peer and individual family income not presented in this table. *** p<0.01, ** p<0.05, * p<0.1